

A methylome-wide association study of major depression with out-of-sample case–control classification and trans-ancestry comparison

Received: 12 January 2024

Accepted: 23 July 2025

Published online: 16 September 2025

 Check for updates

A list of authors and their affiliations appears at the end of the paper

Major depression (MD) is a leading cause of global disease burden, and both experimental and population-based studies suggest that differences in DNA methylation may be associated with the condition. However, previous DNA methylation studies have, so far, not been widely replicated, suggesting a need for larger meta-analysis studies. Here we conducted a meta-analysis of methylome-wide association analysis for lifetime MD across 18 studies of 24,754 European-ancestry participants (5,443 MD cases) and an East Asian sample (243 cases, 1,846 controls). We identified 15 CpG sites associated with lifetime MD with methylome-wide significance. The methylation score created using the methylome-wide association analysis summary statistics was significantly associated with MD status in an out-of-sample classification analysis (area under the curve 0.53). Methylation score was also associated with five inflammatory markers, with the strongest association found with tumor necrosis factor beta. Mendelian randomization analysis revealed 23 CpG sites potentially causally linked to MD, with 7 replicated in an independent dataset. Our study provides evidence that variations in DNA methylation are associated with MD, and further evidence supporting involvement of the immune system.

Major depression (MD) is a common psychiatric disorder arising from a complex combination of genetic and environmental factors that include lifestyle factors such as physical activity, smoking, alcohol consumption and body mass index (BMI)^{1–3}. The heritability of MD estimated from twin studies is 37% (ref. 4), and polygenic risk scores trained on genome-wide association studies (GWASs) currently explain 1.5–3.2% of the variance in MD⁵.

DNA methylation (DNAm) is one of the most studied epigenetic processes and is influenced by both genetic and environmental factors⁵. DNAm is dynamic and is associated with changes in environmental and lifestyle factors, including smoking⁶, alcohol⁷ and BMI⁸, all factors that are implicated in MD³. A recent study⁹ identified associations between methylation scores (MSs), calculated using methylome-wide

association study (MWAS) summary statistics for several relevant lifestyle factors, and MD¹⁰. These environmental MS measures were able to capture additional variation associated with MD when added to direct lifestyle measures, and it is thought this may be due to their ability to act as an archive of environmental exposure.

There is growing evidence from MWAS that DNAm measured in whole blood may be associated with MD. Jovanova et al.¹¹ looked at 11 cohorts comprising 11,256 participants of European and African ancestry and identified 3 cytosine–phosphate–guanine (CpG) sites associated with depressive symptoms, which were annotated to genes implicated in axon guidance¹¹. Starnawska et al.¹² investigated depressive symptomatology in a sample of 724 monozygotic twins, with top findings annotated to genes previously implicated

✉ e-mail: xueyi.shen@ed.ac.uk; andrew.mcintosh@ed.ac.uk

Table 1 | Information for cohorts that participated in the methylome-wide meta-analysis

Cohort name	PMID	N cases	Sample size	MD prevalence (%)	Case phenotype	Mean age	Gender (% male)	Genetic ethnicity	DNAm array	Tissue type
Avon Longitudinal Study of Parents and Children/ Accessible Resource for Integrated Epigenomic Studies	25991711	204	657	31.1	Questionnaire	29.4/45	0	European	450K	Whole blood
The biological classification of mental disorders/ the Optimized Treatment Identification at the Max Planck Institute	32393358; 33054737	222	305	72.8	Clinical interview	37.5	38.4	European	EPIC	Whole blood
BiDirect	24924233	287	561	51.5	Questionnaire	51.6	49.7	European	EPIC	Whole blood
Dublin Cohort	26419460	37	186	19.9	Questionnaire	38.0	52.7	European	450K	PBMCs
Environmental Risk Longitudinal Twin Study	12236608; 12537874	322	1,625	19.8	Clinical interview	18.5	51.0	Predominantly European	450K	Whole blood
EXTEND	29790996	190	1,181	16.1	Self-reported diagnosis/GP visit	56.3	48.0	Predominantly European	450K	Whole blood
Finnish Twin Cohort (450K set)	31796134; 31640839	163	985	16.5	Clinical interview	23.8	45.9	Finnish, European	450K	Whole blood
Finnish Twin Cohort (EPIC set)	31796134; 31640839	93	366	25.4	Clinical interview	25.4	49.5	Finnish, European	EPIC	Whole blood
FOR2107	30267149	340	667	51.0	Clinical interview	35.3	36.7	European	EPIC	Whole blood
GS (sets 1 and 2)	30918249	1,632	9,502	17.2	Clinical interview	49.8	41.0	European	EPIC	Whole blood
Janssen	36319817	191	222	86.0	DSM-criteria questionnaire	42.8	19.7	Predominantly European	EPIC	Whole blood
Biopsychosocial factors of major depression in youth/University Hospital of Munich	30947532	381	633	60.2	Clinical interview	15.1	33.3	European	EPIC	Whole blood
Munich Antidepressant Response Signature/ GlaxoSmithKline	18586274; 16822851	311	497	62.6	DSM-criteria questionnaire	48.1	42.5	European	450K	Whole blood
Netherlands Twin Register	31666148	436	2,701	16.1	Clinical interview	36.5	33.0	European	450K	Whole blood
Study of Health in Pomerania-Trend	35348705	84	492	17.1	DSM-criteria questionnaire	51.1	46.3	European	EPIC	Whole blood
TwinsUK	30760334	201	692	29.0	Clinical interview	59.0	0	European	450K	Whole blood
Understanding Society/UK Household Longitudinal Study Set 1	30401456	72	1,121	6.4	Self-reported diagnosis/GP visit	58.4	41.9	European	EPIC	Whole blood
Understanding Society/UK Household Longitudinal Study Set 2	30401456	277	2,361	11.7	Self-reported diagnosis/GP visit	51.1	45.9	European	EPIC	Whole blood

Mean ages are reported for two subsets of the Avon Longitudinal Study of Parents and Children cohort. *DSM*, *Diagnostic and Statistical Manual of Mental Disorders*⁵⁴; GP, general practitioner; PMID, PubMed identifier.

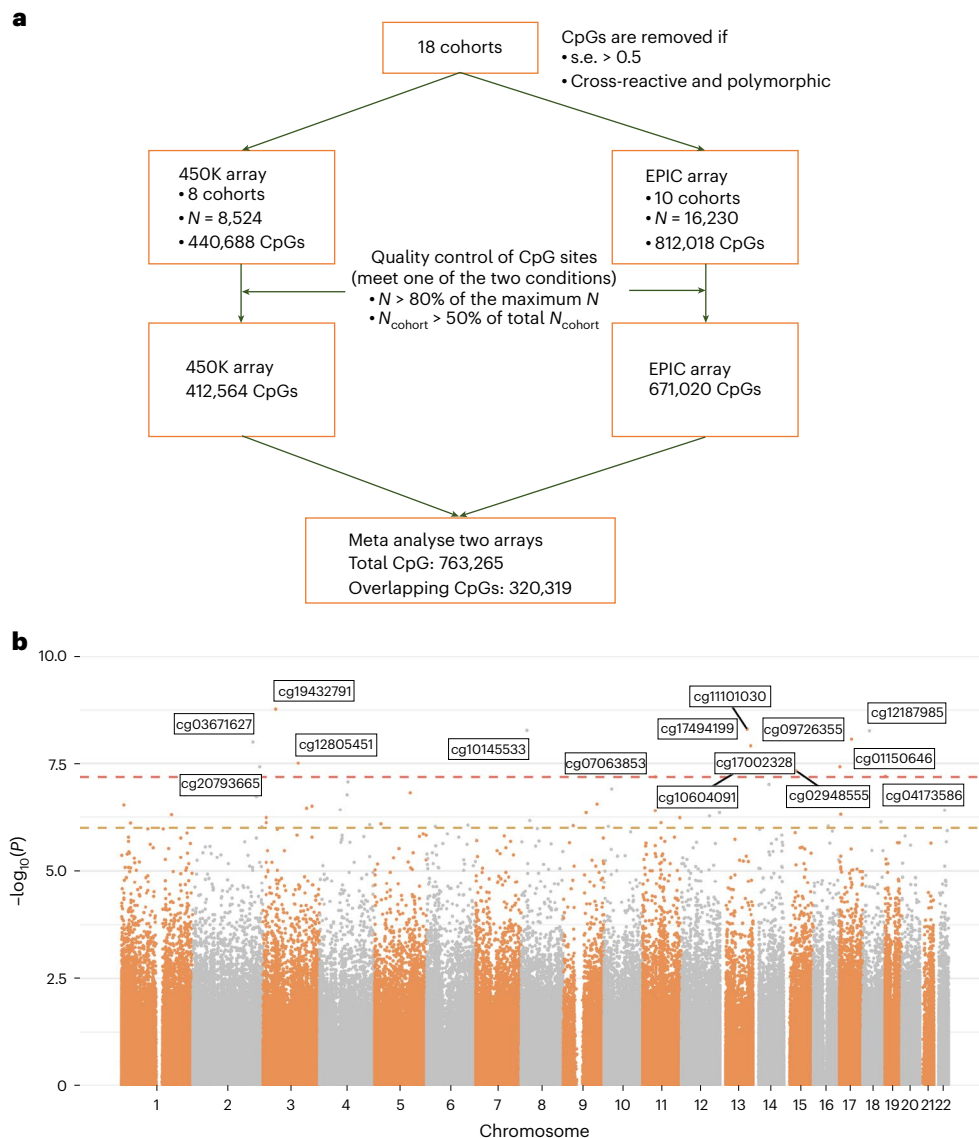


Fig. 1 | Meta-analysis of MWAS for MD. a, Workflow of meta-analysis. **b**, Manhattan plot for the meta-analysis of MWAS for MD. Each dot represents a CpG probe. The x axis represents the relative positions of the probes in the

genome. The y axis represents $-\log_{10}$ -transformed, two-sided P values. The red and yellow dashed lines represent the significance threshold for Bonferroni and FDR correction, respectively.

in depression¹². Finally, Huls et al.¹³ identified DNAm associations with MD in dorsal lateral prefrontal cortex samples from $N = 608$ participants and uncovered CpGs annotated to several genes that are relevant to MD¹³. In addition, MS calculated using penalized regression models have previously been utilized to investigate MD¹⁴. We recently used lasso regression to calculate an MD-MS in $N = 1,780$ participants (cases = 363; controls = 1,417) and found that the scores explain approximately 1.75% of the variance in prevalent MD, acting additively with polygenic risk score¹⁴. Findings from these studies have been somewhat inconsistent, and they have either included only European-ancestry samples^{15,16} or used a combined multi-ancestry sample without considering trans-ancestry differences¹¹. This has limited generalizability of previous findings. Larger MWAS studies may provide more reliable estimates of the differences in DNAm between MD cases and controls and, in doing so, bring insight into the molecular mechanisms associated with MD. We further hypothesized that DNAm associations with MDD would generalize across ancestries and sought to test this in a case-control sample of East Asian ancestry.

Given previous evidence highlighting the role of DNAm in MD, we conducted an MD MWAS meta-analysis using whole-blood DNAm data from 18 cohorts comprising 24,754 individuals (5,443 cases) of European ancestry. We sought to identify whether differences in DNAm were a potential cause or consequence of MD using a two-sample Mendelian randomization framework. Further, we trained a DNAm classifier of MD status from our summary statistics and assessed whether it could classify MD case-control status in an independent testing sample and its association with the abundance of inflammatory protein markers. Finally, we assessed whether significant effect sizes were positively associated with those in an independent East Asian sample.

Results

MD meta-MWAS

A total of 15 CpG sites were significantly associated with MD after Bonferroni correction in the basic model ($P < 6.55 \times 10^{-8}$, $P_{\text{Bonferroni}} < 0.05$; see Table 1 for details of participating studies and Fig. 1 and Table 2 for significant findings).

