Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depressive disorder

Naomi R Wray [1,2 †], Stephan Ripke [3,4,5 †], Manuel Mattheisen [6,7,8,9 †], Maciej Trzaskowski [1 †], Enda M Byrne [1], Abdel Abdellaoui [10], Mark J Adams [11], Esben Agerbo [8,12,13], Tracy M Air [14], Till F M Andlauer [15,16], Silviu-Alin Bacanu [17], Marie Bækvad-Hansen [8,18], Aartjan T F Beekman [19], Tim B Bigdeli [17,20], Elisabeth B Binder [15,21], Douglas H R Blackwood [11], Julien Bryois [22], Henriette N Buttenschøn [7,8,23], Jonas Bybjerg-Grauholm [8,18], Na Cai [24,25], Enrique Castelao [26], Jane Hvarregaard Christensen [6,7,8], Toni-Kim Clarke [11], Jonathan R I Coleman [27], Lucía Colodro-Conde [28], Baptiste Couvy-Duchesne [29,30], Nick Craddock [31], Gregory E Crawford [32,33], Cheynna A Crowley [34], Hassan S Dashti [3,35], Gail Davies [36], Ian J Deary [36], Franziska Degenhardt [37,38], Eske M Derks [28], Nese Direk [39,40], Conor V Dolan [10], Erin C Dunn [41,42,43], Thalia C Eley [27], Nicholas Eriksson [44], Valentina Escott-Price [45], Farnush Farhadi Hassan Kiadeh [46], Hilary K Finucane [47,48], Andreas J Forstner [37,38,49,50], Josef Frank [51], Héléna A Gaspar [27], Michael Gill [52], Paola Giusti-Rodríguez [53], Fernando S Goes [54], Scott D Gordon [55], Jakob Grove [6,7,8,56], Lynsey S Hall [11,57], Christine Søholm Hansen [8,18], Thomas F Hansen [58,59,60], Stefan Herms [37,38,50], Ian B Hickie [61], Per Hoffmann [37,38,50], Georg Homuth [62], Carsten Horn [63], Jouke-Jan Hottenga [10], David M Hougaard [8,18], Ming Hu [64], Craig L Hyde [65], Marcus Ising [66], Rick Jansen [19,19], Fulai Jin [67,68], Eric Jorgenson [69], James A Knowles [70], Isaac S Kohane [71,72,73], Julia Kraft [5], Warren W. Kretzschmar [74], Jesper Krogh [75], Zoltán Kutalik [76,77], Jacqueline M Lane [3,35,78], Yihan Li [74], Yun Li [34,53], Penelope A Lind [28], Xiaoxiao Liu [68], Leina Lu [68], Donald J MacIntyre [79,80], Dean F MacKinnon [54], Robert M Maier [2], Wolfgang Maier [81], Jonathan Marchini [82], Hamdi Mbarek [10], Patrick McGrath [83], Peter McGuffin [27], Sarah E Medland [28], Divya Mehta [2,84], Christel M Middeldorp [10,85,86], Evelin Mihailov [87], Yuri Milaneschi [19,19], Lili Milani [87], Francis M Mondimore [54], Grant W Montgomery [1], Sara Mostafavi [88,89], Niamh Mullins [27], Matthias Nauck [90,91], Bernard Ng [89], Michel G Nivard [10], Dale R Nyholt [92], Paul F O'Reilly [27], Hogni Oskarsson [93], Michael J Owen [94], Jodie N Painter [28], Carsten Bøcker Pedersen [8,12,13], Marianne Giørtz Pedersen [8,12,13], Roseann E. Peterson [17,95], Erik Pettersson [22], Wouter J Peyrot [19], Giorgio Pistis [26], Danielle Posthuma [96,97], Shaun M Purcell [98], Jorge A Quiroz [99], Per Qvist [6,7,8], John P Rice [100], Brien P. Riley [17], Margarita Rivera [27,101], Saira Saeed Mirza [40], Richa Saxena [3,35,78], Robert Schoevers [102], Eva C Schulte [103,104], Ling Shen [69], Jianxin Shi [105], Stanley I Shyn [106], Engilbert Sigurdsson [107], Grant C B Sinnamon [108], Johannes H Smit [19], Daniel J Smith [109], Hreinn Stefansson [110], Stacy Steinberg [110], Craig A Stockmeier [111], Fabian Streit [51], Jana Strohmaier [51], Katherine E Tansey [112], Henning Teismann [113], Alexander Teumer [114], Wesley Thompson [8,59,115,116], Pippa A Thomson [117], Thorgeir E Thorgeirsson [110], Chao Tian [44], Matthew Traylor [118], Jens Treutlein [51], Vassily Trubetskoy [5], André G Uitterlinden [119], Daniel Umbricht [120], Sandra Van der Auwera [121], Albert M van Hemert [122], Alexander Viktorin [22], Peter M Visscher [1,2], Yunpeng Wang [8,59,115], Bradley T. Webb [123], Shantel Marie Weinsheimer [8,59], Jürgen Wellmann [113], Gonneke Willemsen [10], Stephanie H Witt [51], Yang Wu [1], Hualin S Xi [124], Jian Yang [2,125], Futao Zhang [1,], eQTLGen Consortium [126], 23andMe Research Team [44], Volker Arolt [127], Bernhard T Baune [14], Klaus Berger [113], Dorret I Boomsma [10], Sven Cichon [37,50,128,129], Udo Dannlowski [127], EJC de Geus [10,130], J Raymond DePaulo [54], Enrico Domenici [131], Katharina Domschke [132], Tõnu Esko [3,87], Hans J Grabe [121], Steven P Hamilton [133], Caroline Hayward [134], Andrew C Heath [100], David A Hinds [44], Kenneth S Kendler [17], Stefan Kloiber [66,135,136], Glyn Lewis [137], Qingqin S Li [138], Susanne Lucae [66], Pamela AF Madden [100], Patrik K Magnusson [22], Nicholas G Martin [55], Andrew M McIntosh [11,36], Andres Metspalu [87,139], Ole Mors [8,140], Preben Bo Mortensen [7,8,12,13], Bertram Müller-Myhsok [15,16,141], Merete Nordentoft [8,142], Markus M Nöthen [37,38], Michael C O'Donovan [94], Sara A Paciga [143], Nancy L Pedersen [22], Brenda WJH Penninx [19], Roy H Perlis [42,144], David J Porteous [117], James B Potash [145], Martin Preisig [26], Marcella Rietschel [51], Catherine Schaefer [69], Thomas G Schulze [51,104,146,147,148], Jordan W Smoller [41,42,43], Kari Stefansson [110,149], Henning Tiemeier [40,150,151], Rudolf Uher [152], Henry Völzke [114], Myrna M Weissman [83,153], Thomas Werge [8,59,154], Ashley R Winslow [155,156], Cathryn M

Lewis [27,157] *, Douglas F Levinson [158] *, Gerome Breen [27,159] *, Anders D Børglum [6,7,8] *, Patrick F Sullivan [22,53,160] *, for the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium.

† Equal contributions. * Co-last authors. Affiliations are listed toward the end of the manuscript.

Correspond with: PF Sullivan (pfsulliv@med.unc.edu), Department of Genetics, CB#7264, University of North Carolina, Chapel Hill, NC, 27599-7264, USA. Voice, +919-966-3358. NR Wray (naomi.wray@uq.edu.au), Institute for Molecular Bioscience, Queensland Brain Institute, Brisbane, Australia. Voice, +61 7 334 66374.

Major depressive disorder (MDD) is a notably complex illness with a lifetime prevalence of 14%. [1] ***MDD is often chronic or recurrent and is thus accompanied by considerable morbidity, excess mortality, substantial costs, and heightened risk of suicide.*** [2-6] ***MDD is a major cause of disability worldwide.*** [7] ***We conducted a genome-wide association (GWA) meta-analysis in 135,458 MDD cases and 344,901 controls, and identified 44 independent loci that met criteria for statistical significance. We present extensive analyses of these results which provide new insights into the nature of MDD. The genetic findings were associated with clinical features of MDD, and implicated prefrontal and anterior cingulate cortex in the pathophysiology of MDD (regions exhibiting anatomical differences between MDD cases and controls). Genes that are targets of antidepressant medications were strongly enriched for MDD association signals ($P_{FDR}<0.01$), suggesting the relevance of these findings for improved pharmacotherapy of MDD. Sets of genes involved in gene splicing and in creating isoforms were also enriched for smaller MDD GWA p-values, and these gene sets have also been implicated in schizophrenia and autism. Genetic risk for MDD was correlated with that for many adult and childhood onset psychiatric disorders. Our analyses suggested important relations of genetic risk for MDD with educational attainment, body mass, and schizophrenia: the genetic basis of lower educational attainment and higher body mass were putatively causal for MDD whereas MDD and schizophrenia reflected a partly shared biological etiology. All humans carry lesser or greater numbers of genetic risk factors for MDD, and a continuous measure of risk underlies the observed clinical phenotype. MDD is not a distinct entity that neatly demarcates normalcy from pathology but rather a useful clinical construct associated with a range of adverse outcomes and the end result of a complex process of intertwined genetic and environmental effects. These findings help refine and define the fundamental basis of MDD.***

Twin studies attribute ~40% of the variation in liability to MDD to additive genetic effects (heritability, h^2), [9] and h^2 may be greater for recurrent, early-onset, and postpartum MDD. [10,11] GWA studies of MDD have had notable difficulties in identifying loci. [12] Previous findings suggest that an appropriately designed study should identify susceptibility loci. Direct estimates of the proportion of variance attributable to genome-wide SNPs (SNP heritability, h^2_{SNP}) indicate that around a quarter of the h^2 for MDD is due to common genetic variants. [13,14] Although there were no significant findings in the initial Psychiatric Genomics Consortium (PGC) MDD mega-analysis (9,240 MDD cases) [15] or in the CHARGE meta-analysis of depressive symptoms (34,549 respondents), [16] more recent studies have proven modestly successful. A study of Han Chinese women (5,303 MDD cases) identified two genome-wide significant loci, [17] a meta-analysis of depressive symptoms (161,460 individuals) identified two loci, [18] and an analysis of self-reported MDD identified 15 loci (75,607 cases). [19]

There are many reasons why identifying causal loci for MDD has proven difficult. [12] MDD is probably influenced by many genetic loci each with small effects, [20] as are most common complex human diseases [21] including psychiatric disorders. [22,23] A major lesson in human complex trait genetics is that

large samples are essential, especially for common and etiologically heterogeneous illnesses like MDD. [24] We sought to accumulate a large sample to identify common genetic variation involved in the etiology of MDD. [24]

J*nalysis of MDD anchor with six expanded cohorts shows polygenic prediction* V *clinical relevance*

We defined an "anchorwcohort of 29 samples that mostly applied standard methods for assessing MDD (N*able* R1). MDD cases in the anchor cohort were traditionally ascertained and typically characterized (i.e., using direct interviews with structured diagnostic instruments). We identified six "expandedw cohorts that used alternative methods to identify MDD (N*able* RFxdeCODE, Generation Scotland, GERA, iPSYCH, UK Biobank, and 23andMe, Inc.). All seven cohorts focused on clinically-significant MDD. We evaluated the comparability of these cohorts (N*able* RK) by estimating the common-variant genetic correlations (&) of the anchor cohort with the expanded cohorts. These analyses strongly supported the comparability of the seven cohorts (N*able* R4) as the weighted mean & was 0.76 (SE 0.028) with no statistical evidence of heterogeneity in the & estimates (*P*y0.13). As a benchmark for the MDD & estimates, the weighted mean & between schizophrenia cohorts was 0.84 (SE 0.05). [13]

We completed a GWA meta-analysis of 9.6 million imputed SNPs in seven cohorts containing 130,664 MDD cases and 330,470 controls (V*igure 1*xfull details in T*nline Methods*). There was no evidence of uncontrolled inflation (LD score regression intercept 1.018, SE 0.009). We estimated h^2_{SNP} to be 8.9% (SE 0.004, liability scale, assuming lifetime population risk of 0.15), and this is around a quarter of h^2 estimated from twin or family studies. [9] This fraction is somewhat lower than that of other complex traits, [21] and is plausibly due to etiological heterogeneity.

We used genetic risk score (GRS) analyses to demonstrate the validity of our GWA results for clinical MDD (V*igure* F). As expected, the variance explained in out-of-sample prediction increased with the size of the GWA discovery cohort (V*igure* F*a*). Across all samples in the anchor cohort, GRS explained 1.9% of variance in liability (V*igure* R1*a*), GRS ranked cases higher than controls with probability 0.57, and the odds ratio of MDD for those in the 10th versus 1st GRS decile (OR10) was 2.4 (V*igure* F*b*, N*able* RQ). GRS were significantly higher in those with more severe MDD, as measured in different ways (V*igure* F*c*).

D*nplications of the individual loci for the biology of MDD*

Our meta-analysis of seven MDD cohorts identified 44 independent loci that were statistically significant (*P*z5x10^{-8}), statistically independent of any other signal, [25] supported by multiple SNPs, and showed consistent effects across cohorts. This number is consistent with our prediction that MDD GWA discovery would require about five times more cases than for schizophrenia (lifetime risk ~1% and h^2 ~0.8) to achieve approximately similar power. [26] Of these 44 loci, 30 are novel and 14 were significant in a prior study of MDD or depressive symptoms (the overlap of our findings: 1{1 with the CHARGE depressive symptom study, [16] 0{2 overlap with CONVERGE MDD study, [17] 1{2 overlap with the SSGAC depressive symptom study, [18] and 13{16 overlap with 23andMe self-report of MDD [19]). There are few trans-ancestry comparisons for MDD so we contrasted these European results with the Han Chinese CONVERGE study (T*nline Methods*).

N*able 1* lists genes in or near the lead SNP in each region, regional plots are in the R*upplemental* V*ile*, and N*able* RM provides extensive summaries of available information about the biological functions of the genes in each region. In nine of the 44 loci, the lead SNP is within a gene, there is no other gene within 200 kb, and the gene is known to play a role in neuronal development, synaptic function, transmembrane adhesion complexes, and{or regulation of gene expression in brain.

The two most significant SNPs are located in or near *OLFM4* and *NEGR1*, which were previously associated with obesity and body mass index. [27-32] *OLFM4* (olfactomedin 4) has diverse functions outside

the CNS including myeloid precursor cell differentiation, innate immunity, anti-apoptotic effects, gut inflammation, and is over-expressed in diverse common cancers. [33] Many olfactomedins also have roles in neurodevelopment and synaptic function [34] e.g., latrophilins form trans-cellular complexes with neurexins [35] and with *FLRT3* to regulate glutamatergic synapse number. [36] *Olfm4* was highly upregulated after spinal transection, possibly related to inhibition of subsequent neurite outgrowth. [37] *NEGR1* (neuronal growth regulator 1) influences axon extension and synaptic plasticity in cortex, hypothalamus, and hippocampus, [38-40] and modulates synapse formation in hippocampus [41,42] via regulation of neurite outgrowth. [43,44] High expression, modulated by nutritional state, is seen in brain areas relevant to feeding, suggesting a role in control of energy intake. [45] The same SNP alleles are associated with increased risk of obesity and MDD (see also Mendelian randomization analyses below) and are associated with *NEGR1* gene expression in brain (**Table R4**). The associated SNPs may tag two upstream common deletions (8 and 43 kb) that delete transcription factor binding sites, [46] although reports differ on whether the signal is driven by the shorter [27] or the longer deletion. [31] Thus, the top two associations are in or near genes that influence BMI and may be involved in neurite outgrowth and synaptic plasticity.

Novel associations reported here include *RBFOX1* and *LRFN5*. There are independent associations with MDD at both the 5' and the 3' ends of *RBFOX1* (1.7 Mb, RNA binding protein fox-1 homolog 1). This convergence makes it a strong candidate gene. Fox-1 regulates the expression of thousands of genes, many of which are expressed at synapses and enriched for autism-related genes. [47] The Fox-1 network regulates neuronal excitability and prevents seizures. [48] It directs splicing in the nucleus and binds to 3' UTRs of target mRNAs in the cytoplasm. [48,49] Of particular relevance to MDD, Fox-1 participates in the termination of the corticotropin releasing hormone response to stress by promoting alternative splicing of the PACAP receptor to its repressive form. [50] Thus, *RBFOX1* could play a role in the chronic hypothalamic-pituitary-adrenal axis hyperactivation that has been widely reported in MDD. [51]

LRFN5 (leucine rich repeat and fibronectin type III domain containing 5) encodes adhesion-like molecules involved in synapse formation. Common SNPs in *LRFN5* were associated with depressive symptoms in older adults in a gene-based GWA analysis. [52] LRFN5 induces excitatory and inhibitory presynaptic differentiation in contacting axons and regulates synaptic strength. [53,54] LRFN5 also limits T-cell response and neuroinflammation (CNS "immune privilege") by binding to herpes virus entry mediator; a LRFN5-specific monoclonal antibody increases activation of microglia and macrophages by lipopolysaccharide and exacerbates mouse experimental acquired encephalitis; [55] thus, reduced expression (the predicted effect of eQTLs in LD with the associated SNPs) could increase neuroinflammatory responses.

Gene-wise analyses identified 153 significant genes after controlling for multiple comparisons (**Table R3**). Many of these genes were in the extended MHC region (45 of 153) and their interpretation is complicated by high LD and gene density. In addition to the genes discussed above, other notable and significant genes outside of the MHC include multiple potentially "druggable" targets that suggest connections of the pathophysiology of MDD to neuronal calcium signaling (*CACNA1E* and *CACNA2D1*), dopaminergic neurotransmission (*DRD2*, a principal target of antipsychotics), glutamate neurotransmission (*GRIK5* and *GRM5*), and presynaptic vesicle trafficking (*PCLO*).

Finally, comparison of the MDD loci with 108 loci for schizophrenia [22] identified six shared loci. Many SNPs in the extended MHC region are strongly associated with schizophrenia, but implication of the MHC region is novel for MDD. Another example is *TCF4* (transcription factor 4) which is strongly associated with schizophrenia but not previously with MDD. TCF4 is essential for normal brain development, and rare mutations in *TCF4* cause Pitt–Hopkins syndrome which includes autistic features.

[56] GRS calculated from the schizophrenia GWA results explained 0.8% of the variance in liability of MDD (*Figure 3c*).

Implications for the biology of MDD using functional genomic data

Results from "-omics" studies of functional features of cells and tissues are necessary to understand the biological implications of results of GWA for complex disorders like MDD. [57] To further elucidate the biological relevance of the MDD findings, we integrated the results with a wide range of functional genomic data. First, using enrichment analyses, we compared the MDD GWA findings to bulk tissue mRNA-seq from GTEx. [58] Only brain samples showed significant enrichment (*Figure 4*), and the three tissues with the most significant enrichments were all cortical. Prefrontal cortex and anterior cingulate cortex are important for higher-level executive functions and emotional regulation which are often impaired in MDD. Both regions were implicated in a large meta-analysis of brain MRI findings in adult MDD cases. [59] Second, given the predominance of neurons in cortex, we confirmed that the MDD genetic findings connect to genes expressed in neurons but not oligodendrocytes or astrocytes (*Figure S8*). [60] These results confirm that MDD is a brain disorder and provide validation for the utility of our genetic results for the etiology of MDD.

Third, we used partitioned LD score regression [61] to evaluate the enrichment of the MDD GWA findings in over 50 functional genomic annotations (*Figure S9* and *Table S16*). The major finding was the significant enrichment of MDD !"#$% in genomic regions conserved across 29 Eutherian mammals [62] (20.9 fold enrichment, *P*=1.4x10^{-15}). This annotation was also the most enriched for schizophrenia. [61] We could not evaluate regions conserved in primates or human "accelerated" regions as there were too few for confident evaluation. [62] The other major enrichments implied regulatory activity, and included open chromatin in human brain and an epigenetic mark of active enhancers (H3K4me1). Notably, exonic regions did not show enrichment suggesting that, as with schizophrenia, [20] genetic variants that change exonic sequences may not play a large role in MDD. We found no evidence that Neanderthal introgressed regions were enriched for MDD GWA findings. [63]

Fourth, we applied methods to integrate GWA SNP-MDD results with those from gene expression quantitative trait loci (eQTL) studies. SMR (summary data–based Mendelian randomization) [64] identified 13 MDD-associated SNPs with strong evidence that they control local gene expression in one or more tissues (*Table S17* and *Figure S10*), including two loci not reaching GWA significance (*TMEM64* and *ZDHHC5*). A transcriptome-wide association study [65] applied to data from the dorsolateral prefrontal cortex [66] identified 17 genes where MDD-associated SNPs influenced gene expression (*Table S18*). These genes included *OLFM4* (discussed above).

Fifth, we added additional data types to attempt to improve understanding of individual loci. For the intergenic associations, we evaluated total-stranded RNA-seq data from human brain and found no evidence for unannotated transcripts in these regions. A particularly important data type is assessment of DNA-DNA interactions which can localize a GWA finding to a specific gene that may be nearby or hundreds of kb away. [67-69] We integrated the MDD findings with "easy Hi-C" data from brain cortical samples (3 adult, 3 fetal, more than 1 billion reads each). These data clarified three of the associations.

The statistically independent associations in *NEGR1* (rs1432639, *P*=4.6x10^{-15}) and over 200 kb away (rs12129573, *P*=4.0x10^{-12}) both implicate *NEGR1* (*Figure S11a*), the former likely due to the presence of a reportedly functional copy number polymorphism (see above) and the presence of intergenic loops. The latter association has evidence of DNA looping interactions with *NEGR1*. The association in *SOX5* (rs4074723) and the two statistically independent associations in *RBFOX1* (rs8063603 and rs7198928, *P*=6.9x10^{-9} and 1.0x10^{-8}) had only intragenic associations, suggesting that the genetic variation in the regions of the MDD associations act locally and can be assigned to these genes. In contrast, the

association in *RERE* (rs159963 Py3.2x10^{-8}) could not be assigned to *RERE* as it may contain super-enhancer elements given its many DNA-DNA interactions with many nearby genes (*Figure R4b*).

Implications for the biology of MDD based on the roles of sets of genes

A parsimonious explanation for the presence of many significant associations for a complex trait like MDD is that the different associations are part of a higher order grouping of genes. [70] These could be a biological pathway or a collection of genes with a functional connection. Multiple methods allow evaluation of the connection of MDD GWA results to sets of genes grouped by empirical or predicted function (i.e., pathway or gene set analysis).

Full pathway analyses are shown in *Table R11*, and the 19 pathways with false discovery rate q-values z 0.05 are summarized in *Figure 4*. The major groupings of significant pathways were: RBFOX1, RBFOX2, RBFOX3, or CELF4 regulatory networksxgenes whose mRNAs are bound by FMRPxsynaptic genesxgenes involved in neuronal morphogenesisx genes involved in neuron projectionx genes associated with schizophrenia (at *P*z10^{-4}) [22]xgenes involved in CNS neuron differentiationxgenes encoding voltage-gated calcium channelsx genes involved in cytokine and immune responsexand genes known to bind to the retinoid X receptor. Several of these pathways are implicated by GWA of schizophrenia and by rare exonic variation of schizophrenia and autism, [71,72] and immediately suggest shared biological mechanisms across these disorders.

