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

Major depressive disorder (MDD) is a common illness accompanied by considerable morbidity, mortality, costs, and heightened risk of suicide. We conducted a genome-wide association meta-analysis based in 135,458 cases and 344,901 controls and identified 44 independent and significant loci. The genetic findings were associated with clinical features of major depression and implicated brain regions exhibiting anatomical differences in cases. Targets of antidepressant medications and genes involved in gene splicing were enriched for smaller association signal. We found important relationships of genetic risk for major depression with educational attainment, body mass, and schizophrenia: lower educational attainment and higher body mass were putatively causal, whereas major depression and schizophrenia reflected a partly shared biological etiology. All humans carry lesser or greater numbers of genetic risk factors for major depression. These findings help refine the basis of major depression and imply that a continuous measure of risk underlies the clinical phenotype.

MDD is a notably complex and common illness[1]. It is often chronic or recurrent and is thus accompanied by considerable morbidity, disability, excess mortality, substantial costs, and heightened risk of suicide[2–8]. Twin studies attribute approximately 40% of the variation in liability to MDD to additive genetic effects (phenotype heritability, h^2)[9], and h^2 may be greater for recurrent, early-onset, and postpartum MDD[10,11]. Genome-wide association studies (GWAS) of MDD have had notable difficulties in identifying individual associated loci[12]. For example, there were no significant findings in the initial Psychiatric Genomics Consortium (PGC) MDD mega-analysis (9,240 cases)[13] or in the CHARGE meta-analysis of depressive symptoms ($n = 34,549$)[14]. More recent studies have proven modestly successful. A study of Han Chinese women (5,303 recurrent MDD cases) identified significant loci[15], a meta-analysis of depressive symptoms (161,460 individuals) identified 2 loci[16], and an analysis of self-reported major depression identified 15 loci (75,607 cases).

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[17], as are most common diseases[18], including psychiatric disorders[19,20]. Estimates of the proportion of variance attributable to genome-wide SNPs (SNP heritability, h^2_{SNP}) indicate that around one-quarter of the h^2 for MDD is due to common genetic variants[21,22] and demonstrate that a genetic signal is detectable in genome-wide association data, implying that larger sample sizes are needed to detect specific loci given their effect sizes. Such a strategy has been proven in studies of schizophrenia, the flagship adult psychiatric disorder in genomics research. We thus accumulated clinical, population, and volunteer cohorts[23]. This pragmatic approach takes the view that sample size can overcome heterogeneity to identify risk alleles that are robustly associated with major depression. Potential concerns about combining carefully curated research cohorts with volunteer cohorts were considered; multiple lines of evidence suggest that the results are likely to be applicable to clinical MDD. As discussed below, our analyses have neurobiological, clinical, and therapeutic relevance for major depression.
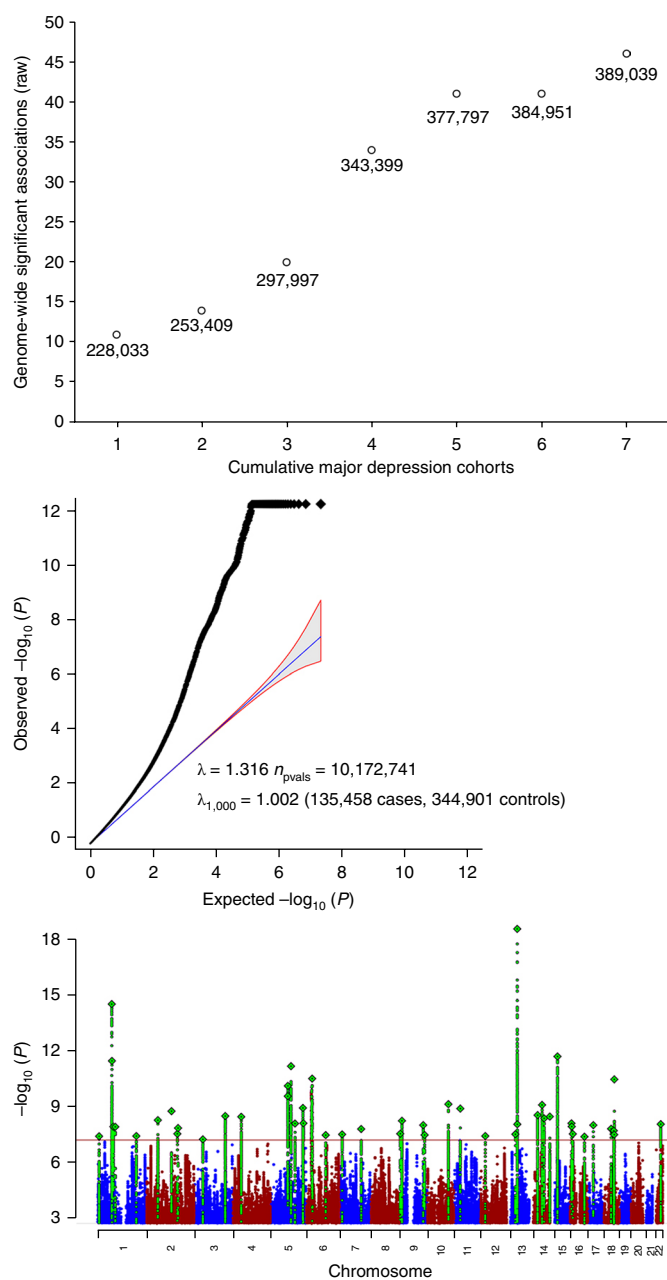
Results

Cohort analyses: phenotype validation. We identified seven cohorts that used a range of methods to ascertain cases with major depression (Table 1 and Supplementary Tables 1–3). The methods used by these cohorts were thoroughly reviewed, drawing on the breadth of expertise in the PGC, and we assessed the comparability of the cohorts using genomic data. We use 'MDD' to refer to directly evaluated subjects meeting standard criteria for major depressive disorder and use 'major depression' where case status was determined using alternative methods as well as to the phenotype from the full meta-analysis.

We evaluated the comparability of the seven cohorts by estimating the common variant genetic correlations (r_g) between them. These analyses supported the comparability of the seven cohorts (Supplementary Table 3), as the weighted mean r_g was 0.76 (s.e. = 0.03). The high genetic correlations between the 23andMe and other cohorts are notable. While there was no statistical evidence of heterogeneity in the r_g estimates for pairs of cohorts ($P = 0.13$), the estimate was statistically different from 1, which may reflect etiological heterogeneity. This estimate can be benchmarked against the slightly larger weighted mean r_g between schizophrenia cohorts of 0.84 (s.e. = 0.05)[21].

Given the positive evidence of the genetic comparability of these cohorts, we completed a genome-wide association meta-analysis of 9.6 million imputed SNPs in 135,458 MDD and major depression cases and 344,901 controls (Fig. 1). There was no evidence of residual population stratification[24] (LD score regression intercept = 1.018, s.e. = 0.009). We estimated h^2_{SNP} to be 8.7% (s.e. = 0.004, liability scale, assuming lifetime risk of 0.15; Supplementary Fig. 1 and Supplementary Table 3b), and note that this is about one-quarter of the h^2 estimated from twin or family studies[9]. This fraction is somewhat lower than that of other complex traits[18] and is plausibly due to etiological heterogeneity (reflecting the mean $r_g < 1$ between cohorts).

To evaluate the impact of combining major depression cohorts that used different ascertainment methods, we undertook a series of genetic risk score (GRS) prediction analyses to demonstrate the validity of our genome-wide association results for clinical MDD (Fig. 2). Notably, the variance explained in out-of-sample prediction increased with the size of the genome-wide association discovery cohort (Fig. 2a), with the GRS from the full discovery

Fig. 1 | Results of genome-wide association meta-analysis of seven cohorts for major depression. a, Relationship between adding cohorts and the number of genome-wide significant genomic regions (before the vetting used to define the final 44 regions). Beginning with the largest cohort (cohort 1 on the *x* axis), we added the next largest cohort (cohort 2) until all cohorts were included (7 cohorts). The number next to each point is the total effective sample size, equivalent to the sample size where the numbers of cases and controls are equal. **b**, Association test quantile–quantile plot showing a marked departure from the null model of no associations (*y* axis truncated at 10^{-12}). **c**, Manhattan plot for association tests from meta-analysis of 135,458 major depression cases and 344,901 controls, with the *x* axis showing genomic position (chromosomes 1–22 plus the X chromosome) and the *y* axis showing statistical significance as $-\log_{10}(P)$ *z* statistics; the threshold for significance accounting for multiple testing is shown by the red horizontal line ($P = 5 \times 10^{-8}$).

sample meta-analysis explaining 1.9% of variance in liability (Fig. 2a, Supplementary Fig. 2, and Supplementary Table 4). For any randomly selected case and control, GRS ranked cases higher

than controls with probability of 0.57 and the odds ratio of MDD for those in the tenth versus first GRS decile (OR = 10) was 2.4 (Fig. 2b and Supplementary Table 4). GRS analyses in other disorders (for example, schizophrenia[25]) have shown that the mean GRS increases with clinical severity in cases. We found a significantly higher major depression GRS in those with more severe MDD, as measured in different ways (Fig. 2c). Last, because around one-half of the major depression cases were identified by self-report (i.e., diagnosis or treatment for clinical depression by a medical professional), we further evaluated the comparability of the 23andMe cohort with the other cohorts (full meta-analysis excluding 23andMe, 'FMex23andMe') as detailed in Fig. 2c, Supplementary Table 5, and the Supplementary Note. Taken together, we interpret these results as supporting this meta-analysis of GWAS for these seven cohorts.

Implications for the biology of major depression. Our meta-analysis of seven MDD and major depression cohorts identified 44 independent loci that were statistically significant ($P < 5 \times 10^{-8}$), statistically independent of any other signal[26], and supported by multiple SNPs. This number supports our prediction that genome-wide association discovery in major depression would require about five times more cases than for schizophrenia (lifetime risk ~1% and h^2 ~0.8) to achieve approximately similar power[27]. Of these 44 loci, 30 are new and 14 were significant in a prior study of MDD or depressive symptoms. The overlap of our findings with prior reports was as follows: 1 of 1 with CHARGE depressive symptom[14], 1 of 2 overlap with SSGAC depressive symptom[16], and 12 of 15 overlap with Hyde et al.[28]. There are few trans-ancestry comparisons for major depression, so we contrasted these European results with the Han Chinese CONVERGE study[15] (Supplementary Note). The loci identified in CONVERGE are uncommon in Europeans (rs12415800, 0.45 versus 0.02; rs35936514, 0.28 versus 0.06) and were not significant in our analysis.

Table 2 lists genes in or near the lead SNP in each region, regional plots are in Supplementary Data 1, and Supplementary Tables 6 and 7 provide summaries of available information about the biological functions of the genes in each region. In the Supplementary Note, we review four key genes in more detail: *OLFM4* and *NEGR1* (notable for reported associations with obesity and body mass index[29–34]), *RBFOX1* (notable for independent associations at both the 5′ and 3′ ends, a splicing regulator[35,36], with a functional role that may be consistent with chronic hypothalamic–pituitary–adrenal axis hyperactivation reported in MDD[37]), and *LRFN5* (notable for its role in presynaptic differentiation[38,39] and neuroinflammation[40]).

Gene-wise analyses identified 153 significant genes after controlling for multiple comparisons (Supplementary Table 7). Many of these genes were in the extended major histocompatibility complex (MHC) region (45 of 153), and their interpretation is complicated by high linkage disequilibrium (LD) and gene density. In addition to the genes discussed above, other notable and significant genes outside of the MHC region 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 major depression loci with 108 loci for schizophrenia[19] identified 6 shared loci. Many SNPs in the extended MHC region are strongly associated with schizophrenia, but implication of the MHC region is new for major depression. Another example is *TCF4* (transcription factor 4), which is strongly associated with schizophrenia but was not previously associated with MDD. TCF4 is essential for normal brain development, and rare mutations in *TCF4* cause Pitt–Hopkins syndrome, which includes autistic features[41]. The GRS calculated from the schizophrenia

Table 1 | Brief description of the seven MDD or major depression cohorts used in the meta-analysis

Sample	Country	Case ascertainment	Cases	Controls
PGC29[13,a]	Various	Structured diagnostic interviews[b]	16,823	25,632
deCODE[13]	Iceland	National inpatient electronic records	1,980	9,536
GenScotland[78,79]	UK	Structured diagnostic interview	997	6,358
GERA[80]	USA	Kaiser Permanente Northern California Healthcare electronic medical records (1995–2013)	7,162	38,307
iPSYCH[81]	Denmark	National inpatient electronic records	18,629	17,841
UK Biobank[82] (pilot data release)	UK	From self-reported MDD symptoms or treatment or electronic records[69]	14,260	15,480
23andMe[28] (discovery sample)[c]	USA	Self-reported diagnosis or treatment for clinical depression by a medical professional	75,607	231,747
Total			135,458	344,901

[a]Nineteen samples in addition to the ten samples published in the first PGC-MDD paper[13]. [b]One sample used natural language processing of electronic medical records followed by expert diagnostic review. [c]In Hyde et al.[28], SNPs in 15 genomic regions met genome-wide significance in the combined discovery and replication samples and 11 regions achieved genome-wide significance in the discovery sample made available to the research community and used here. More details are provided in Supplementary Tables 1–3.

genome-wide association results explained 0.8% of the variance in liability of MDD (Fig. 2c).