Table 2 | Significant CpG sites in the meta-analysis of MD MWAS

CpG	Chromosome	Base pair location	Beta	s.e.	P	P-adj	N	UCSC gene	Relation to Island	Array	EWAS catalog/ EWAS atlas (PMID)	GWAS catalog gene look-up (PMID)
cg19432791	3	45894556	0.025	0.004	1.70×10^{-9}	0.001	16,545	LZTFL1	OpenSea	EPIC	C-reactive protein (CRP) ^a (38692281); glucose (33151971); schizophrenia (26619358)	Multiple sclerosis (31604244); COVID-19 (35910207); CD3 protein levels (32929287)
cg11101030	13	93744566	-0.035	0.006	4.99×10^{-9}	0.004	15,901	-	OpenSea	EPIC	-	-
cg10145533	8	20350779	0.019	0.003	5.34×10^{-9}	0.004	24,191	-	OpenSea	Both	Crohn's disease ^a (27279921); preterm birth ^a (28428831); autoimmune disease ^a (31024609)	-
cg12187985	18	23654410	0.021	0.004	5.43×10^{-9}	0.004	16,850	SS18	OpenSea	EPIC	CRP ^a (38692281); preterm birth (28428831); rheumatoid arthritis (23334450); LRIG1 protein (35945220)	Neuroticism (30595370); autoimmune thyroid disease (32581359)
cg09726355	17	42544621	0.019	0.003	8.53×10^{-9}	0.007	16,628	GPATCH8	OpenSea	EPIC	Rheumatoid arthritis (23334450); HIV infection (27105112); alcohol consumption (27843151)	Cerebrospinal fluid progranulin levels (33087363); brain morphology (32665545); educational attainment (35361970)
cg03671627	2	209185667	0.026	0.005	9.85×10^{-9}	0.008	16,850	PIKFYVE	OpenSea	EPIC	CRP ^a (38692281); kidney disease ^a (33933144); alcohol consumption (31789449); childhood obesity (28614626)	BMI (36581621); asthma (27611488)
cg17494199	13	106834942	-0.011	0.002	1.22×10^{-8}	0.009	24,748	-	OpenSea	Both	Very early preterm birth ^a (30833390); birth weight ^a (31015461); fetal brain development ^a (25650246)	-
cg17002328	14	91751773	0.016	0.003	2.32×10^{-8}	0.018	24,750	CCDC88C	N_Shelf	Both	CRP ^a (38692281); multiple sclerosis ^a (30479356); chronic fatigue syndrome (30036399); schizophrenia (33279932); autism (27404287); BMI (33517419); ADHD (28785368)	White-matter hyperintensity volume (33293549); subcortical volume (32665545); Autism (35215271); Alzheimer's disease (26830138)
cg12805451	3	123672734	0.024	0.004	3.11×10^{-8}	0.024	16,850	CCDC14	OpenSea	EPIC	-	intelligence (29844566); general cognitive ability (29326435)
cg20793665	2	232549224	0.019	0.003	3.79×10^{-8}	0.029	24,752	-	S_Shelf	Both	CRP ^a (38692281)	-

Table 2 (continued) | Significant CpG sites in the meta-analysis of MD MWAS

CpG	Chromosome	Base pair location	Beta	s.e.	P	P-adj	N	UCSC gene	Relation to Island	Array	EWAS catalog/ EWAS atlas (PMID)	GWAS catalog gene look-up (PMID)
cg01150646	17	1377304	0.013	0.002	3.80×10 ⁻⁸	0.029	16,850	MYO1C	S_Shore	EPIC	Insulin resistance (30792424); BMI (29278407); Alzheimer's disease (33257653); alcohol consumption (27843151); preterm birth (28428831)	Alzheimer's disease (35770850); educational attainment (30038396)
cg02948555	15	42749848	0.036	0.007	4.72×10 ⁻⁸	0.036	15,911	ZNF106	OpenSea	EPIC	Downs syndrome (33547282); early onset intracranial atherosclerotic stenosis (31142690)	CRP (31900758); cholesterol to total lipid ratio in low-density lipoprotein cholesterol (35213538)
cg10604091	13	50650086	0.023	0.004	5.27×10 ⁻⁸	0.040	16,850	DLEU2	OpenSea	EPIC	BMI (29278407); inflammatory bowel disease (32281463); Alzheimer's disease (25129075); birth weight (31015461); insufficient sleep (30718923)	BMI (31669095); creatinine (35710981, 34594039); non-high-density lipoprotein cholesterol levels (34887591)
cg04173586	19	2167496	0.025	0.005	6.30×10 ⁻⁸	0.048	24,043	DOT1L	S_Shore	Both	Schizophrenia ^a (33646943); alcohol consumption ^a (31789449)	BMI-adjusted waist/hip ratio (30575882); gamma-linolenic acid (26584805)
cg07063853	11	43333512	0.020	0.004	6.42×10 ⁻⁸	0.049	23,395	API5	Island	Both	HIV infection ^a (27105112); obesity (29692867); rheumatoid arthritis (27585642); Alzheimer's disease (33257653)	–

EWAS catalog (<http://www.ewascatalog.org/>), EWAS atlas (<https://ngdc.cncb.ac.cn/ewas/atlas>) and GWAS catalog (<https://www.ebi.ac.uk/gwas/>) were used to find associated traits in previous studies (search was based on genes). Beta, standardized regression coefficient; P, double-sided P value; P-adj, Bonferroni-corrected P value. ^aTraits associated with CpG sites.

Five of the significant CpG sites positionally mapped to genes associated with mental health, neurodegenerative and developmental disorders. The gene mapped from cg17002328 (*CCDC88C*) was associated with schizophrenia¹⁷ and attention-deficit/hyperactivity disorder (ADHD)¹⁶ in epigenetic studies and with brain structural measures in GWAS, such as cortical surface area¹⁸ and accumbens volume¹⁹. CpG site cg17494199 was associated with preterm birth²⁰, birth weight²¹ and fetal brain development²² in previous epigenetic studies. Genes mapped from cg01150646 (*MYO1C*), cg10604091 (*DLEU2*) and cg07063853 (*API5*) were associated with Alzheimer's disease in brain-tissue methylation levels^{23–25}.

A total of five significant CpG sites were associated with autoimmune diseases and biomarkers in previous studies, one of which was the most significantly CpG site (cg19432791 on chromosome 3), mapped to gene *LZTFL1*. From the catalog of epigenome-wide association studies (EWAS catalog), the *LZTFL1* gene was associated with biomarkers

related to pain, such as glucose level²⁶, and autoimmune diseases or markers, such as rheumatoid arthritis²⁷ and C9 protein levels²⁸. Four of the significant CpG sites, cg12187985 (*SSI8*), cg10145533, cg09726355 (*GPATCH8*) and cg07063853 (*API5*), map to genes previously found associated with autoimmune markers and diseases (for example, rheumatoid arthritis²⁷) in MWAS studies.

Other CpG sites, cg03671627 and cg02948555 (*ZNF106*), were associated with traits and markers relevant to obesity in both genetic and epigenetic studies. These markers include, for example, BMI²⁹ and the ratio of cholesterol to total lipids³⁰.

A complete list of CpG sites and related genes and traits can be found in Table 2. Quantile–quantile plot of the meta-analysis and the inflation factors of the association statistics for each individual cohort can be found in Supplementary Figs. 3 and 4, respectively. See Supplementary Fig. 5 for effect sizes of each study for the significant CpG sites.

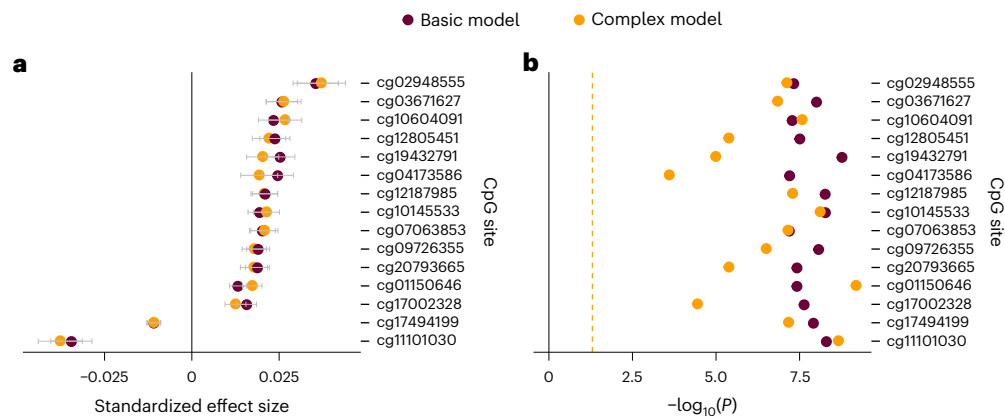


Fig. 2 | Comparison between basic and complex models. a, b. In both comparison of effect sizes indicated by standardized coefficients of linear regression (**a**) and comparison of double-sided P values (**b**), each dot represents a CpG probe. In **a**, the x axis represents effect size, and in **b**, the x axis represents $-\log_{10}$ -transformed, double-sided P value. Error bars represent \pm one standard

error of the mean. Differences in P values reflect both the model used and the sample sizes. The y axis represents individual CpG sites. The yellow dashed line in **b** represents the significance threshold for nominal significance (double-sided $P < 0.05$).

Identification of DMRs

A total of 37 differentially methylated regions (DMRs) were identified to be associated with MD after Bonferroni correction. The largest associated region locates within the major histocompatibility complex region, discoidin domain receptor tyrosine kinase 1 (*DDR1*) gene (chr. 6:30853258–30854233), previously implicated in major psychiatric disorders (for example, schizophrenia, MD and bipolar disorder).

The full list of significant DMRs can be found in Supplementary Table 1.

Pathway enrichment analysis

Following the meta-MWAS, pathway enrichment analysis was conducted to identify potential biological pathways that the significant CpG sites were implicated in. No gene ontology (GO) term or Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway was significantly enriched after false discovery rate (FDR) correction (Supplementary Table 2). The top ten most significantly enriched GO terms included pathways relevant to protein and metabolic processes (for example, negative regulation of protein localization to ciliary membrane) (P ranged from 0.003 to 8.14×10^{-4}). For KEGG, 'transcriptional misregulation in cancer' and 'lysine degradation' were the only pathways that reached nominal significance ($P \leq 0.007$).

The basic model versus complex model

We compared results for the basic model (18 cohorts, $N_{\text{total}} = 24,754$) and complex model (15 cohorts, $N_{\text{total}} = 20,196$; Supplementary Fig. 1) to evaluate potential confounding effects of BMI and alcohol consumption (Fig. 2). Effect sizes of the two models were highly correlated for the significant CpG sites ($r = 0.988$) and for CpG sites across the entire methylome ($r = 0.920$). All significant CpG sites in the basic model remained significant in the fully adjusted model after Bonferroni correction across the 15 significant CpGs found in the basic model, despite a significant reduction in sample size ($P < 2.51 \times 10^{-4}$, $P_{\text{Bonferroni}} < 0.004$ corrected across the significant CpG sites in discovery analysis).

Out-of-sample classification of MD using methylation scores

MD methylation scores (MD–MSs) were used to assess whether the present MD MWAS summary statistics can be used to classify MD case–control status in an independent sample. All five MD–MSs showed positive effect sizes (higher score associated with higher liability of MD). There was an increasing trend of effect sizes as the P -value threshold for MS calculation became increasingly stringent (Fig. 3). Of the five

scores tested, only MS at P -value threshold of $\leq 5 \times 10^{-8}$ was found to be associated with MD diagnosis ($\beta = 0.13$, $P = 0.003$, area under the curve = 0.53; Fig. 3). P values for other MSs ranged from 0.069 to 0.314.

Association between MD–MS and inflammatory protein abundance

Following out-of-sample classification using MD–MS, we sought to identify inflammatory proteins of which the abundance was associated with MD–MS. Five proteins were significantly ($P < 3.5 \times 10^{-3}$) associated with the MD–MS created at the P threshold of $\leq 5 \times 10^{-8}$. The strongest association was found in tumor necrosis factor B ($\beta = -0.15$, $P = 2.37 \times 10^{-5}$). Other proteins were interleukin 6 ($\beta = 0.12$, $P = 5.18 \times 10^{-4}$), transforming growth factor alpha ($\beta = 0.11$, $P = 2.8 \times 10^{-3}$), CD5 ($\beta = -0.11$, $P = 2.1 \times 10^{-3}$) and EN-RAGE ($\beta = 0.11$, $P = 3.5 \times 10^{-3}$; Supplementary Fig. 6).