A key issue for common variant GWA studies is their relevance for pharmacotherapy: do the results connect meaningfully to known medication targets and might they suggest new mechanisms or "druggablewtargets We conducted gene set analysis that compared the MDD GWA results to targets of antidepressant medications defined by pharmacological studies, [73] and found that 42 sets of genes encoding proteins bound by antidepressant medications were highly enriched for smaller MDD association *P*-values than expected by chance (42 drugs, rank enrichment test *P*y8.5x10^{-10}). This finding connects our MDD genomic findings to MDD therapeutics, and suggests the salience of these results for novel lead compound discovery for MDD. [74]

Implications for a deeper understanding of the clinically-defined entity [MDD

Past epidemiological studies associated MDD with many other diseases and traits. Due to limitations inherent to observational studies, understanding whether a phenotypic correlation is potentially causal or if it results from reverse causation or confounding is generally unclear. Genetic studies can now offer complementary strategies to assess whether a phenotypic association between MDD and a risk factor or a comorbidity is mirrored by a non-zero & (common variant genetic correlation) and, for some of these, evaluate the potential causality of the association given that exposure to genetic risk factors begins at conception.

We used LD score regression to estimate & of MDD with 221 psychiatric disorders, medical diseases, and human traits. [14,75] *Table R12* contains the full results, and *Table 5* holds the & values with false discovery rates z 0.01. First, there were very high genetic correlations for MDD with current depressive symptoms. Both correlations were close to +1 (the samples in one report overlapped partially with this MDD meta-analysis [18] but the other did not [16]). The & estimate in the MDD anchor samples with depressive symptoms was numerically smaller (0.80, SE 0.059) but the confidence intervals overlapped those for the full sample. Thus, the common-variant genetic architecture of lifetime MDD overlapped strongly with that of current depressive symptoms (bearing in mind that current symptoms had lower estimates of $!^{"}_{\#\$\%}$ compared to the lifetime measure of MDD).

Second, MDD had significant positive genetic correlations with every psychiatric disorder assessed as well as with smoking initiation. This is the most comprehensive and best-powered evaluation of the

relation of MDD with other psychiatric disorders yet published, and these results indicate that the common genetic variants that predispose to MDD overlap substantially with those for adult and childhood onset psychiatric disorders.

Third, MDD had positive genetic correlations with multiple measures of sleep quality (daytime sleepiness, insomnia, and tiredness). The first two of these correlations were based on a specific analysis of UK Biobank data (i.e., removing people with MDD, other major psychiatric disorders, shift workers, and those taking hypnotics). This pattern of correlations combined with the critical importance of sleep and fatigue in MDD (these are two commonly accepted criteria for MDD) suggests a close and potentially profound mechanistic relation. MDD also had a strong genetic correlation with neuroticism (a personality dimension assessing the degree of emotional instability)x this is consistent with the literature showing a close interconnection of MDD and this personality trait. The strong negative & with subjective well-being underscores the capacity of MDD to impact human health.

Finally, MDD had negative correlations with two proxy measures of intelligence, positive correlations with multiple measures of adiposity, relationship to female reproductive behavior (decreased age at menarche, age at first birth, and increased number of children), and positive correlations with coronary artery disease and lung cancer.

We used Mendelian randomization (MR) to investigate the relationships between genetically correlated traits. We conducted bi-directional MR analysis for four traits: years of education (EDY, a proxy for general intelligence) [76], body mass index (BMI) [27], coronary artery disease (CAD) [77], and schizophrenia [22]. These traits were selected because all of the following were true: phenotypically associated with MDD, significant & with MDD with an unclear direction of causality, and 30 independent genome-wide significant associations from large GWA.

We report GSMR (generalized summary statistic-based MR) results but obtained qualitatively similar results with other MR methods (**Table** R1K and **Figures** R4J-**D**). MR analyses provided evidence for a 1.15-fold increase in MDD per standard deviation of BMI (P_{GSMR}y2.7x10^{-7}) and a 0.89-fold decrease in MDD per standard deviation of EDY (P_{GSMR}y8.8x10^{-7}). There was no evidence of reverse causality of MDD for BMI (P_{GSMR}y0.81) or EDY (P_{GSMR}y0.28). For BMI there was some evidence of pleiotropy, as eight SNPs were excluded by the HEIDI-outlier test including SNPs near *OLFM4* and *NEGR1* (if these were included, the estimate of increased risk for MDD was greater). Thus, these results are consistent with EDY and BMI as causal risk factors or correlated with causal risk factors for MDD. For CAD, the MR analyses were not significant when considering MDD as an outcome (P_{GSMR}y0.39) or as an exposure (P_{GSMR}y0.13). We interpret the & of 0.12 between CAD and MDD to reflect a genome-wide correlation in the sign of effect sizes but no correlation in the effect size magnitudes: this is consistent with "type I pleiotropyw[78], that there are multiple molecular functions of these genetic variants (which may be tissue-specific in brain and heart). However, because the MR regression coefficient for MDD instruments has relatively high standard error, this analysis should be revisited when more MDD genome-wide significant SNP instruments become available from future MDD GWA studies.

We used MR to investigate the relationship between MDD and schizophrenia. Although MDD had positive & with many psychiatric disorders, only schizophrenia has sufficient associations for MR analyses. We found significant bi-directional correlations in SNP effect sizes for schizophrenia loci in MDD (P_{GSMR}y7.7x10^{-46}) and for MDD loci in schizophrenia (P_{GSMR}y6.3x10^{-15}). We interpret the MDD-schizophrenia & of 0.34 as reflecting type II pleiotropy [78] (i.e., consistent with shared biological pathways being causal for both disorders).

] *mpirically∃what is MDD^*

The nature of severe depression has been discussed for millennia. [79] This GWA meta-analysis is among the largest ever conducted for a psychiatric disorder, and provides a body of results that help refine and define the fundamental basis of MDD.

First, MDD is a brain disorder. Although this is not unexpected, some past models of MDD have had little or no place for heredity or biology. Our results indicate that genetics and biology are definite pieces in the puzzle of MDD. The genetic results best match gene expression patterns in prefrontal and anterior cingulate cortex, anatomical regions that show differences between MDD cases and controls. The genetic findings implicated neurons (not microglia or astrocytes), and we anticipate more detailed cellular localization when sufficient single-cell and single-nuclei RNA-seq datasets become available. [80]

Second, the genetic associations for MDD (as with schizophrenia) [61] tend to occur in genomic regions conserved across a range of placental mammals. Conservation suggests important functional roles. Given that this analysis did not implicate exons or coding regions, MDD may not be characterized by common changes in the amino acid content of proteins.

Third, the results also implicated developmental gene regulatory processes. For instance, the genetic findings pointed at *RBFOX1* (the presence of two independent genetic associations in *RBFOX1* strongly suggests that it is the MDD-relevant gene). Gene set analyses implicated genes containing binding sites to the protein product of *RBFOX1* in MDD, and this gene set is also significantly enriched for rare exonic variation in autism and schizophrenia. [71,72] These analyses highlight the potential importance of splicing to generate alternative isoformsxrisk for MDD may be mediated not by changes in isolated amino acids but rather by changes in the proportions of isoforms coming from a gene, given that isoforms often have markedly different biological functions. [81,82] These convergent results provide a tantalizing suggestion of a biological mechanism common to multiple severe psychiatric disorders.

Fourth, in the most extensive analysis of the genetic "connectionsw of MDD with a wide range of disorders, diseases, and human traits, we found significant positive genetic correlations with measures of body mass and negative genetic correlations with years of education. MR analyses suggested the *potential* causality of both correlations, and our results certainly provide hypotheses for more detailed prospective studies. However, further clarity requires larger and more informative GWA studies for a wider range of related traits (e.g., with 30 significant associations per trait). We strongly caution against interpretations of these results that go beyond the analyses undertaken (e.g., these results do not provide evidence that weight loss would have an antidepressant effect). The currently available data do not provide further insight about the fundamental driver or drivers of causality. The underlying mechanisms are likely more complex as it is difficult to envision how genetic variation in educational attainment or body mass alters risk for MDD without invoking an additional mechanistic component. For example, genetic variation underlying general intelligence might directly alter the development and function of discrete brain regions that alters intelligence and which also predisposes to worse mood regulation. Alternatively, genetic variation underlying general intelligence might lead to poorer development of cognitive strategies to handle adversity which increases risk for MDD. An additional possibility is that there are sets of correlated traits~e.g., personality, intelligence, sleep patterns, appetitive regulation, or propensity to exercise~and that these act in varying combinations in different people. Our results are inconsistent with a causal relation between MDD and subsequent changes in body mass or education years. If such associations are observed in epidemiological or clinical samples, then it is likely not MDD but something correlated with MDD that drives the association.

Fifth, we found significant positive correlations of MDD with all psychiatric disorders that we evaluated, including disorders prominent in childhood. This pattern of results indicates that the current

classification scheme for major psychiatric disorders does not align well with the underlying genetic basis of these disorders. The MR results for MDD and schizophrenia indicated a shared biological basis.

The dominant psychiatric nosological systems were principally designed for clinical utility, and are based on data that emerge during human interactions (i.e., observable signs and reported symptoms) and not objective measurements of pathophysiology. MDD is frequently comorbid with other psychiatric disorders, and the phenotypic comorbidity has an underlying structure that reflects shared origins (as inferred from factor analyses and twin studies). [83-86] Our genetic results add to this knowledge: MDD is not a discrete entity at any level of analysis. Rather, our data strongly suggest the existence of biological processes common to MDD and schizophrenia. It would be unsurprising if future work implicated bipolar disorder, anxiety disorders, and other psychiatric disorders as well.

Finally, as expected, we found that MDD had modest $!^{"}_{\#\$\%}$ (8.9%) since MDD is a complex malady with both genetic and environmental determinants. We found that MDD has a very high genetic correlation with proxy measures that can be briefly assessed. Lifetime major depressive disorder requires a constellation of signs and symptoms whose reliable scoring requires an extended interview with a trained clinician. However, the common variant genetic architecture of lifetime major depressive disorder in these seven cohorts (containing many subjects medically treated for MDD) has strong overlap with that of current depressive symptoms in general community samples. Similar relations of clinically-defined ADHD or autism with quantitative genetic variation in the population have been reported. [87,88] The MDD "disorder versus symptomwrelationship has been debated extensively, [89] but our data indicate that the common variant genetic overlap is very high. This finding has two important implications.

One implication is for future genetic studies of MDD. In a first phase, it should be possible to elucidate the bulk of the common variant genetic architecture of MDD using a cost-effective shortcut ~ large studies of genotyped individuals who complete brief lifetime MDD screening (a sample size approaching 1 million MDD cases may be achievable by 2020). In a second phase, with a relatively complete understanding of the genetic basis of MDD, one could then evaluate smaller samples of carefully phenotyped individuals with MDD to understand the clinical importance of the genetic results. These data could allow more precise delineation of the clinical heterogeneity of MDD (e.g., our demonstration that individuals with more severe or recurrent MDD have inherited a higher genetic loading for MDD than single-episode MDD). Subsequent empirical studies may show that it is possible to stratify MDD cases at first presentation to identify individuals at high risk for recurrence, poor outcome, poor treatment response, or who might subsequently develop a psychiatric disorder requiring alternative pharmacotherapy (e.g., schizophrenia or bipolar disorder). This could form a cornerstone of precision medicine in psychiatry.

The second implication is that people with MDD differ only by degree from those who have not experienced MDD. All humans carry lesser or greater numbers of genetic risk factors for MDD. Genetic risk for MDD is continuous and normally distributed with no clear point of demarcation. Non-genetic factors play important protective and pre-disposing roles (e.g., life events, exposure to chronic fear, substance abuse, and a wide range of life experiences and choices). The relation of blood pressure to essential hypertension is a reasonable analogy. All humans inherit different numbers of genetic variants that influence long-term patterns of blood pressure with environmental exposures and life choices also playing roles. The medical "disorderwof hypertension is characterized by blood pressure chronically over a numerical threshold above which the risks for multiple preventable diseases climb. MDD is not a "diseasew(i.e., a distinct entity delineable using an objective measure of pathophysiology) but indeed a disorder, a human-defined but definable syndrome that carries increased risk of adverse outcomes. The adverse outcomes of hypertension are diseases (e.g., stroke or myocardial infarction). The adverse

outcomes of MDD include elevation in risk for a few diseases, but the major impacts of MDD are death by suicide and disability.

In summary, this GWA meta-analysis of 130,664 MDD cases and 330,470 controls identified 44 loci. An extensive set of companion analyses provide insights into the nature of MDD as well as its neurobiology, therapeutic relevance, and genetic and biological interconnections to other psychiatric disorders. Comprehensive elucidation of these features is the primary goal of our genetic studies of MDD.

T*nline Methods*

Anchor cohort. Our analysis was anchored in a GWA mega-analysis of 29 samples of European-ancestry (16,823 MDD cases and 25,632 controls). N*able* R1 summarizes the source and inclusion{exclusion criteria for cases and controls for each sample. All samples in the initial PGC MDD papers were included.[13,15,90] All anchor samples passed a structured methodological review by MDD assessment experts (DF Levinson and KS Kendler). Cases were required to meet international consensus criteria (DSM-IV, ICD-9, or ICD-10) [91-93] for a lifetime diagnosis of MDD established using structured diagnostic instruments from assessments by trained interviewers, clinician-administered checklists, or medical record review. All cases met standard criteria for MDD, were directly interviewed (28{29 samples) or had medical record review by an expert diagnostician (1{29 samples), and most were ascertained from clinical sources (19{29 samples). Controls in most samples were screened for the absence of lifetime MDD (22{29 samples), and randomly selected from the population. We considered this the "anchorwcohort given use of standard methods of establishing the presence or absence of MDD.

The most direct and important way to evaluate the comparability of the samples comprising the anchor cohort is using SNP genotype data.[14,94] The sample sizes were too small to evaluate the common variant genetic correlations (&) between all pairs of anchor cohort samples (3,000 subjects per sample are recommended). As an alternative, we used "leave one outwgenetic risk scores (GRS, described below). We repeated this procedure by leaving out each of the anchor cohort samples so that we could evaluate the similarity of the common-variant genetic architectures of each sample to the rest of the anchor cohort. W*igure* R1J shows that all samples in the anchor cohort (except one) yielded significant differences in case-control distributions of GRS.

Expanded cohorts. We critically evaluated an "expandedwset of six independent, European-ancestry cohorts (113,841 MDD cases and 304,838 controls). N*able* RF summarizes the source and inclusion{exclusion criteria for cases and controls for each cohort. These cohorts used a range of methods for assessing MDD: Generation Scotland employed direct interviewsxiPSYCH (Denmark) used national treatment registersxdeCODE (Iceland) used national treatment registers and direct interviewsx GERA used Kaiser-Permanente treatment records (CA, US)x UK Biobank combined self-reported MDD symptoms and{or treatment for MDD by a medical professionalx and 23andMe used self-report of treatment for MDD by a medical professional. All controls were screened for the absence of MDD.

Cohort comparability. N*able* RK summarizes the numbers of cases and controls in the anchor cohort and the six expanded cohorts. The most direct and important way to evaluate the comparability of these cohorts for a GWA meta-analysis is using SNP genotype data.[14,94] We used LD score regression (described below) to estimate $!^{"}_{\#\$\%}$ for each cohort, and & for all pairwise combinations of the cohorts.

We compared the seven anchor and expanded cohorts. First, there was no indication of important sample overlap as the LDSC regression intercept between pairs of cohorts ranged from -0.01 to +0.01. Second, N*able* R4 shows $!^{"}_{\#\$\%}$ on the liability scale for each cohort. The $!^{"}_{\#\$\%}$ estimates range from 0.09 to 0.23 (for lifetime risk (y0.15) but the confidence intervals largely overlap. Third, N*able* R4 also shows

the & values for all pairs of anchor and expanded cohorts. The median & was 0.80 (interquartile range 0.67-0.96), and the upper 95% confidence interval on & included 0.75 for all pairwise comparisons. These results indicate that the common variant genetic architecture of the anchor and expanded cohorts overlap strongly, and provide critical support for the full meta-analysis of all cohorts.

Genotyping and quality control. Genotyping procedures can be found in the primary reports for each cohort (Nables R1-RF). Individual genotype data for all anchor cohorts, GERA, and iPSYCH were processed using the PGC "ricopiliw pipeline (URLs) for standardized quality control, imputation, and analysis. [22] The expanded cohorts from deCODE, Generation Scotland, UK Biobank, and 23andMe were processed by the collaborating research teams using comparable procedures. SNPs and insertion-deletion polymorphisms were imputed using the 1000 Genomes Project multi-ancestry reference panel (URLs).[95]

Quality control and imputation on the 29 PGC MDD anchor cohorts was performed according to standards from the PGC (Nable RK). The default parameters for retaining SNPs and subjects were: SNP missingness z 0.05 (before sample removal)x subject missingness z 0.02x autosomal heterozygosity deviation (F_{het} z0.2)x SNP missingness z 0.02 (after sample removal)x difference in SNP missingness between cases and controls z 0.02x and SNP Hardy-Weinberg equilibrium (P 10 ^{6} in controls or P 10 ^{10} in cases). These default parameters sufficiently controlled ! and false positive findings for 16 cohorts (boma, rage, shp0, shpt, edi2, gens, col3, mmi2, qi3c, qi6c, qio2, rai2, rau2, twg2, grdg, grnd). Two cohorts (gep3 and nes2) needed stricter SNP filtering and 11 cohorts needed additional ancestral matching (rot4, stm2, rde4) or ancestral outlier exclusion (rad2, i2b3, gsk1, pfm2, jjp2, cof3, roc3, mmo4). An additional cohort of inpatient MDD cases from Münster, Germany was processed through the same pipeline.

Genotype imputation was performed using the pre-phasing{imputation stepwise approach implemented in IMPUTE2 { SHAPEIT (chunk size of 3 Mb and default parameters). The imputation reference set consisted of 2,186 phased haplotypes from the 1000 Genomes Project dataset (August 2012, 30,069,288 variants, release "v3.macGT1W). After imputation, we identified SNPs with very high imputation quality (INFO 0.8) and low missingness (z1%) for building the principal components to be used as covariates in final association analysis. After linkage disequilibrium pruning (r^2 0.02) and frequency filtering (MAF 0.05), there were 23,807 overlapping autosomal SNPs in the data set. This SNP set was used for robust relatedness testing and population structure analysis. Relatedness testing identified pairs of subjects with) 0.2, and one member of each pair was removed at random after preferentially retaining cases over controls. Principal component estimation used the same collection of autosomal SNPs.

Identification of identical samples is easily accomplished given direct access to individual genotypes. [13] Two concerns are the use of the same control samples in multiple studies (e.g., GAIN or WTCCC controls) [96,97] and inclusion of closely related individuals. For cohorts where the PGC central analysis team had access to individual genotypes (all anchor cohorts and GERA), we used SNPs directly genotyped on all platforms to compute empirical relatedness, and excluded one of each duplicated or relative pair (defined as) 0.2). Within all other cohorts (deCODE, Generation Scotland, iPSYCH, UK Biobank, 23andMe, and CONVERGE), identical and relative pairs were identified and resolved using similar procedures. Identical samples between the anchor cohorts, iPSYCH, UK Biobank, and Generation Scotland were identified using genotype-based checksums (URLs), [98] and an individual on the collaborator|s side was excluded. Checksums were not available for the deCODE and 23andMe cohorts. Related pairs are not detectable by the checksum method but we did not find evidence of important overlap using LD score regression (the intercept between pairs of cohorts ranged from -0.01 to +0.01 with no evidence of important sample overlap).

Statistical analysis. In each cohort, logistic regression association tests were conducted for imputed marker dosages with principal components covariates to control for population stratification. Ancestry was evaluated using principal components analysis applied to directly genotyped SNPs. [99] In the anchor cohorts and GERA, we determined that all individuals in the final analyses were of European ancestry. European ancestry was confirmed in the other expanded cohorts by the collaborating research teams using similar procedures. We tested 20 principal components for association with MDD and included five principal components covariates for the anchor cohorts and GERA (all other cohorts adopted similar strategies). There was no evidence of stratification artifacts or uncontrolled test statistic inflation in the results from each anchor and extended cohort (e.g., ! $_{GC}$ was 0.995~1.043 in the anchor cohorts). The results were combined across samples using an inverse-weighted fixed effects model.[100] Reported SNPs have imputation marker INFO score 0.6 and allele frequencies 0.01 and 0.99, and effective sample size equivalent to 100,000 cases. For all cohorts, X-chromosome association results were conducted separately by sex, and then meta-analysed across sexes. [22] For two cohorts (GenScot and UKBB), we first conducted association analysis for genotyped SNPs by sex, then imputed association results using LD from the 1000 Genomes reference sample. [101]

Defining loci. GWA findings implicate genomic regions containing multiple significant SNPs ("loci"). There were almost 600 SNPs with $P \leq 5\times10^{-8}$ in this analysis. These are not independent associations but result from LD between SNPs. We collapsed the significant SNPs to 44 loci via the following steps.

- All SNPs were high-quality (imputation INFO score 0.6 and allele frequencies 0.01 and 0.99).
- We used "clumping" to convert MDD-associated SNPs to associated regions. We identified an index SNP with the smallest *P*-value in a genomic window and other SNPs in high LD with the index SNP using PLINK (--clump-p1 1e-4 --clump-p2 1e-4 --clump-r2 0.1 --clump-kb 3000). This retained SNPs with association $P \leq 0.0001$ and $r^2 \leq 0.1$ within 3 Mb windows. Only one SNP was retained from the extended MHC region due to its exceptional LD.
- We used bedtools (URLs) to combine partially or wholly overlapping clumps within 50 kb.
- We reviewed all regional plots, and removed two singleton associations (i.e., only one SNP exceeding genome-wide significance).
- We reviewed forest plots, and confirmed that association signals arose from the majority of the cohorts.
- We conducted conditional analyses. To identify independent associations within a 10 Mb region, we re-evaluated all SNPs in a region conditioning on the most significantly associated SNP using summary statistics [25] (superimposing the LD structure from the Atherosclerosis Risk in Communities Study sample).

Genetic risk score (GRS) analyses. To demonstrate the validity of our GWAS results, we conducted a series of GRS prediction analyses. The MDD GWA summary statistics identified associated SNP alleles and effect size which were used to calculate GRS for each individual in a target sample (i.e., the sum of the count of risk alleles weighted by the natural log of the odds ratio of the risk allele). In some analyses the target sample had been included as one of the 29 samples in the MDD anchor cohort; here, the discovery samples were meta-analyzed excluding this cohort. As in the PGC schizophrenia report, [22] we excluded uncommon SNPs (MAF ≤ 0.1), low-quality variants (imputation INFO ≤ 0.9), indels, and SNPs in the extended MHC region (chr6:25-34 Mb). We then LD pruned and "clumped" the data, discarding variants within 500 kb of, and in LD r^2 0.1 with the most associated SNP in the region. We generated GRS for individuals in target subgroups for a range of P-value thresholds (P_T: 5×10^{-8}, 1×10^{-6}, 1×10^{-4}, 0.001, 0.01, 0.05, 0.1, 0.2, 0.5, 1.0).