Implications from integration of functional genomic data. Results from 'omic' studies of functional features of cells and tissues are necessary to understand the biological implications of the results of GWAS for complex disorders[42]. To further elucidate the biological relevance of the major depression findings, we integrated the results with functional genomic data. First, using enrichment analyses, we compared the major depression GWAS findings to bulk tissue mRNA-seq from Genotype-Tissue Expression (GTEx) data[43]. Only brain samples showed significant enrichment (Fig. 3a), 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 of these regions were implicated in a large meta-analysis of brain magnetic resonance imaging (MRI) findings in adult MDD cases[44]. Second, given the predominance of neurons in the cortex, we confirmed that the major depression genetic findings connect to genes expressed in neurons but not oligodendrocytes or astrocytes (Fig. 3b)[45]. Given the different methods used by the seven MDD/major depression cohorts in this study, demonstration of enrichment of association signals in the brain regions expected to be most relevant to MDD provides independent support for the validity of our approach.

Third, we used partitioned LD score regression[46] to evaluate the enrichment of the major depression genome-wide association findings in over 50 functional genomic annotations (Fig. 3c and Supplementary Table 8). The major finding was the significant enrichment of h^2_{SNP} in genomic regions conserved across 29 Eutherian mammals[47] (20.9-fold enrichment, $P = 1.4 \times 10^{-15}$). This annotation was also the most enriched for schizophrenia[46]. We could not evaluate regions conserved in primates or human 'accelerated' regions, as there were too few for confident evaluation[47]. The other 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[17], genetic variants that change exonic sequences may not have a large role in major depression. We found no evidence that Neanderthal introgressed regions were enriched for major depression genome-wide association findings[48].

Fourth, we applied methods to integrate genome-wide association SNP results with those from gene expression and methylation quantitative trait locus (eQTL and meQTL) studies. SMR[49] analysis identified 13 major depression–associated SNPs with strong evidence that they control local gene expression in one or more tissues

and 9 with strong evidence that they control local DNA methylation (Supplementary Table 9 and Supplementary Data 2). A transcriptome-wide association study[50] applied to data from the dorsolateral prefrontal cortex[51] identified 17 genes where major depression–associated SNPs influenced gene expression (Supplementary Table 10). 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 genome-wide association finding to a specific gene that may be nearby or hundreds of kilobases away[52–54]. We integrated the major depression results with 'easy Hi-C' data from brain cortical samples (3 adult, 3 fetal, >1 billion reads each). These data clarified three associations. The statistically independent associations in *NEGR1* (rs1432639, $P = 4.6 \times 10^{-15}$) and over 200 kb away (rs12129573, $P = 4.0 \times 10^{-12}$) both implicate *NEGR1* (Supplementary Fig. 3a), the former likely due to the presence of a reportedly functional copy number polymorphism (see Supplementary Note) 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, $P = 6.9 \times 10^{-9}$ and 1.0×10^{-8}) had only intragenic associations, suggesting that the genetic variation in the regions of the major depression associations act locally and can be assigned to these genes. In contrast, the association in *RERE* (rs159963, $P = 3.2 \times 10^{-8}$) could not be assigned to *RERE* as it may contain super-enhancer elements given its many DNA–DNA interactions with many nearby genes (Supplementary Fig. 3b).

Implications based on the roles of sets of genes. A parsimonious explanation for the presence of many significant associations for a complex trait is that the different associations are part of a higher-order grouping of genes[55]. These could be a biological pathway or a collection of genes with a functional connection. Multiple methods allow evaluation of the connection of major depression genome-wide association results to sets of genes grouped by empirical or predicted function (pathway or gene set analysis).

Full pathway analyses are in Supplementary Table 11, and 19 pathways with false discovery rate q values <0.05 are summarized in Fig. 4. The major groupings of significant pathways were as follows: RBFOX1, RBFOX2, RBFOX3, or CELF4 regulatory networks; genes whose mRNAs are bound by FMRP; synaptic genes; genes
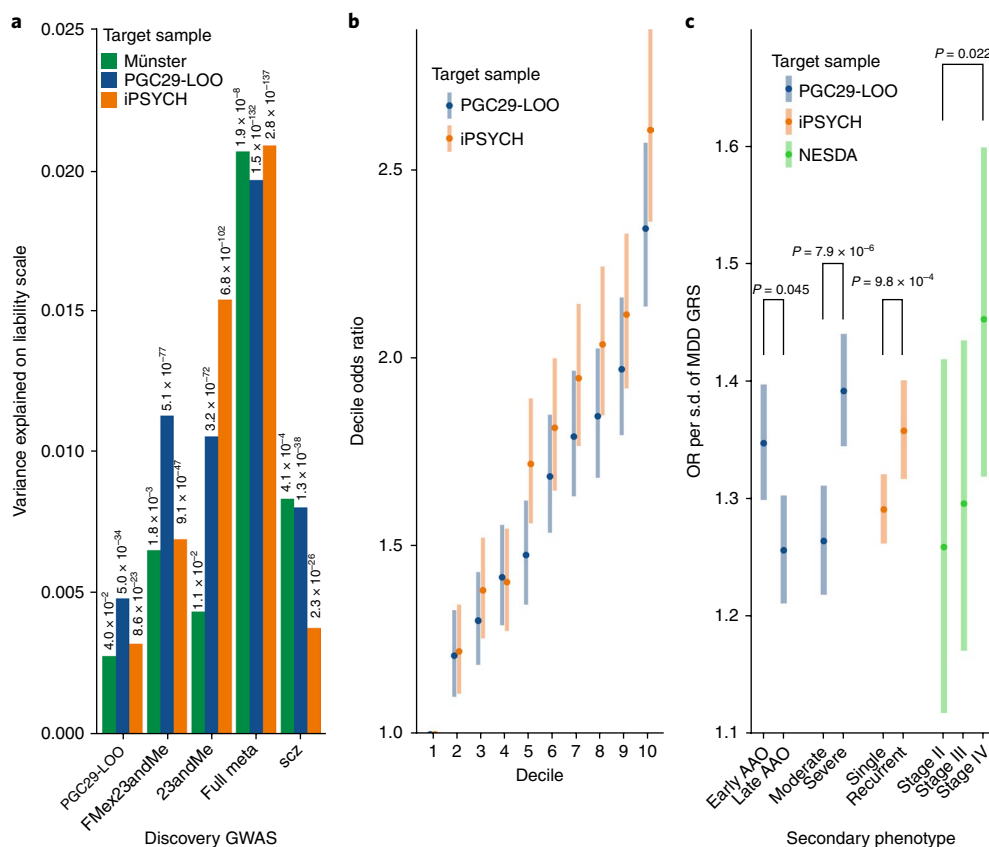
Fig. 2 | Genetic risk score prediction analyses into PGC29 MDD target samples. a, Variance explained (liability scale) based on different discovery samples for three target samples: PGC29 (16,823 cases, 25,632 controls), iPSYCH (a nationally representative sample of 18,629 cases and 17,841 controls), and a clinical cohort from Münster not included in the genome-wide association analysis (845 MDD inpatient cases, 834 controls). For PGC29-LOO, the target sample was one of the PGC29 samples with the discovery sample being the remaining 28 PGC29 samples, hence representing leave-one-out (LOO) analysis. **b,** Odds ratios of major depression per GRS decile relative to the first decile for the iPSYCH and PGC29 target samples. **c,** Odds ratios of major depression in GRS s.d.: PGC29-LOO, 3,950 early-onset versus 3,950 late-onset cases; PGC29-LOO, 4,958 severe versus 3,976 moderate cases defined by count of endorsed MDD symptom criteria; iPSYCH, 5,574 cases of recurrent MDD versus 12,968 single-episode cases; and NESDA from PGC29, severity defined as chronic/unremitting MDD, 610 'stage IV' cases versus 499 'stage II' or 332 first-episode MDD cases[77] . Error bars represent 95% confidence intervals. Logistic regression association test P values are shown in the target sample for the GRS generated from SNPs with $P < 0.05$ in the discovery sample. FMex23andMe, full meta-analysis excluding 23andMe; scz, schizophrenia[19].

involved in neuronal morphogenesis; genes involved in neuron projection; genes associated with schizophrenia (at $P < 10^{-4}$)[19]; genes involved in central nervous system (CNS) neuron differentiation; genes encoding voltage-gated calcium channels; genes involved in cytokine and immune response; and genes known to bind to the retinoid X receptor. Several of these pathways are implicated by GWAS of schizophrenia and by rare exonic variation of schizophrenia and autism[56,57] and immediately suggest shared biological mechanisms across these disorders.

A key issue for common variant GWAS is their relevance for pharmacotherapy. We conducted gene set analysis that compared the major depression genome-wide association results to targets of antidepressant medications defined by pharmacological studies[58] and found that 42 sets of genes encoding proteins bound by antidepressant medications were highly enriched for smaller major depression association P values than expected by chance (42 drugs, rank enrichment test $P = 8.5 \times 10^{-10}$). This finding connects our major depression genomic findings to MDD therapeutics and suggests the salience of these results for new lead compound discovery for MDD[59].

Implications based on relationships with other traits. Prior epidemiological studies associated MDD with many other diseases

and traits. Because of limitations inherent to observational studies, understanding whether a phenotypic correlation is potentially causal or if it results from reverse causation or confounding is generally difficult. Genetic studies now offer complementary strategies to assess whether a phenotypic association between MDD and a risk factor or a comorbidity is mirrored by a nonzero r_g (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 the r_g of major depression with 221 psychiatric disorders, medical diseases, and human traits[22,60]. Supplementary Table 12 contains the full results, and Table 3 shows the r_g values with false discovery rates <0.01. First, the r_g values were very high between our major depression genome-wide association results and those from two studies of current depressive symptoms. Both correlations were close to 1 (the samples in one report overlapped partially with this meta-analysis[16], but the samples from the other did not[14]).

Second, we found significant positive genetic correlations between major depression and every psychiatric disorder assessed along with smoking initiation. This is a comprehensive and well-powered evaluation of the relationship of MDD with other psychiatric disorders, and these results indicate that the common genetic

Table 2 | 44 significantly associated genomic regions in meta-analysis of 135,458 major depression cases and 344,901 controls