Heterogeneity analysis

Leave-one-out analysis. We systematically investigated the heterogeneity of effect sizes in each individual study by conducting leave-one-out (LOO) analysis. For every CpG site, all effect sizes for leave-one-out analyses remained in the same direction as the meta-analysis (Supplementary Fig. 7). All tested CpG sites remained significant after leaving individual studies out except when the largest study, Generation Scotland (GS), was left out ($P < 1.03 \times 10^{-4}$). When GS was left out, 11 out of the 15 tested CpG sites remained nominally significant ($P < 8.91 \times 10^{-3}$). The four sites that became non-significant when GS was omitted were cg02948555 ($\beta_{\text{meta-MWAS}} = 0.036$, $\beta_{\text{LOO}} = 0.008$, $P_{\text{LOO}} = 0.511$), cg11101030 ($\beta_{\text{meta-MWAS}} = -0.035$, $\beta_{\text{LOO}} = -0.01$, $P_{\text{LOO}} = 0.235$), cg07063853 ($\beta_{\text{meta-MWAS}} = 0.02$, $\beta_{\text{LOO}} = 0.01$, $P_{\text{LOO}} = 0.232$) and cg01150646 ($\beta_{\text{meta-MWAS}} = 0.013$, $\beta_{\text{LOO}} = 0.007$, $P_{\text{LOO}} = 0.054$).

Effect of age difference between studies. Due to systematic differences of age range across studies, we conducted meta-regression to investigate the impact of age difference on the findings of meta-analysis. Of the 15 CpG sites significant in the meta-MWAS, 14 sites did not show an effect of age difference across studies (absolute β ranged from 1.11×10^{-6} to 6.09×10^{-4} , $P > 0.234$; Supplementary Fig. 8). One site, cg04173586, showed a significant effect of age ($\beta = -9.25 \times 10^{-4}$, $P = 0.006$). However, leave-one-out analysis showed highly consistent findings for cg04173586 across studies. There were 14 out of 17 participating cohorts that showed effect sizes consistent with the meta-MWAS (Supplementary Fig. 5) and leave-one-out analyses were significant for all iterations (P ranged from 1.03×10^{-4} to 2.38×10^{-8} ; Supplementary Fig. 5 and Supplementary Table 3).

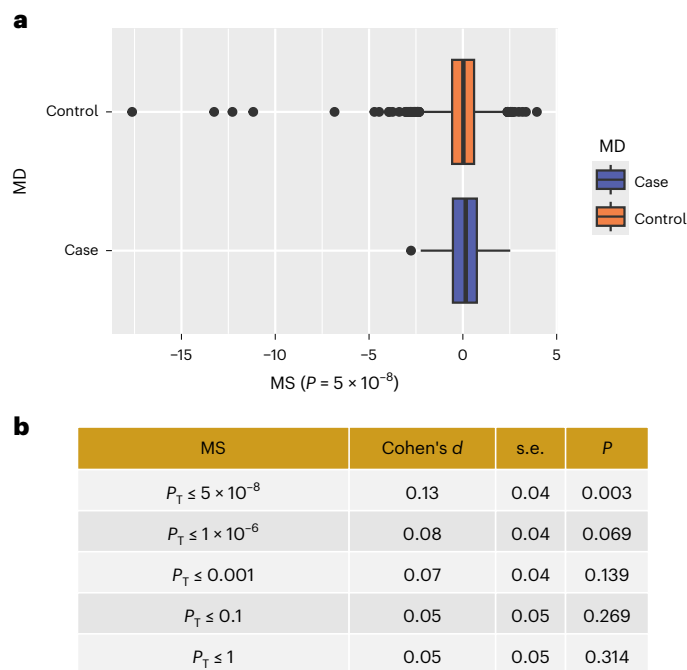


Fig. 3 | Out-of-sample classification of MD using MS. a, Box plot for case–control classification using MS created using P -value threshold at 5×10^{-8} . MS was standardized and residualized against age, sex and values for aryl hydrocarbon receptor repressor (AHRR) probes. Residuals of MSs were used for the plot. The x axis represents MS, and the y axis represents MD case and control groups. Box boundaries are determined by the top and bottom quartiles. Box whiskers are defined by 1.5 interquartile range. Any values outside of the box whiskers are presented as individual data points. Median values per group are indicated as vertical lines within the boxes. **b**, Statistics for MS association tests. Cohen's d was calculated by comparing MD case–control differences in MS. Double-sided P values are reported. P_T , P threshold for creating the MS.

Correlation matrix for effect sizes. Heterogeneity between studies was analyzed by looking at the between-study correlation of effect sizes estimated using the basic model (Supplementary Fig. 9). Correlations between the 18 studies participating in the MD meta-MWAS ranged from -0.19 to 0.31 . The highest positive correlation was found between the Netherlands Twin Register and Janssen ($r = 0.31$). Of the 153 pairwise correlations, 96 were positive (62.7%). Compared with the MD meta-MWAS, BMI MWAS (10 studies) showed higher, positive effect size correlations between studies (r ranged from 0.305 to 0.864 , 100% of the pairs were positively and significantly correlated).

Mendelian randomization

Using *cis* mQTL data from GS and Wald ratio Mendelian randomization (MR) method, we found 23 significant and potentially causal effects of DNA methylation on MD (absolute β ranged from 0.06 to 0.93 , P ranged from 6.88×10^{-3} to 4.58×10^{-6} ; Fig. 4 and Supplementary Data 1). There were 17 CpG sites located in the major histocompatibility complex region (mapping to the *DDR1R* gene) on liability to MD (β ranged from 0.06 to 0.17 , P ranged from 1.1×10^{-3} to 8.05×10^{-5}). See Supplementary Data 1 for the full list of significant causal effects found in the discovery analysis in GS.

Of the 23 potentially causal effects found in the discovery analysis, 4 CpG sites were specific to the EPIC array, while 19 were available on both array types and could, therefore, be included in the replication analysis. Seven CpG sites tested were replicated ($P < 0.003$; Fig. 4). Six of the replicated CpG sites were in the major histocompatibility complex region (β 0.040 to 0.073 , P 1.22×10^{-4} to 1.19×10^{-9}). In addition, a potentially causal effect from cg26829071 to MD (mapping to *GPR133*, $\beta = -0.056$, $P = 0.004$) was found on chromosome 12.

No significant causal effects from MD to DNAm were found (absolute β 1.62×10^{-3} to 0.22 , $P_{\text{uncorrected}} > 0.001$, $P_{\text{FDR}} > 0.66$).

MWAS in East Asian ancestry

We sought to identify DNAm associations with MD in East Asian ancestry using data from Taiwan Biobank. There were no methylome-wide significant findings identified. Correlation between the methylome-wide significant CpGs effect sizes identified in the main results ($N = 15$) and the same CpGs in Taiwan Biobank was $r = 0.48$. The effect direction was consistent between the two cohorts for 11 of the 15 CpGs (73.3% in the same direction; Supplementary Tables 4–6). Effect sizes were significantly and positively correlated when we extended the comparison in the top 100 and top 1,000 CpG sites found within the MWAS from European samples (top 100 CpG sites found in European samples: $r = 0.26$, $P = 0.009$; top 1,000 CpG sites found in European samples: $r = 0.22$, $P = 4.07 \times 10^{-12}$).

Discussion

Our meta-MWAS of 24,754 individuals found 15 CpG sites to be significantly associated with MD, an increase of 12 CpGs on previously reported associations¹¹. Significant findings revealed CpGs mapped to genes associated with autoimmune markers and depression-related traits, such as BMI. Seven potentially causal effects from DNAm to MD were identified and replicated. Leave-one-out analysis showed that effects were highly consistent across studies for the significant sites. However, the correlation in CpG–MD association effect sizes was relatively low and showed considerable between-study heterogeneity for MD, in contrast to BMI, which showed uniformly positive and more homogeneous correlations across all available studies. A positive correlation was observed for top MD CpG sites between European and East Asian samples ($r = 0.48$ for significant CpG sites and $r = 0.26$ for top 100 CpG sites), and effects of 11 of the 15 significant CpG sites found in the European sample remained in the same direction in the East Asian sample.

Five CpG sites mapped to genes previously reported in association with autoimmune disorders and markers relevant to activity of the hypothalamic–pituitary–adrenal (HPA) axis. The top CpG, cg19432791, mapped to *LZTFL1*, a gene that was identified as an effector gene that contributes to severe autoimmune responses and inflammation, such as risk of respiratory failure caused by COVID-19³¹. In addition to its association with inflammation, *LZTFL1* regulates ciliary localization of the Bardet–Biedl syndrome protein complex. The Bardet–Biedl syndrome complex is a well-replicated causative protein marker for obesity³² and highly relevant to HPA-axis activity by being involved in transducing leptin signals in hypothalamic neurons^{33,34}. Several top CpG sites and mapped genes in our MWAS were previously linked to other psychiatric disorders. For example, two CpG sites, cg04173586 and cg19432791, were associated with schizophrenia^{35,36}. Another CpG site, cg17002328, mapped to gene *CCDC88C*, was found to be related to multiple psychiatric disorders, including schizophrenia¹⁷, attention-deficit/hyperactivity disorder¹⁶ and autism³⁷. Four of the 11 genes identified in our MWAS (*LZTFL1*, *SSI8*, *CCDC88C* and *DLEU2*) were found in significant DMRs associated with suicide attempts in adults with bipolar disorder³⁸. Beyond psychiatric disorders, five CpG sites showed associations with CRP^{8,39}. CRP is one of the most well-studied predictors of vulnerability to depression⁴⁰ and persistence of depressive symptoms⁴¹. As a pro-inflammatory cytokine, CRP has been associated with stress-related hyperactivity in the HPA axis⁴². The proteome-wide association analysis study (PWAS) findings provided additional evidence for the association between the top CpG sites and inflammation, for example, in tumor necrosis factor beta and interleukin 6 (IL6). Our findings provide further evidence supporting the association between MD and chronic inflammation, particularly involving the HPA axis. However, no association between these results and MD had been identified in previous MWAS, probably due to the limited sample sizes of MWAS conducted on mental health-related traits.

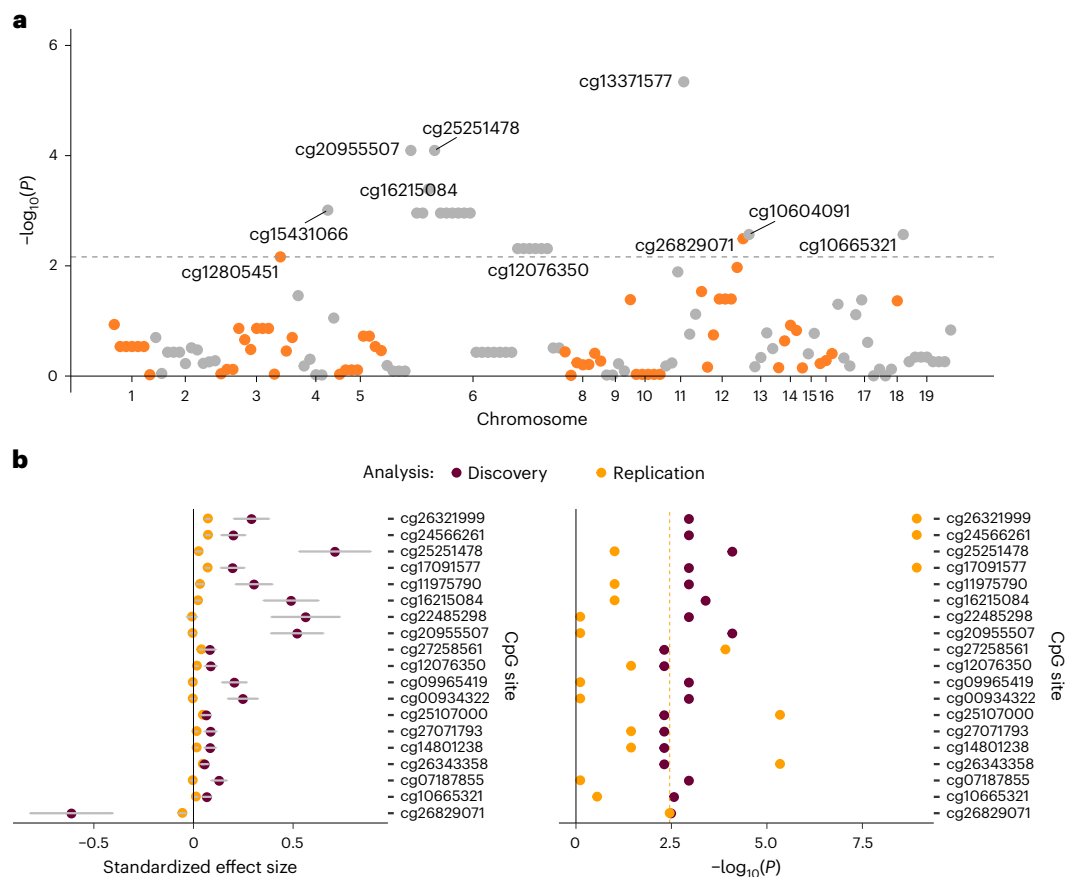


Fig. 4 | MR analysis of the causal effect of DNAm on MD. a, P plot for discovery MR analysis in GS. Each dot represents a CpG site. The x axis represents chromosomes and base pair position. The y axis represents $-\log_{10}$ -transformed P value of Wald ratio MR analysis. The gray dashed line shows the FDR-corrected significance threshold. Discovery MR analysis was performed on CpG sites available in the EPIC array. **b,** Replication analysis in GoDMC for the significant CpGs found in the discovery analysis. Replication analysis was performed on CpG

sites available on the 450K array. Of the 23 significant CpG sites in the discovery analysis, 19 were available on both EPIC and 450K arrays and therefore were included in the replication analysis. The x axes represent effect size and $-\log_{10}$ -transformed P value on the left and right panels. The y axis represents individual CpG sites. The yellow dashed line in the right panel represents FDR-significance threshold.