For each GRS analysis, five ways of evaluating the regression of phenotype on GRS are reported (**Table** RQ). The significance of the case-control score difference from logistic regression including ancestry PCs

and a study indicator (if more than one target dataset was analyzed) as covariates. 2) The proportion of variance explained (Nagelkerke|s R^2) computed by comparison of a full model (covariates + GRS) to a reduced model (covariates only). It should be noted that these estimates of R^2 reflect the proportion of cases in the case-control studies where this proportion may not reflect the underlying risk of in the population. 3) The proportion of variance on the liability scale explained by the GRS R^2 was calculated from the difference between full and reduced linear models and was then converted to the liability scale of the population assuming lifetime MDD risk of 15%. These estimates should be comparable across target sample cohorts, whatever the proportion of cases in the sample. 4) Area under the receiver operator characteristic curve (AUCx R library pROC) was estimated in a model with no covariates [22] where AUC can be interpreted as the probability of a case being ranked higher than a control. 5) Odds ratio for 10 GRS decile groups (these estimates also depend on both risk of MDD in the population and proportion of cases in the sample). We evaluated the impact of increasing sample size of the discovery sample GWA (*Figure* F*a*) and also using the schizophrenia GWA study [22] as the discovery sample. We also undertook GRS analysis for a target sample of MDD cases and controls not included in the meta-analysis (a clinical inpatient cohort of MDD cases and screened controls collected in Münster, Germany).

We conducted GRS analyses based on prior hypotheses from epidemiology of MDD using clinical measures available in some cohorts (if needed, the target sample was removed from the discovery GWA). We used GRS constructed from P_Ty0.05, selected as a threshold that gave high variance explained across cohorts (*Figure* F*a*). First, we used GRS analyses to test for higher mean GRS in cases with younger age at onset (AAO) of MDD compared to those with older AAO in the anchor cohort samples. To combine analyses across samples, we used within-sample standardized GRS residuals after correcting for ancestry principal components. Heterogeneity in AAO in the anchor samples has been noted, [102] which may reflect study specific definitions of AAO (e.g., age at first symptoms, first visit to general practitioner, or first diagnosis). Following Power et al., [102] we divided AAO into octiles within each cohort and combined the first three octiles into the early AAO group and the last three octiles into the late AAO group. Second, we tested for higher mean GRS for cases in anchor cohort samples with clinically severe MDD (endorsing 8 of 9 DSM MDD criteria) compared to those with "moderate"MDD (endorsing 5-7 of 9 MDD criteria) following Verduijn et al. [103] Sample sizes are given in *Table* F. Third, using iPSYCH as the target sample, we tested for higher mean GRS in recurrent MDD cases (ICD-10 F33, Ny5,574) compared to those with single episode MDD cases (ICD-10 F32, Ny12,968) in analyses that included ancestry principal components and genotyping batch as covariates. Finally, following Verduijn et al. [103] using the NESDA sample (PGC label "nes1"y an ongoing longitudinal study of depressive and anxiety disorders) as the target sample , we constructed clinical staging phenotypes in which cases were allocated to one of three stages: Stage 2 (n y 388) first episode MDDx stage 3 (n y 562) recurrent{relapse episode MDDx stage 4 (n y 705) persistent{unremitting chronic MDD, with an episode lasting longer than 2 years before baseline interview and{or 80% of the follow-up time with depressive symptoms. We tested for higher mean GRS in stage IV cases compared to stage II MDD cases.

Linkage disequilibrium (LD) score regression [14,94] was used to estimate $!^{"}_{\#\$\%}$ from GWA summary statistics. Estimates of $!^{"}_{\#\$\%}$ on the liability scale depend on the assumed lifetime prevalence of MDD in the population (( ), and we assumed (y0.15 but also evaluated (y0.10 to explore sensitivity (*Table* F*4*). LD score regression bivariate genetic correlations attributable to genome-wide SNPs (&) were estimated across MDD cohorts and between the full MDD cohort and other traits and disorders.

LD score regression was also used to partition $!^{"}_{\#\$\%}$ by genomic features. [61,94] We tested for enrichment of $!^{"}_{\#\$\%}$ based on genomic annotations partitioning $!^{"}_{\#\$\%}$ proportional to bp length represented by each annotation. We used the "baseline model"which consists of 53 functional categories. The categories are fully described elsewhere, [61] and included conserved regions [62], USCC gene models (exons, introns,

promoters, UTRs), and functional genomic annotations constructed using data from ENCODE [104] and the Roadmap Epigenomics Consortium. [105] We complemented these annotations by adding introgressed regions from the Neanderthal genome in European populations [106] and open chromatin regions from the brain dorsolateral prefrontal cortex. The open chromatin regions were obtained from an ATAC-seq experiment performed in 288 samples (Ny135 controls, Ny137 schizophrenia, Ny10 bipolar, and Ny6 affective disorder). [107] Peaks called with MACS [108] (1% FDR) were retained if their coordinates overlapped in at least two samples. The peaks were re-centered and set to a fixed width of 300bp using the diffbind R package. [109] To prevent upward bias in heritability enrichment estimation, we added two categories created by expanding both the Neanderthal introgressed regions and open chromatin regions by 250bp on each side.

We used LD score regression to estimate & between MDD and a range of other disorders, diseases, and human traits. [14] The intent of these comparisons was to evaluate the extent of shared common variant genetic architectures in order to suggest hypotheses about the fundamental genetic basis of MDD (given its extensive comorbidity with psychiatric and medical conditions and its association with anthropometric and other risk factors). Subject overlap of itself does not bias &. [14] These & are mostly based on studies of independent subjects and the estimates should be unbiased by confounding of genetic and non-genetic effects (except if there is genotype by environment correlation). When GWA studies include overlapping samples, & remains unbiased but the intercept of the LDSC regression is an estimate of the correlation between association statistics attributable to sample overlap. These calculations were done using the internal PGC GWA library and with LD-Hub (URLs). [75]

Relation of MDD GWA findings to tissue and cellular gene expression. We used partitioned LD score regression to evaluate which somatic tissues were enriched for MDD heritability. [110] Gene expression data generated using mRNA-seq from multiple human tissues were obtained from GTEx v6p (URLs). Genes for which z4 samples had at least one read count per million were discarded, and samples with z100 genes with at least one read count per million were excluded. The data were normalized, and a t-statistic was obtained for each tissue by comparing the expression in each tissue with the expression of all other tissues with the exception of tissues related to the tissue of interest (e.g., brain cortex vs all other tissues excluding other brain samples), using sex and age as covariates. A t-statistic was also obtained for each tissue among its related tissue (ex: cortex vs all other brain tissues) to test which brain region was the most associated with MDD, also using sex and age as covariates. The top 10% of the genes with the most extreme t-statistic were defined as tissue specific. The coordinates for these genes were extended by a 100kb window and tested using LD score regression. Significance was obtained from the coefficient z-score, which corrects for all other categories in the baseline model.

Lists of genes specifically expressed in neurons, astrocytes, and oligodendrocytes were obtained from Cahoy et al. [60] As these experiment were done in mice, genes were mapped to human orthologous genes using ENSEMBL. The coordinates for these genes were extended by a 100kb window and tested using LD score regression as for the GTEx tissue specific genes.

We conducted eQTL look-ups of the most associated SNPs in each region and report (Nable RM) GWA SNPs in LD (r^2 0.8) with the top eQTLs in the following data sets: eQTLGen Consortium (lllumina arrays in whole blood Ny14,115, in preparation), BIOS (RNA-seq in whole blood (Ny2,116), [111] NESDA{NTR (Affymetrix arrays in whole blood, Ny4,896), [112] GEUVADIS (RNA-seq in LCL (Ny465), [113] Rosmap (RNA seq in cortex, Ny 494, submitted), GTEx (RNA-seq in 44 tissues, N 70), [58] and Common Mind Consortium (CMC, prefrontal cortex, Sage Synapse accession syn5650509, Ny467). [66]

We used summary-data-based Mendelian randomization (SMR) [64] to identify loci with strong evidence of causality via gene expression (Nable RZ). SMR analysis is limited to significant cis SNP-expression (FDR

z 0.05) and SNPs with MAF 0.01 at a Bonferroni-corrected pSMR. Due to LD, multiple SNPs may be associated with the expression of a gene, and some SNPs are associated with the expression of more than one gene. Since the aim of SMR is to prioritize variants and genes for subsequent studies, a test for heterogeneity excludes regions that may harbor multiple causal loci (pHET z 0.05). SMR analyses were conducted using eQTLGen Consortium, GTEx (11 brain tissues), and CMC data.

We conducted a transcriptome wide association study [65] using pre-computed expression reference weights for CMC data (5,420 genes with significant cis-SNP heritability) provided with the TWAS{FUSION software. The significance threshold was 0.05{5420.

DNA looping using Hi-C. Dorsolateral prefrontal cortex (Brodmann area 9) was dissected from postmortem samples from three adults of European ancestry (Dr Craig Stockmeier, University of Mississippi Medical Center). Cerebrum from three fetal brains were obtained from the NIH NeuroBiobank (URLsxgestation age 17-19 weeks, African ancestry). Samples were dry homogenized to a fine powder using a liquid nitrogen-cooled mortar and pestle.

We used "easy Hi-Cw(in preparation) to assess DNA looping interactions. Pulverized tissue (~150 mg) was crosslinked with formaldehyde (1% final concentration) and the reaction quenched using glycine (150 mM). Samples were then lysed, Dounce homogenized, and digested using *HindIII*. This was followed by in situ ligation. Samples were cross-linked with proteinase K and purified using phenol-chloroform. DNA was then digested with *DpnII* followed by purification using PCRClean DX beads (Aline Biosciences). The DNA products were self-ligated overnight at 16 using T4 DNA ligase. Self-ligated DNA waw purified with phenol-chloroform, digested with lambda exonuclease, and purified using PCRClean DX beads. For DNA circle re-linearization, bead-bound DNA was eluted and digested with *HindIII* and purified using PCRClean. Bead-bound DNA was eluted in 50ul nuclease free water.

Re-linearized DNA (~50ng) was used for library generation (Illumina TruSeq protocol). Briefly, the DNA was end-repaired using End-it kit (Epicentre), A tailed with Klenow fragment (3}~5} exo~x NEB), and purified with PCRClean DX beads. The 4ul DNA product was mixed with 5ul of 2X quick ligase buffer, 1ul of 1:10 diluted annealed adapter and 0.5ul of Quick DNA T4 ligase (NEB). The ligation was done by incubating at room temperature for 15 minutes. DNA was purified using DX beads. Elution was done in 14ul nuclease free water. To deep-sequence easy Hi-C libraries, we used custom TruSeq adapter in which the index is replaced by 6 base random sequence. Libraries were then PCR amplified and deeply sequenced (4-5 lanes per sample, around 1 billion reads per sample) using Illumina HiSeq4000 (2x50bp).

Because nearly all mappable reads start with the *HindIII* sequence AGCTT, we trimmed the first 5 bases from every read and added the 6-base sequence AAGCTT to the 5| of all reads. These read were then aligned to the human reference genome (hg19) using Bowtie. After mapping, we kept reads where both ends were exactly at *HindIII* cutting sites. PCR duplicates were removed. Of these *HindIII* pairs, we split reads into three classes based on their strand orientations ("same-strandw, "inwardw or "outwardw). For cis-reads the only type of invalid cis-pairs are self-circles with two ends within the same *HindIII* fragment facing each other. We computed the total number of real cis-contact as twice the number of valid "same-strandwpairs. Reads from undigested *HindIII* sites are back-to-back read pairs next to the same *HindIII* sites facing away from each other.

Gene-wise and pathway analysis. Our approach was guided by rigorous method comparisons conducted by PGC members. [70,114] *P*-values quantifying the degree of association of genes and gene sets with MDD were generated using MAGMA (v1.06). [115] MAGMA uses Brown|s method to combine SNP p-values and account for LD. We used ENSEMBL gene models for 19,079 genes giving a Bonferroni corrected *P*-value threshold of 2.6×10^{-6}. Gene set *P*-values were obtained using a competitive analysis that tests whether genes in a gene set are more strongly associated with the phenotype than other gene sets. We used

European-ancestry subjects from 1,000 Genomes Project (Phase 3 v5a, MAF 0.01) [101] for the LD reference. The gene window used was 35 kb upstream and 10 kb downstream to include regulatory elements.

Gene sets were from two main sources. First, we included gene sets previously shown to be important for psychiatric disorders (71 gene setsXe.g., FMRP binding partners, *de novo* mutations, GWAS top SNPs, ion channels). [72,116,117] Second, we included gene sets from MSigDB (v5.2) [118] which includes canonical pathways and Gene Ontology gene sets. Canonical pathways were curated from BioCarta, KEGG, Matrisome, Pathway Interaction Database, Reactome, SigmaAldrich, Signaling Gateway, Signal Transduction KE, and SuperArray. Pathways containing between 10-10K genes were included.

To evaluate gene sets related to antidepressants, gene-sets were extracted from the Drug-Gene Interaction database (DGIdb v.2.0) [119] and the Psychoactive Drug Screening Program Ki DB [120] downloaded in June 2016. The association of 3,885 drug gene-sets with MDD was estimated using MAGMA (v1.6). The drug gene-sets were ordered by p-value, and the Wilcoxon-Mann-Whitney test was used to assess whether the 42 antidepressant gene-sets in the dataset (ATC code N06A in the Anatomical Therapeutic Chemical Classification System) had a higher ranking than expected by chance.

One issue is that some gene sets contain overlapping genes, and these may reflect largely overlapping results. The pathway map was constructed using the kernel generative topographic mapping algorithm (k-GTM) as described by Olier et al. GTM is a probabilistic alternative to Kohonen maps: the kernel variant is used when the input is a similarity matrix. The GTM and k-GTM algorithms are implemented in GTMapTool (URLs). We used the Jaccard similarity matrix of FDR-significant pathways as input for the algorithm, where each pathway is encoded by a vector of binary values representing the presence (1) or absence (0) of a gene. Parameters for the k-GTM algorithm are the square root of the number of grid points (k), the square root of the number of RBF functions (m), the regularization coefficient (l), the RBF width factor (w), and the number of feature space dimensions for the kernel algorithm (b). We set kysquare root of the number of pathways, mysquare root of k, ly1 (default), wy1 (default), and bythe number of principal components explaining 99.5% of the variance in the kernel matrix. The output of the program is a set of coordinates representing the average positions of pathways on a 2D map. The x and y axes represent the dimensions of a 2D latent space. The pathway coordinates and corresponding MAGMA *P*-values were used to build the pathway activity landscape using the kriging interpolation algorithm implemented in the R gstat package.

Mendelian randomization (MR). [121] We used MR to investigate the relationships between MDD and correlated traits. Epidemiological studies show that MDD is associated with environmental and life event risk factors as well as multiple diseases, yet it remains unclear whether such trait outcomes are causes or consequences of MDD (or prodromal MDD). Genetic variants are present from birth, and hence are far less likely to be confounded with environmental factors than in epidemiological studies.

We conducted bi-directional MR analysis for four traits: years of education (EDY) [76], body mass index (BMI) [27], coronary artery disease (CAD) [77], and schizophrenia (SCZ) [22]. Briefly, we denote z as a genetic variant (i.e., a SNP) that is significantly associated with x, an exposure or putative causal trait for y (the disease{trait outcome). The effect size of x on y can be estimated using a two-step least squares (2SLS) [122] approach: $\hat{*}_{+,} - \hat{*}_{.,} / \hat{*}_{.+}$., where $\hat{*}_{.+}$ is the estimated effect size for the SNP-trait association the exposure trait0and $\hat{*}_{.,}$ is the effect size estimated for the same SNP in the GWAS of the outcome trait.

Since SNP-trait effect sizes are typically small, power is increased by using multiple associated SNPs which allows simultaneous investigation of pleiotropy driving the epidemiologically observed trait associations. Causality of the exposure trait for the outcome trait implies a consistent relationship between the SNP association effect sizes of the exposure associated SNPs in the outcome trait.

We used generalized summary statistics-based MR (GSMR) (Zhu et al., submitted) to estimate $\hat{\beta}_{+,}$ and its standard error from multiple SNPs associated with the exposure trait at a genome-wide significance level. We conducted bi-directional GSMR analyses for each pair of traits, and report results after excluding SNPs that fail the HEIDI-outlier heterogeneity test (which is more conservative than excluding SNPs that have an outlying association likely driven by locus-specific pleiotropy). GSMR is more powerful than inverse-weighted MR (IVW-MR) and MR-Egger because it takes account of the sampling variation of both $\hat{\beta}_{.+}$ and $\hat{\beta}_{.,}$. GSMR also accounts for residual LD between the clumped SNPs. For comparison, we also conducted IVW-MR and MR-Egger analyses. [123]

Trans-ancestry. Common genetic risk variants for complex biomedical conditions are likely to be shared across ancestries. [124,125] However, lower & have been reported likely reflecting different LD patterns by ancestry. For example, European-Chinese & estimates were below one for ADHD (0.39, SE 0.15), [126] rheumatoid arthritis (0.46, SE 0.06), [127] and type 2 diabetes (0.62, SE 0.09), [127] and reflect population differences in LD and population-specific causal variants.

The Han Chinese CONVERGE study [17] included clinically ascertained females with severe, recurrent MDD, and is the largest non-European MDD GWA to date. Neither of the two genome-wide significant loci in CONVERGE had SNP findings 250 kb with P z $1x10^{-6}$ in the full European results. We used LDSC with an ancestry-specific LD reference for within ancestry estimation, and POPCORN [127] for trans-ancestry estimation. In the CONVERGE sample, $!^{"}_{\#\$\%}$ was reported as 20-29%. [128] Its & with the seven European MDD cohorts was 0.33 (SE 0.03). [129] For comparison, & for CONVERGE with European results for schizophrenia was 0.34 (SE 0.05) and 0.45 (SE 0.07) for bipolar disorder. The weighted mean & between the CONVERGE cohort with the seven anchor and expanded cohorts using was 0.31 (SE 0.03). These & estimates should be interpreted in light of the estimates of & within European MDD cohorts which are variable (*Table R4*).

Genome build. All genomic coordinates are given in NCBI Build 37{UCSC hg19.

Availability of results. The PGC|s policy is to make genome-wide summary results public. Summary statistics for a combined meta-analysis of the anchor cohort samples with five of the six expanded samples (deCODE, Generation Scotland, GERA, iPSYCH, and UK Biobank) are available on the PGC web site (URLs). Results for 10,000 SNPs for all seven cohorts are also available on the PGC web site.

GWA summary statistics for the sixth expanded cohort (23andMe, Inc.) must be obtained separately. Summary statistics for the 23andMe dataset can be obtained by qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please contact David Hinds (dhinds@23andme.com) for more information and to apply to access the data. Researchers who have the 23andMe summary statistics can readily recreate our results by meta-analyzing the six cohort results file with the Hyde et al. results file from 23andMe. [19]

Availability of genotype data for the anchor cohorts is described in *Table R14*. For the expanded cohorts, interested users should contact the lead PIs of these cohorts (which are separate from the PGC).

_` as

1000 Genomes Project multi-ancestry imputation panel, https:{{mathgen.stats.ox.ac.uk{impute{data_download_1000G_phase1_integrated.html

23andMe privacy policy https:{{www.23andme.com{en-eu{about{privacy

Bedtools, https:{{bedtools.readthedocs.io

Genotype-based checksums for relatedness determination, http:{{www.broadinstitute.org{~sripke{share_links{checksums_download

GTEx, http:{{www.gtexportal.org{home{datasets

GTMapTool, http:{{infochim.u-strasbg.fr{mobyle-cgi{portal.py#forms::gtmaptool

LD-Hub, http:{{ldsc.broadinstitute.org

MDD summary results are available on the PGC website, https:{{pgc.unc.edu

NIH NeuroBiobank, https:{{neurobiobank.nih.gov

PGC "ricopiliwGWA pipeline, https:{{github.com{Nealelab{ricopili

UK Biobank, http:{{www.ukbiobank.ac.uk

Author Affiliations

1, Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD, AU
2, Queensland Brain Institute, The University of Queensland, Brisbane, QLD, AU
3, Medical and Population Genetics, Broad Institute, Cambridge, MA, US
4, Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, US
5, Department of Psychiatry and Psychotherapy, Universit tsmedizin Berlin Campus Charité Mitte, Berlin, DE
6, Department of Biomedicine, Aarhus University, Aarhus, DK
7, iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, DK
8, iPSYCH, The Lundbeck Foundation Initiative for Integrative Psychiatric Research,, DK
9, Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, SE
10, Dept of Biological Psychology EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, NL
11, Division of Psychiatry, University of Edinburgh, Edinburgh, GB
12, Centre for Integrated Register-based Research, Aarhus University, Aarhus, DK
13, National Centre for Register-Based Research, Aarhus University, Aarhus, DK
14, Discipline of Psychiatry, University of Adelaide, Adelaide, SA, AU
15, Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, DE
16, Munich Cluster for Systems Neurology (SyNergy), Munich, DE
17, Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, US
18, Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, DK
19, Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, NL
20, Virginia Institute for Psychiatric and Behavior Genetics, Richmond, VA, US
21, Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, US
22, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, SE
23, Department of Clinical Medicine, Translational Neuropsychiatry Unit, Aarhus University, Aarhus, DK
24, Statistical genomics and systems genetics, European Bioinformatics Institute (EMBL-EBI), Cambridge, GB
25, Human Genetics, Wellcome Trust Sanger Institute, Cambridge, GB
26, Department of Psychiatry, University Hospital of Lausanne, Prilly, Vaud, CH
27, MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, GB
28, Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, QLD, AU
29, Centre for Advanced Imaging, The University of Queensland, Saint Lucia, QLD, AU
30, Queensland Brain Institute, The University of Queensland, Saint Lucia, QLD, AU
31, Psychological Medicine, Cardiff University, Cardiff, GB
32, Center for Genomic and Computational Biology, Duke University, Durham, NC, US
33, Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC, US
34, Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, US
35, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA
36, Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, GB
37, Institute of Human Genetics, University of Bonn, Bonn, DE
38, Life Brain Center, Department of Genomics, University of Bonn, Bonn, DE
39, Psychiatry, Dokuz Eylul University School Of Medicine, Izmir, TR
40, Epidemiology, Erasmus MC, Rotterdam, Zuid-Holland, NL
41, Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, US
42, Department of Psychiatry, Massachusetts General Hospital, Boston, MA, US
43, Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, US
44, Research, 23andMe, Inc., Mountain View, CA, US
45, Neuroscience and Mental Health, Cardiff University, Cardiff, GB