Chr.	Region (Mb)	SNP	Location (bp)	P	A1/A2	OR (A1)	s.e. (log(OR))	Freq.	Prev.	Gene context
1	8.390–8.895	rs159963	8,504,421	3.2×10^{-8}	A/C	0.97	0.0049	0.56	H,S	[RERE]; SLC45A1, 100,194
1	72.511–73.059	rs1432639	72,813,218	4.6×10^{-15}	A/C	1.04	0.0050	0.63	H	NEGR1, –64,941
1	73.275–74.077	rs12129573	73,768,366	4.0×10^{-12}	A/C	1.04	0.0050	0.37	S	LINC01360, –3,486
1	80.785–80.980	rs2389016	80,799,329	1.0×10^{-8}	T/C	1.03	0.0053	0.28	H	
1	90.671–90.966	rs4261101	90,796,053	1.0×10^{-8}	A/G	0.97	0.0050	0.37		
1	197.343–197.864	rs9427672	197,754,741	3.1×10^{-8}	A/G	0.97	0.0058	0.24		DENND1B, –10,118
2	57.765–58.485	rs11682175	57,987,593	4.7×10^{-9}	T/C	0.97	0.0048	0.52	H,S	VRK2, –147,192
2	156.978–157.464	rs1226412	157,111,313	2.4×10^{-8}	T/C	1.03	0.0059	0.79		[LINC01876]; NR4A2, 69,630; GPD2, –180,651
3	44.222–44.997	chr3_44287760_I	44,287,760	4.6×10^{-8}	I/D	1.03	0.0051	0.34	T	[TOPAZ1]; TCAIM, –91,850; ZNF445, 193,501
3	157.616–158.354	rs7430565	158,107,180	2.9×10^{-9}	A/G	0.97	0.0048	0.58	H	[RSRC1]; LOC100996447, 155,828; MLF1, –181,772
4	41.880–42.189	rs34215985	42,047,778	3.1×10^{-9}	C/G	0.96	0.0063	0.24		[SLC30A9]; LINC00682, –163,150; DCAF4L1, 59,294
5	87.443–88.244	chr5_87992715_I	87,992,715	7.9×10^{-11}	I/D	0.97	0.0050	0.58	H	LINC00461, –12,095; MEF2C, 21,342
5	103.672–104.092	chr5_103942055_D	103,942,055	7.5×10^{-12}	I/D	1.03	0.0048	0.48	C	
5	124.204–124.328	rs116755193	124,251,883	7.0×10^{-9}	T/C	0.97	0.0050	0.38		LOC101927421, –120,640
5	164.440–164.789	rs11135349	164,523,472	1.1×10^{-9}	A/C	0.97	0.0048	0.48	H	
5	166.977–167.056	rs4869056	166,992,078	6.8×10^{-9}	A/G	0.97	0.0050	0.63		[TENM2]
6	27.738–32.848	rs115507122	30,737,591	3.3×10^{-11}	C/G	0.96	0.0063	0.18	S	Extended MHC
6	99.335–99.662	rs9402472	99,566,521	2.8×10^{-8}	A/G	1.03	0.0059	0.24		FBXL4, –170,672; C6orf168, 154,271
7	12.154–12.381	rs10950398	12,264,871	2.6×10^{-8}	A/G	1.03	0.0049	0.41		[TMEM106B]; VWDE, 105,637
7	108.925–109.230	rs12666117	109,105,611	1.4×10^{-8}	A/G	1.03	0.0048	0.47		
9	2.919–3.009	rs1354115	2,983,774	2.4×10^{-8}	A/C	1.03	0.0049	0.62	H	PUM3, –139,644; LINC01231, –197,814
9	11.067–11.847	rs10959913	11,544,964	5.1×10^{-9}	T/G	1.03	0.0057	0.76		
9	119.675–119.767	rs7856424	119,733,595	8.5×10^{-9}	T/C	0.97	0.0053	0.29		[ASTN2]
9	126.292–126.735	rs7029033	126,682,068	2.7×10^{-8}	T/C	1.05	0.0093	0.07		[DENND1A]; LHX2, –91,820
10	106.397–106.904	rs61867293	106,563,924	7.0×10^{-10}	T/C	0.96	0.0061	0.20	H	[SORCS3]
11	31.121–31.859	rs1806153	31,850,105	1.2×10^{-9}	T/G	1.04	0.0059	0.22		[DKFZp686K1684]; [PAUPAR]; ELP4, 44,032; PAX6, –10,596
12	23.924–24.052	rs4074723	23,947,737	3.1×10^{-8}	A/C	0.97	0.0049	0.41		[SOX5]
13	44.237–44.545	rs4143229	44,327,799	2.5×10^{-8}	A/C	0.95	0.0091	0.92		[ENOX1]; LACC1, –125,620; CCDC122, 82,689
13	53.605–54.057	rs12552	53,625,781	6.1×10^{-19}	A/G	1.04	0.0048	0.44	H	[OLFM4]; LINC01065, 80,099
14	41.941–42.320	rs4904738	42,179,732	2.6×10^{-9}	T/C	0.97	0.0049	0.57		[LRFN5]
14	64.613–64.878	rs915057	64,686,207	7.6×10^{-10}	A/G	0.97	0.0049	0.42		[SYNE2]; MIR548H1, –124,364; ESR2, 7,222
14	75.063–75.398	chr14_75356855_I	75,356,855	3.8×10^{-9}	D/I	1.03	0.0049	0.49		[DLST]; PROX2, –26,318; RPS6KL1, 13,801
14	103.828–104.174	rs10149470	104,017,953	3.1×10^{-9}	A/G	0.97	0.0049	0.49	S	BAG5, 4,927; APOPT1, –11,340
15	37.562–37.929	rs8025231	37,648,402	2.4×10^{-12}	A/C	0.97	0.0048	0.57	H	

Table 2 | 44 significantly associated genomic regions in meta-analysis of 135,458 major depression cases and 344,901 controls (continued)

Chr.	Region (Mb)	SNP	Location (bp)	P	A1/A2	OR (A1)	s.e. (log(OR))	Freq.	Prev.	Gene context
16	6.288–6.347	rs8063603	6,310,645	6.9×10^{-9}	A/G	0.97	0.0053	0.65		[RBFOX1]
16	7.642–7.676	rs7198928	7,666,402	1.0×10^{-8}	T/C	1.03	0.0050	0.62		[RBFOX1]
16	13.022–13.119	rs7200826	13,066,833	2.4×10^{-8}	T/C	1.03	0.0055	0.25		[SHISA9]; CPPED1, –169,089
16	71.631–72.849	rs11643192	72,214,276	3.4×10^{-8}	A/C	1.03	0.0049	0.41		PMFBP1, –7,927; DHX38, 67,465
17	27.345–28.419	rs17727765	27,576,962	8.5×10^{-9}	T/C	0.95	0.0088	0.92		[CRYBA1]; MYO18A, –69,555; NUFIP2, 5,891
18	36.588–36.976	rs62099069	36,883,737	1.3×10^{-8}	A/T	0.97	0.0049	0.42		[MIR924HG]
18	50.358–50.958	rs11663393	50,614,732	1.6×10^{-8}	A/G	1.03	0.0049	0.45	O	[DCC]; MIR4528, –148,738
18	51.973–52.552	rs1833288	52,517,906	2.6×10^{-8}	A/G	1.03	0.0054	0.72		[RAB27B]; CCDC68, 50,833
18	52.860–53.268	rs12958048	53,101,598	3.6×10^{-11}	A/G	1.03	0.0051	0.33	S	[TCF4]; MIR4529, –44,853
22	40.818–42.216	rs5758265	41,617,897	7.6×10^{-9}	A/G	1.03	0.0054	0.28	H,S	[L3MBTL2]; EP300-AS1, –24,392; CHADL, 7,616

Chr. (chromosome) and region (boundaries in Mb; hg19) are shown, defined by the locations of SNPs with $P < 1 \times 10^{-5}$ and LD $r^2 > 0.1$ with the most strongly associated SNP (logistic regression; the lowest P value in the region listed is not corrected for multiple testing), whose location is given in base pairs. In three regions, a second SNP fulfilled the filtering criteria and was followed up with conditional analyses: chr. 1, conditional analysis selected only rs1432639 as significant, with $P = 2.0 \times 10^{-4}$ for rs12134600 after fitting rs1432639; chr. 5, conditional analysis showed two independent associations selecting rs247910 and rs10514301 as the most strongly associated SNPs; chr. 10, conditional analysis selected only rs61867293 with $P = 8.6 \times 10^{-5}$ for rs1021363 after conditioning on rs61867293. For each of the 47 SNPs, there was at least one additional genome-wide significant SNP in the cluster of surrounding SNPs with low P values. Chromosome X was analyzed but had no findings that met genome-wide significance. Entries in the "Prev." column indicate which of four previous studies identified genome-wide significant associations in a region: H, Hyde et al.[28], 23andMe genome-wide association of self-reported clinical depression (the discovery sample overlaps with this paper); O, Okbay et al.[16], meta-analysis of genome-wide association of MDD, depressive symptoms, psychological well-being, and neuroticism (includes many PGC29 samples); S, PGC report on 108 schizophrenia-associated loci[19]; C, CHARGE Consortium meta-analysis of depressive symptoms[14]. The "Gene context" column shows the distances between the peak SNP and the closest genes; brackets indicate that the peak SNP was within that gene. The closest genes upstream (taking strand into account; a negative number indicates distance in base pairs between the peak SNP and the nearest gene boundary) and downstream (positive distance in base pairs) are also shown if there was a flanking gene within 200 kb. The name of the closest gene is in brackets. Note that it is generally not known whether associated SNPs have biological effects on these or other more distant genes. A1/A2, the two alleles (or insertion–deletion) at each SNP; A1 was tested for association and its OR (odds ratio) and s.e. are shown; Freq., the frequency of A1 in controls across all cohorts.

variants that predispose to MDD overlap substantially with those for adult- and childhood-onset psychiatric disorders, although they remain substantially distinct as well.

Third, the common variant genetic architecture of major depression was positively correlated with multiple measures of sleep quality (daytime sleepiness, insomnia, and tiredness). The first two of these correlations used UK Biobank data with people endorsing major depression or other major psychiatric disorders, shift workers, and those taking hypnotics excluded. This pattern of correlations combined with the importance of sleep and fatigue in major depression (two criteria for MDD) suggests a close and potentially profound mechanistic relationship. Major depression also had a strong genetic correlation with neuroticism (a personality dimension assessing the degree of emotional instability); this is consistent with the literature showing a close interconnection of MDD and this personality trait. The strong negative r_g with subjective well-being underscores the capacity of major depression to impact human health.

Finally, major depression had significant negative genetic correlations with data from two studies of educational attainment, which while often considered at the genetic level as proxy measures of intelligence also likely includes more complex personality constructs. With this in mind, it is relevant to note that the r_g between major depression and IQ[61] was not significantly different from zero, despite the r_g between years of education and IQ being 0.7, implying complex relationships between these traits worthy of future investigation. We also found significant 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 bidirectional Mendelian randomization (MR) to investigate the relationships between four traits genetically correlated with major depression: years of education (EDY)[62], body mass index (BMI)[29], coronary artery disease (CAD)[63], and schizophrenia[19]. These traits were selected because all of the following were true: they were phenotypically associated with MDD, had significant r_g with MDD, and had >30 independent genome-wide significant associations from large GWAS. We report GSMR[64] results but obtained qualitatively similar results with other MR methods (Supplementary Fig. 4 and Supplementary Table 13). MR analyses provided evidence for a 1.12-fold increase in major depression per s.d. of BMI ($P_{GSMR} = 1.2 \times 10^{-7}$) and a 0.84-fold decrease in major depression per s.d. of EDY ($P_{GSMR} = 2.3 \times 10^{-6}$). There was no evidence of reverse causality of major depression for BMI ($P_{GSMR} = 0.53$) or EDY ($P_{GSMR} = 0.11$). For BMI, there was some evidence of pleiotropy, as six BMI-associated SNPs were excluded by the HEIDI outlier test including SNPs near OLFM4 and NEGR1. Thus, these results are consistent with EDY and BMI as either causal risk factors or correlated with causal risk factors for major depression. These results provide hypotheses for future research to understand these potentially directional relationships.

For CAD, the MR analyses were not significant when considering major depression as an outcome ($P_{GSMR} = 0.30$) or as an exposure ($P_{GSMR} = 0.12$); however, the high standard error of the estimates using MDD SNP instruments implies that this analysis should be revisited when more major genome-wide significant SNP instruments for depression become available from future GWAS.

We used MR to investigate the relationship between major depression and schizophrenia. Although major depression had positive r_g with many psychiatric disorders, only schizophrenia had

a

b

c
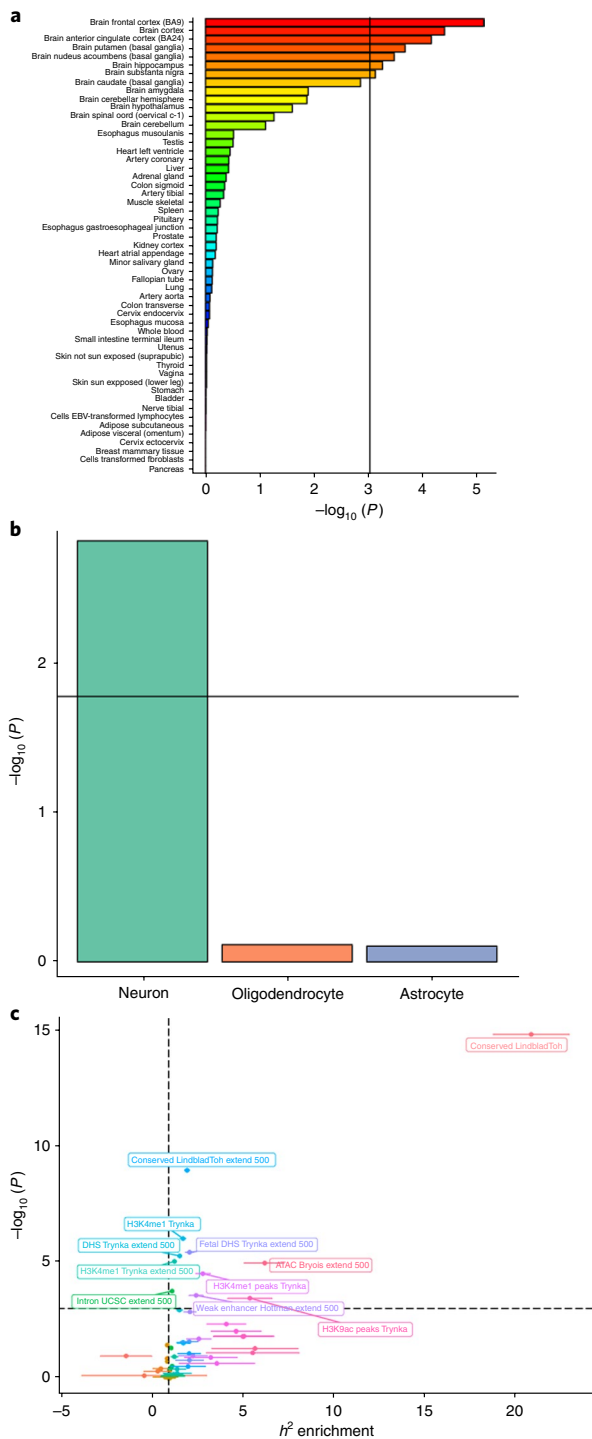
Fig. 3 | Analyses exploring enrichment of major depression association results based on different SNP annotations. a, Enrichment in bulk tissue mRNA-seq data from GTEx shown as t statistics; sample sizes in the GTEx database range from $n = 75$–564. The threshold for significance accounting for multiple testing is shown by the vertical line. **b**, Major depression results and enrichment in three major brain cell types shown as t statistics; the threshold for significance accounting for multiple testing is shown by the horizontal line. Sample sizes vary as these data were aggregated from many different sources. **c**, Partitioned LDSC to evaluate enrichment of the major depression genome-wide association findings in over 50 functional genomic annotations (Supplementary Table 8) shown as enrichment statistics; the threshold for significance accounting for multiple testing is shown by the horizontal dashed line. Sample sizes vary as these data were aggregated from many different sources.