Three CpG sites were associated with traits (for example, blood cholesterol levels, waist-to-hip ratio and pulse pressure) related to BMI. Obesity has been repeatedly associated with depression, and MR analyses from previous studies have indicated that this may be a causal association^{3,43}. Within the genes in the vicinity of the four CpG sites, *ZNF106* is involved in the regulation of calcium homeostasis that is crucial for cell survival and death⁴⁴. Other genes, including *MYO1C* and *FNDC3B*, are understood to be involved in energy metabolism and homeostasis and highly sensitive to stress and inflammation^{45,46}. Although the previous findings relevant to the BMI-related genes were based on GWAS studies, one of the genes, *FNDC3B*, was found in an earlier MD MWAS study based on brain-tissue samples¹³. Considering the dynamic nature of DNA methylation and its high sensitivity to environmental stressors, our findings suggest metabolic processes may play a potentially crucial role in depression, which may be exacerbated by adverse environmental factors and dysfunctional stress coping mechanisms.

We found stronger evidence for DNAm causing MD than for the reverse direction. This aligns with the broader investigation by Min et al., which showed that DNAm more frequently affects complex traits than vice versa¹⁵. However, previous research has shown MD can influence DNAm at CpG sites linked to MD genetic risk loci⁴⁷. While our MWAS identified a dominant causal pathway from DNAm to MD in top CpG sites identified in our MWAS, reverse causation may occur in other genomic regions, particularly at MD genetic risk loci⁴⁷.

We observed varying results from different cohorts in our study, with little evidence of systematic differences in study age ranges contributing to the heterogeneity. The higher degree of effect size heterogeneity in the MD versus BMI MWAS meta-analyses suggests that phenotype may be a reason for differences between the studies and not other methodological factors, such as sample processing or covariate adjustment. In the MD meta-analysis, larger studies ($N > 1,000$) showed stronger correlation for the top associations, suggesting that sufficient statistical power may help overcome the issue of phenotyping inconsistency. This suggests that future individual MWAS studies of MD should be larger. The considerable costs of DNAm profiling, compared with genotyping, act as a barrier to achieving these aims.

Although there was no significant finding in the MWAS conducted in non-European studies, it is notable that effects were consistent with the significant findings on European samples. Top pathways associated with MD also showed convergent results with the main analysis (for example pathways relevant to inflammation and immune processing). Correlations for effect sizes between European and East Asian samples were positive with the South Asian sample, consistent with previous evidence showing relatively high genetic consistency between European and East Asian groups⁴⁸. Developing larger non-European DNA methylation samples will be crucial to provide a statistically balanced comparison between ancestry groups and to identify ancestry-group-specific DNA methylation sites for MD⁴⁹.

Our study combined many studies with widely varying sample sizes, compared results between two ancestry groups and replicated MR findings using two large mQTL datasets. We provided a comprehensive evaluation of sampling and analytic strategies to guide future large-scale meta-MWAS for mental health disorders. While the blood draw was not timed to coincide with the onset of a depressive episode, limiting causal inferences based on the temporal order of DNAm exposure and MD onset, the MR analyses helped to address this potential limitation. Further, we appreciate that factors, such as antidepressant use, could potentially influence the relationships observed⁵⁰. Future research should consider environmental factors as potential moderating factors to better understand the epigenetic variation in individuals with MD. In addition, there is a lack of replication in other tissue types that are directly relevant to mood regulation, such as brain tissue. Studies have shown that the genetic drivers of DNAm have similar effects across multiple cell types⁵¹. Future clinical applications and larger sample sizes make whole-blood DNAm data more feasible than post-mortem tissue samples. However, to ensure the validity of the findings, future studies should broaden their scope by encompassing additional cell and tissue types.

Methods

Participants

A total of 24,754 European-ancestry participants (5,443 MD cases) from 18 studies were included in the meta-analysis. The mean age of participants in each study ranged from 15 to 59 years. Details for each individual study can be found in Table 1 and Supplementary Information, 'Cohort details' section. Written consent was obtained from all participants. The study was approved by the NHS Tayside Research Ethics committee (05/s1401/89). Ethics declaration and participant compensation for each participating cohort can be found in the Supplementary Information, 'Cohort details' section.

DNAm data preparation and quality check

DNAm data were obtained from DNA extracted from whole blood. Eight studies used the Infinium Human Methylation 450 (450K) BeadChip array (Illumina Inc.; number of CpG sites ranged from 275,868 to 438,752 after quality check), and the other ten studies used the Illumina Infinium Methylation EPIC array (Illumina Inc., number of CpG sites ranged from 673,085 to 809,447 after quality check). Quality checks and data normalization were conducted by each individual study team. Details are provided in the protocol papers for each individual study (Table 1 and Supplementary Information, 'Cohort details' section). In brief, the majority of studies used functional normalization for methylation data preprocessing, unless stated otherwise in the Supplementary Information, 'Cohort details' section⁵². Similar quality check procedures were used, including removing probes with outlying signal intensity, bead count and detection *P* values, removing participants with mismatched sex prediction from DNA methylation data, and removing cross-hybridizing probes that map to common genetic variants (at minor allele frequency > 0.05) or polymorphic probes⁵². *M* values were used for the association analysis⁵³.

Diagnosis of MD

Lifetime diagnosis of MD was derived on the basis of structured clinical interview or self-reported symptoms. Those studies that derived diagnoses of MD on the basis of structured clinical interviews used criteria from the *DSM* Fifth/Fourth Edition⁵⁴. Self-declared MD was based on depressive symptoms lasting for more than 2 weeks and help-seeking due to depression. Studies that derived MD diagnosis on the basis of multiple time points defined cases as those who experienced any depressive episodes during their lifetime, and controls were those who did not declare MD throughout. A total of 9 studies defined MD cases using structural clinical interview (*N* cases = 3,790), 3 studies used *DSM*-criteria questionnaires (*N* cases = 586), 3 studies used self-administered

questionnaires for depressive symptoms (*N* cases = 528), and 3 studies defined MD cases on the basis of self-declared visits to a general practitioner (*N* cases = 539). Details for MD diagnosis for each cohort can be found in Table 1 and the Supplementary Information, 'Cohort details'.

Additional exclusion criteria per study are stated in the Supplementary Information, 'Cohort details' section.

Association analysis

MWAS was conducted in each individual study before the meta-analysis. Linear regression was used to test the association between DNAm (*M* values, dependent variable) and MD diagnosis (independent variable) using a pipeline available at the URL <https://github.com/psychiatric-genomics-consortium/mdd-mwas>. Those cohorts that used their own specific pipelines are specified in the Supplementary Information, 'Cohort details' section. The pipeline uses the R package 'limma' (version 3.60.6) for linear regression on large omic data⁵⁵. Three models representing increasingly rigorous correction for potential confounders were estimated. Covariates for the simplest model were age, sex, batch, the first 20 methylation principal components (PCs) or surrogate variables⁵⁶ based on the study protocol for each individual cohort, and white-blood-cell proportions estimated from DNAm data of CD8⁺ T cells, CD4⁺ T cells, natural killer cells, B cells and granulocytes⁵⁶. The AHRR-adjusted model had an additional covariate that adjusted for smoking status by including the *M* values for the AHRR probe (cg05575921), due to its known accuracy in predicting smoking⁵⁷ and its consistent availability in all studies. Finally, a third model with additional covariates (referred to as the 'complex model') was fitted that contained BMI and alcohol consumption in addition to all the other covariates included in the previous models.

Results for the AHRR-adjusted model (referred to as the 'main model') are reported as the main findings. Standardized Cohen's *d* between MD cases and controls were reported as effect sizes.

Meta-analysis

Meta-analysis of cohort-level MWAS was conducted using METAL (version 2011)⁵⁸ in a two-stage process. First, meta-analysis was performed on studies that used 450K and EPIC arrays separately, due to the difference in CpG sites available for each array (Fig. 1).

Those CpG sites that either were available for more than half of the studies using the given array or had a total sample size over 80% of the max sample size were kept for further analysis. CpG sites with excessive standard errors (s.e. > 0.5; Supplementary Fig. 1) were removed from analysis. Second, the summary statistics for 450K and EPIC array were meta-analyzed. A fixed-effect inverse-variance model was used without genomic control correction. *P* values were Bonferroni corrected (*P*-value threshold = 6.55×10^{-8} to reach Bonferroni-corrected significance) for all 763,265 CpGs included in the analysis. CpGs were mapped to genes using an annotation object generated by the 'IlluminaHumanMethylationEPICanno.ilm10b4.hg19' R package (version 3.13)⁵⁹. The annotation object was created on the basis of the product file provided by Illumina for the Infinium MethylationEPIC v1.0 Beadchip⁶⁰, with the UCSC gene names provided as the target gene regions of the assay. We searched the EWAS Atlas (<https://ngdc.cncb.ac.cn/ewas/atlas>) and EWAS Catalog (<http://www.ewascatalog.org/>) for significantly associated CpGs and genes and the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) for annotated genes.

Pathway enrichment analysis

Following the meta-MWAS, we conducted pathway enrichment analysis to identify the biological pathways in which the significant CpG sites may be implicated. We used the 'gometh' function from the 'missMethyl' R package (version 1.38.0)⁶¹ for pathway analyses using the results of the AHRR-adjusted model. Significant CpGs found in the MWAS after Bonferroni correction were selected, and the rest of the CpGs profiled in the EPIC array were included as the background list. GO terms and

KEGG pathways were analyzed separately. *P* values for both enrichment analyses were FDR corrected.

Identification of differentially methylated regions

DMRs associated with MD were identified using the 'dmrff' R package (version 1.1.2)⁶². A DMR was identified if it contained at least two CpG sites within a maximum of 500-bp window, showed consistent direction of effect and both/all had meta-MWAS $P < 0.05$. Statistics were meta-analyzed within the identified region. *P* values were Bonferroni corrected for all regions (>2 CpG sites) and single CpG sites altogether. Significant DMRs were identified if the two-sided, Bonferroni-corrected $P < 0.05$.

Analysis of the confounding effect of BMI and alcohol consumption

BMI and alcohol consumption are risk factors for MD and are known to have widespread associations with DNAm^{7,29}. To investigate whether the signals found in the main model were due to the effects of BMI and alcohol consumption, we conducted an additional meta-analysis of a complex model for the 14 cohorts that had BMI and alcohol consumption data available ($N_{\text{total}} = 20,196$; see Supplementary Fig. 2 for sample sizes of individual studies). BMI and alcohol consumption were added as additional covariates for the complex model. We compared effect sizes and *P* values between the basic and complex models for significant associations found in the main model.

Out-of-sample classification of MD using MD-MS

Calculation of MD-MS and statistical model. We created MD-MS in an independent testing sample by calculating the weighted sum of *M* values. Effect sizes from the MD meta-MWAS summary statistics were used as weights. Five *P*-value thresholds were used to select the CpG sites— $P \leq 1$, $P \leq 0.01$, $P \leq 0.001$, $P \leq 1 \times 10^{-6}$ and $P \leq 5 \times 10^{-8}$ —resulting in five MD-MSs.

MD diagnosis was set as the independent variable and MS as the dependent variable in the logistic regression model using the 'glm' function in R. Covariates were age, sex and *M* values of AHRR probe. DNAm-estimated cell proportions were not associated with any of the MD-MSs ($P > 0.5$) and were therefore not included as covariates. Batch and genomic relationship matrix were pre-corrected by residualizing *M* values against these covariates.