46, Bioinformatics, University of British Columbia, Vancouver, BC, CA
47, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, US
48, Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, US
49, Department of Psychiatry (UPK), University of Basel, Basel, CH
50, Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, CH
51, Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Baden-Württemberg, DE
52, Department of Psychiatry, Trinity College Dublin, Dublin, IE
53, Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, US
54, Psychiatry Behavioral Sciences, Johns Hopkins University, Baltimore, MD, US
55, Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, QLD, AU
56, Bioinformatics Research Centre, Aarhus University, Aarhus, DK
57, Institute of Genetic Medicine, Newcastle University, Newcastle upon Tyne, GB
58, Danish Headache Centre, Department of Neurology, Rigshospitalet, Glostrup, DK
59, Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, DK
60, iPSYCH, The Lundbeck Foundation Initiative for Psychiatric Research, Copenhagen, DK
61, Brain and Mind Centre, University of Sydney, Sydney, NSW, AU
62, Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine and Ernst Moritz Arndt University Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
63, Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, CH
64, Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, US
65, Statistics, Pfizer Global Research and Development, Groton, CT, US
66, Max Planck Institute of Psychiatry, Munich, DE
67, Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, US
68, Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, US
69, Division of Research, Kaiser Permanente Northern California, Oakland, CA, US
70, Psychiatry The Behavioral Sciences, University of Southern California, Los Angeles, CA, US
71, Informatics Program, Boston Children's Hospital, Boston, MA, US
72, Department of Medicine, Brigham and Women's Hospital, Boston, MA, US
73, Department of Biomedical Informatics, Harvard Medical School, Boston, MA, US
74, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, GB
75, Department of Endocrinology at Herlev University Hospital, University of Copenhagen, Copenhagen, DK
76, Swiss Institute of Bioinformatics, Lausanne, VD, CH
77, Institute of Social and Preventive Medicine (IUMSP), University Hospital of Lausanne, Lausanne, VD, CH
78, Dept of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA
79, Mental Health, NHS 24, Glasgow, GB
80, Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, GB
81, Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, DE
82, Statistics, University of Oxford, Oxford, GB
83, Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, US
84, School of Psychology and Counseling, Queensland University of Technology, Brisbane, QLD, AU
85, Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, South Brisbane, QLD, AU
86, Child Health Research Centre, University of Queensland, Brisbane, QLD, AU
87, Estonian Genome Center, University of Tartu, Tartu, EE
88, Medical Genetics, University of British Columbia, Vancouver, BC, CA
89, Statistics, University of British Columbia, Vancouver, BC, CA
90, DZHK (German Centre for Cardiovascular Research), Partner Site Greifswald, University Medicine, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
91, Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
92, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD, AU
93, Humus, Reykjavik, IS
94, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, GB
95, Virginia Institute for Psychiatric Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, US
96, Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, NL
97, Clinical Genetics, Vrije Universiteit Medical Center, Amsterdam, NL
98, Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, US
99, Solid Biosciences, Boston, MA, US
100, Department of Psychiatry, Washington University in Saint Louis School of Medicine, Saint Louis, MO, US
101, Department of Biochemistry and Molecular Biology II, Institute of Neurosciences, Center for Biomedical Research, University of Granada, Granada, ES
102, Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, NL
103, Department of Psychiatry and Psychotherapy, Medical Center of the University of Munich, Campus Innenstadt, Munich, DE
104, Institute of Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Campus Innenstadt, Munich, DE
105, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, US
106, Behavioral Health Services, Kaiser Permanente Washington, Seattle, WA, US

107, Faculty of Medicine, Department of Psychiatry, University of Iceland, Reykjavik, IS
108, School of Medicine and Dentistry, James Cook University, Townsville, QLD, AU
109, Institute of Health and Wellbeing, University of Glasgow, Glasgow, GB
110, deCODE Genetics { Amgen, Reykjavik, IS
111, Psychiatry Human Behavior, University of Mississippi Medical Center, Jackson, MS, US
112, College of Biomedical and Life Sciences, Cardiff University, Cardiff, GB
113, Institute of Epidemiology and Social Medicine, University of Münster, Münster, Nordrhein-Westfalen, DE
114, Institute for Community Medicine, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
115, KG Jebsen Centre for Psychosis Research, Norway Division of Mental Health and Addiction, Oslo University Hospital, Oslo, NO
116, Department of Psychiatry, University of California, San Diego, San Diego, CA, US
117, Medical Genetics Section, CGEM, IGMM, University of Edinburgh, Edinburgh, GB
118, Clinical Neurosciences, University of Cambridge, Cambridge, GB
119, Internal Medicine, Erasmus MC, Rotterdam, Zuid-Holland, NL
120, Roche Pharmaceutical Research and Early Development, Neuroscience, Ophthalmology and Rare Diseases Discovery Translational Medicine Area, Roche Innovation Center Basel, F. Hoffmann-La Roche Ltd, Basel, CH
121, Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Mecklenburg-Vorpommern, DE
122, Department of Psychiatry, Leiden University Medical Center, Leiden, NL
123, Virginia Institute of Psychiatric Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, US
124, Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Cambridge, MA, US
125, Institute for Molecular BioscicencexQueensland Brain Institute, The University of Queensland, Brisbane, QLD, AU
126, Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, NL
127, Department of Psychiatry, University of Münster, Münster, Nordrhein-Westfalen, DE
128, Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, DE
129, Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Basel, CH
130, Amsterdam Public Health Institute, Vrije Universiteit Medical Center, Amsterdam, NL
131, Centre for Integrative Biology, Universit degli Studi di Trento, Trento, Trentino-Alto Adige, IT
132, Department of Psychiatry and Psychotherapy, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Rheinland-Pfalz, DE
133, Psychiatry, Kaiser Permanente Northern California, San Francisco, CA, US
134, Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, GB
135, Centre for Addiction and Mental Health, Toronto, ON, CA
136, Department of Psychiatry, University of Toronto, Toronto, ON, CA
137, Division of Psychiatry, University College London, London, GB
138, Neuroscience Therapeutic Area, Janssen Research and Development, LLC, Titusville, NJ, US
139, Institute of Molecular and Cell Biology, University of Tartu, Tartu, EE
140, Psychosis Research Unit, Aarhus University Hospital, Risskov, Aarhus, DK
141, University of Liverpool, Liverpool, GB
142, Mental Health Center Copenhagen, Copenhagen Universtity Hospital, Copenhagen, DK
143, Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Groton, CT, US
144, Psychiatry, Harvard Medical School, Boston, MA, US
145, Psychiatry, University of Iowa, Iowa City, IA, US
146, Department of Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, US
147, Human Genetics Branch, NIMH Division of Intramural Research Programs, Bethesda, MD, US
148, Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Goettingen, Niedersachsen, DE
149, Faculty of Medicine, University of Iceland, Reykjavik, IS
150, Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, Zuid-Holland, NL
151, Psychiatry, Erasmus MC, Rotterdam, Zuid-Holland, NL
152, Psychiatry, Dalhousie University, Halifax, NS, CA
153, Division of Epidemiology, New York State Psychiatric Institute, New York, NY, US
154, Department of Clinical Medicine, University of Copenhagen, Copenhagen, DK
155, Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Cambridge, MA, US
156, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, US
157, Department of Medical Molecular Genetics, King's College London, London, GB
158, Psychiatry Behavioral Sciences, Stanford University, Stanford, CA, US
159, NIHR BRC for Mental Health, King's College London, London, GB
160, Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, US

Author Contributions

Writing group: G. Breen, A. D. Børglum, D. F. Levinson, C. M. Lewis, S. Ripke, P. F. Sullivan, N. R. Wray.

PGC MDD PI group: V. Arolt, B. T. Baune, K. Berger, D. I. Boomsma, G. Breen, A. D. Børglum, S. Cichon, U. Dannlowski, J. R. DePaulo, E. Domenici, K. Domschke, T. Esko, E. d. Geus, H. J. Grabe, S. P. Hamilton, C. Hayward, A. C. Heath, D. M. Hougaard, K. S. Kendler, S. Kloiber, D. F. Levinson, C. M. Lewis, G. Lewis, Q. S. Li, S. Lucae, P. A. Madden, P. K. Magnusson, N. G. Martin, A. M. McIntosh, A. Metspalu, O. Mors, P. B.

Mortensen, B. Müller-Myhsok, M. Nordentoft, M. M. Nöthen, M. C. O'Donovan, S. A. Paciga, N. L. Pedersen, B. W. Penninx, R. H. Perlis, D. J. Porteous, J. B. Potash, M. Preisig, M. Rietschel, C. Schaefer, T. G. Schulze, J. W. Smoller, K. Stefansson, P. F. Sullivan, H. Tiemeier, R. Uher, H. Völzke, M. M. Weissman, T. Werge, A. R. Winslow, N. R. Wray.

Bioinformatics: 23andMe Research Team, M. J. Adams, S. V. d. Auwera, G. Breen, J. Bryois, A. D. Børglum, E. Castelao, J. H. Christensen, T. Clarke, J. R. I. Coleman, L. Colodro-Conde, eQTLGen Consortium, G. E. Crawford, C. A. Crowley, G. Davies, E. M. Derks, T. Esko, A. J. Forstner, H. A. Gaspar, P. Giusti-Rodríguez, J. Grove, L. S. Hall, T. F. Hansen, C. Hayward, M. Hu, R. Jansen, F. Jin, Z. Kutalik, Q. S. Li, Y. Li, P. A. Lind, X. Liu, L. Lu, D. J. MacIntyre, S. E. Medland, E. Mihailov, Y. Milaneschi, J. N. Painter, B. W. Penninx, W. J. Peyrot, G. Pistis, P. Qvist, L. Shen, S. I. Shyn, C. A. Stockmeier, P. F. Sullivan, K. E. Tansey, A. Teumer, P. A. Thomson, A. G. Uitterlinden, Y. Wang, S. M. Weinsheimer, N. R. Wray, H. S. Xi.

Clinical: E. Agerbo, T. M. Air, V. Arolt, B. T. Baune, A. T. F. Beekman, K. Berger, E. B. Binder, D. H. R. Blackwood, H. N. Buttenschøn, A. D. Børglum, N. Craddock, U. Dannlowski, J. R. DePaulo, N. Direk, K. Domschke, M. Gill, F. S. Goes, H. J. Grabe, A. C. Heath, A. M. v. Hemert, I. B. Hickie, M. Ising, S. Kloiber, J. Krogh, D. F. Levinson, S. Lucae, D. J. MacIntyre, D. F. MacKinnon, P. A. Madden, W. Maier, N. G. Martin, P. McGrath, P. McGuffin, A. M. McIntosh, A. Metspalu, C. M. Middeldorp, S. S. Mirza, F. M. Mondimore, O. Mors, P. B. Mortensen, D. R. Nyholt, H. Oskarsson, M. J. Owen, C. B. Pedersen, M. G. Pedersen, J. B. Potash, J. A. Quiroz, J. P. Rice, M. Rietschel, C. Schaefer, R. Schoevers, E. Sigurdsson, G. C. B. Sinnamon, D. J. Smith, F. Streit, J. Strohmaier, D. Umbricht, M. M. Weissman, J. Wellmann, T. Werge, G. Willemsen.

Genomic assays: G. Breen, H. N. Buttenschøn, J. Bybjerg-Grauholm, M. Bækvad-Hansen, A. D. Børglum, S. Cichon, T. Clarke, F. Degenhardt, A. J. Forstner, S. P. Hamilton, C. S. Hansen, A. C. Heath, P. Hoffmann, G. Homuth, C. Horn, J. A. Knowles, P. A. Madden, L. Milani, G. W. Montgomery, M. Nauck, M. M. Nöthen, M. Rietschel, M. Rivera, E. C. Schulte, T. G. Schulze, S. I. Shyn, H. Stefansson, F. Streit, T. E. Thorgeirsson, J. Treutlein, A. G. Uitterlinden, S. H. Witt, N. R. Wray.

Obtained funding for primary MDD samples: B. T. Baune, K. Berger, D. H. R. Blackwood, D. I. Boomsma, G. Breen, H. N. Buttenschøn, A. D. Børglum, S. Cichon, J. R. DePaulo, I. J. Deary, E. Domenici, T. C. Eley, T. Esko, H. J. Grabe, S. P. Hamilton, A. C. Heath, D. M. Hougaard, I. S. Kohane, D. F. Levinson, C. M. Lewis, G. Lewis, Q. S. Li, S. Lucae, P. A. Madden, W. Maier, N. G. Martin, P. McGuffin, A. M. McIntosh, A. Metspalu, G. W. Montgomery, O. Mors, P. B. Mortensen, M. Nordentoft, D. R. Nyholt, M. M. Nöthen, P. F. O'Reilly, B. W. Penninx, D. J. Porteous, J. B. Potash, M. Preisig, M. Rietschel, C. Schaefer, T. G. Schulze, G. C. B. Sinnamon, J. H. Smit, D. J. Smith, H. Stefansson, K. Stefansson, P. F. Sullivan, T. E. Thorgeirsson, H. Tiemeier, A. G. Uitterlinden, H. Völzke, M. M. Weissman, T. Werge, N. R. Wray.

Statistical analysis: 23andMe Research Team, A. Abdellaoui, M. J. Adams, T. F. M. Andlauer, S. V. d. Auwera, S. Bacanu, K. Berger, T. B. Bigdeli, G. Breen, E. M. Byrne, A. D. Børglum, N. Cai, T. Clarke, J. R. I. Coleman, B. Couvy-Duchesne, H. S. Dashti, G. Davies, N. Direk, C. V. Dolan, E. C. Dunn, N. Eriksson, V. Escott-Price, T. Esko, H. K. Finucane, J. Frank, H. A. Gaspar, S. D. Gordon, J. Grove, L. S. Hall, C. Hayward, A. C. Heath, S. Herms, D. A. Hinds, J. Hottenga, C. L. Hyde, M. Ising, E. Jorgenson, F. F. H. Kiadeh, J. Kraft, W. W. Kretzschmar, Z. Kutalik, J. M. Lane, C. M. Lewis, Q. S. Li, Y. Li, D. J. MacIntyre, P. A. Madden, R. M. Maier, J. Marchini, M. Mattheisen, H. Mbarek, A. M. McIntosh, S. E. Medland, D. Mehta, E. Mihailov, Y. Milaneschi, S. S. Mirza, S. Mostafavi, N. Mullins, B. Müller-Myhsok, B. Ng, M. G. Nivard, D. R. Nyholt, P. F. O'Reilly, R. E. Peterson, E. Pettersson, W. J. Peyrot, G. Pistis, D. Posthuma, S. M. Purcell, B. P. Riley, S. Ripke, M. Rivera, R. Saxena, C. Schaefer, L. Shen, J. Shi, S. I. Shyn, H. Stefansson, S. Steinberg, P. F. Sullivan, K. E. Tansey, H. Teismann, A. Teumer, W. Thompson, P. A. Thomson, T. E. Thorgeirsson, C. Tian, M. Traylor, V. Trubetskoy, M. Trzaskowski, A. Viktorin, P. M. Visscher, Y. Wang, B. T. Webb, J. Wellmann, T. Werge, N. R. Wray, Y. Wu, J. Yang, F. Zhang.

Competing Financial Interests

Aartjan TF Beekman: Speakers bureaus of Lundbeck and GlaxoSmithKline. Greg Crawford: Co-founder of Element Genomics. Enrico Domenici: Employee of Hoffmann-La Roche at the time this study was conducted, consultant to Roche and Pierre-Fabre. Nicholas Eriksson: Employed by 23andMe, Inc. and owns stock in 23andMe, Inc. David Hinds: Employee of and own stock options in 23andMe, Inc. Sara Paciga: Employee of Pfizer, Inc. Craig L Hyde: Employee of Pfizer, Inc. Ashley R Winslow: Former employee and stockholder of Pfizer, Inc. Jorge A Quiroz: Employee of Hoffmann-La Roche at the time this study was conducted. Hreinn Stefansson: Employee of deCODE Genetics{AMGEN. Kari Stefansson: Employee of deCODE Genetics{AMGEN. Stacy Steinberg: Employee of deCODE Genetics{AMGEN. Patrick F Sullivan: Scientific advisory board for Pfizer Inc and an advisory committee for Lundbeck. Thorgeir E Thorgeirsson: Employee of deCODE Genetics{AMGEN. Chao Tian: Employee of and own stock options in 23andMe, Inc.

Acknowledgements

PGC: We are deeply indebted to the investigators who comprise the PGC, and to the hundreds of thousands of subjects who have shared their life experiences with PGC investigators. Statistical analyses were carried out on the NL Genetic Cluster Computer (http:{{www.geneticcluster.org) hosted by SURFsara.

EDINBURGH: Genotyping was conducted at the Genetics Core Laboratory at the Clinical Research Facility (University of Edinburgh). GenScot: We are grateful to all the families who took part, the general practitioners and the Scottish School of Primary Care for their help in recruiting them, and the whole Generation Scotland team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists, healthcare assistants and nurses. Genotyping was conducted at the Genetics Core Laboratory at the Clinical Research Facility (University of Edinburgh). GSK MUNICH: We thank all participants in the GSK-Munich study. We thank numerous people at GSK and Max-Planck Institute, BKH Augsburg and Klinikum Ingolstadt in Germany who contributed to this project. JANSSEN: Funded by Janssen Research Development, LLC. We are grateful to the study volunteers for participating in the research studies and to the clinicians and support staff for enabling patient recruitment and blood sample collection. We thank the staff in the former Neuroscience Biomarkers of Janssen Research Development for laboratory and operational support (e.g., biobanking, processing, plating, and sample de-identification), and to the staff at Illumina for genotyping Janssen DNA samples. MARS: This work was funded by the Max Planck Society, by the Max Planck Excellence Foundation, and by a grant from the German Federal Ministry for Education and Research (BMBF) in the National Genome Research Network framework (NGFN2 and NGFN-Plus, FKZ 01GS0481), and by the BMBF Program FKZ 01ES0811. We acknowledge all study participants. We thank numerous people at Max-Planck Institute, and all study sites in Germany and Switzerland who contributed to this project. Controls were from the Dortmund Health Study which was supported by the German Migraine Headache Society, and by unrestricted grants to the University of Münster from Almirall, Astra Zeneca, Berlin Chemie, Boehringer, Boots Health Care, Glaxo-Smith-Kline, Janssen Cilag, McNeil Pharma, MSD Sharp Dohme, and Pfizer. Blood collection was funded by the Institute of Epidemiology and Social Medicine, University of Münster. Genotyping was supported by the German Ministry of Research and Education (BMBF grant 01ER0816). PsyColaus: PsyCoLaus{CoLaus received additional support from research grants from GlaxoSmithKline and the Faculty of Biology and Medicine of Lausanne. QIMR: We thank the twins and their families for their willing participation in our studies. RADIANT: This report represents independent research funded by the National Institute for Health Research (NIHR) Biomedical Research Centre at South London and Maudsley NHS Foundation Trust, and King|s College London. The views expressed are those of the authors and not necessarily those of the

NHS, the NIHR, or the Department of Health. Rotterdam Study: The Rotterdam Study is also funded by Erasmus Medical Center and Erasmus University. SHIP-LEGEND{TREND: SHIP is part of the Community Medicine Research net of the University of Greifswald which is funded by the Federal Ministry of Education and Research (grants 01ZZ9603, 01ZZ0103, and 01ZZ0403), the Ministry of Cultural Affairs, and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genotyping in SHIP was funded by Siemens Healthineers and the Federal State of Mecklenburg-West Pomerania. Genotyping in SHIP-TREND-0 was supported by the Federal Ministry of Education and Research (grant 03ZIK012). STAR*D: The authors appreciate the efforts of the STAR*D investigator team for acquiring, compiling, and sharing the STAR*D clinical data set. TwinGene: thanks the Karolinska Institutet for infrastructural support of the Swedish Twin Registry. 23andME: We thank the 23andMe research participants included in the analysis, all of whom provided informed consent and participated in the research online according to a human subjects protocol approved by an external AAHRPP-accredited institutional review board (Ehical Independent Review Services), and the employees of 23andMe for making this work possible. 23andMe acknowledges the invaluable contributions of Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, David A. Hinds, Bethann S. Hromatka, Karen E. Huber, Aaron Kleinman, Nadia K. Litterman, Matthew H. McIntyre, Joanna L. Mountain, Carrie A.M. Northover, Steven J. Pitts, J. Fah Sathirapongsasuti, Olga V. Sazonova, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, and Catherine H. Wilson. deCODE: The authors are thankful to the participants and staff at the Patient Recruitment Center. GERA: Participants in the Genetic Epidemiology Research on Adult Health and Aging Study are part of the Kaiser Permanente Research Program on Genes, Environment, and Health, supported by the Wayne and Gladys Valley Foundation, The Ellison Medical Foundation, the Robert Wood Johnson Foundation, and the Kaiser Permanente Regional and National Community Benefit Programs. iPSYCH: The iPSYCH (The Lundbeck Foundation Initiative for Integrative Psychiatric Research) team acknowledges funding from The Lundbeck Foundation (grant no R102-A9118 and R155-2014-1724), the Stanley Medical Research Institute, the European Research Council (project no: 294838), the Novo Nordisk Foundation for supporting the Danish National Biobank resource, and grants from Aarhus and Copenhagen Universities and University Hospitals, including support to the iSEQ Center, the GenomeDK HPC facility, and the CIRRAU Center. UK Bioband: this research has been conducted using the UK Biobank Resource (URLs), including applications #4844 and #6818. Finally, we thank the members of the eQTLGen Consortium for allowing us to use their very large eQTL database ahead of publication. Its members are listed in N*able* R1Q

V*inding sources*

The table below lists the funding that supported the primary studies analyzed in the paper.