Table 3 | LDSC genetic correlations of MDD with other disorders, diseases, and human traits

Trait	r_g	s.e.	FDR	h^2_{SNP}	Ref.
Depressive symptoms, CHARGE	0.91	0.123	3.2×10^{-12}	0.04	[14]
Depressive symptoms, SSGAC	0.98	0.034	1.3×10^{-176}	0.05	[16]
ADHD (iPSYCH-PGC)	0.42	0.033	6.1×10^{-36}	0.24	[83]
Anorexia nervosa	0.13	0.028	7.1×10^{-5}	0.55	[84]
Anxiety disorders	0.80	0.140	2.0×10^{-7}	0.06	[85]
Autism spectrum disorders (iPSYCH-PGC)	0.44	0.039	8.4×10^{-28}	0.20	[86]
Bipolar disorder	0.32	0.034	3.3×10^{-19}	0.43	[20]
Schizophrenia	0.34	0.025	7.7×10^{-40}	0.46	[19]
Smoking, ever versus never	0.29	0.038	7.0×10^{-13}	0.08	[87]
Daytime sleepiness[a]	0.19	0.048	5.7×10^{-4}	0.05	–
Insomnia[a]	0.38	0.038	4.0×10^{-22}	0.13	–
Tiredness	0.67	0.037	6.2×10^{-72}	0.07	[88]
Subjective well-being	−0.65	0.035	7.5×10^{-76}	0.03	[16]
Neuroticism	0.70	0.031	2.5×10^{-107}	0.09	[16]
College completion	−0.17	0.034	6.7×10^{-6}	0.08	[89]
Years of education	−0.13	0.021	1.6×10^{-8}	0.13	[62]
Body fat	0.15	0.038	6.5×10^{-4}	0.11	[90]
Body mass index	0.09	0.026	3.6×10^{-3}	0.19	[32]
Obesity class 1	0.11	0.029	1.6×10^{-3}	0.22	[30]
Obesity class 2	0.12	0.033	3.0×10^{-3}	0.18	[30]
Obesity class 3	0.20	0.053	1.6×10^{-3}	0.12	[30]
Overweight	0.13	0.030	1.4×10^{-4}	0.11	[30]
Waist circumference	0.11	0.024	8.2×10^{-5}	0.12	[91]
Waist-to-hip ratio	0.12	0.030	2.9×10^{-4}	0.11	[91]
Triglycerides	0.14	0.028	1.0×10^{-5}	0.17	[92]
Age at menarche	−0.14	0.023	6.3×10^{-8}	0.20	[93]
Age of first birth	−0.29	0.029	6.1×10^{-22}	0.06	[94]
Father's age at death	−0.28	0.058	3.0×10^{-5}	0.04	[95]
Number of children ever born	0.13	0.036	2.4×10^{-3}	0.03	[94]
Coronary artery disease	0.12	0.027	8.2×10^{-5}	0.08	[63]
Squamous cell lung cancer	0.26	0.075	3.6×10^{-3}	0.04	[96]

All genetic correlations (r_g) estimated using bivariate LDSC applied to major depression genome-wide association results are in Supplementary Table 12. Shown above are the r_g values of major depression with false discovery rate (FDR) < 0.01 (FDR estimated for 221 genetic correlations; H_0: $r_g = 0$). Thematically related traits are indicated by bold font. iPSYCH is a nationally representative cohort based on blood spots collected at birth. Within iPSYCH, r_g with ADHD was 0.58 (s.e. = 0.050) and with ASD was 0.51 (s.e. = 0.07); these values are larger than those listed above and are inconsistent with artifactual correlations. h^2_{SNP} is shown to aid interpretation as high r_g in the context of high h^2_{SNP} is more noteworthy than when h^2_{SNP} is low. [a]Self-reported daytime sleepiness and insomnia from UK Biobank excluding subjects with major depression, other psychiatric disorders (bipolar disorder, schizophrenia, autism, intellectual disability), shift workers, and those taking hypnotics.

sufficient associations for MR analyses. We found significant bidirectional correlations in SNP effect sizes for schizophrenia loci in major depression ($P_{GSMR} = 1.1 \times 10^{-40}$) and for major depression loci in schizophrenia ($P_{GSMR} = 1.5 \times 10^{-11}$). These results suggest that the major depression–schizophrenia r_g of 0.34 is consistent with
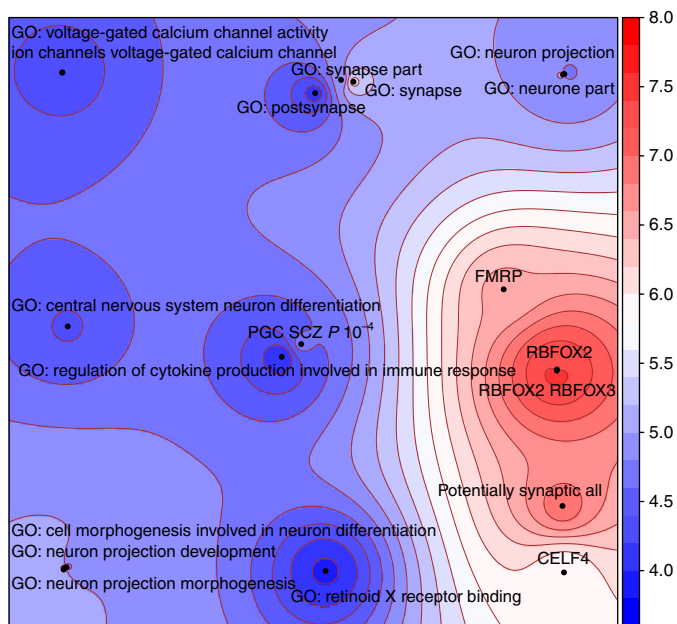
Fig. 4 | Generative topographic mapping of the 19 significant pathway results. The average position of each pathway on the map is represented by a point. The map is colored by the $-\log_{10} P$ value obtained using MAGMA. The x and y coordinates were obtained from a kernel generative topographic mapping algorithm (GTM) that reduces high-dimensional gene sets to a 2D scatterplot by accounting for gene overlap between gene sets. Each point represents a gene set. Nearby points are more similar in gene overlap than more distant points. The color surrounding each point (gene set) corresponds to significance according to the scale on the right. The significant pathways (Supplementary Table 11) fall into nine main clusters as described in the text.

partially shared biological pathways being causal for both disorders. Although it is plausible that diagnostic misclassification/ambiguity (for example, misdiagnosis of MDD as schizoaffective disorder) could contaminate these analyses, levels of misclassification would need to be implausibly high (30% unidirectional, 15% bidirectional) to result in an r_{g} of ~0.3[65].

All MR analyses were repeated after excluding the 23 and Me cohort, and the pattern of the results was the same (Supplementary Table 13).

Discussion

The nature of severe depression has been discussed for millennia[66]. This genome-wide association meta-analysis is among the largest ever conducted in psychiatric genetics and provides a body of results that helps refine the fundamental basis of major depression.

In conducting this meta-analysis of major depression, we employed a pragmatic approach by including cohorts that met empirical criteria for sufficient genetic and phenotypic similarity. Our approach was cautious, clinically informed, guided by empirical data, and selective (for example, we did not include cohorts with bipolar disorder (which requires MDD), depressive symptoms, neuroticism, or well-being). Approximately 44% of all major depression cases were assessed using traditional methods (PGC29, GenScot), treatment registers (iPSYCH, GERA; such approaches have been extensively used to elucidate the epidemiology of major depression), or a combination of methods (deCODE, UK Biobank), whereas ~56% of cases were from 23andMe (via self-report)[28]. Multiple lines of genetic evidence supported conducting meta-analysis of these seven cohorts (for example, out-of-sample prediction, sign tests, and genetic correlations).

However, our approach may be controversial to some readers given the unconventional reliance on self-report of major depression. We would reframe the issue: we hypothesize that brief methods of assessing major depression are informative for the genetics of MDD. We present a body of results that is consistent with this hypothesis. Even if unconventional, our hypothesis is testable and falsifiable, and we invite and welcome empirical studies to further support or refute this hypothesis.

Our results lead us to draw some broad conclusions. First, major depression is a brain disorder. Although this is not unexpected, some past models of MDD have had little or no place for heredity or biology. The genetic results best match gene expression patterns in the 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-nucleus RNA-seq datasets become available[67].

Second, the genetic associations for major depression (as with schizophrenia)[46] tend to occur in genomic regions conserved across a range of placental mammals. Conservation suggests important functional roles. Notably, our analyses did not implicate exons or coding regions.

Third, the results also implicated developmental gene regulatory processes. For instance, the genetic findings pointed at the splicing regulator *RBFOX1* (the presence of two independent genetic associations in *RBFOX1* strongly suggests that it is the relevant gene). Gene set analyses implicated genes containing binding sites to the protein product of *RBFOX1*, and this gene set is also significantly enriched for rare exonic variation in autism and schizophrenia[56,57]. These analyses highlight the potential importance of splicing to generate alternative isoforms; risk for major depression 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[68,69]. These convergent results provide possible clues to a biological mechanism common to multiple severe psychiatric disorders that merits future research.

Fourth, in the most extensive analysis of the genetic 'connections' of major depression 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, while showing no evidence of genetic correlation with IQ. MR analysis results are consistent with both BMI and years of education being causal, or correlated with causal, risk factors for major depression, and our results provide hypotheses and motivation for more detailed prospective studies, as currently available data may not provide 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. While the significant MR analyses need further investigations to fully understand, the negative MR results provide important evidence that there is not a direct causal relationship between MDD and subsequent changes in body mass or education years. If such associations are observed in epidemiological or clinical samples, then other factors must drive the association.

Fifth, we found significant positive correlations of major depression 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. Currently, only schizophrenia has a sufficient number of genome-wide significant loci to conduct MR analysis, but the bidirectionally significant MR results are consistent with a shared biological basis for major depression and schizophrenia.

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)[70–73]. Our genetic results add to this knowledge: major depression is not a discrete entity at any level of analysis. Rather, our data strongly suggest the existence of biological processes common to major depression and schizophrenia (and likely other psychiatric disorders).

Finally, as expected, we found that major depression had modest h^2_{SNP} (8.7%), as it is a complex malady with both genetic and environmental determinants. We found that major depression 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 depression 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 relationships of clinically defined ADHD or autism with quantitative genetic variation in the population have been reported[74,75]. The 'disorder-versus-symptom' relationship has been debated extensively[76], but our data indicate that the common variant genetic overlap is very high. This finding has important implications.

One implication is for future genetic studies. 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 online self-report assessments of lifetime MDD (a sample size approaching 1 million MDD cases may be achievable by 2020). Use of online assessment could allow for recording of a broad range of phenotypes including comorbidities and putative environmental exposures, but with the key feature being large samples with consistently assessed measures. In a second phase, with a relatively complete understanding of the genetic basis of major depression, one could then evaluate smaller samples of carefully phenotyped individuals with MDD to understand the clinical importance of the genetic results. 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 (for example, schizophrenia or bipolar disorder). This could form a cornerstone of precision medicine in psychiatry.

In summary, this genome-wide association meta-analysis of 135,438 MDD and major depression cases and 344,901 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.

URLs. 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; dbGaP, https://www.ncbi.nlm.nih.gov/gap; genotype-based checksums for relatedness determination, http://www.broadinstitute.org/~sripke/share_links/checksums_download; GSMR, http://cnsgenomics.com/software/gsmr/; 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/; PGC, http://www.med.unc.edu/pgc; NIH NeuroBiobank, https://neurobiobank.nih.gov/;

PGC ricopili genome-wide association analysis pipeline, https://github.com/Nealelab/ricopili; SMR, http://cnsgenomics.com/software/smr/#Overview; TWAS, http://gusevlab.org/projects/fusion/; UK Biobank, http://www.ukbiobank.ac.uk/.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-018-0090-3.