Testing sample. GS DNAm set 3 data were used for out-of-sample classification (GS DNAm sets 1 and 2 were included in the meta-MWAS). GS DNAm set 3 data used the Illumina Infinium Methylation EPIC array, had no overlap with the GS DNAm data used in the meta-MWAS, and relatedness within set 3 and with the rest of the GS sample was removed by regressing the *M* values against the genomic relationship matrix. Quality check and preprocessing were kept consistent with the GS sample used in the meta-MWAS (sets 1 and 2).

MD diagnosis was derived using electronic health records (EHRs) from GP diagnosis⁶³. Details for the EHRs were explained in the protocol paper by Kerr et al.⁶³. In brief, a subset of participants ($N = 20,032$) of GS with genotyping data gave consent to link their data to EHRs. All Read codes from the PH1021^{64–66} and PH149^{67,68} inventories of the Health Data Research UK Phenotype Library for primary care data of depression were used to identify cases of lifetime MD. Participants with ≥ 1 entry of diagnosis of depression were classified as cases, and those with no entry of any diagnosis for depression or no data to indicate depressive status were controls. The final testing sample contained 504 cases and 8,372 controls.

Association between MD-MS and inflammatory protein markers

A previous study demonstrated widespread association between protein abundance and DNAm²⁸. We conducted a PWAS for MD-MS seeking

to reveal the potentially downstream proteomic consequences of the measured DNAm differences.

Lothian Birth Cohort 1936 (LBC1936), a cohort independent of the MD MWAS, was used for the PWAS analysis. LBC1936 is a community-based birth cohort of participants born in 1936, recruited in Scotland. The sample used in the PWAS analysis had 875 people with both DNAm and proteome data collected at mean age 69.8 ± 0.8 years. DNAm in LBC1936 was profiled in whole-blood samples using the HumanMethylation450 BeadChip Kit (Illumina). Sample preparation and quality check were kept consistent with previously published work⁶⁹. Proteomic data were profiled using lithium heparin collected plasma samples analyzed using a 92-plex proximity extension assay (inflammation panel; Olink Bioscience). The Olink team performed preprocessing using NPX Manager software. For 22 proteins, over 40% of samples fell below the lowest limit of detection, leaving 70 post-quality-check proteins.

General linear models were used to test the association between relative abundance across all 70 proteins and MD-MS created at $P_{\text{threshold}} \leq 5 \times 10^{-8}$. Protein abundance levels were rank-based-inverse normalized and residualized against age, sex, first four genetic PCs and array for proteomic data before entering association analysis. Residual scores of protein abundance were set as the dependent variable. Array for DNAm data and the AHRR probe were included as covariates in the GLM. Two-sided *P* values were corrected using FDR correction.

Heterogeneity analysis

Leave-one-out analysis. To investigate whether a particular study had disproportionate influence on any meta-analytic association, we conducted leave-one-out meta-analyses. For each significant CpG, 18 iterations of meta-analysis were conducted leaving each individual study out, regardless of profiling arrays.

Meta-regression for age. For the CpGs that were significant in the MWAS meta-analysis, we used meta-regression to analyze the potential effect of age across studies. A mixed-effect model with the 'metareg' R package⁷⁰ was used for the meta-regression analysis. Mean age of each individual cohort was set as a random effect. Standardized regression coefficients from summary statistics of each individual study were set as estimates of treatment effects (the 'TE' option). Between-study variance was estimated using the restricted maximum likelihood method. The standardized regression coefficient of the random effect of age was reported as effect sizes for the meta-regression analysis.

Comparison between MD and BMI associations. To further evaluate the heterogeneity of MD-DNAM associations between studies, we looked at the correlation of effect sizes for the summary statistics of the MD MWAS for individual studies. GS that was included in the MWAS, being the largest study sample in the meta-analysis ($N = 9,502$), was used to select a list of CpG sites of interest. The 1,000 most significantly MD-associated CpG sites in GS were selected. Effect sizes for these CpG sites were extracted for all other studies. Correlation analysis was conducted on the effect sizes. We also performed an MWAS of BMI and conducted a similar analysis, for comparison with MD, to assess whether BMI-DNAM associations were similarly heterogeneous.

MR

To identify potentially causal CpG sites to MD, we conducted MR analyses using Genetics of DNA Methylation Consortium (GoDMC; discovery analysis) and GS data (replication analysis).

MD GWAS summary statistics. GWAS summary statistics were obtained from the Howard et al. meta-analysis for MD GWAS from the PGC, 23andMe and UK Biobank³. A total of 807,553 individuals (246,363 cases and 561,190 controls) of European ancestry were included in the MD meta-GWAS.

GoDMC and GS methylation quantitative trait loci. Quantitative trait loci associated with DNAm (mQTL) summary statistics were obtained from GS and GoDMC. For GS, full mQTL summary statistics ($N = 17,620$) were obtained without any P -value thresholding. OmicS-data-based complex trait analysis was used for mQTL estimation^{47,71}. Covariates were kept consistent with the main model for the MD MWAS, except for using self-reported smoking behavior (current smoker, past smoker or non-smoker) and pack years (quantity of smoking) to control for smoking and adding ten genetic PCs as covariates. GoDMC mQTL data were obtained through the consortium website (<http://www.godmc.org.uk/>)^{15,47} using the same pipeline described in the GoDMC protocol paper by Min et al.¹⁵. The mQTL data contains 32 cohorts with 25,561 participants of European ancestry. The mQTL meta-analysis from GoDMC adopted a two-stage approach. First, a truncated set of mQTL data that reached the threshold of P value $< 1 \times 10^{-5}$ were obtained from participating cohorts. This initial stage created a candidate list of mQTL associations ($n = 120,212,413$). Meta-analyses for mQTL were then conducted on these candidate associations.

Samples for GS mQTL analysis, GoDMC mQTL analysis and MD GWAS were mutually exclusive.

Selection of CpG list. A list of CpG sites that were either (1) significant in the MD MWAS or (2) within the identified DMRs were selected as CpG sites of interest. We extracted *cis* mQTL for further MR analysis¹⁵.

In the GoDMC dataset, a total of 156 CpG sites that met the preceding criteria had at least one *cis* mQTL. The *cis* mQTL summary statistics for these CpG sites were extracted from the GoDMC dataset. For those CpG sites that had more than one *cis* mQTL after clumping (1 Mb window, $P < 5 \times 10^{-8}$), the most significant mQTL with the lowest P value was selected for analysis. Those CpG sites that showed significant causal association with MD were selected for replication analysis using the GS mQTL summary statistics.

MR methods. MR analysis was conducted using the ‘TwoSampleMR’ R package (version 0.5.6)⁷². To identify causal effects of DNAm on MD, we used the Wald ratio MR method⁷³ to analyze causal effects on MD using *cis* mQTLs (within a 1 Mb window in vicinity of the chosen CpG site). The most significant mQTL for each CpG site that reached the threshold of $P < 5 \times 10^{-8}$ was selected. Causal effects from DNAm to MD were tested using both GoDMC and GS mQTL data. For the causal effect in the reverse direction (from MD to DNAm), MD GWAS summary statistics were clumped at $P \leq 5 \times 10^{-8}$, with a 1 Mb window and $r^2 = 0.001$. Causal effects from MD to DNAm were tested using mQTL data from the entire GS sample (sets 1, 2 and 3).

MWAS in East Asian ancestry

We sought to investigate MD associations with DNAm in participants of East Asian (Taiwan Biobank) ancestry. Demographic and descriptive statistics are included in the Supplementary Information, ‘Cohort details’ section. As in the main meta-analysis, biological and technical covariates, as well as age, sex and smoking (indexed by AHR probe cg05575921), were included as covariates. Evidence of trans-ancestry transferability of MD CpG effects was investigated by testing for the correlation of effect sizes across both ancestries. We then used the function ‘gometh’ in package ‘missMethyl’ to assess ontology and pathway enrichment (GO and KEGG) for differentially methylated CpG sites at a threshold of $P < 1 \times 10^{-5}$ ($N_{\text{CpG}} = 24$), as used in previous studies⁷⁴.

Reporting summary

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

Data availability

Summary statistics for MD were obtained from Psychiatric Genomics Consortium (<https://doi.org/10.6084/m9.figshare.27061255>) (ref. 75).

GoDMC mQTL data were obtained from <http://www.godmc.org.uk/>. DNAm reference data were obtained from the ‘IlluminaHumanMethylationEPICanno.ilm10b4.hg19’ R package (version 3.13). The annotation object was created on the basis of the product file provided by Illumina for the Infinium MethylationEPIC v1.0 Beadchip (<https://support.illumina.com/downloads/infinium-methylationepic-v1-0-product-files.html>), with the UCSC gene names provided as the target gene regions of the assay. We searched the EWAS Atlas (<https://ngdc.cncb.ac.cn/ewas/atlas>) and EWAS Catalog (<http://www.ewascatalog.org/>) for significantly associated CpGs and genes and the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) for annotated genes. According to the terms of consent, access to any form of individual-level data requires application for each individual cohort. Summary statistics of the MWAS meta-analysis are available at <https://doi.org/10.7488/ds/7929> and via figshare at <https://doi.org/10.6084/m9.figshare.29940299> (ref. 76).

Code availability

The pipeline used for cohort MWAS analysis is publicly available via GitHub (<https://github.com/psychiatric-genomics-consortium/mdd-mwas>). Scripts used for meta-analysis and downstream analyses can be found at https://github.com/xshen796/MD_metaMWAS_PGCMDW1.

References

- 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).
- McIntosh, A. M., Sullivan, P. F. & Lewis, C. M. Uncovering the genetic architecture of major depression. *Neuron* **102**, 91–103 (2019).
- Howard, D. M. et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nat. Neurosci.* **22**, 343–352 (2019).
- Petterson, E. et al. Genetic influences on eight psychiatric disorders based on family data of 4,408,646 full and half-siblings, and genetic data of 333,748 cases and controls. *Psychol. Med.* **49**, 1166–1173 (2019).
- Zeng, Y. et al. Parent of origin genetic effects on methylation in humans are common and influence complex trait variation. *Nat. Commun.* **10**, 1383 (2019).
- Joeanes, R. et al. Epigenetic signatures of cigarette smoking. *Circ. Cardiovasc. Genet.* **9**, 436–447 (2016).
- Liu, C. et al. A DNA methylation biomarker of alcohol consumption. *Mol. Psychiatry* **23**, 422–433 (2018).
- Wahl, S. et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541**, 81–86 (2017).
- Barbu, M. C. et al. Complex trait methylation scores in the prediction of major depressive disorder. *eBioMedicine* **79**, 104000 (2022).
- Nabais, M. F. et al. An overview of DNA methylation-derived trait score methods and applications. *Genome Biol.* **24**, 28 (2023).
- Story Jovanova, O. et al. DNA methylation signatures of depressive symptoms in middle-aged and elderly persons: meta-analysis of multiethnic epigenome-wide studies. *JAMA Psychiatry* **75**, 949–959 (2018).
- Starnawska, A. et al. Epigenome-wide association study of depression symptomatology in elderly monozygotic twins. *Transl. Psychiatry* **9**, 214 (2019).
- Huls, A. et al. Association between DNA methylation levels in brain tissue and late-life depression in community-based participants. *Transl. Psychiatry* **10**, 262 (2020).
- Barbu, M. C. et al. Epigenetic prediction of major depressive disorder. *Mol. Psychiatry* **26**, 5112–5123 (2021).
- Min, J. L. et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nat. Genet.* **53**, 1311–1321 (2021).