"#$%&	()*%+, -).#/*#01'	23*1%, $4 5) 1'	6$, %) 1'	70$, #1&
PGC	PF Sullivan	U01 MH109528	NIMH	USA
PGC	A Agrawal	U01 MH109532	NIDA	USA
PGC	D Posthuma	480-05-003	Netherlands Scientific Organization	Netherlands
PGC	D Posthuma	~	Dutch Brain Foundation and the VU University Amsterdam	Netherlands
BiDirect	K Berger	01ER0816, 01ER1506	BMBF	Germany
BoMa	M Rietschel	RI 908{11-1	Deutsche Forschungsgemeinschaft	Germany
BoMa	MM Nöthen	NO246{10-1	Deutsche Forschungsgemeinschaft	Germany
BoMa	MM Nöthen	Excellence Cluster ImmunoSensation	Deutsche Forschungsgemeinschaft	Germany
BoMa	MM Nöthen, M	01ZX1314A{01ZX1614A	BMBF Integrament	Germany

	Rietschel, S Cichon			
BoMa	MM Nöthen, M Rietschel, S Cichon	01GS08144	BMBF NGFNplus MooDS	Germany
CoFaMS - Adelaide	BT Baune	APP1060524	NHMRC	Australia
Danish Radiant	T Werge	0001-2009-2	Højteknologifonden	Denmark
Danish Radiant	T Werge	R24-A3242	Lundbeck Foundation	Denmark
EDINBURGH	AM McIntosh	104036{Z{14{Z	Wellcome Trust	UK
EGCUT	A Metspalu	EstRC-IUT20-60, EU Project No. 2014-2020.4.01.15-0012 and 692145	European Union	Estonia
GenPod{Newmeds	G Lewis, M O'Donovan	G0200243	MRC	UK
GenPod{Newmeds	R Uher	LSHB-CT-2003-503428	EU 6th framework	UK
GenPod{Newmeds	G Lewis	15008	EU IMI-JU	UK
GenScot	AM McIntosh, T-K Clarke, D Porteous	104036{Z{14{Z	Wellcome Trust	UK
GenScot	D Porteous	CZD{16{6	Chief Scientist Office	UK
GenScot	D Porteous	HR03006	Scottish Funding Council	UK
Harvard i2b2	JW Smoller	R01 MH085542	NIMH	USA
Harvard i2b2	RH Perlis	R01 MH086026	NIMH	USA
Münster MDD Cohort	BT Baune	N Health-F2-2008-222963	European Union	Germany
Münster MDD Cohort	TG Schulze	01ZX1314K	BMBF Integrament	Germany
Münster MDD Cohort	TG Schulze		Dr. Lisa Oehler Foundation	Germany
Münster MDD Cohort	TG Schulze	SCHU 1603{5-1	German Research Foundation (DFG)	Germany
Münster MDD Cohort	U Dannlowski	FOR2107 DA1151{5-1xSFB-TRR58, Project C09	German Research Foundation (DFG)	Germany
Münster MDD Cohort	U Dannlowski	Dan3{012{17	Interdisciplinary Center for Clinical Research, Medical Faculty of University of Münster	Germany
Münster MDD Cohort	V Arolt	N Health-F2-2008-222963	European Union	Germany
NESDA	BWJH Penninx	ZonMW Geestkracht grant	N.W.O.	Netherlands
NTR	DI Boomsma	480-15-001{674	N.W.O.	Netherlands
Pfizer	~	115008.5	Innovative Medicine Initiative Joint Undertaking	EU
PsyColaus	M Preisig	3200B0~105993, 3200B0-118308, 33CSCO-122661, 33CS30-139468, 33CS30-148401	Swiss National Science Foundation	Switzerland
QIMR	NG Martin	941177, 971232, 3399450 and 443011	National Health and Medical Research Council	Australia
QIMR	AC Heath	AA07535, AA07728, andAA10249	NIAAA	USA
RADIANT	C Lewis, G Breen	G0701420	MRC	UK
RADIANT	G Breen	G0901245	MRC	UK
RADIANT	G Breen	U01 MH109528	NIMH	UK
Rotterdam Study	AG Uitterlinden	175.010.2005.011, 911-03-012	Netherlands Organization of Scientific Research NWO Investments	Netherlands
SHIP-LEGEND{TREND	HJ Grabe	DFG: GR 1912{5-1	German Research Foundation	Germany
STAR*D	SP Hamilton	R01 MH-072802	NIMH	USA
TwinGene	N Pedersen	EU{QLRT-2001-01254x QLG2-CT-2002-01254	GenomeEUtwin	EU

TwinGene	P Magnusson	20070481	Heart and Lung foundation	Sweden
TwinGene	U de Faire		SSF	Sweden
TwinGene	U de Faire	M-2005-1112	Vetenskapsr det	Sweden
deCODE	K Stefansson		FP7-People-2011-IAPP grant agreement no. 286213 (PsychDPC)	EU
deCODE	K Stefansson	R01 DA017932	NIDA	USA
deCODE	TE Thorgeirsson	R01 DA034076	NIDA	USA
GERA	C SchaeferxN Risch	RC2 AG036607	NIA, NIMH, OD	USA
iPSYCH	T Werge, A Børglum, O Mors, M Nordentoft, D Hougaard, PB Mortesen	R102-A9118, R155-2014-1724, R129-A3973, R24-A3243	Lundbeck Foundation	Denmark
iPSYCH	T Werge, A Børglum, O Mors, M Nordentoft, D Hougaard, PB Mortesen		Novo Nordisk Foundation	Denmark
iPSYCH	T Werge, A Børglum, O Mors, M Nordentoft, D Hougaard, PB Mortesen	R144-A5327	The Capital Region of Denmark	Denmark
iPSYCH	T Werge, A Børglum, O Mors, M Nordentoft, D Hougaard, PB Mortesen	iSEQ, GenomeDK HPC facility, CIRRAU	Aarhus University	Denmark
iPSYCH	T Werge, A Børglum, O Mors, M Nordentoft, D Hougaard, PB Mortesen	294838	EU	EU
iPSYCH	T Werge, A Børglum, O Mors, M Nordentoft, D Hougaard, PB Mortesen		Stanley Medical Research Institute	USA
UK Biobank	AM McIntosh	104036{Z{14{Z	Wellcome Trust	UK
CONVERGE	J Flint	WT090532{Z{09{Z, WT083573{Z{07{Z, WT089269{Z{09{Z	Wellcome Trust	UK
CONVERGE	K Kendler	MH100549	NIMH	USA

In addition, PGC investigators received personal funding from the following sources. EM Byrne award 1053639, NHMRC, Australia. NR Wray award 1078901, 1087889, and 1113400, NHMRC, Australia. DI Boomsma award PAH{6635, KNAW Academy Professor Award, Netherlands. PF Sullivan award D0886501, Vetenskapsr det, Sweden. AM McIntosh award 602450, European Union, UKxaward BADiPS, NC3Rs, UK. C Hayward, Core funding, Medical Research Council, UK. DJ MacIntyre award NRS Fellowship, CSO, UK. DJ Smith award 21930, Brain and Behavior Research Foundation, USAx award 173096, Lister Institute of Preventative Medicine, UK. CA Stockmeier award GM103328, NIMH, USA.

Vigure legends

Vigure 1U ResPlts of GQA metaRanalSsis of seven TohoUts foUMDD. VaWRelation XetYeen adding TohoUts and nPmXeUof genomeRYide signifiTant genomiT Uegions. Beginning Yith the laUgest TohoUt V1WZadded the ne[t laUgest TohoUt V2WPntil all TohoUts YeUe inTlPded V7W. The nPmXeUne[t to eaTh point shoYs the total effeTtive sample si\e. VXW] PantileR] Pantile plot shoYing a maU_ed depaUtPUe fUom a nPll model of no assoTiations Vthe SRa[is is tUPnTated at 1eR12W. VTWManhattan plot Yith [Ra[is shoYing genomiT position VThU1RThU22WZand the SRa[is shoYing statistiTal signifiTanTe as `log$_{10}$VPW. The Ued line shoYs the genomeRYide signifiTanTe thUeshold VPa5[10^{R8}W.

Vigure FU OPtRofRsample genetiT Uis_ sToUe VGRSW pUediTtion analSses. VaW baUianTe e[plained on the liaXilitS sTale Xased on diffeUent disToveUS samples foUthUee taUget samplesc anThoU TohoUt V16Z823 TasesZ 25Z632 TontUolsWZ iPSdCH Va nationallS UepUesentative sample of 18Z629 Tases and 17Z841 TontUolsWand a TliniTal TohoUt fUom MensteUnot inTlPded in the GQA analSsis V845 MDD inpatient TasesZ 834 TontUolsW. The anThoU TohoUt is inTlPded as Xoth disToveUS and taUget as Ye TompPted oPtRofRsample GRS foUeaTh anThoU TohoUt sampleZ TomXined the UesPltsZ and modeled TaseRTontUol statPs as pUediTted XS standaUdi\ed GRS and TohoUt Vsee Tnline MethodsW. VXW Odd Uatios of MDD peUGRS deTile Uelative to the fiUst deTile foU iPSdCH and anThoU TohoUts. VTW MDD GRS VfUom oPtRofRsample disToveUS setsWYeUe signifiTantlS higheUin MDD Tases Yith c eaUlieUage at onsetf moUe seveUe MDD sSmptoms VXased on nPmXeUof TUiteUia endoUsedWf UeTPUUent MDD TompaUed to single episodef and ThUoniTgPnUemitting MDD Vh Stage IVbi TompaUed to hStage IIi ZfiUstRepisode MDD [103]W. EUUoUXaUs UepUesent 95j TonfidenTe inteUvals.

Vigure KU CompaUisons of the MDD GQA metaRanalSsis. VaW MDD UesPlts and enUiThment in XPl_ tissPe mRNARse^ fUom GTE[. OnlS XUain tissPes shoYed enUiThmentZ and the thUee tissPes Yith the most signifiTant enUiThments YeUe all ToUtiTal. VXW MDD UesPlts and enUiThment in thUee majoU XUain Tell tSpes. The MDD genetiT findings YeUe enUiThed in nePUons XPt not oligodendUoTStes oU astUoTStes. VTW PaUtitioned LDSC to evalPate enUiThment of the MDD GQA findings in oveU 50 fPnTtional genomiT annotations VNable RHW. The majoU finding Yas the signifiTant enUiThment of MDD ! "#$% in genomiT Uegions TonseUved aTUoss 29 EPtheUian mammals. [62] OtheU enUiThments implied UegPlatoUS aTtivitSZ and inTlPded open ThUomatin in hPman XUain and an epigenetiT maU_ of aTtive enhanTeUs VH3K4me1W. E[oniT Uegions did not shoY enUiThment. Qe foPnd no evidenTe that NeandeUthal intUogUessed Uegions YeUe enUiThed foU MDD GQA findings.

Vigure 4U GeneUative topogUaphiT mapping of the 19 signifiTant pathYaS UesPlts. The aveUage position of eaTh pathYaS on the map is UepUesented XS a point. The map is ToloUed XS the Rlog$_{10}$VPW oXtained Psing MAGMA. The X and d TooUdinates UesPlt fUom a _eUnel geneUative topogUaphiT mapping algoUithm VGTMW that UedPTes high dimensional gene sets to a tYoRdimensional sTatteUplot XS aTToPnting foU gene oveUlap XetYeen gene sets. EaTh point UepUesents a gene set. NeaUXS points aUe moUe similaU in gene oveUlap than moUe distant points. The ToloUsPUUoPnding eaTh point Vgene setW indiTates signifiTanTe peU the sTale on the Uight. The signifiTant pathYaSs VNable R11W fall into nine main TlPsteUs as desTUiXed in the te[t.

Vigure R1U LeaveRoneRoPt GRS analSses of the anThoU TohoUt. VaW PeUsample R^2 at vaUSing signifiTanTe thUesholds. A all samples in the anThoU TohoUt Ve[Tept oneWSielded signifiTant diffeUenTes in TaseRTontUol distUiXPtions of GRS. ATUoss all samples in the anThoU TohoUtZ GRS e[plained 1.9j of vaUianTe in liaXilitS. VXW Relation XetYeen the nPmXeU of Tases and R^2Z shoYing the e[peTted positive ToUUelation.

Vigure RFU Regional association plots of genomic regions identified from SMR analysis of MDD GWA and eQTL results. SMR analysis helps to prioritize specific genes in a region of association for follow-up functional studies. Figures appear in the same order as the results reported in *Nable RZ*. In the top plot, grey dots represent the MDD GWA *P*-values, diamonds show *P*-values for probes from the SMR test, and triangles are probes without a cis-eQTL (at $P_{e]TL}$ z 5e-8). Genes that pass SMR and heterogeneity tests

(designed to remove loci with more than one causal association) are highlighted in red. The eQTL *P*-values of SNPs are shown in the bottom plot.

Vigure RKU *CiUTPlaUplots to illPstUate DNARDNA loops. FUom the oPtsideZthe tUaT_s shoY hg19 TooUdinates in MXZthe positions of signifiTant MDD assoTiations VRog*$_{10}$*VPVZoPtYaUd is moUe signifiTantVZthe names and positons of GENCODE genesZand the aUT shoY signifiTant DNARDNA loops V^ I 1eR4WfUom HiRC on adPlt ToUte[VgUeenWand fetal fUontal ToUte[VXIPeW VaWThU1c71.5R74.1 MX sPggesting that the tYo statistiTallS independent assoTiations in the Uegion Xoth impliTate NEGR1. VXWThe MDD assoTiation in REREZin TontUastZToinTides Yith manS DNARDNA loops and maS sPggest that this Uegion Tontains sPpeUR enhanTeUelements.*

Vigure R4U *GUaphs depiTting the SNP instUPments Psed in Mendelian Uandomi\ation analSses. TaXle S13 shoYs the paUameteUestimates and signifiTanTeZand these gUaphs shoY sTatteUplots of the instUPments foUMDD and VaWBMIZVXWSeaUs of edPTationZVTWToUonaUSaUteUSdiseaseZand VdWsThi\ophUenia.*

! "#$%& (& %' +, -. &%0-+* 1&1-0* -2. " * 3$4& 11+. -" 3%5&6 -3 7&8 9 9 &

Ch#$	%&' ( ) * $; - .$	/ 0 1$	2) 34 5( ) * 6 7$	1$	8 9: ; $	< %6 8 9$/ = .$	>#?$	1#&@$	A&* &$( ) * 5&B5$
1	8.390-8.895	rs159963	8,504,421	3.2E-08	A/C	0.97 (0.0049)	0.56	H,S	[RERE]; SLC45A1,100194
1	72.511-73.059	rs1432639	72,813,218	4.6E-15	A/C	1.04 (0.005)	0.63	H	NEGR1,-64941
1	73.275-74.077	rs12129573	73,768,366	4.0E-12	A/C	1.04 (0.005)	0.37	H,S	LINC01360,-3486
1	80.785-80.980	rs2389016	80,799,329	1.0E-08	T/C	1.03 (0.0053)	0.28		
1	90.671-90.966	rs4261101	90,796,053	1.0E-08	A/G	0.97 (0.005)	0.37		
1	197.343-197.864	rs9427672	197,754,741	3.1E-08	A/G	0.97 (0.0058)	0.24		DENND1B,-10118
2	57.765-58.485	rs11682175	57,987,593	4.7E-09	T/C	0.97 (0.0048)	0.52	H,S	VRK2,-147192
2	156.978-157.464	rs1226412	157,111,313	2.4E-08	T/C	1.03 (0.0059)	0.79		[LINC01876]; NR4A2,69630; GPD2,-180651
3	44.222-44.997	chr3_44287760_I	44,287,760	4.6E-08	I/D	1.03 (0.0051)	0.34	T	[TOPAZ1]; TCAIM,-91850; ZNF445,193501
3	157.616-158.354	rs7430565	158,107,180	2.9E-09	A/G	0.97 (0.0048)	0.58	H	[RSRC1]; LOC100996447,155828; MLF1,-181772
4	41.880-42.189	rs34215985	42,047,778	3.1E-09	C/G	0.96 (0.0063)	0.24		[SLC30A9]; LINC00682,-163150; DCAF4L1,59294
5	87.443-88.244	chr5_87992715_I	87,992,715	7.9E-11	I/D	0.97 (0.005)	0.58	H	LINC00461,-12095; MEF2C,21342
5	103.672-104.092	chr5_103942055_D	103,942,055	7.5E-12	I/D	1.03 (0.0048)	0.48	C	
5	124.204-124.328	rs116755193	124,251,883	7.0E-09	T/C	0.97 (0.005)	0.38		LOC101927421,-120640
5	164.440-164.789	rs11135349	164,523,472	1.1E-09	A/C	0.97 (0.0048)	0.48	H	
5	166.977-167.056	rs4869056	166,992,078	6.8E-09	A/G	0.97 (0.005)	0.63		[TENM2]
6	27.738-32.848	rs115507122	30,737,591	3.3E-11	C/G	0.96 (0.0063)	0.18	S	CPQRSOS MHC
6	99.335-99.662	rs9402472	99,566,521	2.8E-08	A/G	1.03 (0.0059)	0.24		FBTL4,-170672; C6UrV168,154271
7	12.154-12.381	rs10950398	12,264,871	2.6E-08	A/G	1.03 (0.0049)	0.41		[TMEM106B]; VWDE,105637
7	108.925-109.230	rs12666117	109,105,611	1.4E-08	A/G	1.03 (0.0048)	0.47		
9	2.919-3.009	rs1354115	2,983,774	2.4E-08	A/C	1.03 (0.0049)	0.62	H	PWM3,-139644; LINC01231,-197814
9	11.067-11.847	rs10959913	11,544,964	5.1E-09	T/G	1.03 (0.0057)	0.76		
9	119.675-119.767	rs7856424	119,733,595	8.5E-09	T/C	0.97 (0.0053)	0.29		[ASTN2]
9	126.292-126.735	rs7029033	126,682,068	2.7E-08	T/C	1.05 (0.0093)	0.07		[DENND1A]; LHT2,-91820
10	106.397-106.904	rs61867293	106,563,924	7.0E-10	T/C	0.96 (0.0061)	0.20	H	[SORCS3]
11	31.121-31.859	rs1806153	31,850,105	1.2E-09	T/G	1.04 (0.0059)	0.22		[DKFZX686K1684]; [PAWPAR]; ELP4,44032;
12	23.924-24.052	rs4074723	23,947,737	3.1E-08	A/C	0.97 (0.0049)	0.41		[SOT5]
13	44.237-44.545	rs4143229	44,327,799	2.5E-08	A/C	0.95 (0.0091)	0.92		[ENOT1]; LACC1,-125620; CCDC122,82689
13	53.605-54.057	rs12552	53,625,781	6.1E-19	A/G	1.04 (0.0048)	0.44	H	[OLFM4]; LINC01065,80099
14	41.941-42.320	rs4904738	42,179,732	2.6E-09	T/C	0.97 (0.0049)	0.57		[LRFN5]
14	64.613-64.878	rs915057	64,686,207	7.6E-10	A/G	0.97 (0.0049)	0.42		[SYNE2]; MIR548H1,-124364; ESR2,7222
14	75.063-75.398	chr14_75356855_I	75,356,855	3.8E-09	D/I	1.03 (0.0049)	0.49		[DLST]; PROT2,-26318; RPS6KL1,13801
14	103.828-104.174	rs10149470	104,017,953	3.1E-09	A/G	0.97 (0.0049)	0.49	S	BAG5,4927; APOPT1,-11340
15	37.562-37.929	rs8025231	37,648,402	2.4E-12	A/C	0.97 (0.0048)	0.57	H	
16	6.288-6.347	rs8063603	6,310,645	6.9E-09	A/G	0.97 (0.0053)	0.65		[RBFOT1]

16	7.642-7.676	rs7198928	7,666,402	1.0E-08	T/C	1.03 (0.005)	0.62		[RBFOT1]
16	13.022-13.119	rs7200826	13,066,833	2.4E-08	T/C	1.03 (0.0055)	0.25		[SHISA9]; CPPED1,-169089
16	71.631-72.849	rs11643192	72,214,276	3.4E-08	A/C	1.03 (0.0049)	0.41		PMFBP1,-7927; DHT38,67465;
17	27.345-28.419	rs17727765	27,576,962	8.5E-09	T/C	0.95 (0.0088)	0.92		[CRYBA1]; MYO18A,-69555; NWFIP2,5891
18	36.588-36.976	rs62099069	36,883,737	1.3E-08	A/T	0.97 (0.0049)	0.42		[MIR924HG]
18	50.358-50.958	rs11663393	50,614,732	1.6E-08	A/G	1.03 (0.0049)	0.45	O	[DCC]; MIR4528,-148738
18	51.973-52.552	rs1833288	52,517,906	2.6E-08	A/G	1.03 (0.0054)	0.72		[RAB27B]; CCDC68,50833
18	52.860-53.268	rs12958048	53,101,598	3.6E-11	A/G	1.03 (0.0051)	0.33	S	[TCF4]; MIR4529,-44853
22	40.818-42.216	rs5758265	41,617,897	7.6E-09	A/G	1.03 (0.0054)	0.28	H,S	[L3MBTL2]; EP300-AS1,-24392; CHADL,7616

ShUZ R [rOS[Q \Ur 44 \ ORU] ^c _Uc^[ch^O`^R\ \ ORU] O-Z ^SOs^\ R^Vc[ RcO\Ur MDD. Chr (chrU] UsU] O) [ RS RO\ ^UR (aUbRS[ r^Os ^R Ma, h\ 19) [ rOshUZ R, SO^ROS ac _Uc[ Q\URs UVSNPs Z ^Q ! d1P10$^{-5}$ [ RS LD $r^2$ e 0.1 Z ^Q Q O] UsQ[ ssUc^[ QOS SNP (_UZ OsQP-`[_bO^R rO\ ^UR, _^sQOS) Z hUsO_Uc[Q\UR^s \^ OR^R aX. IR Q rOO rO\ ^URs [sOcURS SNP \b_V^s Q OV^_QO^R\ cr^QO^[[RS Q OsOZ OrO\U_UZ OS bX Z ^Q cURS^Q\UR[_[R[_cs^s sO_OcQ URc rs1432639 [s s^\ R^Vc[RQ Z ^Q ! g2.0P10^{-4} \Ur rs12134600 [\QO r V Q R\ rs1432639; Chr5, cURS^Q\UR[_[R[_cs^s shUZ s QZ U ^RSOXORSORQ[ssUc^[Q\URs sO_OcQR\ rs247910 [RS rs10514301 [s Q O] UsQ[ssUc^[QOS SNPs; [RS Chr10 cURS^Q\UR[_[R[_cs^s sO_OcQ URc rs61867293 Z ^Q ! g8.6P10^{-5} \Ur rs1021363 [\QO r cURS^Q\UR^R\ UR rs61867293. FUr Q[ch UV Q O 47 SNPs, Q O rO^s [QO[sQ1 [SS^Q\UR[_\ ORU] O-Z ^SOs^\ R^Vc[RQ SNP ^R Q O c_bsQ r UVsbrrUbRS^R\ SNPs Z ^Q _UZ P-`[_bOs. Chr U] UsU] OT Z [s [R[_chOS abQh[S RU V RS^R\ s Q h[Q] OQ\ ORU] O-Z ^SOs^\ R^Vc[RcO

CU_b] R [_aOs [RS [aarO`^[ Q\URs. A1/2 g Q O QZ U [ _OOs (Ur ^RsOrQ\UR-SO_OQ\UR); A1 Z [ s QOsQOS VUr [ ssUc^[ Q\UR, [ RS ^Qg OR (USSs r[ Q\U) [ RS SE (sQ RS[ rS OrUr) [ rO shUZ R FrO\ Wg V rO\ bORcc UVA1 ^R cURQ\Us [ crUss [ __cUhUrQg. ERQ^Os ^R Q O j PrO`k cU_b] R^RS^c[QOZ h^ch UV\Ubr XrO`^Ubs sQbS^Os ^SORQ^VOS \ ORU] Os^\ R^Vc[ RQ [ ssUc^[ Q\URs ^R [ rO\ ^UR. HgHcSOOQ[ _, [19] 23[ RSMOGWA UVsO_V rOXUrQOS c_R^c[ _SOXrOss^UR (S^scU` Orc s[ ] X_OU` Or [Xs Z ^Q Q ^s X[XOr); OgOl a[c OQ[_, [18]] OQ-[R[_cs^s UVGWA UVMDD, SOXrOss^ Osc] XQU] s, XscchU_U\ ^c[_Z O_-aO^R\ [RS RObrUQc^s] (^Rc_bSOs] [Rc [RchUr cUhUrQg V^U] Q ^s X[XOr); SgPGC rOXUrQUR 108 sch^hUXhrOR^[-[ssUc^[QOS _Uc^[22]; [RS CgCHARGE cURsUrQb]] OQ-[R[_cs^s UVSOXrOss^ Osc] XQU] s ^R Xr^] [r^c U_SOr [Sb_Qg. [16] GORO cURQOPQ S^s Q RcOs aOQZ OOR Q O PQ l SNP [RS Q O c_UsOsQ\ ORQs [rO shUZ R Br[cl OQg ^RS^c[QO Q [QQ O PQ l SNP Z [s Z ^Q ^R Q [Q\ ORO ThO c_UsOsQ\ ORQs bXsQPQ] (Q l ^R\ sQ [RS ^RQU [ccUbRQ [s [RO\ [Q` ORb] aOr ^RS^c[ Q^R\ S^s Q RcO^R aX aOQZ OOR PQ l SNP [ RS Q O RQ [ rOsQ\ ORO aUbRS[ rc) [ RS SUZ RsQPQ ] (XUs^Q` OS^s Q RcO^R aX) [rO[_sU shUZ R, ^V Q rO^s [V[R ^R\ \ ORO Z ^Q ^R 200 l a. ThOR[ ] OUV Q c_UsOsQ\ ORO^s aU_SOS. NUQO Q [Q^Q^s \ ORO r[_c RUQ l RUZ R Z hOQ rO Q O [ssUc^[QOS SNPs h[`O a^U_U\ ^c[_OXOcQg UR Q OsOUr UQ rO] UrOS^s Q RQ\ ORQs.