References

1. Kessler, R. C. & Bromet, E. J. The epidemiology of depression across cultures. *Annu. Rev. Public Health* **34**, 119–138 (2013).
2. Judd, L. L. The clinical course of unipolar major depressive disorders. *Arch. Gen. Psychiatry* **54**, 989–991 (1997).
3. Lopez, A. D., Mathers, C. D., Ezzati, M., Jamison, D. T. & Murray, C. J. Global and regional burden of disease and risk factors, 2001: systematic analysis of population health data. *Lancet* **367**, 1747–1757 (2006).
4. Wittchen, H. U. et al. The size and burden of mental disorders and other disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 655–679 (2011).
5. Ferrari, A. J. et al. Burden of depressive disorders by country, sex, age, and year: findings from the global burden of disease study 2010. *PLoS Med.* **10**, e1001547 (2013).
6. Angst, F., Stassen, H. H., Clayton, P. J. & Angst, J. Mortality of patients with mood disorders: follow-up over 34–38 years. *J. Affect. Disord.* **68**, 167–181 (2002).
7. Gustavsson, A. et al. Cost of disorders of the brain in Europe 2010. *Eur. Neuropsychopharmacol.* **21**, 718–779 (2011).
8. Murray, C. J. et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2197–2223 (2012).
9. Sullivan, P. F., Neale, M. C. & Kendler, K. S. Genetic epidemiology of major depression: review and meta-analysis. *Am. J. Psychiatry* **157**, 1552–1562 (2000).
10. Rice, F., Harold, G. & Thapar, A. The genetic aetiology of childhood depression: a review. *J. Child Psychol. Psychiatry* **43**, 65–79 (2002).
11. Viktorin, A. et al. Heritability of perinatal depression and genetic overlap with nonperinatal depression. *Am. J. Psychiatry* **73**, 158–165 (2016).
12. Levinson, D. F. et al. Genetic studies of major depressive disorder: why are there no GWAS findings, and what can we do about it. *Biol. Psychiatry* **76**, 510–512 (2014).
13. Major Depressive Disorder Working Group of the PGC. A mega-analysis of genome-wide association studies for major depressive disorder. *Mol. Psychiatry* **18**, 497–511 (2013).
14. Hek, K. et al. A genome-wide association study of depressive symptoms. *Biol. Psychiatry* **73**, 667–678 (2013).
15. CONVERGE Consortium. Sparse whole-genome sequencing identifies two loci for major depressive disorder. *Nature* **523**, 588–591 (2015).
16. Okbay, A. et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat. Genet.* **48**, 624–633 (2016).
17. Sullivan, P. F. et al. Psychiatric genomics. An update and an agenda. *Am. J. Psychiatry* **175**, 15–27 (2018).
18. Visscher, P. M., Brown, M. A., McCarthy, M. I. & Yang, J. Five years of GWAS discovery. *Am. J. Hum. Genet.* **90**, 7–24 (2012).
19. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421–427 (2014).
20. Psychiatric GWAS Consortium Bipolar Disorder Working Group. Large-scale genome-wide association analysis of bipolar disorder identifies a new susceptibility locus near *ODZ4*. *Nat. Genet.* **43**, 977–983 (2011).
21. Cross-Disorder Group of the Psychiatric Genomics Consortium. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat. Genet.* **45**, 984–994 (2013).
22. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236–1241 (2015).
23. Wray, N. R. et al. Genome-wide association study of major depressive disorder: new results, meta-analysis, and lessons learned. *Mol. Psychiatry* **17**, 36–48 (2012).

24. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

25. Meier, S. M. et al. High loading of polygenic risk in cases with chronic schizophrenia. *Mol. Psychiatry* **21**, 969–974 (2016).

26. Yang, J. et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44**, 369–375 (2012).

27. Wray, N. R. & Maier, R. Genetic basis of complex genetic disease: the contribution of disease heterogeneity to missing heritability. *Curr. Epidemiol. Rep.* **1**, 220–227 (2014).

28. Hyde, C. L. et al. Identification of 15 genetic loci associated with risk of major depression in individuals of European descent. *Nat. Genet.* **48**, 1031–1036 (2016).

29. Locke, A. E. et al. Genetic studies of body mass index yield new insights for obesity biology. *Nature* **518**, 197–206 (2015).

30. Berndt, S. I. et al. Genome-wide meta-analysis identifies 11 new loci for anthropometric traits and provides insights into genetic architecture. *Nat. Genet.* **45**, 501–512 (2013).

31. Bradfield, J. P. et al. A genome-wide association meta-analysis identifies new childhood obesity loci. *Nat. Genet.* **44**, 526–531 (2012).

32. Speliotes, E. K. et al. Association analyses of 249,796 individuals reveal 18 new loci associated with body mass index. *Nat. Genet.* **42**, 937–948 (2010).

33. Willer, C. J. et al. Six new loci associated with body mass index highlight a neuronal influence on body weight regulation. *Nat. Genet.* **41**, 25–34 (2009).

34. Thorleifsson, G. et al. Genome-wide association yields new sequence variants at seven loci that associate with measures of obesity. *Nat. Genet.* **41**, 18–24 (2009).

35. Fogel, B. L. et al. RBFOX1 regulates both splicing and transcriptional networks in human neuronal development. *Hum. Mol. Genet.* **21**, 4171–4186 (2012).

36. Gehman, L. T. et al. The splicing regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. *Nat. Genet.* **43**, 706–711 (2011).

37. Pariante, C. M. & Lightman, S. L. The HPA axis in major depression: classical theories and new developments. *Trends Neurosci.* **31**, 464–468 (2008).

38. Choi, Y. et al. SALM5 trans-synaptically interacts with LAR-RPTPs in a splicing-dependent manner to regulate synapse development. *Sci. Rep.* **6**, 26676 (2016).

39. Mah, W. et al. Selected SALM (synaptic adhesion-like molecule) family proteins regulate synapse formation. *J. Neurosci.* **30**, 5559–5568 (2010).

40. Zhu, Y. et al. Neuron-specific SALM5 limits inflammation in the CNS via its interaction with HVEM. *Sci. Adv.* **2**, e1500637 (2016).

41. Amiel, J. et al. Mutations in *TCF4*, encoding a class I basic helix-loop-helix transcription factor, are responsible for Pitt–Hopkins syndrome, a severe epileptic encephalopathy associated with autonomic dysfunction. *Am. J. Hum. Genet.* **80**, 988–993 (2007).

42. Akbarian, S. et al. The PsychENCODE project. *Nat. Neurosci.* **18**, 1707–1712 (2015).

43. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).

44. Schmaal, L. et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. *Mol. Psychiatry* **22**, 900–909 (2017).

45. Cahoy, J. D. et al. A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).

46. Finucane, H. K. et al. Partitioning heritability by functional category using GWAS summary statistics. *Nat. Genet.* **47**, 1228–1235 (2015).

47. Lindblad-Toh, K. et al. A high-resolution map of human evolutionary constraint using 29 mammals. *Nature* **478**, 476–482 (2011).

48. Simonti, C. N. et al. The phenotypic legacy of admixture between modern humans and Neandertals. *Science* **351**, 737–741 (2016).

49. Zhu, Z. et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* **48**, 481–487 (2016).

50. Gusev, A. et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat. Genet.* **48**, 245–252 (2016).

51. Fromer, M. et al. Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat. Neurosci.* **19**, 1442–1453 (2016).

52. Smemo, S. et al. Obesity-associated variants within *FTO* form long-range functional connections with *IRX3*. *Nature* **507**, 371–375 (2014).

53. Won, H. et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523–527 (2016).

54. Martin, J. S. et al. HUGIn. Hi-C Unifying Genomic Interrogator. *Bioinformatics* **33**, 3793–3795 (2017).

55. Network and Pathway Analysis Subgroup of Psychiatric Genomics Consortium. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. *Nat. Neurosci.* **18**, 199–209 (2015).

56. De Rubeis, S. et al. Synaptic, transcriptional and chromatin genes disrupted in autism. *Nature* **515**, 209–215 (2014).

57. Genovese, G. et al. Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia. *Nat. Neurosci.* **19**, 1433–1441 (2016).

58. Gaspar, H. A. & Breen, G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Sci. Rep.* **7**, 12460 (2017).

59. Breen, G. et al. Translating genome-wide association findings into new therapeutics for psychiatry. *Nat. Neurosci.* **19**, 1392–1396 (2016).

60. Zheng, J. et al. LD Hub: a centralized database and web interface to perform LD score regression that maximizes the potential of summary level GWAS data for SNP heritability and genetic correlation analysis. *Bioinformatics* **33**, 272–279 (2017).

61. Sniekers, S. et al. Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nat. Genet.* **49**, 1107–1112 (2017).

62. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).

63. Nikpay, M. et al. A comprehensive 1000 Genomes–based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).

64. Zhu, Z. et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* **9**, 224 (2018).

65. Wray, N. R., Lee, S. H. & Kendler, K. S. Impact of diagnostic misclassification on estimation of genetic correlations using genome-wide genotypes. *Eur. J. Hum. Genet.* **20**, 668–674 (2012).

66. Hippocrates. *The Aphorisms of Hippocrates* (Collins & Co., New York, 1817).

67. Skene, N. G. et al. Genetic identification of brain cell types underlying schizophrenia. *Nat. Genet.* (in the press).

68. Yang, X. et al. Widespread expansion of protein interaction capabilities by alternative splicing. *Cell* **164**, 805–817 (2016).

69. Zhang, X. et al. Cell-type-specific alternative splicing governs cell fate in the developing cerebral cortex. *Cell* **166**, 1147–1162 (2016).

70. Kessler, R. C. et al. The epidemiology of major depressive disorder: results from the National Comorbidity Survey Replication (NCS-R). *J. Am. Med. Assoc.* **289**, 3095–3105 (2003).

71. Hasin, D. S., Goodwin, R. D., Stinson, F. S. & Grant, B. F. Epidemiology of major depressive disorder: results from the National Epidemiologic Survey on Alcoholism and Related Conditions. *Arch. Gen. Psychiatry* **62**, 1097–1106 (2005).

72. Kendler, K. S. et al. The structure of genetic and environmental risk factors for syndromal and subsyndromal common DSM-IV axis I and all axis II disorders. *Am. J. Psychiatry* **168**, 29–39 (2011).

73. Kendler, K. S., Prescott, C. A., Myers, J. & Neale, M. C. The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. *Arch. Gen. Psychiatry* **60**, 929–937 (2003).

74. Robinson, E. B. et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat. Genet.* **48**, 552–555 (2016).

75. Middeldorp, C. M. et al. A genome-wide association meta-analysis of attention-deficit/hyperactivity disorder symptoms in population-based pediatric cohorts. *J. Am. Acad. Child. Adolesc. Psychiatry* **55**, 896–905 (2016).

76. Kendell, R. E. The classification of depressions: a review of contemporary confusion. *Br. J. Psychiatry* **129**, 15–28 (1976).

77. Verduijn, J. et al. Using clinical characteristics to identify which patients with major depressive disorder have a higher genetic load for three psychiatric disorders. *Biol. Psychiatry* **81**, 316–324 (2017).

78. Smith, B. H. et al. Cohort profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int. J. Epidemiol.* **42**, 689–700 (2013).

79. Fernandez-Pujals, A. M. et al. Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in Generation Scotland: Scottish Family Health Study (GS:SFHS). *PLoS One* **10**, e0142197 (2015).

80. Banda, Y. et al. Characterizing race/ethnicity and genetic ancestry for 100,000 subjects in the Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort. *Genetics* **200**, 1285–1295 (2015).

81. Pedersen, C. B. et al. The iPSYCH2012 case–cohort sample: new directions for unravelling genetic and environmental architectures of severe mental disorders. *Mol. Psychiatry* **23**, 6–14 (2018).

82. Allen, N. E., Sudlow, C., Peakman, T., Collins, R. & UK Biobank. UK Biobank data: come and get it. *Sci. Transl. Med.* **6**, 224ed4 (2014).

83. Demontis, D. et al. Discovery of the first genome-wide significant risk loci for ADHD. Preprint at *bioRxiv* https://www.biorxiv.org/content/early/2017/06/03/145581 (2017).