16. Gervin, K., Nordeng, H., Ystrom, E., Reichborn-Kjennerud, T. & Lyle, R. Long-term prenatal exposure to paracetamol is associated with DNA methylation differences in children diagnosed with ADHD. *Clin. Epigenetics* **9**, 77 (2017).
17. Li, M. et al. Genome-wide DNA methylation analysis of peripheral blood cells derived from patients with first-episode schizophrenia in the Chinese Han population. *Mol. Psychiatry* **26**, 4475–4485 (2021).
18. Shadrin, A. A. et al. Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *Neuroimage* **244**, 118603 (2021).
19. Zhao, B. et al. Genome-wide association analysis of 19,629 individuals identifies variants influencing regional brain volumes and refines their genetic co-architecture with cognitive and mental health traits. *Nat. Genet.* **51**, 1637–1644 (2019).
20. Knijnenburg, T. A. et al. Genomic and molecular characterization of preterm birth. *Proc. Natl Acad. Sci. USA* **116**, 5819–5827 (2019).
21. Kupers, L. K. et al. Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nat. Commun.* **10**, 1893 (2019).
22. Spiers, H. et al. Methylomic trajectories across human fetal brain development. *Genome Res.* **25**, 338–352 (2015).
23. De Jager, P. L. et al. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nat. Neurosci.* **17**, 1156–1163 (2014).
24. Lunnon, K. et al. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* **17**, 1164–1170 (2014).
25. Zhang, L. et al. Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease. *Nat. Commun.* **11**, 6114 (2020).
26. Antoun, E. et al. Maternal dysglycaemia, changes in the infant's epigenome modified with a diet and physical activity intervention in pregnancy: secondary analysis of a randomised control trial. *PLoS Med.* **17**, e1003229 (2020).
27. Liu, Y. et al. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nat. Biotechnol.* **31**, 142–147 (2013).
28. Gadd, D. A. et al. Integrated methylome and phenome study of the circulating proteome reveals markers pertinent to brain health. *Nat. Commun.* **13**, 4670 (2022).
29. Geurts, Y. M. et al. Novel associations between blood DNA methylation and body mass index in middle-aged and older adults. *Int. J. Obes.* **42**, 887–896 (2018).
30. Richardson, T. G. et al. Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. *PLoS Biol.* **20**, e3001547 (2022).
31. Downes, D. J. et al. Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. *Nat. Genet.* **53**, 1606–1615 (2021).
32. Guo, D. F. & Rahmouni, K. Molecular basis of the obesity associated with Bardet-Biedl syndrome. *Trends Endocrinol. Metab.* **22**, 286–293 (2011).
33. Guo, D. F. et al. The BBSome controls energy homeostasis by mediating the transport of the leptin receptor to the plasma membrane. *PLoS Genet.* **12**, e1005890 (2016).
34. Lopez, M. Hypothalamic leptin resistance: from BBB to BBSome. *PLoS Genet.* **12**, e1005980 (2016).
35. Hannon, E. et al. DNA methylation meta-analysis reveals cellular alterations in psychosis and markers of treatment-resistant schizophrenia. *eLife* **10**, e58430 (2021).
36. Jaffe, A. E. et al. Mapping DNA methylation across development, genotype and schizophrenia in the human frontal cortex. *Nat. Neurosci.* **19**, 40–47 (2016).
37. Homs, A. et al. Genetic and epigenetic methylation defects and implication of the ERMN gene in autism spectrum disorders. *Transl. Psychiatry* **6**, e855 (2016).
38. Mirza, S. et al. Blood epigenome-wide association studies of suicide attempt in adults with bipolar disorder. *Transl. Psychiatry* **14**, 70 (2024).
39. Koskeridis, F. et al. Pleiotropic genetic architecture and novel loci for C-reactive protein levels. *Nat. Commun.* **13**, 6939 (2022).
40. Schrepf, A., Markon, K. & Lutgendorf, S. K. From childhood trauma to elevated C-reactive protein in adulthood: the role of anxiety and emotional eating. *Psychosom. Med.* **76**, 327–336 (2014).
41. Iob, E., Kirschbaum, C. & Steptoe, A. Persistent depressive symptoms, HPA-axis hyperactivity, and inflammation: the role of cognitive-affective and somatic symptoms. *Mol. Psychiatry* **25**, 1130–1140 (2020).
42. Haapakoski, R., Mathieu, J., Ebmeier, K. P., Alenius, H. & Kivimaki, M. Cumulative meta-analysis of interleukins 6 and 1beta, tumour necrosis factor alpha and C-reactive protein in patients with major depressive disorder. *Brain Behav. Immun.* **49**, 206–215 (2015).
43. Wray, N. R. et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat. Genet.* **50**, 668–681 (2018).
44. Joyce, P. I. et al. Deficiency of the zinc finger protein ZFP106 causes motor and sensory neurodegeneration. *Hum. Mol. Genet.* **25**, 291–307 (2016).
45. Arif, E. et al. The motor protein Myo1c regulates transforming growth factor- β -signaling and fibrosis in podocytes. *Kidney Int.* **96**, 139–158 (2019).
46. Wang, X., Huang, Y., Li, S. & Zhang, H. Integrated machine learning methods identify FNDC3B as a potential prognostic biomarker and correlated with immune infiltrates in glioma. *Front. Immunol.* **13**, 1027154 (2022).
47. Shen, X. et al. DNA methylome-wide association study of genetic risk for depression implicates antigen processing and immune responses. *Genome Med.* **14**, 36 (2022).
48. Lam, M. et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat. Genet.* **51**, 1670–1678 (2019).
49. Nievergelt, C. M. et al. International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nat. Commun.* **10**, 4558 (2019).
50. Davyson, E. et al. Insights from a methylome-wide association study of antidepressant exposure. *Nat. Commun.* **16**, 1908 (2025).
51. Braun, P. R. et al. Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Transl. Psychiatry* **9**, 47 (2019).
52. Walker, R. M. et al. Assessment of dried blood spots for DNA methylation profiling. *Wellcome Open Res.* **4**, 44 (2019).
53. Du, P. et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinf.* **11**, 587 (2010).
54. *Diagnostic and Statistical Manual of Mental Disorders* 5th edn (APA, 2013).
55. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
56. Hattab, M. W. et al. Correcting for cell-type effects in DNA methylation studies: reference-based method outperforms latent variable approaches in empirical studies. *Genome Biol.* **18**, 24 (2017).
57. Philibert, R., Dogan, M., Beach, S. R. H., Mills, J. A. & Long, J. D. AHRH methylation predicts smoking status and smoking intensity in both saliva and blood DNA. *Am. J. Med. Genet. B* **183**, 51–60 (2020).

58. Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190–2191 (2010).
59. Fortin, J. P. & Hansen, K. D. *IlluminaHumanMethylationEPICanno.ilm10b4.hg19: Annotation for Illumina's EPIC Methylation Arrays*. R package version 0.6.0 (2017).
60. Infinium MethylationEPIC v1.0 Product Files (Legacy BeadChip). *Illumina* <https://support.illumina.com/downloads/infinium-methylationepic-v1-0-product-files.html> (2017).
61. Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* **32**, 286–288 (2016).
62. Suderman, M. et al. dmrff: identifying differentially methylated regions efficiently with power and control. Preprint at *bioRxiv* <https://doi.org/10.1101/508556> (2018).
63. Kerr, S. M. et al. Electronic health record and genome-wide genetic data in Generation Scotland participants. *Wellcome Open Res.* **2**, 85 (2017).
64. Cornish, R. P., John, A., Boyd, A., Tilling, K. & Macleod, J. Defining adolescent common mental disorders using electronic primary care data: a comparison with outcomes measured using the CIS-R. *BMJ Open* **6**, e013167 (2016).
65. John, A. et al. Recent trends in primary-care antidepressant prescribing to children and young people: an e-cohort study. *Psychol. Med.* **46**, 3315–3327 (2016).
66. John, A. et al. *PH1021/2204—Depression—Primary Care* (HDRUK Phenotype Library, 2022); <http://phenotypes.healthdatagateway.org/phenotypes/PH1021/version/2204/detail/>
67. Kuan, V. et al. A chronological map of 308 physical and mental health conditions from 4 million individuals in the English National Health Service. *Lancet Digit. Health* **1**, e63–e77 (2019).
68. Kuan, V. et al. *PH149/298—Depression* (HDRUK Phenotype Library, 2021); <http://phenotypes.healthdatagateway.org/phenotypes/PH149/version/298/detail/>
69. Hillary, R. F. et al. Multi-method genome- and epigenome-wide studies of inflammatory protein levels in healthy older adults. *Genome Med.* **12**, 60 (2020).
70. Schwarzer, G., C. J. R. & Gucker, G. *Meta-Analysis with R* (Springer, 2015).
71. Zhang, F. et al. OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol.* **20**, 107 (2019).
72. Hemani, G. et al. The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, e34408 (2018).
73. Lawlor, D. A., Harbord, R. M., Sterne, J. A. C., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
74. Chu, A. Y. et al. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat. Commun.* **8**, 1286 (2017).
75. Adams, M. J., Lewis, C. & McIntosh, A. GWAS summary statistics for major depression (PGC MDD2025). *figshare* <https://doi.org/10.6084/m9.figshare.27061255.v4> (2025).
76. Shen, X. et al. Summary statistics for the MAS meta-analysis for major depression (AHRM model). *figshare* <https://doi.org/10.6084/m9.figshare.29940299> (2025).

Acknowledgements

Generation Scotland (GS): this work is supported by Wellcome Trust grants (226770/Z/22/Z, 220857/Z/20/Z, 216767/Z/19/Z supporting A.M.M., X.S.) and UKRI award (ImmunoMIND, MR/Z50354X/1; A.M.M., X.S.). For the purpose of open access, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. DNAm profiling was supported by

funding from NARSAD (ref 27404; D.M.H.). Genotyping of the GS samples was funded by the MRC and Wellcome Trust (104036/Z/14/Z). GS also receives support from the Chief Scientist Office of the Scottish Government Health Directorates (CZD/16/6) and the Scottish Funding Council (HRO3006). E. Davyson was supported by the United Kingdom Research and Innovation (grant EP/SO2431X/1), UKRI Centre of Doctoral Training in Biomedical AI at the University of Edinburgh, School of Informatics. P.N. is funded by BBSRC grant BBS/E/RL/230001A and acknowledges support from the MRC Human Genetics Unit program grant 'Quantitative traits in health and disease' (U. MC_UU_00007/10) and grant MC_PC_U127592696. Avon Longitudinal Study of Parents and Children (ALSPAC) / Accessible Resource for Integrated Epigenomic Studies (ARIES): the UK Medical Research Council and Wellcome (grant ref 217065/Z/19/Z) and the University of Bristol provide core support for ALSPAC. A comprehensive list of grants funding is available on the ALSPAC website (<http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf>). Part of this data was collected using REDCap; see the REDCap website for details <https://projectredcap.org/resources/citations/>. We thank all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. BeCOME/OPTIMA: OPTIMA and BeCOME are funded by the Max Planck Society. We thank all members of the BeCOME and OPTIMA Working Groups, contributing contributors to the research project, including physicians, psychologists, study nurses, researchers and research assistants, and of course patients of the hospital of the Max Planck Institute of Psychiatry in Munich. BiDirect: the BiDirect Study is supported by grants of the German Ministry of Research and Education (BMBF) to the University of Münster (01ER0816 and 01ER1506). Biopsychosocial factors of major depression in youth (BioMD-Y/LMU): BioMD-Y was supported by the Randebrock Stiftung associated with the kbo-Heckscher-Klinikum, Munich, Germany, during the years 2009–2011 upon its inception, and has since been financed through intramural funds. DNA extraction, genome-wide genotyping and DNA methylation analyses were supported by the Max Planck Society. DNA Methylation of the BioMD-Y/LMU cohort was funded in part by S. Schaefer, private donor of the Max Planck Foundation. Dublin cohort: this research was supported by a grant from AbbVie (number 10118) to H.M., T.M.M. and E.M. E.M. was the recipient of the Boston Scientific Newman Fellowship awarded by the UCD Foundation. The authors thank all the participants in the Dublin Cohort and all researchers involved in the original IBD Study. Environmental Risk (E-Risk) Longitudinal Twin Study: the E-Risk Study is funded by grants from the UK Medical Research Council (G1002190; MR/X010791/1). Additional support was provided by the US National Institute of Child Health and Human Development (HD077482) and the Jacobs Foundation. This project represents independent research partly funded by the NIHR Maudsley Biomedical Research Centre at South London and Maudsley NHS Foundation Trust and King's College London (NIHR203318). H.L.F. was supported by the UK ESRC Centre for Society and Mental Health at King's College London (ES/S012567/1). C.C.Y.W. receives salary support from the National Institute for Health and Care Research (NIHR) Biomedical Research Centre for Mental Health, South London and Maudsley National Health Service (NHS) Foundation Trust and Institute of Psychiatry, Psychology, and Neuroscience, Kings College London. The views expressed are those of the authors and not necessarily those of the ESRC, NIHR, Department of Health and Social Care, King's College London, or the UK Department of Health. We are grateful to the E-Risk study mothers and fathers, the twins, and the twins' teachers for their participation. Our thanks to the E-Risk team for their dedication, hard work and insights. EXTEND: this study was funded by the National Institute for