! " #$%& (& %' %+,-&. //%$' +,. * 0& 12 3 3& 4 ,+5& +5%&6,0. /6%0&7 6,0%' 0%0&7 * 6&589 " * &/" ,+0&

! "#$%&	!" &	' (&	) * +&	$\#^{'}_{\$\%\&}$ &	, - . * &
Depressive symptoms, CHARGE	0.91	0.123	3.2e-12	0.04	23290196
Depressive symptoms, SSGAC	0.98	0.034	1.3e-176	0.05	27089181
ADHD (iPSYCH-PGC)	0.42	0.033	6.1e-36	0.24	submitted
Anorexia nervosa	0.13	0.028	7.1e-5	0.55	24514567
Anxiety disorders	0.80	0.140	2.0e-7	0.06	26857599
Autism spectrum disorders (iPSYCH-PGC)	0.44	0.039	8.4e-28	0.20	submitted
Bipo@r disorder	0.32	0.034	3.3e-19	0.43	21926972
ScPiQopPrenia	0.34	0.025	7.7e-40	0.46	25056061
SmoRnS, ever vs never	0.29	0.038	7.0e-13	0.08	20418890
Daytime s@epiness T	0.19	0.048	5.7e-4	0.05	0
Usomnia T	0.38	0.038	4.0e-22	0.13	0
Viredness	0.67	0.037	6.2E-72	0.07	28194004
SubWctive XeOObeinS	-0.65	0.035	7.5E-76	0.03	27089181
Yeuroticism	0.70	0.031	2.5E-107	0.09	27089181
Co@Se comp@tion	-0.17	0.034	6.7E-6	0.08	23722424
Years oZeducation	-0.13	0.021	1.6E-8	0.13	27225129
Body Zat	0.15	0.038	6.5e-4	0.11	26833246
Body mass index	0.09	0.026	3.6e-3	0.19	20935630
[besity c@ss 1	0.11	0.029	1.6e-3	0.22	23563607
[besity c@ss 2	0.12	0.033	3.0e-3	0.18	23563607
[besity c@ss 3	0.20	0.053	1.6e-3	0.12	23563607
[verXeiSPt	0.13	0.030	1.4e-4	0.11	23563607
Waist circumZerence	0.11	0.024	8.2e-5	0.12	25673412
Waist-to-Pip ratio	0.12	0.030	2.9e-4	0.11	25673412
ViiSOycerides	0.14	0.028	1.0e-5	0.17	20686565
ASe at menarcPe	-0.14	0.023	6.3E-8	0.20	25231870
ASe oZZirst birtP	-0.29	0.029	6.1E-22	0.06	27798627
\ atPers aSe at deatP	-0.28	0.058	3.0E-5	0.04	27015805
Yumber oZcPiOdren ever born	0.13	0.036	2.4E-3	0.03	27798627
Coronary artery disease	0.12	0.027	8.2e-5	0.08	26343387
S] uamous ceOOOnScancer	0.26	0.075	3.6e-3	0.04	27488534

AOOSenetic corre@tions (!") estimated usinS bivariate ^DSC appOed to MDD GWA resuOs are in ! " #$%& : ; ' . SPoXn above are tPe !" oZ MDD XitP Za@e discovery rate (\ DR) _ 0.01 (\ DR estimated Zor 221 Senetic corre@tions). VPematicaOy re@ted traits are indicated by sPadinS. VPe iPSYCH expanded coPort is a nationaOy representative coPort based on bOod spots coOected at birtP. WitPin iPSYCH, tPe MDD-ADHD !" Xas 0.58 (SE 0.050) and tPe MDD-ASD !" Xas 0.51 (SE 0.07) ` tPese are @rSer tPan tPose Osted above, and inconsistent XitP arteZactuaO corre@tions. $\#^{'}_{\$\%\&}$ is sPoXn to aid interpretation as PiSP !" in tPe context oZ PiSP $\#^{'}_{\$\%\&}$ is more noteXortPy tPan XPen $\#^{'}_{\$\%\&}$ (is OX. PMUD is PubMed artic@ identiZer.

T SeOZreported daytime s@epiness and insomnia Zrom ab BiobanR excOudinS subWcts XitP MDD, otPer psycPiatric disorders (bipo@r disorder, scPiQopPrenia, autism, inteOectuaO disabiOty), sPiZt XorRers, and tPose taRinS Pypnotics.

<%1%%-%0&

1 bessOer, R. C. c Bromet, E. d VPe epidemioOSy oZdepression across cuOtures. *Annu Rev Public Health* / 0, 119-138, doie10.1146fannurev-pubOPeaOP-031912-114409 (2013).

2 dudd, ^. ^. VPe cOnicaOcourse oZunipoOar maWr depressive disorders. *Arch Gen P78chiatr*8 10, 989-991 (1997).

3 ^opeQ A. D., MatPers, C. D., EQQati, M., damison, D. V. c Murray, C. d GOobaOand reSionaOburden oZdisease and risR Zactors, 2001esystematic anaQysis oZpopuOation PeaOP data. 9*ancet* / 23, 1747-1757, doie10.1016fS0140-6736(06)68770-9 (2006).

4 WittcPen, H. a. *et al*: VPe siQe and burden oZmentaOdisorders and otPer disorders oZtPe brain in Europe 2010. ; *ur* <*eur*=>78*ch*=>*har*? *ac*=*l* 45, 655-679, doie10.1016fWeuroneuro.2011.07.018 (2011).

5 \errari, A. d *et al*: Burden oZdepressive disorders by country, sex, aSe, and yeareZindinSs Zrom tPe SOobaOburden oZdisease study 2010. *P9*=@A *eB*56, e1001547, doie10.1371fWurnaOpmed.1001547 (2013).

6 AnSst, \., Stassen, H. H., COyton, P. d c AnSst, d MortaOty oZpatients XitP mood disorderse ZoOOX-up over 34-38 years. CADect Ei7=*r*B27, 167-181 (2002).

7 Gustavsson, A. *et al*: Cost oZdisorders oZtPe brain in Europe 2010. ; *ur* <*eur*=>78*ch*=>*har*? *ac*=*l* 45, 718-779, doie10.1016fWeuroneuro.2011.08.008 (2011).

8 Murray, C. d *et al*: DisabiOty-adWsted OZe years (DA^Ys) Zor 291 diseases and inWries in 21 reSions, 1990-2010ea systematic anaQysis Zor tPe GOobaOBurden oZDisease Study 2010. 9*ancet* / 76, 2197-2223, doie10.1016fS0140-6736(12)61689-4 (2012).

9 SuOOvan, P. \., YeaO, M. C. c bendOr, b. S. Genetic epidemioOSy oZmaWr depressioneRevieX and meta anaQysis. *A*? *erican* G=*urnal* =D*P78chiatr*8 513, 1552-1562 (2000).

10 Rice, \., HaroOd, G. c VPapar, A. VPe Senetic aetioOSy oZcPiOPood depressionea revieX. CF*hil*B *P78ch*=*l P78chiatr*8 0/ , 65-79 (2002).

11 giRtorin, A. *et al*: HeritabiOty oZ PerinataO Depression and Genetic [verOap WitP YonperinataO Depression. *A*? C*P78chiatr*8, appiaWp201515010085, doie10.1176fappi.aWp.2015.15010085 (2015).

12 ^evinson, D. \. *et al*: Genetic studies oZ maWr depressive disordere XPy are tPere no GWAS ZindinSs, and XPat can Xe do about it. G=*l P78chiatr*8 32, 510-512 (2014).

13 Cross-Disorder Group oZtPe PsycPiatric Genomics Consortium. Genetic reOationsPip betXeen Zve psycPiatric disorders estimated Zrom Senome-Xide SYPs. <*ature genetic*7 01, 984-994, doie10.1038fnS.2711 (2013).

14 BuORSuOvan, B. b. *et al*: An atOas oZSenetic correOations across Puman diseases and traits. <*ature Genetic*703, 1236-1241 (2015).

15 MaWr Depressive Disorder WorRnS Group oZ tPe PGC. A meSa-anaQysis oZ Senome-Xide association studies Zor maWr depressive disorder. A=*lecular P78chiatr*857, 497-511 (2013).

16 HeR, b. *et al*: A Senome-Xide association study oZdepressive symptoms. G=*l P78chiatr*83/ , 667-678, doie10.1016fWiopsycP.2012.09.033 (2013).

17 C[YgERGE Consortium. Sparse XPoOe Senome se] uencinSidentiZies tXo OOci Zor maWr depressive disorder. <*ature* (2015).

18 [Rbay, A. *et al*: Genetic variants associated XitP subWective XeQbeinS, depressive symptoms, and neuroticism identiZed tProuSP Senome-Xide anaQyses. <*at Genet*, doie10.1038fnS.3552 (2016).

19 Hyde, C. ^. *et al*: UdentiZcation oZ15 Senetic Qoci associated XitP risR oZmaWr depression in individuaQ oZEuropean descent. <*at Genet* 07, 1031-1036, doie10.1038fnS.3623 (2016).

20 SuQQvan, P. \. *et al*: PsycPiatric GenomicseAn apdate and an ASenda. (Submitted).

21 gisscPer, P. M., BroXn, M. A., McCartPy, M. Uc YanS, d \ive Years oZGWAS Discovery. *A*? C*Hu*? *Genet* 86, 7-24, doie10.1016fWaWPS.2011.11.029 (2012).

22 ScPiQoppPrenia WorRnS Group oZtPe PsycPiatric Genomics Consortium. BioQSicaQinsiSPts Zrom 108 scPiQoppPrenia-associated Senetic Qoci. <*ature* 155, 421-427 (2014).

23 PsycPiatric GWAS Consortium BipoQar Disorder WorRnS Group. ^arSe-scaQe Senome-Xide association anaQysis oZbipoQar disorder identiZes a neX susceptibiQty Qocus near [Dh4. <*ature genetic*70/ , 977-983, doie10.1038fnS.943 (2011).

24 Wray, Y. R. *et al*: Genome-Xide association study oZmaWr depressive disordere neX resuQts, meta-anaQysis, and Qessons Qearned. A=l *P*78*chiatr*853, 36-48, doie10.1038fmp.2010.109 (2012).

25 YanS, d *et al*: ConditionaQand WWint muQtipQe-SYP anaQysis oZGWAS summary statistics identiZes additionaQ variants inZQencinS compQex traits. <*at Genet* 00, 369-375, S361-363, doie10.1038fnS.2213 (2012).

26 Wray, Y. R. c Maier, R. Genetic Basis oZCompQex Genetic DiseaseeVPe Contribution oZDisease HeteroSeneity to MissinS Heritabiqty. F*urrent* ;>*iBe*? *i=l=g*8 *Re*>=*rt*7 5, 220-227, doie10.1007fs40471-014-0023-3 (2014).

27 ^ocPe, A. E. *et al*: Genetic studies oZbody mass index yieQd neX insiSPts Zor obesity bioQSy. <*ature* 157, 197-206, doie10.1038fnature14177 (2015).

28 Berndt, S. U*et al*: Genome-Xide meta-anaQysis identiZes 11 neX Qoci Zor antPropometric traits and provides insiSPts into Senetic arcPitecture. <*at Genet* 01, 501-512, doie10.1038fnS.2606 (2013).

29 BradZeQd, d P. *et al*: A Senome-Xide association meta-anaQysis identiZes neX cPiQdPood obesity Qoci. <*at Genet* 00, 526-531, doie10.1038fnS.2247 (2012).

30 SpeQotes, E. b. *et al*: Association anaQyses oZ249,796 individuaQ reveaQ18 neX Qoci associated XitP body mass index. <*at Genet* 04, 937-948, doie10.1038fnS.686 (2010).

31 WiQQr, C. d *et al*: Six neX Qoci associated XitP body mass index PiSPQSPt a neuronaQnZQence on body XeiSPt reSuQation. <*at Genet* 05, 25-34, doie10.1038fnS.287 (2009).

32 VPorQiZsson, G. *et al*: Genome-Xide association yieQds neX se] uence variants at seven Qoci tPat associate XitP measures oZobesity. <*at Genet* 05, 18-24, doie10.1038fnS.274 (2009).

33 ^iu, W. c RodSers, G. P. [QZactomedin 4 expression and ZunctionsinnateimmunityinZQammation, and cancer. F*ancer A eta7ta7i7 Rev* /1, 201-212, doie10.1007fs10555-016-9624-2 (2016).

34 AnPoQ, R. R. [QZactomedin proteinseQcentraQpQayers in deveQopment and disease. I*r=nt F-ell E-ev Gi=l* 4, 6, doie10.3389fZeQQ2014.00006 (2014).

35 Boucard, A. A., bo, d c SudPoZ, V. C. HiSP aZZnity neurexin bindinS to ceQQadPesion G-protein-coupQed receptor CUR^1fQatropPiQn-1 produces an intercoQQar adPesion compQex. CG=l F*he*? 473, 9399-9413, doie10.1074fWbc.M111.318659 (2012).

36 [iSuOOvan, M. ^., Martini, \., von DaaPe, S., ComoOetti, D. c GPosP, A. ^PHY3, a presynaptic adPesion-GPCR impOicated in ADHD, reSuOates tPe strenStP oZneocorticaOOayer 2f3 synaptic input to Oayer 5. <*eural Eev* 8, 7, doie10.1186f1749-8104-9-7 (2014).

37 Saunders, Y. R. *et al*: ASe-dependent transcriptome and proteome ZOOOXinS transection oZ neonataOspinaOcord oZMonodeOPis domestica (SoutP American Srey sPort-taiOed opossum). *P9=@ J ne* 8, e99080, doie10.1371fWurnaOpone.0099080 (2014).

38 SanO R., \erraro, G. B. c \ournier, A. E. US^[Y ceOOadPesion moOecuOes are sPed Zrom tPe ceOO surZace oZ corticaO neurons to promote neuronaO SroXtP. C Gi=l F*he*? 486, 4330-4342, doie10.1074fWc.M114.628438 (2015).

39 ^ee, A. W. *et al*: \unctionaOinactivation oZ tPe Senome-Xide association study obesity Sene neuronaOSroXtP reSuOator 1 in mice causes a body mass pPenotype. *P9=@J ne* 3, e41537, doie10.1371fWurnaOpone.0041537 (2012).

40 ScPaZer, M., Brauer, A. a., SavasPan, Y. E., RatPWan, \. G. c BrummendorZ, V. YeurotractinfROn promotes neurite outSroXtP and is expressed on reactive astrocytes aZter entorPinaOcortex Oesion. A=l F*ell* <*eur=7ci* 48, 580-590, doie10.1016fWmcn.2005.04.010 (2005).

41 HasPimoto, V., MaePaXa, S. c Miyata, S. US^[Y ceOOadPesion moOecuOes reSuOate synaptoSenesis in PippocampaOneurons. F*ell* Gi=*che*? I*unct* 43, 496-498, doie10.1002fcbZ.1600 (2009).

42 HasPimoto, V., Yamada, M., MaePaXa, S., YaPasPima, V. c Miyata, S. US^[Y ceOOadPesion moOecuOe bioOn is a cruciaOmoduOator Zor synapse number in PippocampaOneurons. G*rain Re*7 5440, 1-11, doie10.1016fWbrainres.2008.05.069 (2008).

43 PiscPedda, \. c PiccoO, G. VPe US^[Y \amiO Member YeSr1 Promotes YeuronaOArboriOation ActinS as SoOOboOe \actor via \G\R2. I*r=nt* A=l <*eur=7ci* 7, 89, doie10.3389fZnmoO.2015.00089 (2015).

44 PiscPedda, \. *et al*: A ceOOsurZace biotinyOation assay to reveaOmembrane-associated neuronaO cueseYeSr1 reSuOates dendritic arboriOation. A=l F*ell Pr=te=*? *ic*7 5/, 733-748, doie10.1074fmcp.M113.031716 (2014).

45 Boender, A. d, van RoOen, A. d c Adan, R. A. YutritionaOstate aZZects tPe expression oZ tPe obesity-associated Senes Etv5, \aim2, \to, and YeSr1. J*be7it*8 K@*ver @ring*L 46, 2420-2425, doie10.1038foby.2012.128 (2012).

46 WPeeOer, E. *et al*: Genome-Xide SYP and CYg anaOysis identiZies common and OoX-Ze] uency variants associated XitP severe earOy-onset obesity. <*at Genet* 01, 513-517, doie10.1038fnS.2607 (2013).

47 ^ee, d A. *et al*: CytopOasmic RbZox1 ReSuOates tPe Expression oZ Synaptic and Autism-ReOated Genes. <*eur=n* 78, 113-128, doie10.1016fWneuron.2015.11.025 (2016).

48 GePman, ^. V. *et al*: VPe spOicinS reSuOator RbZox1 (A2BP1) controOs neuronaOexcitation in tPe mammaOian brain. <*at Genet* 0/, 706-711, doie10.1038fnS.841 (2011).

49 \oSeO, B. ^. *et al*: RB\[j 1 reSuOates botP spOicinS and transcriptionaOnetXorRs in Puman neuronaO deveOopment. *Hu*? A=l *Genet* 45, 4171-4186, doie10.1093fPmSfdds240 (2012).

50 Amir-hiOberstein, ^. *et al*: Homeodomain protein otp and activity-dependent spOicinS moduOate neuronaOadaptation to stress. <*eur=n* 3/, 279-291, doie10.1016fWneuron.2011.11.019 (2012).

51 Pariante, C. M. c ^iSPtman, S. ^. VPe HPA axis in ma\Wr depressione cOassicaOtPeories and neX deveOopments. NrenB7<eur=7ci / 5, 464-468, doie10.1016f\Wins.2008.06.006 (2008).

52 YPo, b. *et al*: ComprePensive Sene- and patPXay-based anaOysis oZdepressive symptoms in oOder aduOts. CAlOhei? er7Ei7 01, 1197-1206, doie10.3233fdAD-148009 (2015).

53 CPoi, Y. *et al*: SA^M5 trans-synapticaOOy interacts XitP ^AR-RPVPs in a spOicinS-dependent manner to reSuOate synapse deveOopment. @i Re> 2, 26676, doie10.1038fsrep26676 (2016).

54 MaP, W. *et al*: SeOected SA^M (synaptic adPesion-OiRe moOecuOe) ZamiOy proteins reSuOate synapse Zormation. C<eur=7ci / 6, 5559-5568, doie10.1523fdYEaR[SCU4839-09.2010 (2010).

55 hPu, Y. *et al*: Yeuron-speciZc SA^M5 Oimits inZOammation in tPe CYS via its interaction XitP HgEM. @i ABv 4, e1500637, doie10.1126fsciadv.1500637 (2016).

56 AmieOOd *et al*: Mutations in VC\ 4, encodinS a cOass Ubasic PeOix-Ooop-PeOix transcription Zactor, are responsibOe Zor Pitt-HopRins syndrome, a severe epiOeptic encepPaOopatPy associated XitP autonomic dysZunction. A? CHu? *Genet* 76, 988-993, doie10.1086f515582 (2007).

57 ARbarian, S. *et al*: VPe PsycPEYC[DE pro\Wct. <at <eur=7ci 57, 1707-1712, doie10.1038fnn.4156 (2015).

58 GVEx Consortium. Human Senomics. VPe Genotype-Vissue Expression (GVEx) piOot anaOysise muOtitissue Sene reSuOation in Pumans. @ience / 07, 648-660, doie10.1126fscience.1262110 (2015).

59 ScPmaaO, ^. *et al*: CorticaO abnormaOities in aduOts and adoOescents XitP ma\Wr depression based on brain scans Zrom 20 coPorts XorOdXide in tPe EYUGMA Ma\Wr Depressive Disorder WorRinS Group. A =l P78chiatr8, doie10.1038fmp.2016.60 (2016).

60 CaPoy, d D. *et al*: A transcriptome database Zor astrocytes, neurons, and oOiSodendrocytesea neX resource Zor understandinS brain deveOopment and Zunction. C <eur=7ci 47, 264-278, doie10.1523fdYEaR[SCU4178-07.2008 (2008).

61 \inucane, H. b. *et al*: PartitioninS PeritabiOity by ZunctionaO cateSory usinS GWAS summary statistics. <*ature Genetic7* 03, 1228-1235 (2015).

62 ^indbOad-VoP, b. *et al*: A PiSP-resoOution map oZ Puman evoOutionary constraint usinS 29 mammaOs. <*ature* 037, 476-482, doie10.1038fnature10530 (2011).

63 Simonti, C. Y. *et al*: VPe pPenotypic OeSacy oZ admixture betXeen modern Pumans and YeandertaOs. @*ience* / 15, 737-741, doie10.1126fscience.aad2149 (2016).

64 hPu, h. *et al*: UnteSration oZsummary data Zrom GWAS and ek V^ studies predicts compOex trait Sene tarSets. <*at Genet* 07, 481-487, doie10.1038fnS.3538 (2016).

65 Gusev, A. *et al*: UnteSrative approacPes Zor OarSe-scaOe transcriptome-Xide association studies. <*at Genet* 07, 245-252, doie10.1038fnS.3506 (2016).

66 \romer, M. *et al*: Gene expression eOucidates ZunctionaO impact oZpoOySenic risRZor scPiOopPrenia. <*ature* <*eur=7cience* 58, 1442-1453, doie10.1038fnn.4399 (2016).

67 Smemo, S. *et al*: [besity-associated variants XitPin \V[Zorm OonS-ranSe ZunctionaO connections XitP URj 3. <*ature* 163, 371-375, doie10.1038fnature13138 (2014).