84. Boraska, V. et al. A genome-wide association study of anorexia nervosa. *Mol. Psychiatry* **19**, 1085–1094 (2014).
85. Otowa, T. et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Mol. Psychiatry* **21**, 1391–1399 (2016).
86. Grove, J. et al. Common risk variants identified in autism spectrum disorder. Preprint at *bioRxiv* https://www.biorxiv.org/content/early/2017/11/27/224774 (2017).
87. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat. Genet.* **42**, 441–447 (2010).
88. Deary, V. et al. Genetic contributions to self-reported tiredness. *Mol. Psychiatry* **23**, 609–620 (2018).
89. Rietveld, C. A. et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science* **340**, 1467–1471 (2013).
90. Lu, Y. et al. New loci for body fat percentage reveal link between adiposity and cardiometabolic disease risk. *Nat. Commun.* **7**, 10495 (2016).
91. Shungin, D. et al. New genetic loci link adipose and insulin biology to body fat distribution. *Nature* **518**, 187–196 (2015).
92. Teslovich, T. M. et al. Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
93. Perry, J. R. et al. Parent-of-origin-specific allelic associations among 106 genomic loci for age at menarche. *Nature* **514**, 92–97 (2014).
94. Barban, N. et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
95. Pilling, L. C. et al. Human longevity is influenced by many genetic variants: evidence from 75,000 UK Biobank participants. *Aging* **8**, 547–560 (2016).
96. Patel, Y. M. et al. Novel association of genetic markers affecting CYP2A6 activity and lung cancer risk. *Cancer Res.* **76**, 5768–5776 (2016).

Author contributions

Writing group: G.B., A.D.B., D.F.L., C.M.L., S.R., P.F.S., N.R.W. **PGC MDD PI group:** V.A., B.T.B., K.B., D.I.B., G.B., A.D.B., S.C., U.D., J.R.D., E.D., K.D., T.E., E.J.C.d.G., H.J.G., S.P.H., C. Hayward, A.C.H., D.M.H., K.S.K., S.K., D.F.L., C.M.L., G.L., Q.S.L., S.L., P.A.F.M., P.K.M., N.G.M., A.M.M., A.M., O.M., P.B.M., B.M.-M., M. Nordentoft, M.M.N., M.C.O'D., S.A.P., N.L.P., B.W.P., R.H.P., D.J.P., J.B.P., M.P., M. Rietschel, C.S., T.G.S., J.W.S., K.S., P.F.S., H. Tiemeier, R.U., H.V., M.M.W., T.W., A.R.W., N.R.W. **Bioinformatics:** 23andMe Research Team, M.J.A., S.V.d.A., G.B., J.B., A.D.B., E.C., J.H.C., T.-K.C., J.R.I.C., L.C.-C., eQTLGen Consortium, G.E.C., C.A.C., G.D., E.M.D., T.E., A.J.F., H.A.G., P.G.-R., J.G., L.S.H., E.H., T.F.H., C. Hayward, M.H., R.J., F.J., Z.K., Q.S.L., Yihan Li, P.A.L., X.L., L.L., D.J.M., S.E.M., E.M., Y.M., J. Mill, J.N.P., B.W.P., W.J.P., G.P., P.Q., L.S., S.I.S., C.A.S., P.F.S., K.E.T., A.T., P.A.T., A.G.U., Y. Wang, S.M.W., N.R.W., H.S.X. **Clinical:** E.A., T.M.A., V.A., B.T.B., A.T.F.B., K.B., E.B.B., D.H.R.B., H.N.B., A.D.B., N. Craddock, U.D., J.R.D., N.D., K.D., M.G., F.S.G., H.J.G., A.C.H., A.M.v.H., I.B.H., M.I., S.K., J. Krogh, D.F.L., S.L., D.J.M., D.F.M., P.A.F.M., W.M., N.G.M., P. McGrath, P. McGuffin, A.M.M., A.M., C.M.M., S.S.M., F.M.M., O.M., P.B.M., D.R.N., H.O., M.J.O., C.B.P., M.G.P., J.B.P., J.A.Q., J.P.R., M. Rietschel, C.S., R. Schoevers, E.S., G.C.B.S., D.J.S., F.S., J. Strohmaier, D.U., M.M.W., J.W., T.W., G.W. **Genomic assays:** G.B., H.N.B., J.B.-G., M.B.-H., A.D.B., S.C., T.-K.C., F.D., A.J.F., S.P.H., C.S.H., A.C.H., P.H., G.H., C. Horn, J.A.K., P.A.F.M., L.M., G.W.M., M. Nauck, M.M.N., M. Rietschel, M. Rivera, E.C.S., T.G.S., S.I.S., H.S., F.S., T.E.T., J.T., A.G.U., S.H.W. **Obtained funding for primary MDD samples:** B.T.B., K.B., D.H.R.B., D.I.B., G.B., H.N.B., A.D.B., S.C., J.R.D., I.J.D., E.D., T.C.E., T.E., H.J.G., S.P.H., A.C.H., I.S.K., D.F.L., C.M.L., G.L., Q.S.L., S.L., P.A.F.M., W.M., N.G.M., P. McGuffin, A.M.M., A.M., G.W.M., O.M., P.B.M., M. Nordentoft, D.R.N., M.M.N., P.F.O'R., B.W.P., D.J.P., J.B.P., M.P., M. Rietschel, C.S., T.G.S., G.C.B.S., J.H.S., D.J.S., H.S., K.S., P.F.S., T.E.T., H. Tiemeier, A.G.U., H.V., M.M.W., T.W., N.R.W. **Statistical analysis:** 23andMe Research Team, A.A., M.J.A., T.F.M.A., S.V.d.A., S.-A.B., K.B., T.B.B., G.B., E.M.B., A.D.B., N. Cai, T.-K.C., J.R.I.C., B.C.-D., H.S.D., G.D., N.D., C.V.D., E.C.D., N.E., V.E.-P., T.E., H.K.F., J.F., H.A.G., S.D.G., J.G., L.S.H., C. Hayward, A.C.H., S.H., D.A.H., J.-J.H., C.L.H., M.I., E.J., F.F.H.K., J. Kraft, W.W.K., Z.K., J.M.L., C.M.L., Q.S.L., Yun Li, D.J.M., P.A.F.M., R.M.M., J. Marchini, M.M., H.M., A.M.M., S.E.M., D.M., E.M., Y.M., S.S.M., S.M., N.M., B.M.-M., B.N., M.G.N., D.R.N., P.F.O'R., R.E.P., E.P., W.J.P., G.P., D.P., S.M.P., B.P.R., S.R., M. Rivera, R. Saxena, C.S., L.S., J. Shi, S.I.S., H.S., S.S., P.F.S., K.E.T., H. Teismann, A.T., W.T., P.A.T., T.E.T., C.T., M. Traylor, V.T., M. Trzaskowski, A.V., P.M.V., Y. Wang, B.T.W., J.W., T.W., N.R.W., Y. Wu, J.Y., F.Z.

Competing interests

A.T.F.B. is on speaker's bureaus for Lundbeck and GlaxoSmithKline. G.C. is a cofounder of Element Genomics. E.D. was an employee of Hoffmann–La Roche at the time this study was conducted and a consultant to Roche and Pierre-Fabre. N.E. is employed by 23andMe, Inc., and owns stock in 23andMe, Inc. D.H. is an employee of and owns stock options in 23andMe, Inc. S.P. is an employee of Pfizer, Inc. C.L.H. is an employee of Pfizer, Inc. A.R.W. was a former employee and stockholder of Pfizer, Inc. J.A.Q. was an employee of Hoffmann–La Roche at the time this study was conducted. H.S. is an employee of deCODE Genetics/Amgen. K.S. is an employee of deCODE Genetics/Amgen. S.S. is an employee of deCODE Genetics/Amgen. P.F.S. is on the scientific advisory board for Pfizer, Inc., and the advisory committee for Lundbeck. T.E.T. is an employee of deCODE Genetics/Amgen. C.T. is an employee of and owns stock options in 23andMe, Inc.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41588-018-0090-3.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to N.R.W. or P.F.S.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Naomi R. Wray [1,2,160]*, Stephan Ripke[3,4,5,160], Manuel Mattheisen[6,7,8,9,160], Maciej Trzaskowski[1,160], Enda M. Byrne[1], Abdel Abdellaoui[10], Mark J. Adams[11], Esben Agerbo[8,12,13], Tracy M. Air[14], Till M. F. Andlauer[15,16], Silviu-Alin Bacanu[17], Marie Bækvad-Hansen[8,18], Aartjan F. T. Beekman[19], Tim B. Bigdeli[17,20], Elisabeth B. Binder[15,21], Douglas R. H. 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 I. R. Coleman[27], Lucía Colodro-Conde[28], Baptiste Couvy-Duchesne[2,29], Nick Craddock[30], Gregory E. Crawford[31,32], Cheynna A. Crowley[33], Hassan S. Dashti[3,34], Gail Davies[35], Ian J. Deary[35], Franziska Degenhardt[36,37], Eske M. Derks[28], Nese Direk[38,39], Conor V. Dolan[10], Erin C. Dunn[40,41,42], Thalia C. Eley[27], Nicholas Eriksson[43], Valentina Escott-Price[44], Farnush Hassan Farhadi Kiadeh[45], Hilary K. Finucane[46,47], Andreas J. Forstner[36,37,48,49], Josef Frank[50], Héléna A. Gaspar[27], Michael Gill[51], Paola Giusti-Rodríguez[52], Fernando S. Goes[53], Scott D. Gordon[54], Jakob Grove[6,7,8,55], Lynsey S. Hall[11,56], Eilis Hannon[57], Christine Søholm Hansen[8,18], Thomas F. Hansen[58,59,60], Stefan Herms[36,37,49], Ian B. Hickie[61], Per Hoffmann[36,37,49], 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], 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[3,34,78], Yihan Li[74], Yun Li[33,52], Penelope A. Lind[28], Xiaoxiao Liu[68], Leina Lu[68], Donald J. MacIntyre[79,80], Dean F. MacKinnon[53], 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], Lili Milani[87], Jonathan Mill[57], Francis M. Mondimore[53], 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[39], Richa Saxena[3,34,78], Robert Schoevers[102], Eva C. Schulte[103,104], Ling Shen[69], Jianxin Shi[105], Stanley I. Shyn[106], Engilbert Sigurdsson[107], Grant B. C. Sinnamon[108], Johannes H. Smit[19], Daniel J. Smith[109], Hreinn Stefansson[110], Stacy Steinberg[110], Craig A. Stockmeier[111], Fabian Streit[50], Jana Strohmaier[50], 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[43], Matthew Traylor[118], Jens Treutlein[50], 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[50], Yang Wu[1], Hualin S. Xi[124], Jian Yang[1,2], Futao Zhang[1], eQTLGen[125], 23andMe, Volker Arolt[126], Bernhard T. Baune[14], Klaus Berger[113], Dorret I. Boomsma[10], Sven Cichon[36,49,127,128], Udo Dannlowski[126], E. C. J. de Geus[10,129], J. Raymond DePaulo[53], Enrico Domenici[130], Katharina Domschke[131], Tõnu Esko[3,87], Hans J. Grabe[121], Steven P. Hamilton[132], Caroline Hayward[133], Andrew C. Heath[100], David A. Hinds[43], Kenneth S. Kendler[17], Stefan Kloiber[66,134,135], Glyn Lewis[136], Qingqin S. Li[137], Susanne Lucae[66], Pamela F. A. Madden[100], Patrik K. Magnusson[22], Nicholas G. Martin[54], Andrew M. McIntosh[11,35], Andres Metspalu[87,138], Ole Mors[8,139], Preben Bo Mortensen[7,8,12,13], Bertram Müller-Myhsok[15,16,140], Merete Nordentoft[8,141], Markus M. Nöthen[36,37], Michael C. O'Donovan[94], Sara A. Paciga[142], Nancy L. Pedersen[22], Brenda W. J. H. Penninx[19], Roy H. Perlis[41,143], David J. Porteous[117],

James B. Potash[144], Martin Preisig[26], Marcella Rietschel[50], Catherine Schaefer[69], Thomas G. Schulze[50,53,104,145,146], Jordan W. Smoller[40,41,42], Kari Stefansson[110,147], Henning Tiemeier[39,148,149], Rudolf Uher[150], Henry Völzke[114], Myrna M. Weissman[83,151], Thomas Werge[8,59,152], Ashley R. Winslow[153,154], Cathryn M. Lewis[27,155,160], Douglas F. Levinson[156,160], Gerome Breen[27,157,160], Anders D. Børglum[6,7,8,160], Patrick F. Sullivan [22,52,158,160]* and the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium

[1]Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. [2]Queensland Brain Institute, University of Queensland, Brisbane, Queensland, Australia. [3]Medical and Population Genetics, Broad Institute, Cambridge, MA, USA. [4]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA, USA. [5]Department of Psychiatry and Psychotherapy, Universitätsmedizin Berlin Campus Charité Mitte, Berlin, Germany. [6]Department of Biomedicine, Aarhus University, Aarhus, Denmark. [7]iSEQ, Centre for Integrative Sequencing, Aarhus University, Aarhus, Denmark. [8]iPSYCH, Lundbeck Foundation Initiative for Integrative Psychiatric Research, Aarhus, Denmark. [9]Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden. [10]Department of Biological Psychology and EMGO+ Institute for Health and Care Research, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [11]Division of Psychiatry, University of Edinburgh, Edinburgh, UK. [12]Centre for Integrated Register-Based Research, Aarhus University, Aarhus, Denmark. [13]National Centre for Register-Based Research, Aarhus University, Aarhus, Denmark. [14]Discipline of Psychiatry, University of Adelaide, Adelaide, South Australia, Australia. [15]Department of Translational Research in Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany. [16]Munich Cluster for Systems Neurology (SyNergy), Munich, Germany. [17]Department of Psychiatry, Virginia Commonwealth University, Richmond, VA, USA. [18]Center for Neonatal Screening, Department for Congenital Disorders, Statens Serum Institut, Copenhagen, Denmark. [19]Department of Psychiatry, Vrije Universiteit Medical Center and GGZ inGeest, Amsterdam, The Netherlands. [20]Virginia Institute for Psychiatric and Behavior Genetics, Richmond, VA, USA. [21]Department of Psychiatry and Behavioral Sciences, Emory University School of Medicine, Atlanta, GA, USA. [22]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [23]Department of Clinical Medicine, Translational Neuropsychiatry Unit, Aarhus University, Aarhus, Denmark. [24]Statistical Genomics and Systems Genetics, European Bioinformatics Institute (EMBL-EBI), Cambridge, UK. [25]Human Genetics, Wellcome Trust Sanger Institute, Cambridge, UK. [26]Department of Psychiatry, University Hospital of Lausanne, Prilly, Switzerland. [27]MRC Social Genetic and Developmental Psychiatry Centre, King's College London, London, UK. [28]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Herston, Queensland, Australia. [29]Centre for Advanced Imaging, University of Queensland, Brisbane, Queensland, Australia. [30]Psychological Medicine, Cardiff University, Cardiff, UK. [31]Center for Genomic and Computational Biology, Duke University, Durham, NC, USA. [32]Department of Pediatrics, Division of Medical Genetics, Duke University, Durham, NC, USA. [33]Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [34]Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. [35]Centre for Cognitive Ageing and Cognitive Epidemiology, University of Edinburgh, Edinburgh, UK. [36]Institute of Human Genetics, University of Bonn, Bonn, Germany. [37]Life & Brain Center, Department of Genomics, University of Bonn, Bonn, Germany. [38]Psychiatry, Dokuz Eylul University School of Medicine, Izmir, Turkey. [39]Epidemiology, Erasmus MC, Rotterdam, The Netherlands. [40]Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [41]Department of Psychiatry, Massachusetts General Hospital, Boston, MA, USA. [42]Psychiatric and Neurodevelopmental Genetics Unit (PNGU), Massachusetts General Hospital, Boston, MA, USA. [43]Research, 23andMe, Inc., Mountain View, CA, USA. [44]Neuroscience and Mental Health, Cardiff University, Cardiff, UK. [45]Bioinformatics, University of British Columbia, Vancouver, British Columbia, Canada. [46]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA, USA. [47]Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA. [48]Department of Psychiatry (UPK), University of Basel, Basel, Switzerland. [49]Human Genomics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland. [50]Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. [51]Department of Psychiatry, Trinity College Dublin, Dublin, Ireland. [52]Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [53]Psychiatry and Behavioral Sciences, Johns Hopkins University, Baltimore, MD, USA. [54]Genetics and Computational Biology, QIMR Berghofer Medical Research Institute, Brisbane, Queensland, Australia. [55]Bioinformatics Research Centre, Aarhus University, Aarhus, Denmark. [56]Institute of Genetic Medicine, Newcastle University, Newcastle-upon-Tyne, UK. [57]University of Exeter Medical School, Exeter, UK. [58]Danish Headache Centre, Department of Neurology, Rigshospitalet, Glostrup, Denmark. [59]Institute of Biological Psychiatry, Mental Health Center Sct. Hans, Mental Health Services Capital Region of Denmark, Copenhagen, Denmark. [60]iPSYCH, Lundbeck Foundation Initiative for Psychiatric Research, Copenhagen, Denmark. [61]Brain and Mind Centre, University of Sydney, Sydney, New South Wales, Australia. [62]Interfaculty Institute for Genetics and Functional Genomics, Department of Functional Genomics, University Medicine and Ernst Moritz Arndt University Greifswald, Greifswald, Germany. [63]Roche Pharmaceutical Research and Early Development, Pharmaceutical Sciences, Roche Innovation Center Basel, F. Hoffmann–La Roche, Ltd, Basel, Switzerland. [64]Quantitative Health Sciences, Cleveland Clinic, Cleveland, OH, USA. [65]Statistics, Pfizer Global Research and Development, Groton, CT, USA. [66]Max Planck Institute of Psychiatry, Munich, Germany. [67]Case Comprehensive Cancer Center, Case Western Reserve University, Cleveland, OH, USA. [68]Department of Genetics and Genome Sciences, Case Western Reserve University, Cleveland, OH, USA. [69]Division of Research, Kaiser Permanente Northern California, Oakland, CA, USA. [70]Psychiatry and Behavioral Sciences, University of Southern California, Los Angeles, CA, USA. [71]Informatics Program, Boston Children's Hospital, Boston, MA, USA. [72]Department of Medicine, Brigham and Women's Hospital, Boston, MA, USA. [73]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [74]Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford, UK. [75]Department of Endocrinology at Herlev University Hospital, University of Copenhagen, Copenhagen, Denmark. [76]Swiss Institute of Bioinformatics, Lausanne, Switzerland. [77]Institute of Social and Preventive Medicine (IUMSP), University Hospital of Lausanne, Lausanne, Switzerland. [78]Department of Anesthesia, Critical Care and Pain Medicine, Massachusetts General Hospital, Boston, MA, USA. [79]Mental Health, NHS 24 Glasgow, UK. [80]Division of Psychiatry, Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, UK. [81]Department of Psychiatry and Psychotherapy, University of Bonn, Bonn, Germany. [82]Statistics, University of Oxford, Oxford, UK. [83]Psychiatry, Columbia University College of Physicians and Surgeons, New York, NY, USA. [84]School of Psychology and Counseling, Queensland University of Technology, Brisbane, Queensland, Australia. [85]Child and Youth Mental Health Service, Children's Health Queensland Hospital and Health Service, South Brisbane, Queensland, Australia. [86]Child Health Research Centre, University of Queensland, Brisbane, Queensland, Australia. [87]Estonian Genome Center, University of Tartu, Tartu, Estonia. [88]Medical Genetics, University of British Columbia, Vancouver, British Columbia, Canada. [89]Statistics, University of British Columbia, Vancouver, British Columbia, Canada. [90]DZHK (German Centre for Cardiovascular Research), partner site Greifswald, University Medicine, University Medicine Greifswald, Greifswald, Germany. [91]Institute of Clinical Chemistry and Laboratory Medicine, University Medicine Greifswald, Greifswald, Germany. [92]Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, Queensland, Australia.

[93]Humus, Reykjavik, Iceland. [94]MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK. [95]Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [96]Complex Trait Genetics, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands. [97]Clinical Genetics, Vrije Universiteit Medical Center, Amsterdam, The Netherlands. [98]Department of Psychiatry, Brigham and Women's Hospital, Boston, MA, USA. [99]Solid Biosciences, Boston, MA, USA. [100]Department of Psychiatry, Washington University in St. Louis School of Medicine, St. Louis, MO, USA. [101]Department of Biochemistry and Molecular Biology II, Institute of Neurosciences, Center for Biomedical Research, University of Granada, Granada, Spain. [102]Department of Psychiatry, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands. [103]Department of Psychiatry and Psychotherapy, Medical Center of the University of Munich, Campus Innenstadt, Munich, Germany. [104]Institute of Psychiatric Phenomics and Genomics (IPPG), Medical Center of the University of Munich, Campus Innenstadt, Munich, Germany. [105]Division of Cancer Epidemiology and Genetics, National Cancer Institute, Bethesda, MD, USA. [106]Behavioral Health Services, Kaiser Permanente Washington, Seattle, WA, USA. [107]Faculty of Medicine, Department of Psychiatry, University of Iceland, Reykjavik, Iceland. [108]School of Medicine and Dentistry, James Cook University, Townsville, Queensland, Australia. [109]Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK. [110]deCODE Genetics/Amgen, Inc., Reykjavik, Iceland. [111]Psychiatry and Human Behavior, University of Mississippi Medical Center, Jackson, MS, USA. [112]College of Biomedical and Life Sciences, Cardiff University, Cardiff, UK. [113]Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. [114]Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. [115]KG Jebsen Centre for Psychosis Research, Norway Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway. [116]Department of Psychiatry, University of California, San Diego, La Jolla, CA, USA. [117]Medical Genetics Section, CGEM, IGMM, University of Edinburgh, Edinburgh, UK. [118]Clinical Neurosciences, University of Cambridge, Cambridge, UK. [119]Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. [120]Roche Pharmaceutical Research and Early Development, Neuroscience, Ophthalmology and Rare Diseases Discovery and Translational Medicine Area, Roche Innovation Center Basel, F. Hoffmann–La Roche, Ltd, Basel, Switzerland. [121]Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. [122]Department of Psychiatry, Leiden University Medical Center, Leiden, The Netherlands. [123]Virginia Institute of Psychiatric and Behavioral Genetics, Virginia Commonwealth University, Richmond, VA, USA. [124]Computational Sciences Center of Emphasis, Pfizer Global Research and Development, Cambridge, MA, USA. [125]A full list of members and affiliations appears in Supplementary Table 14. [126]Department of Psychiatry, University of Münster, Munster, Germany. [127]Institute of Neuroscience and Medicine (INM-1), Research Center Juelich, Juelich, Germany. [128]Institute of Medical Genetics and Pathology, University Hospital Basel, University of Basel, Basel, Switzerland. [129]Amsterdam Public Health Institute, Vrije Universiteit Medical Center, Amsterdam, The Netherlands. [130]Centre for Integrative Biology, Università degli Studi di Trento, Trento, Italy. [131]Department of Psychiatry and Psychotherapy, Medical Center, Faculty of Medicine, University of Freiburg, Freiburg, Germany. [132]Psychiatry, Kaiser Permanente Northern California, San Francisco, CA, USA. [133]MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK. [134]Centre for Addiction and Mental Health, Toronto, Ontario, Canada. [135]Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada. [136]Division of Psychiatry, University College London, London, UK. [137]Neuroscience Therapeutic Area, Janssen Research and Development, LLC, Titusville, NJ, USA. [138]Institute of Molecular and Cell Biology, University of Tartu, Tartu, Estonia. [139]Psychosis Research Unit, Aarhus University Hospital, Risskov, Aarhus, Denmark. [140]Institute of Translational Medicine, University of Liverpool, Liverpool, UK. [141]Mental Health Center Copenhagen, Copenhagen University Hospital, Copenhagen, Denmark. [142]Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Groton, CT, USA. [143]Psychiatry, Harvard Medical School, Boston, MA, USA. [144]Psychiatry, University of Iowa, Iowa City, IA, USA. [145]Human Genetics Branch, NIMH Division of Intramural Research Programs, Bethesda, MD, USA. [146]Department of Psychiatry and Psychotherapy, University Medical Center Göttingen, Göttingen, Germany. [147]Faculty of Medicine, University of Iceland, Reykjavik, Iceland. [148]Child and Adolescent Psychiatry, Erasmus MC, Rotterdam, The Netherlands. [149]Psychiatry, Erasmus MC, Rotterdam, The Netherlands. [150]Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada. [151]Division of Epidemiology, New York State Psychiatric Institute, New York, NY, USA. [152]Department of Clinical Medicine, University of Copenhagen, Copenhagen, Denmark. [153]Human Genetics and Computational Biomedicine, Pfizer Global Research and Development, Cambridge, MA, USA. [154]Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [155]Department of Medical and Molecular Genetics, King's College London, London, UK. [156]Psychiatry and Behavioral Sciences, Stanford University, Stanford, CA, USA. [157]NIHR BRC for Mental Health, King's College London, London, UK. [158]Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [159]These authors contributed equally: Naomi R. Wray, Stephan Ripke, Manuel Mattheisen, Maciej Trzaskowski. [160]These authors jointly directed this work: Cathryn M. Lewis, Douglas F. Levinson, Gerome Breen, Anders D. Børglum, Patrick F. Sullivan. *e-mail: naomi.wray@uq.edu.au; pfsulliv@med.unc.edu

Methods

PGC29 cohort. Our analysis was anchored in a genome-wide association mega-analysis of 29 samples of European ancestry (16,823 MDD cases and 25,632 controls). Supplementary Table 1 summarizes the source and inclusion/exclusion criteria for cases and controls for each sample. All PGC29 samples passed a structured methodological review by MDD assessment experts (D.F.L. and K.S.K.). Cases were required to meet international consensus criteria (DSM-IV, ICD-9, or ICD-10)[97–99] 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.