Health and Care Research Exeter Clinical Research Facility. This study was supported by the National Institute for Health and Care Research Exeter Biomedical Research Centre. The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care. Finnish Twin Cohort: phenotype and genotype data collection in FinnTwin12 and FinnTwin16 studies of the Finnish twin cohort has been supported by the Wellcome Trust Sanger Institute, the Broad Institute, ENGAGE—European Network for Genetic and Genomic Epidemiology, FP7-HEALTH-F4-2007, grant agreement number 201413, National Institute of Alcohol Abuse and Alcoholism (grants AA-12502, AA-00145 and AA-09203 to R.J. Rose; AA15416 and KO2AA018755 to D.M. Dick; R01AA015416 to J. Salvatore) and the Academy of Finland (grants 100499, 205585, 118555, 141054, 264146, 308248 to J.K., 328685, 307339, 297908 and 251316 to M.O., 309119 to T. Korhonen) and Sigrid Juselius Foundation to M.O. and J.K. and the Centre of Excellence in Complex Disease Genetics (grants 312073, 336823, and 352792 to J.K.). J.K. acknowledges support by the Academy of Finland (grants 265240, 263278). T. Korhonen acknowledges support from the Juho Vainio Foundation. We thank the active participation of the twins, which has made the studies possible. FOR2107: the German multicenter consortium ‘Neurobiology of Affective Disorders. A translational perspective on brain structure and function’ is funded by the German Research Foundation (research unit FOR2107). Principal investigators are T. Kircher (KI588/14-1, KI588/14-2, KI588/20-1, KI588/22-1), U.D. (DA1151/5-1, DA1151/5-2, DA1151/9-1, DA1151/10-1, DA1151/11-1; SFB/TRR 393, project grant number 521379614), A. Krug (KR3822/5-1, KR3822/7-2), I. Nenadic (NE2254/1-2, NE2254/3-1, NE2254/4-1), C. Konrad (KO4291/3-1), M. Rietschel (RI 908/11-1, RI 908/11-2), M. Nöthen (NO 246/10-1, NO 246/10-2), S. Witt (WI 3439/3-1, WI 3439/3-2). Personnel is also in part funded by the DFG SFB/TRR 393 (project grant 521379614). This work was in part supported by the DFG SFB/TRR 393 consortium (<https://www.uni-marburg.de/en/trr-393>), project number 521379614. We are deeply indebted to all study participants and staff. A list of acknowledgements can be found here: www.for2107.de/acknowledgements. Janssen: the Janssen cohort was funded by Janssen Research & Development, LLC. We thank the clinical investigators and research coordinators who ran the clinical study and collected the blood samples used in this study, as well as study participants and their families, whose help and participation made this work possible. We thank the staff from Cancer Genetic Institute and HD Bioscience for performing the DNA extraction from whole-blood samples, plating, QC, and the staff from Illumina to perform the epigenetic and genotyping assays to enable the data generation. Munich Antidepressant Response Signature (MARS/GSK): the MARS cohort was sponsored by the Max Planck Society. The UniDep cohort was funded by the Bavarian Ministry of Commerce and by the Federal Ministry of Education and Research in the framework of the National Genome Research Network, Foerderkennzeichen O1GS0481 and the Bavarian Ministry of Commerce. DNA methylation analysis of a subset of both cohorts was financed by ERA-NET NEURON. We thank all contributors to the research project, including physicians, psychologists, study nurses, researchers and research assistants, and of course patients of the hospital of the Max Planck Institute of Psychiatry in Munich and psychiatric hospitals in Augsburg and Ingolstadt. Lothian Birth Cohort 1936 (LBC1936): the LBC1936 is supported by the Biotechnology and Biological Sciences Research Council and the Economic and Social Research Council (BB/W008793/1) (which supports S.E.H.), Age UK (Disconnected Mind project), the Medical Research Council (G0701120, G1001245, MR/M013111/1, MR/R024065/1), the Milton Damerel Trust, and the University of Edinburgh. S.R.C. is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (221890/Z/20/Z). Methylation typing was supported by Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age

UK, the Wellcome Trust Institutional Strategic Support Fund, the University of Edinburgh, and the University of Queensland. Netherlands Twin Register (NTR): we warmly thank all twin families of the Netherlands Twin Register who make this research possible. This work was supported by the Royal Dutch Academy for Arts and Science (KNAW) Academy Professor Award (PAH/6635) to D.I.B.; the Netherlands Organization for Scientific Research (NWO 480-15-001/674) and Biobanking and Biomolecular Resources Research Infrastructure (BBMRI-NL: 184.021.007; 184.033.111). Genetics of Mental Illness, a lifespan approach to the genetics of childhood and adult neuropsychiatric disorders and comorbid conditions (ERC-230374). J.v.D. was supported by NWO Large Scale infrastructures, X-omics (184.034.019). Psychiatric Genomics Consortium (PGC): N.R.W. acknowledges funding from the Australian National Health and Medical Research Council 1173790 and the Michael Davys Trust. C.M.L., A.M.M., N.R.W. and X.S. are supported by a Wellcome Trust Mental Health award (226770/Z/22/Z). A.M.M., N.R.W. and X.S. are supported by a UKRI award (ImmunoMIND, MR/Z50354X/1). Study of Health in Pomerania (SHIP): SHIP is part of the Community Medicine Research net of the University of Greifswald, which is funded by the Federal Ministry of Education and Research (01ZZ9603, 01ZZ0103 and 01ZZ0403), the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. DNA methylation data have been supported by the DZHK (grant 81X3400104). The University of Greifswald is a member of the Caché Campus program of the InterSystems GmbH. A.T. has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, grant 542489987). The SHIP authors are grateful to P. S. DeVries for his support with the EWAS pipeline. Taiwan Biobank: this work is supported by Population Health Research Center from Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education in Taiwan (grant number NTU-112L9004). We thank all the participants and investigators of the Taiwan Biobank. TwinsUK: The study received support from the ESRC (ES/N000404/1 to J.T.B.) and European HDHL Joint Programming Initiative funding scheme DIMENSION project (BBSRC BB/S020845/1 and BB/T019980/1 to J.T.B.). The TwinsUK study is funded by the Wellcome Trust, Medical Research Council, Versus Arthritis, European Union Horizon 2020, Chronic Disease Research Foundation (CDRF), ZOE LIMITED and the National Institute for Health Research (NIHR) Clinical Research Network (CRN) and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust in partnership with King’s College London. Understanding Society/UK Household Longitudinal Study (UKHLS): We acknowledge the School of Life Sciences, University of Essex for QC of the methylation data. DNA methylation in UKHLS was funded through enhancements to the Economic and Social Research Council (ESRC) grants ES/K005146/1 and ES/N00812X/1. M.K. and A.D. are partially supported by the ESRC (ES/S012486/1). E. Dilkes is a Soc-B student (ESRC project reference number 2604212). We also acknowledge the summary statistics provided by the Psychiatric Genomics Consortium. In addition, we thank the research participants and employees of 23andMe, Inc. for making this work possible.

Author contributions

X.S., M.B., N.R.W. and A.M.M. were closely involved in conceptualizing and designing the study. L.A., J.T.B., K.B., E.B., D.I.B., D.M.H., S.R.C., U.D., K.L.E., H.L.F., A.J.F., H.J.G., J.K., T. Kircher, J.K.B., M.K., P.-H.K., Q.S.L., T.E.M., H.M., T.M.M., G.S.-K., J. Mill, C.M.L., N.R.W. and A.M.M. were involved in data and funding acquisition and overseeing cohort data preparation. X.S. and M.B. developed the MWAS pipeline with support from M.J.A. D. Caramaschi kindly tested the pipeline and provided feedback to optimize it. Analysts in each individual cohort made substantial contribution to data analysis of cohort MWAS. They are D. Caramaschi, R.A., D. Czamara, F.S.D., A.D., E. Dilkes, M.H.-R., F.H.,

L.K., K.-C.L., T.P., A.M.S., G.S., A.W., C.C.Y.W., E. Davyson, H.C., A.-K.A., M.J.A., M.B., J.B., A.C., R.C., E.C.D., L.F., J.F., F.J.F., D.A.G., E.G., E.H., S.E.H., G.H., D.M.H., S.I., T. Korhonen, T.-P.L., J. Martins, E.M., S.M., P.N., M.O., V.P., C.P., A.D.S., F. Stein, F. Streit, A.T., H.V., R.M.W., J.v.D. and N.Y., who contributed substantially to data acquisition, management and clinical assessments for individual cohorts or provided comprehensive comments to study design and the manuscript. All authors contributed to and critically reviewed the manuscript.

Competing interests

Q.S.L. is an employee of Janssen Research & Development, LLC. Q.S.L. owns stock and/or stock options in Johnson & Johnson. H.J.G. has received travel grants and speakers' honoraria from Indorsia, Neuraxpharm, Servier and Janssen Cilag. T. Kircher received unrestricted educational grants from Servier, Janssen, Recordati, Aristo, Otsuka, neuraxpharm. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s44220-025-00486-4>.

Correspondence and requests for materials should be addressed to Xueyi Shen or Andrew M. McIntosh.

Peer review information *Nature Mental Health* thanks Maggie Davis, Mary-Ellen Lynall and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

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

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

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025

Xueyi Shen^{1,63}✉, Miruna Barbu^{1,63}, Doretta Caramaschi², Ryan Arathimos³, Darina Czamara⁴, Friederike S. David⁵, Anna Dearman⁶, Evelyn Dilkes⁶, Marisol Herrera-Rivero^{7,8}, Floris Huider⁹, Luise Kühn¹⁰, Kuan-Chen Lu¹¹, Teemu Palviainen¹², Alicia M. Schowe^{4,13}, Gemma Shireby^{14,15}, Antoine Weihs^{10,16}, Chloe C. Y. Wong³, Eleanor Davyson¹⁷, Hannah Casey¹, Mark J. Adams¹, Antje-Kathrin Allgaier¹⁸, Michael Barber¹⁹, Joe Burrage²⁰, Avshalom Caspi^{3,21}, Ricardo Costeira²², Erin C. Dunn^{10,23,24}, Lisa Feldmann^{25,26}, Josef Frank²⁷, Franz J. Freisleder²⁸, Danni A. Gadd²⁹, Ellen Greimel^{10,25,26}, Eilis Hannon²⁰, Sarah E. Harris³⁰, Georg Homuth³¹, David M. Howard³, Stella Iurato^{4,32}, Tellervo Korhonen¹², Tzu-Pin Lu³³, Nicholas G. Martin³⁴, Jade Martins³⁵, Edel McDermott³⁶, Susanne Meinert^{37,38}, Pau Navarro^{19,39}, Miina Ollikainen^{12,40}, Verena Pehl²⁵, Charlotte Piechaczek²⁵, Aline D. Scherff^{25,26}, Frederike Stein^{41,42}, Fabian Streit^{27,43,44,45}, Alexander Teumer^{10,46}, Henry Völzke^{46,47}, Jenny van Dongen⁹, Rosie M. Walker², Natan Yusupov^{4,48}, Louise Arseneault³, Jordana T. Bell²², Klaus Berger⁸, Elisabeth Binder⁴, Dorret I. Boomsma⁴⁹, Simon R. Cox³⁰, Udo Dannlowski³⁷, Kathryn L. Evans⁵⁰, Helen L. Fisher^{3,51}, Andreas J. Forstner^{5,52,53}, Hans J. Grabe^{10,16}, Jaakko Kaprio¹², Tilo Kircher^{41,42}, Johannes Kopf-Beck^{54,55}, Meena Kumari⁶, Po-Hsiu Kuo^{56,57}, Qingqin S. Li⁵⁸, Terrie E. Moffitt^{3,21}, Hugh Mulcahy^{36,59}, Therese M. Murphy⁶⁰, Gerd Schulte-Körne^{25,26}, Jonathan Mill^{14,20}, Cathryn M. Lewis³, PGC MDD Working Group*, Naomi R. Wray^{61,62} & Andrew M. McIntosh¹✉