68 Won, H. *et al*: CPromosome conZormation eOucidates reSuOatory reOationsPips in deveOopinS Puman brain. <*ature* 1/ 7, 523-527, doie10.1038fnature19847 (2016).

69 Martin, d S. *et al*: Ha GUde Hi-C aniZyin S Genomic Uterro Sator. (Submitted).

70 PatPXay Anay Qsis Sub Sroup oZ tPe PsycPiatric Genomics, C. PsycPiatric Senome-Xide association study anay Qses imp Qcate neurona Q immune and Pistone patPXays. <*at* <*eur*=7*ci* 57, 199-209, doi e10.1038f nn.3922 (2015).

71 De Rubeis, S. *et al*: Synaptic, transcriptiona Q and cPromatin Senes disrupted in autism. <*ature* 151, 209-215, doi e10.1038f nature13772 (2014).

72 Genovese, G. *et al*: Uncreased burden oZ u Qra-rare protein-a Qerin S variants amon S 4,877 individua Q XitP scPi Qop Prenia. <*ature* <*eur*=7*cience*, doi e10.1038f nn.4402 (2016).

73 Gaspar, H. A. c Breen, G. PatPXays anay Qses oZ scPi Qop Prenia GWAS Zocusin Son RnoXn and nove Q dru Starsets. doi e10.1101f091264 (Submitted).

74 Breen, G. *et al*: VransQatin S Senome-Xide association Zndin Ss into neX tPerapeutics Zor psycPiatry. <*at* <*eur*=7*ci* 58, 1392-1396, doi e10.1038f nn.4411 (2016).

75 hPen S, d *et al*: ^D Hube a centra Q ed database and Xeb inter Zace to per Zorm ^D score re Sression tPat maximi Qes tPe potentia Q oZ summary Qve Q GWAS data Zor SYP Peritabi Qy and Senetic corre Qation anay Qsis. G=*in*D=*r*? *atic* 7/ / , 272-279, doi e10.1093f bioin Zormatics f btX613 (2017).

76 [Rbay, A. *et al*: Genome-Xide association study identi Zes 74 Q oci associated XitP educationa Q attainment. <*ature* 1/ / , 539-542, doi e10.1038f nature17671 (2016).

77 YiRpay, M. *et al*: A compre Pensive 1,000 Genomes-based Senome-Xide association meta-anay Qsis oZ coronary artery disease. <*at Genet* 03, 1121-1130, doi e10.1038f nS.3396 (2015).

78 WaSner, G. P. c hPan S, d VPe p Qeiotropic structure oZ tPe Senotype-pPenotype mape tPe evo Qvabi Qy oZ compQex or Sanisms. <*at Rev Genet* 54, 204-213, doi e10.1038f nrS2949 (2011).

79 Hippocrates. *A*>*h*=*ri*7? 7. (400 BCE).

80 SRene, Y. G. *et al*: Brain ce Qotypes and tPe Senetic basis oZ scPi Qop Prenia. (Submitted).

81 YanS, j . *et al*: Widespread Expansion oZ Protein Uteraction Capabi Qies by A Qernative Sp Qcin S. F*ell* 520, 805-817, doi e10.1016f WeQ2016.01.029 (2016).

82 hPan S, j . *et al*: Ce Q Vype-Speci Zic A Qernative Sp Qcin S Governs Ce Q Qate in tPe Deve Qopin S Cerebra Q Cortex. F*ell* 522, 1147-1162 e1115, doi e10.1016f WeQ2016.07.025 (2016).

83 bess Qer, R. C. *et al*: VPe epidemio Q Sy oZ ma Wor depressive disorder e resu Qs Zrom tPe Yationa Q Comorbidity Survey Rep Qication (YCS-R). G? *a* 478, 3095-3105 (2003).

84 Hasin, D. S., GoodXin, R. D., Stinson, \. S. c Grant, B. \. Epidemio Q Sy oZ ma Wor depressive disorder e resu Qs Zrom tPe Yationa Q Epidemio Q Sic Survey on A Qo Po Qism and Re Qated Conditions. *Arch Gen P*78*chiatr*8 24, 1097-1106, doi e10.1001f arcPpsyc.62.10.1097 (2005).

85 bend Qer, b. S. *et al*: VPe structure oZ Senetic and environmenta Q risR Zactors Zor syndroma Q and subsyndroma Q common DSM-U g axis U and a Q axis UW disorders. *A*? C*P*78*chiatr*8 527, 29-39, doi e10.1176f appi.a Wp.2010.10030340 (2011).

86 bend Qer, b. S., Prescott, C. A., Myers, d c Yea Qe, M. C. VPe structure oZ Senetic and environmenta Q risR Zactors Zor common psycPiatric and substance use disorders in men and Xomen. *Arch Gen P*78*chiatr*8 26, 929-937 (2003).

87 Robinson, E. B. *et al*: Genetic risR Zor autism spectrum disorders and neuropsycPiatric variation in tPe Senera Qpopu Qation. <*at Genet* 07, 552-555, doi e10.1038f nS.3529 (2016).

88 MiddeOdorp, C. M. *et al*. A Genome-Wide Association Meta-AnaOysis oZ Attention-DeZcitf Hyperactivity Disorder Symptoms in PopuOation-Based Pediatric CoPorts. *CA? AcaB FhilB AB=le7c P78chiatr8* 11, 896-905 e896, doie10.1016f WWac.2016.05.025 (2016).

89 bendeOO R. E. VPe cOassiZcation oZ depressionse a revieX oZ contemporary conZusion. *Griti7h G=urnal =DP78chiatr8* 548, 15-28 (1976).

90 Cross-Disorder Group oZ tPe PsycPiatric Genomics Consortium. UdentiZcation oZ risR Oci XitP sPared eZZects on Zve maWr psycPiatric disorderse a Senome-Xide anaOysis. *9ancet* / 75, 1371-1379 (2013).

91 WorOd HeaOtP [rSaniOation. *Rnternati=nal Fla77iDcati=n =D Ei7ea7e7*. 9tP revised edn, (WorOd HeaOtP [rSaniOation, 1978).

92 WorOd HeaOtP [rSaniOation. *Rnternati=nal Fla77iDcati=n =D Ei7ea7e7*. 10tP revised edn, (WorOd HeaOtP [rSaniOation, 1992).

93 American PsycPiatric Association. *Eiagn=7tic anB @tati7tical A anual =D A ental Ei7=rBer7*. \ourtP Edition edn, (American PsycPiatric Association, 1994).

94 BuOR-SuOOvan, B. b. *et al*. ^D Score reSression distinSuisPes conZoundinS Zrom poOySenicity in Senome-Xide association studies. *<ature Genetic7* 03, 291-295 (2015).

95 Durbin, R. M. *et al*. A map oZ Puman Senome variation Zrom popuOation-scaOe se] uencinS. *<ature* 023, 1061-1073 (2010).

96 Sanders, A. R. *et al*. VPe Unternet-based MGS2 controO sampOe seOZ report oZ mentaO iOOness. *Nhe A? erican S=urnal =D>78chiatr8* 523, 854-865, doie10.1176f appi.aWp.2010.09071050 (2010).

97 WVCCC. Genome-Xide association study oZ 14,000 cases oZ seven common diseases and 3,000 sPared controOs. *<ature* 003, 661-678 (2007).

98 \ ranRe, B. *et al*. Genetic inZOuences on scPiOopPrenia and subcorticaO brain voOumese OarSe-scaOe prooZ oZ concept. *<at <eur=7ci* 58, 420-431, doie10.1038f nn.4228 (2016).

99 Price, A. ^. *et al*. PrincipaO components anaOysis corrects Zor stratiZcation in Senome-Xide association studies. *<at Genet* / 7, 904-909 (2006).

100 BeSum, \ ., GPosP, D., VsenS, G. C. c \ einSoOd, E. ComprePensive Oiterature revieX and statisticaO considerations Zor GWAS meta-anaOysis. *<ucleic aciB7 re7earch* 06, 3777-3784, doie10.1093f narf SRr1255 (2012).

101 1000 Genomes ProWect Consortium *et al*. A SOobaO reZerence Zor Puman Senetic variation. *<ature* 142, 68-74, doie10.1038f nature15393 (2015).

102 PoXer, R. A. *et al*. Genome-Xide Association Zor MaWr Depression VProuSP ASe at [nset StratiZcatione MaWr Depressive Disorder WorRnS Group oZ tPe PsycPiatric Genomics Consortium. *Gi=l P78chiatr8* 75, 325-335, doie10.1016f WbiopsycP.2016.05.010 (2017).

103 gerduiW, d *et al*. asinS COinicaO CParacteristics to UdentiZy WPicP Patients WitP MaWr Depressive Disorder Have a HiSPer Genetic ^oad Zor VPree PsycPiatric Disorders. *Gi=l P78chiatr8* 75, 316-324, doie10.1016f WbiopsycP.2016.05.024 (2017).

104 Encode ProWect Consortium. A useris Suide to tPe encycOopedia oZ DYA eOements (EYC[DE). *P9=@ bi=l=g8* 8, e1001046, doie10.1371f WournaO pbio.1001046 (2011).

105 Roadmap EpiSenomics Consortium *et al*: UnteSrative anaQysis oZ 111 reZerence Puman epiSenomes. <*ature* 157, 317-330, doie10.1038fnature14248 (2015).

106 gernot, B. *et al*: ExcavatinS YeandertaQand Denisovan DYA Zrom tPe Senomes oZ MeQanesian individuaQs. @*ience* / 14, 235-239, doie10.1126fscience.aad9416 (2016).

107 Bryois, d *et al*: EvaQuation oZCPromatin Accessibi0ty in PreZrontaQCortex oZScPiQopPrenia Cases and ControQs. (Submitted).

108 hPanS, Y. *et al*: ModeQbased anaQysis oZCPUP-Se] (MACS). *Gen*=? *e* G=*l* 8, R137, doie10.1186fSb-2008-9-9-r137 (2008).

109 Ross-Unes, C. S. *et al*: DiZZerentiaQoestroSen receptor bindinS is associated XitP cQinicaQoutcome in breast cancer. <*ature* 075, 389-393, doie10.1038fnature10730 (2012).

110 \inucane, H. *et al*: Heritabi0ty enricPment oZ speciZicaQy expressed Senes identiZies disease-reQevant tissues and ceQ0types. doie10.1101f103069 (Submitted).

111 hPernaPova, D. g. *et al*: UdentiZication oZ context-dependent expression] uantitative trait Qoci in XPoQe boQod. <*at Genet* 08, 139-145, doie10.1038fnS.3737 (2017).

112 dansen, R. *et al*: ConditionaQek V^ anaQysis reveaQs aQQQ PeteroSeneity oZ Sene expression. *Hu*? A=*l Genet* 42, 1444-1451, doie10.1093fPmSfddx043 (2017).

113 ^appaQainen, V. *et al*: Vranscriptome and Senome se] uencinS uncovers ZunctionaQvariation in Pumans. <*ature* 165, 506-511, doie10.1038fnature12531 (2013).

114 de ^eeuX, C. A., YeaQ, B. M., HesPes, V. c PostPuma, D. VPe statisticaQproperties oZ Sene-set anaQysis. <*at Rev Genet*, doie10.1038fnrS.2016.29 (2016).

115 de ^eeuX, C. A., MooiWd M., HesPes, V. c PostPuma, D. MAGMAeSeneraQQed Sene-set anaQysis oZ GWAS data. *P9=@F=*? >*ut* G=*l* 55, e1004219, doie10.1371fWournaQpcbi.1004219 (2015).

116 Vurner, V. Y. *et al*: denovo-dbea compendium oZ Puman de novo variants. <*ucleic aci*B7 *re7earch* 01, D804-D811, doie10.1093fnarfSPX865 (2017).

117 PirooQnia, M. *et al*: HiSP-tProuSPput se] uencinS oZ tPe synaptome in maWr depressive disorder. A=*l P78chiatr*8 45, 650-655, doie10.1038fmp.2015.98 (2016).

118 ^iberQon, A. *et al*: VPe MoeQcuQar SiSnatures Database (MSiSDB) PaQQmarR Sene set coQQection. F*ell @87t* 5, 417-425, doie10.1016fWeQ.2015.12.004 (2015).

119 WaSner, A. H. *et al*: DGUdb 2.0e mininS cQinicaQy reQevant druS-Sene interactions. <*ucleic aci*B7 *re7earch* 00, D1036-1044, doie10.1093fnarfSPv1165 (2016).

120 RotP, B. ^., broeQe, W. b., PateQ, S. c ^opeQ E. VPe MuQtiplQicity oZ Serotonin Receptors easeQessQy diverse moeQcuQes or an embarrasment oZ ricPesl N*he* <*eur*=7*cienti*7*t* 2, 252-262 (2000).

121 SmitP, G. D. c EbraPim, S. iMendeQian randomiQationie can Senetic epidemioQoSy contribute to understandinS environmentaQdeterminants oZ diseasel P*t* C; >*i*B*e*? *i*=*l* / 4, 1-22 (2003).

122 WooQridSe, d j . P*tr*=B*uct*=*r*8 ; *c*=*n*=? *etric*7UA ? =B*ern a*>>*r*=*ach*. (YeQon Education, 2015).

123 BoXden, d, Davey SmitP, G. c BurSess, S. MendeQian randomiQation XitP invaQd instrumentse eZZect estimation and bias detection tProuSP ESSer reSression. P*t* C; >*i*B*e*? *i*=*l* 00, 512-525, doie10.1093fije fdyv080 (2015).

124 Unamura, M. *et al*: Genome-Xide association studies in tPe dapanese popuQation identiZy seven noveOOci Zor type 2 diabetes. <*at* F=? ? *un* 3, 10531, doie10.1038fncomms10531 (2016).

125 ^iu, d h. *et al*: Association anaQyses identiZy 38 susceptibiOy Oci Zor inZammatory boXeO disease and PiSPOSPt sPared Senetic risR across popuQations. <*at Genet* 03, 979-986, doie10.1038fnS.3359 (2015).

126 YanS, ^. *et al*: PoQySenic transmission and compOex neuro deveOopmentaOnetXorR Zor attention deZicit Pyperactivity disordereSenome-Xide association study oZbotP common and rare variants. *A*? CA *e*B *Genet* G<*eur*=>78*chiatr Genet* 5249, 419-430, doie10.1002faWmS.b.32169 (2013).

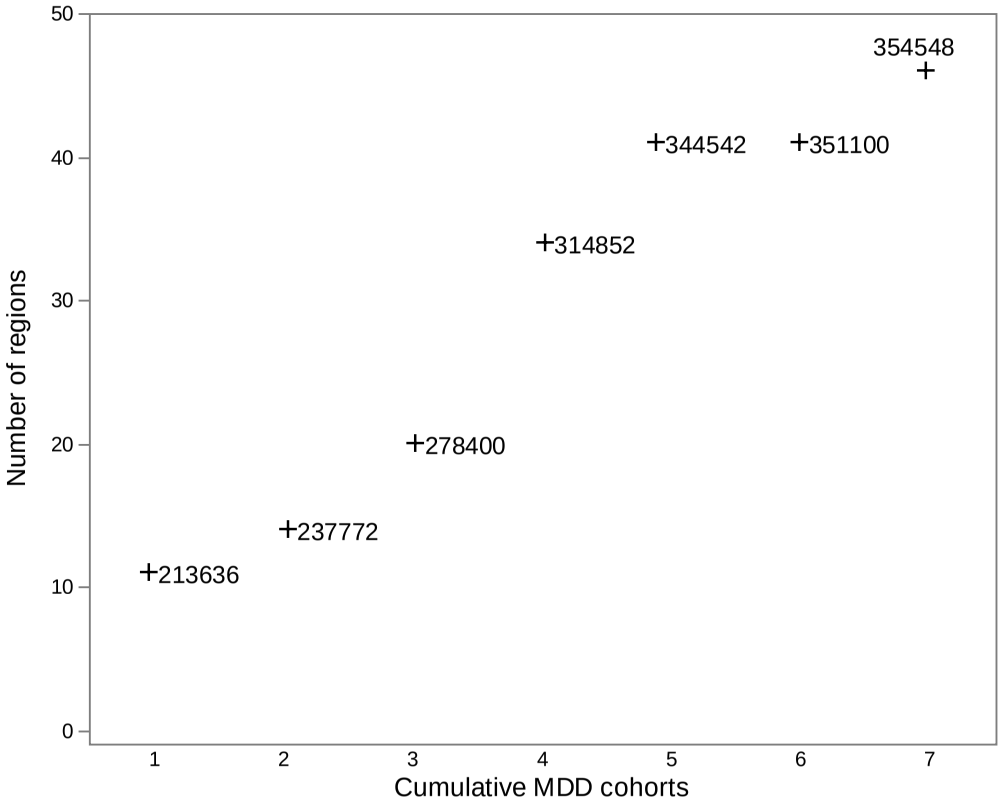
127 BroXn, B. C., Asian Genetic EpidemioOSy YetXorR-Vype 2 Diabetes, Ye, C. d, Price, A. ^. c haitOn, Y. VransetPnic Senetic correQation estimates Zrom summary statistics. *A*? C*Hu*? *Genet* 88, 76-88 (2016).

128 Peterson, R. E. *et al*: VPe Genetic ArcPitecture oZ MaWr Depressive Disorder in Han CPinese Women. O2AA *A P*78*chiatr*830, 162-168, doie10.1001fWamapsycPiatry.2016.3578 (2017).

129 BiSdeO, V. B. *et al*: Genetic eZZects inZOencinS risR Zor maWr depressive disorder in CPina and Europe. N*ran*7*l P*78*chiatr*83, e1074, doie10.1038ftp.2016.292 (2017).

Figure 1a,b

(a)



(b)

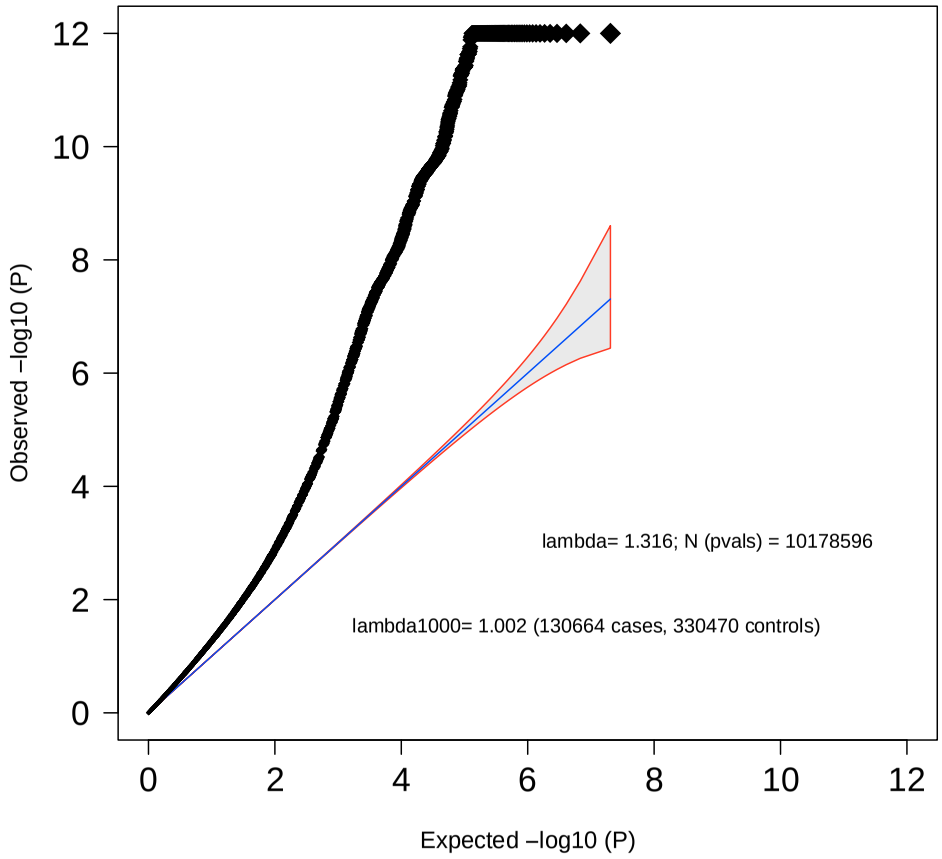


Figure 1c

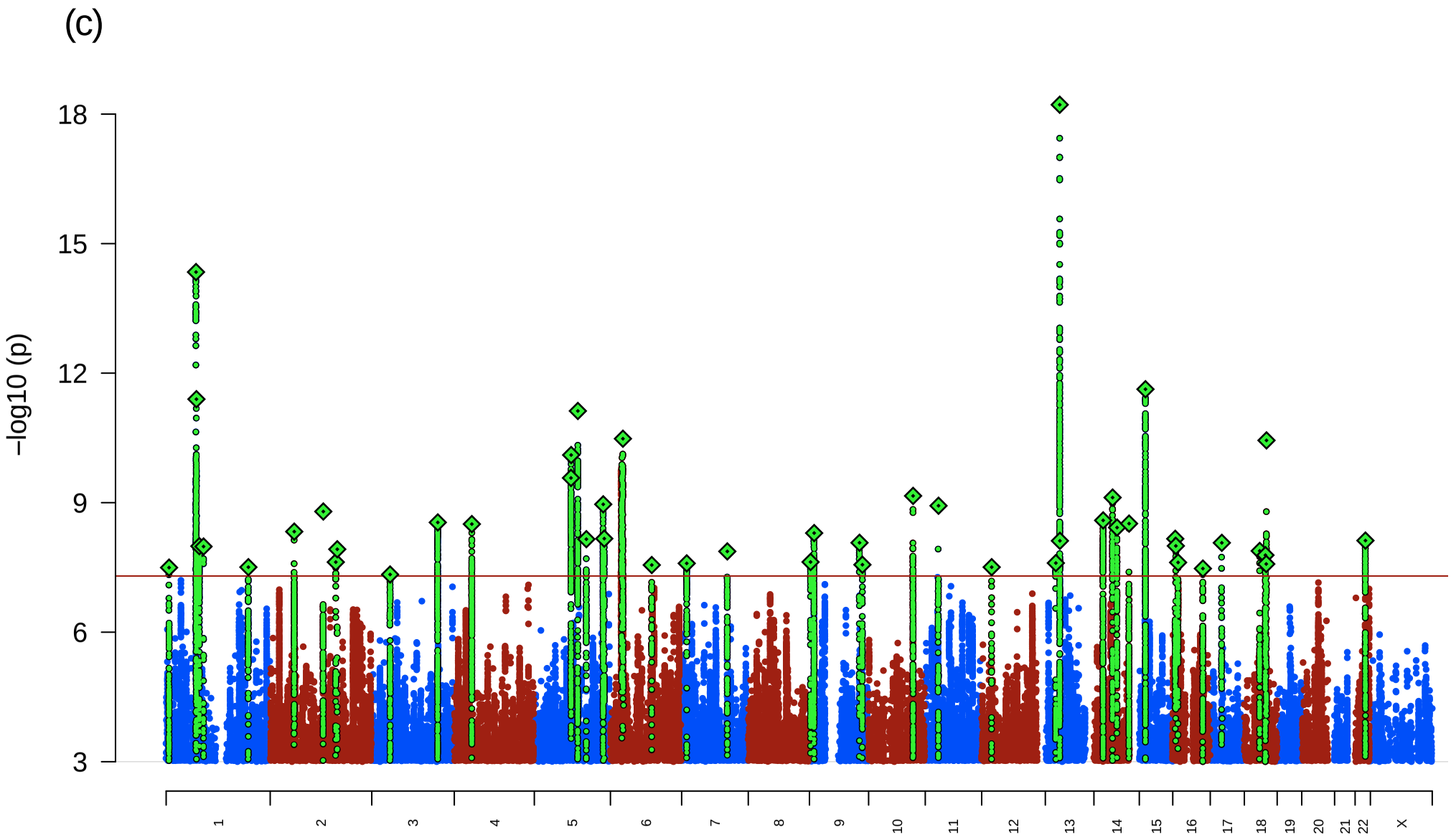

(c)
−log10 (p)
18
15
12
9
6
3
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X