Additional cohorts. We critically evaluated six independent European-ancestry cohorts (118,635 cases and 319,269 controls). Supplementary Table 2 summarizes the source and inclusion/exclusion criteria for cases and controls for each cohort. These cohorts used a range of methods for assessing MDD or major depression. Most studies included here applied otherwise typical inclusion and exclusion criteria for both cases and controls (for example, excluding cases with lifetime bipolar disorder or schizophrenia and excluding controls with major depression).

Cohort comparability. Supplementary Table 3 summarizes the numbers of cases and controls in PGC29 and the six additional cohorts. The most direct and important way to evaluate the comparability of these cohorts for a genome-wide association meta-analysis is using SNP genotype data[22,24]. We used LD score (LDSC) regression (described below) to estimate h^2_{SNP} for each cohort (Supplementary Fig. 1 and Supplementary Table 3) and r_g for all pairwise combinations of the cohorts (Supplementary Table 3b) and to demonstrate no evidence of sample overlap. We used leave-one-sample-out GRSs, finding significant differences in case–control GRS distributions for the left-out sample for all but one PGC29 sample (Supplementary Table 4). For full details of the cohort comparability analyses including GRS analyses, see the Supplementary Note. In GRS analyses, the discovery sample was the genome-wide association sample that provided the allelic weightings for each SNP, used to generate a sum score for each individual in the independent target sample.

Genotyping and quality control. Genotyping procedures can be found in the primary reports for each cohort (summarized in Supplementary Table 3). Individual genotype data for all PGC29 samples, GERA, and iPSYCH were processed using the PGC ricopili pipeline (see URLs) for standardized quality control, imputation, and analysis[19]. The 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 (see URLs)[100]. More detailed information on sample quality control is provided in the Supplementary Note.

LD score regression. LDSC was used to estimate h^2_{SNP} from genome-wide association summary statistics. Estimates of h^2_{SNP} on the liability scale depend on the assumed lifetime prevalence of MDD in the population (K), and we assumed $K = 0.15$ but also evaluated a range of estimates of K to explore sensitivity, including 95% confidence intervals (Supplementary Fig. 1). LDSC bivariate genetic correlations attributable to genome-wide SNPs (r_g) were estimated across all MDD and major depression cohorts and between the full cohort subjected to meta-analysis and other traits and disorders.

LDSC was also used to partition h^2_{SNP} by genomic features[24,46]. We tested for enrichment of h^2_{SNP} based on genomic annotations, partitioning h^2_{SNP} proportional to the base-pair length represented by each annotation. We used the 'baseline model', which consists of 53 functional categories. The categories are fully described elsewhere[46] and included conserved regions[47], USCC gene models (exons, introns, promoters, UTRs), and functional genomic annotations constructed using data from ENCODE[101] and the Roadmap Epigenomics Consortium[102]. We complemented these annotations by adding introgressed regions from the Neanderthal genome in European populations[103] 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 ($n = 135$ controls, 137 schizophrenia cases, 10 bipolar cases, and 6 affective disorder cases)[104]. Peaks called with MACS[105] (1% FDR) were retained if their coordinates overlapped in at least two samples. The peaks were recentered and set to a fixed width of 300 bp using the diffbind R package[106]. 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 250 bp on each side.

We used LDSC to estimate r_g between major depression and a range of other disorders, diseases, and human traits[22]. The intent of these comparisons was to evaluate the extent of shared common variant genetic architectures to suggest hypotheses about the fundamental genetic basis of major depression (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 r_g. These r_g values were 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 GWAS include overlapping samples, r_g remains unbiased but the intercept of the LDSC regression is an estimate of the correlation between the association statistics attributable to sample overlap. These calculations were done using the internal PGC genome-wide association library and with LD-Hub (see URLs)[60].

Integration of genome-wide association findings to tissue and cellular gene expression. We used partitioned LDSC to evaluate which somatic tissues were enriched for major depression heritability[107]. Gene expression data generated using mRNA-seq from multiple human tissues were obtained from GTEx v6p (see URLs). Genes for which <4 samples had at least one read count per million were discarded, and samples with <100 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 (for example, brain cortex versus 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 tissues (for example, cortex versus all other brain tissues) to test which brain region was the most associated with major depression, 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 100-kb 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.[45]. As these experiments were done in mice, genes were mapped to human orthologous genes using Ensembl. The coordinates for these genes were extended by a 100-kb window and tested using LD score regression as for the GTEx tissue-specific genes.

We conducted eQTL lookups of the most associated SNPs in each region and report genome-wide association SNPs in LD ($r^2 > 0.8$) with the top eQTLs in the following datasets: eQTLGen Consortium (Illumina arrays in whole blood, $n = 14,115$), BIOS (RNA-seq in whole blood, $n = 2,116$)[108], NESDA/NTR (Affymetrix arrays in whole blood, $n = 4,896$)[109], GEUVADIS (RNA-seq in LCLs, $n = 465$)[110], Rosmap (RNA-seq in cortex, $n = 494$)[111], GTEx (RNA-seq in 44 tissues, $n > 70$)[43], and Common Mind Consortium (CMC; prefrontal cortex, Sage Synapse accession syn5650509, $n = 467$)[51].

We used SMR[49] to identify loci with strong evidence of causality via gene expression and DNA methylation (eQTLs and meQTLs). SMR analysis is limited to significant cis-SNP expression (FDR < 0.05) and SNPs with MAF >0.01 at a Bonferroni-corrected pSMR. Owing 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. The aim of SMR is to prioritize variants and genes for subsequent studies, and a test for heterogeneity excludes regions that may harbor multiple causal loci (pHET < 0.05; a very conservative threshold). SMR analyses were conducted using eQTLs from the eQTLGen Consortium (whole blood), GTEx (11 brain tissues), and the CMC[43,51] as well as meQTLs from whole blood[112].

We conducted a transcriptome-wide association study[50] using precomputed expression reference weights for CMC data (5,420 genes with significant cis-SNP heritability) provided with the TWAS/FUSION software. The significance threshold was 0.05/5,420.

DNA looping using Hi-C. Dorsolateral prefrontal cortex (Brodmann area 9) was dissected from post-mortem samples from three adults of European ancestry (C.S.). Cerebra from three fetal brains were obtained from the NIH NeuroBiobank (see URLs; gestation age 17–19 weeks, African ancestry). We used 'easy Hi-C' to assess DNA chromatin (looping) interactions (Supplementary Note).

Gene-wise and pathway analyses. Our approach was guided by rigorous method comparisons conducted by PGC members[55,113]. P values quantifying the degree of association of genes and gene sets with MDD were generated using MAGMA (v1.06)[114]. 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 the 1000 Genomes Project (Phase 3 v5a, MAF ≥ 0.01)[115] 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 sets; for example, FMRP binding partners, de novo mutations, GWAS top SNPs, ion channels)[57,116,117]. 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, the Pathway Interaction Database, Reactome, Sigma-Aldrich, Signaling Gateway, Signal Transduction KE, and SuperArray. Pathways containing from 10–10,000 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 NIMH Psychoactive Drug Screening Program Ki DB[120] downloaded in June 2016. The association of 3,885 drug gene sets with major depression 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.[121]. 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 (see 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. The 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 k equal to the square root of the number of pathways, m equal to the square root of k, $l = 1$ (default), $w = 1$ (default), and b equal to the 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. We conducted bidirectional MR[122] analysis for four traits: years of education (EDY)[62], body mass index (BMI)[29], coronary artery disease (CAD)[63], and schizophrenia (SCZ)[19]. We denote z as a genetic variant (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)[123] approach: $\hat{b}_{xy} = \hat{b}_{zy}/\hat{b}_{zx}$, where $\hat{b}_{zx}$ is the estimated effect size for the SNP–trait association for the exposure trait and $\hat{b}_{zy}$ is the effect size estimated for the same SNP in the GWAS of the outcome trait.

We used generalized summary statistics–based MR (GSMR)[64] to estimate $\hat{b}_{xy}$ and its standard error from multiple SNPs associated with the exposure trait at a genome-wide significance level. We conducted bidirectional GSMR analyses for each pair of traits and report results after excluding SNPs that failed 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 $\hat{b}_{zx}$ and $\hat{b}_{zy}$. GSMR also accounts for residual LD between the clumped SNPs. For comparison, we also conducted IVW-MR and MR-Egger analyses[124]. More details are provided in the Supplementary Note.

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

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

Data availability. The PGC's policy is to make genome-wide summary results public. Summary statistics for a combined meta-analysis of PGC29 with five of the six expanded samples (deCODE, Generation Scotland, GERA, iPSYCH, and UK Biobank) are available on the PGC website (see URLs). Results for 10,000 SNPs for all seven cohorts are also available on the PGC website.

Genome-wide association summary statistics for the Hyde et al. cohort (23andMe, Inc.) must be obtained separately. These can be obtained by qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Contact David Hinds (dhinds@23andme.com) to apply for access to the data. Researchers who have the 23andMe summary statistics can readily recreate our results by performing meta-analysis of the six-cohort results file with the Hyde et al. results file from 23andMe[28].

The availability of genotype data for PGC29 is described in Supplementary Table 15. For the expanded cohorts, interested users should contact the lead principal investigators of these cohorts (which are separate from the PGC).

References

97. World Health Organization. *International Classification of Diseases* (World Health Organization, Geneva, 1978).
98. World Health Organization. *International Classification of Diseases* (World Health Organization, Geneva, 1992).
99. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders* (American Psychiatric Association, Washington, DC, 1994).
100. Abecasis, G. R. et al. A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
101. ENCODE Project Consortium. A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
102. Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
103. Vernot, B. et al. Excavating Neandertal and Denisovan DNA from the genomes of Melanesian individuals. *Science* **352**, 235–239 (2016).
104. Bryois, J. et al. Evaluation of chromatin accessibility in prefrontal cortex of schizophrenia cases and controls. *Nat. Commun.* (in the press).
105. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
106. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
107. Finucane, H. et al. Heritability enrichment of specifically expressed genes identifies disease-relevant tissues and cell types. *Nat. Genet.* https://dx.doi.org/10.1038/s41588-018-0081-4 (2018).
108. Zhernakova, D. V. et al. Identification of context-dependent expression quantitative trait loci in whole blood. *Nat. Genet.* **49**, 139–145 (2017).
109. Jansen, R. et al. Conditional eQTL analysis reveals allelic heterogeneity of gene expression. *Hum. Mol. Genet.* **26**, 1444–1451 (2017).
110. Lappalainen, T. et al. Transcriptome and genome sequencing uncovers functional variation in humans. *Nature* **501**, 506–511 (2013).
111. Ng, B. et al. An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
112. Hannon, E., Weedon, M., Bray, N., O'Donovan, M. & Mill, J. Pleiotropic effects of trait-associated genetic variation on DNA methylation: utility for refining GWAS loci. *Am. J. Hum. Genet.* **100**, 954–959 (2017).
113. de Leeuw, C. A., Neale, B. M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. *Nat. Rev. Genet.* **17**, 353–364 (2016).
114. de Leeuw, C. A., Mooij, J. M., Heskes, T. & Posthuma, D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput. Biol.* **11**, e1004219 (2015).
115. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
116. Turner, T. N. et al. denovo-db: a compendium of human de novo variants. *Nucleic Acids Res.* **45** (D1), D804–D811 (2017).
117. Pirooznia, M. et al. High-throughput sequencing of the synaptome in major depressive disorder. *Mol. Psychiatry* **21**, 650–655 (2016).
118. Liberzon, A. et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
119. Wagner, A. H. et al. DGIdb 2.0: mining clinically relevant drug–gene interactions. *Nucleic Acids Res.* **44** (D1), D1036–D1044 (2016).
120. Roth, B. L., Kroeze, W. K., Patel, S. & Lopez, E. The multiplicity of serotonin receptors: uselessly diverse molecules or an embarrasment of riches? *Neuroscientist* **6**, 252–262 (2000).
121. Olier, I., Vellido, A. & Giraldo, J. Kernel generative topographic mapping. in *ESANN 2010 Proc.* 28–30 (2010).
122. Smith, G. D. & Ebrahim, S. 'Mendelian randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int. J. Epidemiol.* **32**, 1–22 (2003).
123. Wooldridge, J. M. *Introductory Econometrics: A Modern Approach* (Cengage Learning, Boston, MA, 2015).
124. Bowden, J., Davey Smith, G. & Burgess, S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int. J. Epidemiol.* **44**, 512–525 (2015).

nature research

Corresponding author(s):   Naomi Wray and Patrick Sullivan

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	This is fully described in the Online Methods and associated Supplemental Tables
Data analysis	Software used is described in the Online Methods.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

See the last two paragraphs in the Online Methods and Table S14.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Variable by analysis but given in the manuscript
Data exclusions	Exclusions done during GWAS quality control; described in detail in the paper
Replication	NA
Randomization	NA
Blinding	This was not practical given our design.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	This is fully described in the Online Methods and associated Supplemental Tables
Recruitment	This is fully described in the Online Methods and associated Supplemental Tables