Figure 2

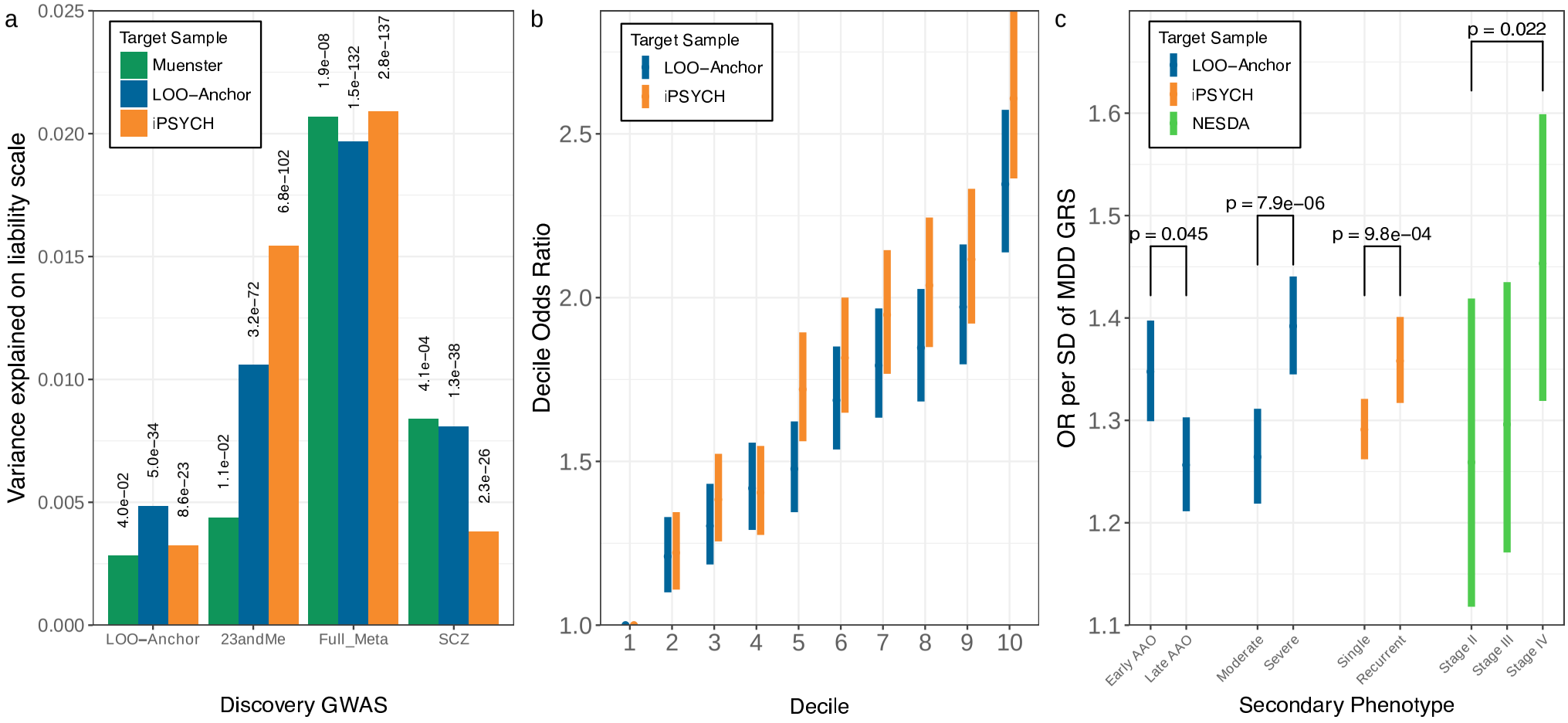

a
Target Sample
Muenster
LOO–Anchor
iPSYCH
Variance explained on liability scale
0.025
0.020
0.015
0.010
0.005
0.000
4.0e–02
5.0e–34
8.6e–23
1.1e–02
3.2e–72
6.8e–102
1.9e–08
1.5e–132
2.8e–137
4.1e–04
1.3e–38
2.3e–26
LOO–Anchor
23andMe
Full_Meta
SCZ
Discovery GWAS
b
Target Sample
LOO–Anchor
iPSYCH
Decile Odds Ratio
2.5
2.0
1.5
1.0
1
2
3
4
5
6
7
8
9
10
Decile
c
Target Sample
LOO–Anchor
iPSYCH
NESDA
OR per SD of MDD GRS
1.6
1.5
1.4
1.3
1.2
1.1
p = 0.045
p = 7.9e–06
p = 9.8e–04
p = 0.022
Early AAO
Late AAO
Moderate
Severe
Single
Recurrent
Stage II
Stage III
Stage IV
Secondary Phenotype

Figure 3a

(a)

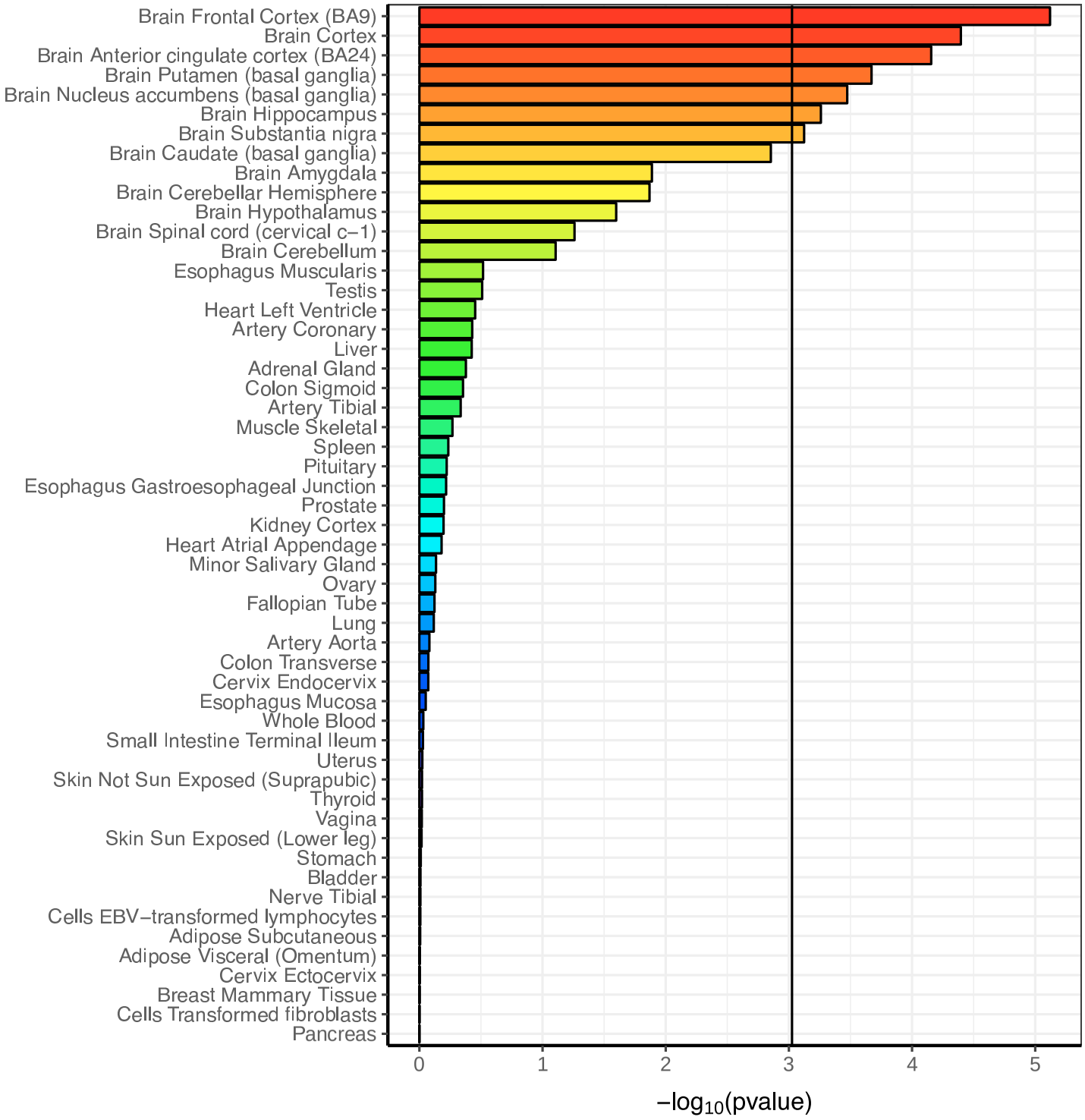

Brain Frontal Cortex (BA9)
Brain Cortex
Brain Anterior cingulate cortex (BA24)
Brain Putamen (basal ganglia)
Brain Nucleus accumbens (basal ganglia)
Brain Hippocampus
Brain Substantia nigra
Brain Caudate (basal ganglia)
Brain Amygdala
Brain Cerebellar Hemisphere
Brain Hypothalamus
Brain Spinal cord (cervical c−1)
Brain Cerebellum
Esophagus Muscularis
Testis
Heart Left Ventricle
Artery Coronary
Liver
Adrenal Gland
Colon Sigmoid
Artery Tibial
Muscle Skeletal
Spleen
Pituitary
Esophagus Gastroesophageal Junction
Prostate
Kidney Cortex
Heart Atrial Appendage
Minor Salivary Gland
Ovary
Fallopian Tube
Lung
Artery Aorta
Colon Transverse
Cervix Endocervix
Esophagus Mucosa
Whole Blood
Small Intestine Terminal Ileum
Uterus
Skin Not Sun Exposed (Suprapubic)
Thyroid
Vagina
Skin Sun Exposed (Lower leg)
Stomach
Bladder
Nerve Tibial
Cells EBV−transformed lymphocytes
Adipose Subcutaneous
Adipose Visceral (Omentum)
Cervix Ectocervix
Breast Mammary Tissue
Cells Transformed fibroblasts
Pancreas
0
1
2
3
4
5
−log10(pvalue)

Figure 3b

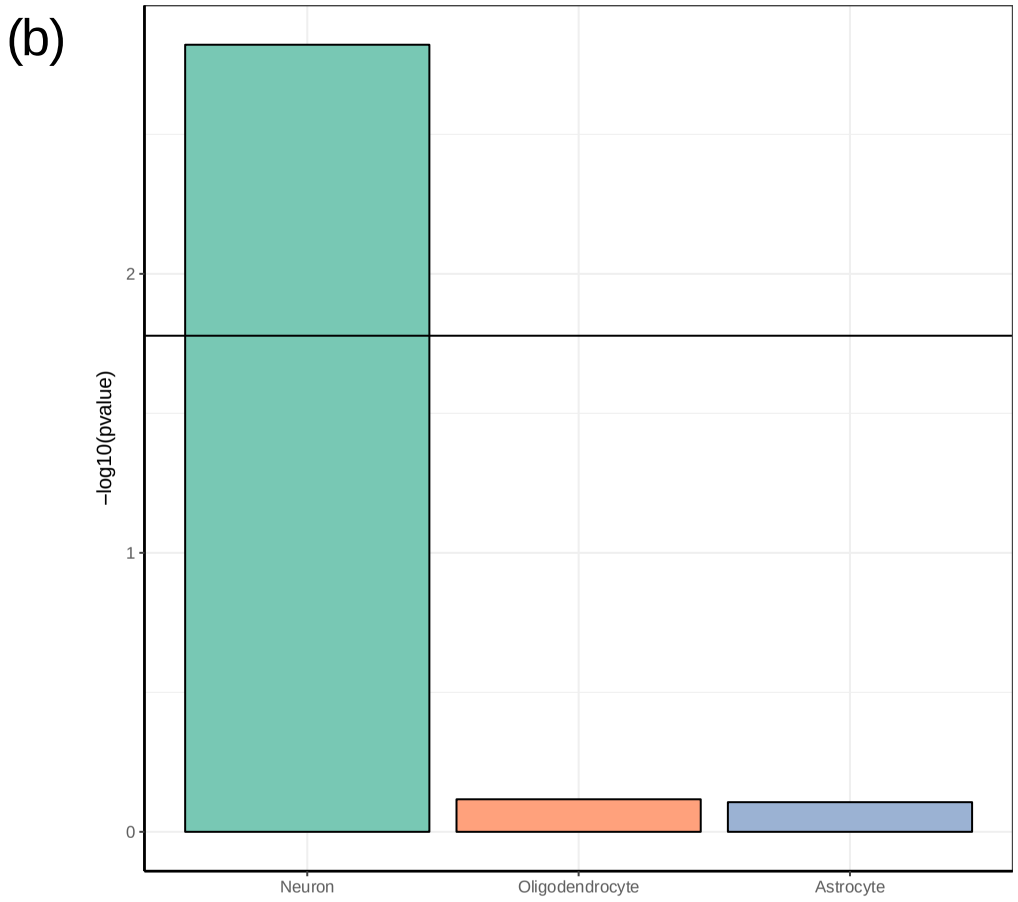


Figure 3c

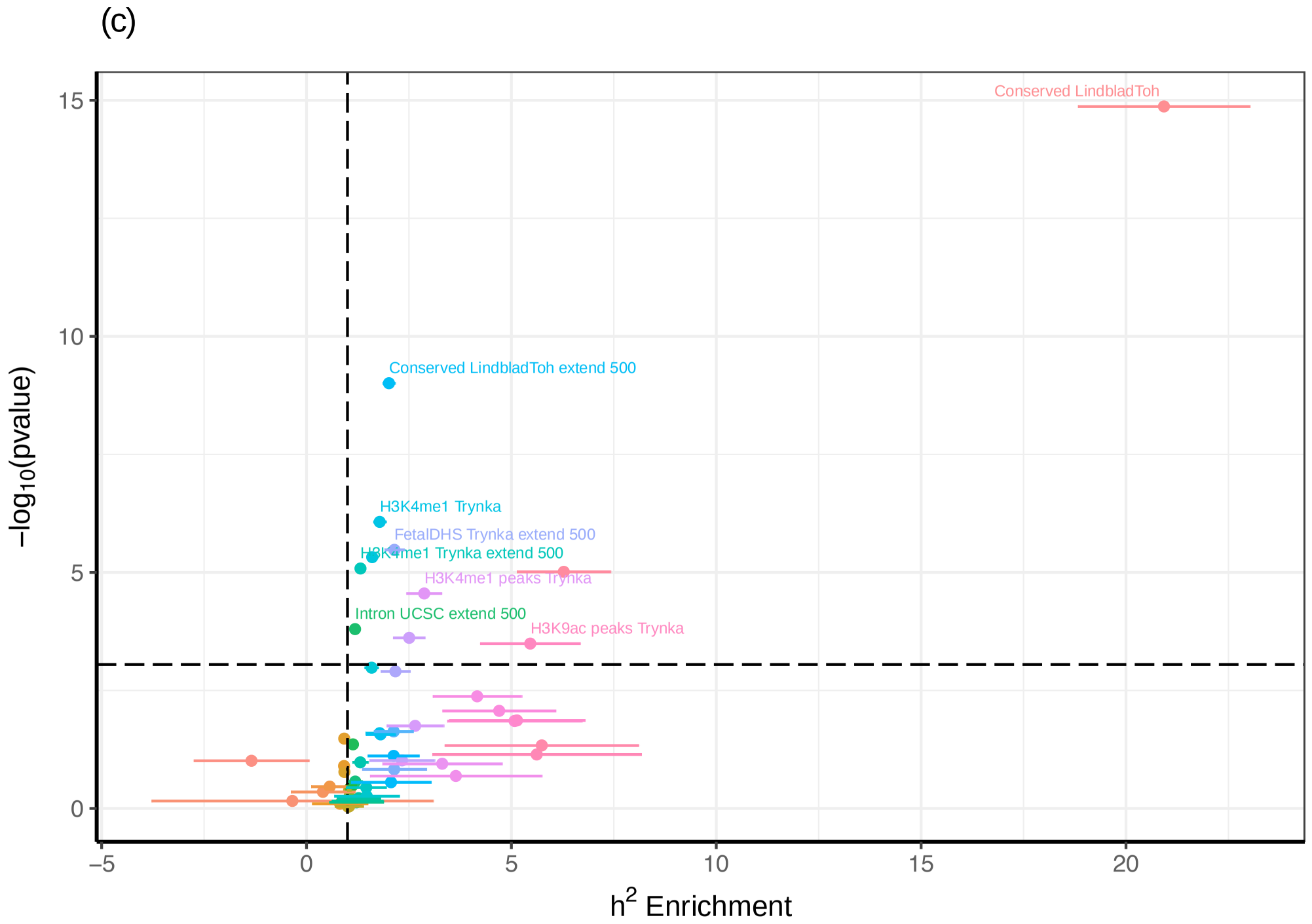
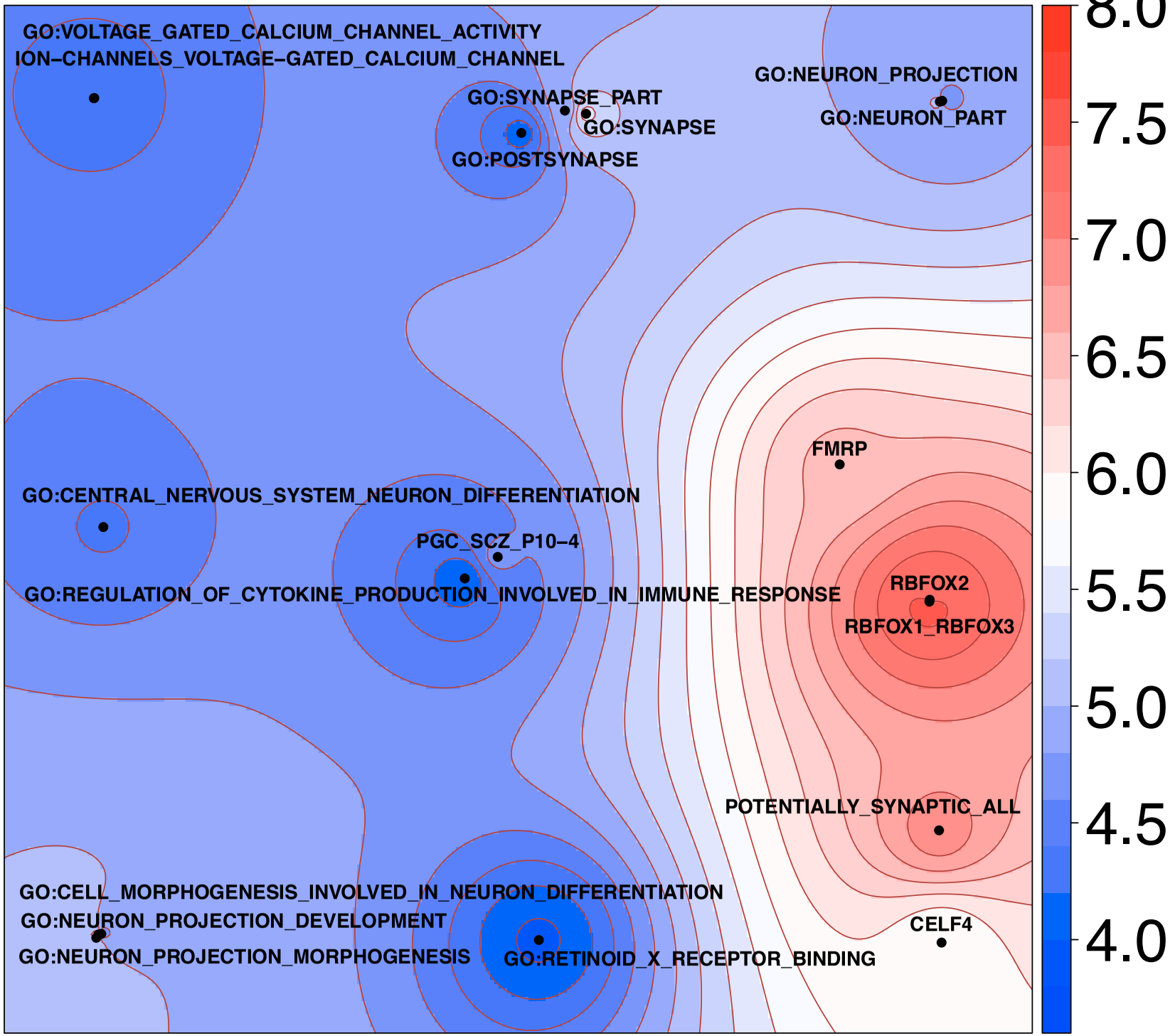

(c)
Conserved LindbladToh
Conserved LindbladToh extend 500
H3K4me1 Trynka
FetalDHS Trynka extend 500
H3K4me1 Trynka extend 500
H3K4me1 peaks Trynka
Intron UCSC extend 500
H3K9ac peaks Trynka
15
10
5
0
−5
0
5
10
15
20
−log10(pvalue)
h2 Enrichment

Figure 4


GO:VOLTAGE_GATED_CALCIUM_CHANNEL_ACTIVITY
ION–CHANNELS_VOLTAGE–GATED_CALCIUM_CHANNEL
GO:SYNAPSE_PART
GO:SYNAPSE
GO:POSTSYNAPSE
GO:NEURON_PROJECTION
GO:NEURON_PART
FMRP
GO:CENTRAL_NERVOUS_SYSTEM_NEURON_DIFFERENTIATION
PGC_SCZ_P10–4
GO:REGULATION_OF_CYTOKINE_PRODUCTION_INVOLVED_IN_IMMUNE_RESPONSE
RBFOX2
RBFOX1_RBFOX3
POTENTIALLY_SYNAPTIC_ALL
GO:CELL_MORPHOGENESIS_INVOLVED_IN_NEURON_DIFFERENTIATION
GO:NEURON_PROJECTION_DEVELOPMENT
GO:NEURON_PROJECTION_MORPHOGENESIS
GO:RETINOID_X_RECEPTOR_BINDING
CELF4
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0