¹Institute of Neuroscience and Cardiovascular Research, University of Edinburgh, Edinburgh, UK. ²Faculty of Health and Life Sciences, Department of Psychology, University of Exeter, Exeter, UK. ³Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK. ⁴Department Genes and Environment, Max-Planck-Institute of Psychiatry, Munich, Germany. ⁵Institute of Human Genetics, University of Bonn, School of Medicine and University Hospital Bonn, Bonn, Germany. ⁶Institute for Social and Economic Research, University of Essex, Colchester, UK. ⁷Department of Genetic Epidemiology, Institute of Human Genetics, University of Münster, Münster, Germany. ⁸Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany. ⁹Biological Psychology, Amsterdam Public Health Institute, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. ¹⁰Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, Germany. ¹¹Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan. ¹²Institute for Molecular Medicine Finland—FIMM, University of Helsinki, Helsinki, Finland. ¹³Ludwig-Maximilians-Universität, Graduate School of Systemic Neuroscience, Munich, Germany. ¹⁴University of Exeter Medical School, University of Exeter, Exeter, UK. ¹⁵Institute of Education, University College London, London, UK. ¹⁶German Centre for Neurodegenerative Diseases (DZNE), Greifswald, Germany. ¹⁷School of Informatics, University of Edinburgh, Edinburgh, UK. ¹⁸Department of Human Sciences, University of the Bundeswehr Munich, Munich, Germany. ¹⁹MRC Human Genetics Unit, University of Edinburgh, Edinburgh, UK. ²⁰Department of Clinical and Biomedical Sciences, University of Exeter, Exeter, UK. ²¹Departments of Psychology and Neuroscience, Psychiatry and Behavioral Sciences, and Institute for Genome Sciences and Policy, Duke University, Durham, NC, USA. ²²Department of Twin Research and Genetic Epidemiology, King's College London, London, UK. ²³Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, MA, USA. ²⁴Department of Psychiatry, Harvard Medical School, Boston, MA, USA. ²⁵Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, LMU University Hospital, Ludwig-Maximilians-University (LMU) Munich, Munich, Germany. ²⁶German Center for Mental Health (DZPG), Partner Site Munich-Augsburg, Munich, Germany. ²⁷Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ²⁸kbo-Heckscher-Klinikum gGmbH, Munich, Germany. ²⁹Institute of Genetics and Cancer, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK. ³⁰Lothian Birth Cohorts group,

Department of Psychology, University of Edinburgh, Edinburgh, UK. ³¹Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, Germany. ³²Roche Diagnostics GmbH, Munich, Germany. ³³Department of Public Health and Institute of Health Data Analytics and Statistics, National Taiwan University, Taipei, Taiwan. ³⁴Queensland Institute of Medical Research Berghofer, Brisbane, Queensland, Australia. ³⁵Decision Intelligence, Knorr-Bremse Services GmbH, Munich, Germany. ³⁶Centre for Colorectal Disease, St Vincent's University Hospital, Dublin, Ireland. ³⁷Institute for Translational Psychiatry, University of Münster, Münster, Germany. ³⁸Institute for Translational Neuroscience, University of Münster, Münster, Germany. ³⁹The Roslin Institute, University of Edinburgh, Edinburgh, UK. ⁴⁰Minerva Foundation Institute for Medical Research, Helsinki, Finland. ⁴¹Department of Psychiatry and Psychotherapy, University of Marburg, Marburg, Germany. ⁴²Center for Mind, Brain and Behavior (CMBB), University of Marburg, Marburg, Germany. ⁴³Hector Institute for Artificial Intelligence in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ⁴⁴Department of Psychiatry and Psychotherapy, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. ⁴⁵German Center for Mental Health (DZPG), Partner Site Mannheim-Heidelberg-Ulm, Mannheim, Germany. ⁴⁶German Centre for Cardiovascular Research (DZHK), Greifswald, Germany. ⁴⁷Institute for Community Medicine, University Medicine Greifswald, Greifswald, Germany. ⁴⁸International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, Germany. ⁴⁹Complex Trait Genetics, Center for Neurogenomics and Cognitive Research, Vrije Universiteit Amsterdam, Amsterdam, the Netherlands. ⁵⁰Institute for Genetics and Cancer, University of Edinburgh, Edinburgh, UK. ⁵¹ESRC Centre for Society and Mental Health, King's College London, London, UK. ⁵²Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Jülich, Germany. ⁵³Centre for Human Genetics, University of Marburg, Marburg, Germany. ⁵⁴Clinical Department, Max Planck Institute of Psychiatry, Munich, Germany. ⁵⁵Department of Psychology, LMU Munich, Munich, Germany. ⁵⁶Department of Public Health & Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, Taiwan. ⁵⁷Department of Psychiatry, National Taiwan University Hospital, Taipei, Taiwan. ⁵⁸Neuroscience, Janssen Research and Development, Titusville, NJ, USA. ⁵⁹School of Medicine, University College Dublin, Dublin, Ireland. ⁶⁰School of Biological, Health and Sports Sciences, Technological University Dublin, Dublin, Ireland. ⁶¹Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia. ⁶²Department of Psychiatry, University of Oxford, Oxford, UK. ⁶³These authors jointly supervised this work: Xueyi Shen, Miruna Barbu. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: xueyi.shen@ed.ac.uk; andrew.mcintosh@ed.ac.uk

PGC MDD Working Group

Xueyi Shen^{1,63}, Doretta Caramaschi², Ryan Arathimos³, Darina Czamara⁴, Friederike S. David⁵, Anna Dearman⁶, Evelyn Dilkes⁶, Marisol Herrera-Rivero^{7,8}, Floris Huiders⁹, Luise Kühn¹⁰, Kuan-Chen Lu¹¹, Teemu Palviainen¹², Alicia M. Schowe^{4,13}, Gemma Shireby^{14,15}, Antoine Weihs^{10,16}, Chloe C. Y. Wong³, Mark J. Adams¹, Antje-Kathrin Allgaier¹⁸, Michael Barber¹⁹, Joe Burrage²⁰, Avshalom Caspi^{3,21}, Ricardo Costeira²², Erin C. Dunn^{23,24}, Lisa Feldmann^{25,26}, Josef Frank²⁷, Franz J. Freisleder²⁸, Ellen Greimel^{25,26}, Eilis Hannon²⁰, Georg Homuth³¹, David M. Howard³, Stella Iurato^{4,32}, Tellervo Korhonen¹², Tzu-Pin Lu³³, Nicholas G. Martin³⁴, Jade Martins³⁵, Edel McDermott³⁶, Susanne Meinert^{37,38}, Miina Ollikainen^{12,40}, Verena Pehl²⁵, Charlotte Piechaczek²⁵, Aline D. Scherff^{25,26}, Frederike Stein^{41,42}, Fabian Streit^{27,43,44,45}, Alexander Teumer^{10,46}, Henry Völzke^{46,47}, Jenny van Dongen⁹, Natan Yusupov^{4,48}, Louise Arseneault³, Jordana T. Bell²², Klaus Berger⁸, Elisabeth Binder⁴, Dorret I. Boomsma⁴⁹, Udo Dannlowski³⁷, Helen L. Fisher^{3,51}, Andreas J. Forstner^{5,52,53}, Hans J. Grabe^{10,16}, Jaakko Kaprio¹², Tilo Kircher^{41,42}, Johannes Kopf-Beck^{54,55}, Meena Kumari⁶, Po-Hsiu Kuo^{56,57}, Qingqin S. Li⁵⁸, Terrie E. Moffitt^{3,21}, Hugh Mulcahy^{36,59}, Therese M. Murphy⁶⁰, Gerd Schulte-Körne^{25,26}, Cathryn M. Lewis³, Naomi R. Wray^{61,62} & Andrew M. McIntosh¹

Reporting Summary

Nature Portfolio 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 Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the 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
 - ☒ A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Summary statistics of MWAS for Major Depression were obtained from all 18 cohorts participated in the study. Details for each individual study can be found in Table 1 and Supplementary Information, Cohort details section. Written consents were obtained from all participants. In brief, DNAm data were obtained from DNA extracted from whole blood. Eight studies used the Infinium Human Methylation 450 (450K) BeadChip array (Illumina Inc., number of CpG sites ranged from 275,868 to 438,752 after quality check) and the other ten studies used the Illumina Infinium Methylation EPIC array (Illumina Inc., number of CpG sites ranged from 673,085 to 809,447 after quality check). Life-time diagnosis of MD was derived based on structured clinical interview or self-reported symptoms.
Data analysis	<p>MWAS was conducted in each individual study before the meta-analysis. Linear regression was used to test the association between DNA methylation (M-values, dependent variable) and MD diagnosis (independent variable) using a pipeline available at the URL: https://github.com/psychiatric-genomics-consortium/mdd-mwas. Those cohorts that used their own specific pipelines were specified in the Supplementary Information, Cohort details section.</p> <p>The pipeline uses the R package 'limma' (version 3.60.6) for linear regression on large omic data. Three models representing increasingly rigorous correction for potential confounders were estimated. Covariates for the simplest model were age, sex, batch, the first 20 methylation principal components (PCs) or surrogate variables (SVs) based on the study protocol for each individual cohort, and white-blood cell proportions estimated from DNA methylation data of CD8+T, CD4+T, natural killer cells, B cells and granulocytes. The AHRR (aryl hydrocarbon receptor repressor)-adjusted model had an additional covariate that adjusted for smoking status by including the M-values for the AHRR probe (cg05575921), due to its known accuracy in predicting smoking and its consistency and availability in all studies. Finally, a third model with additional covariates (referred to as the 'complex model') was fitted that contained body mass index (BMI) and alcohol consumption in addition to all the other covariates included in the previous models. Results for the AHRR model were reported as the main findings.</p>

Meta-analysis of cohort-level MWAS was conducted using METAL (version 2011)⁵⁹ in a two-stage process. First, meta-analysis was performed on studies that used 450K and EPIC arrays separately, due to the difference of CpG sites available for each array (Figure 1).

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 and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

Summary statistics for MD was obtained from Psychiatric Genomics Consortium (<https://doi.org/10.6084/m9.figshare.27061255>). GoDMC mQTL data was obtained from the URL: <http://www.godmc.org.uk/>. DNAm reference data was obtained from the 'IlluminaHumanMethylationEPICanno.ilm10b4.hg19' R package (version 3.13). The annotation object was created based on the product file provided by Illumina for the Infinium MethylationEPIC v1.0 Beadchip (<https://support.illumina.com/downloads/infinium-methylationepic-v1-0-product-files.html>), with the UCSC gene names provided as the target gene regions of the assay. We searched the EWAS Atlas (<https://ngdc.cncb.ac.cn/ewas/atlas>) and EWAS Catalog (<http://www.ewascatalog.org/>) for significantly associated CpGs and genes and the GWAS Catalog (<https://www.ebi.ac.uk/gwas/>) for annotated genes. According to the terms of consent, access to any form of individual-level data requires application for each individual cohort.

Summary statistics of the MWAS meta-analysis is available at the URL: <https://doi.org/10.7488/ds/7929>.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender

We report biological sex throughout in our main text and Supplementary information.

Reporting on race, ethnicity, or other socially relevant groupings

Ethnicity was reported as genetic ancestry in the main text. In summary, we conducted an MWAS meta-analysis for Major Depression using DNA extracted from whole blood in 18 cohorts comprising 24,754 individuals of European ancestry. We also assessed whether effect sizes for significant findings in European samples were positively associated with those in an independent East Asian sample.

Population characteristics

A total of 24,754 European-ancestry participants (5,443 cases) from 18 studies were included in the meta-analysis. The mean age of participants in each study ranged from 15 to 59 years. Details for each individual study can be found in Table 1 and Supplementary information, Cohort details.

Recruitment

Twelve of the cohorts included in our study are population cohorts and self-selection bias may present in these cohorts.

Ethics oversight

The study was approved by the NHS Tayside Research Ethics committee (05/s1401/89). Ethics declaration for each participating cohort can be found in the Supplementary information.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size

Sample sizes were not predetermined before the study due to a lack of findings in previous studies leading to insufficient information to determine sample sizes. However, the present study comprises 18 studies and a total of 24,754 individuals, which is the largest MWAS for Major Depression to date and over twice the size of any previous MWAS for depression or depressive symptoms.

Data exclusions

MWAS meta-analysis was conducted on summary statistics shared by each individual study. Data exclusion criteria were described in detail in Supplementary information, Cohort details. In brief, similar quality check procedures were used for all studies, including removing probes with outlying signal intensity, bead count and detection p-values, removing participants with mismatched sex prediction from DNA methylation data and removing cross-hybridising probes that map to common genetic variants (at MAF > 0.05) and polymorphic probes. At the meta-analysis stage, those CpG sites that were either available for more than half of the studies using the given array or had a total sample size over 80% of the max sample size were kept for further analysis. CpG sites with excessive standard errors (SE > 0.5, see Supplementary Figure 1) were removed from analysis.

Replication	Replication of meta-MWAS was conducted via leave-one-out meta-analysis and out-of-sample methylation score classification. Replication for Mendelian Randomisation was conducted by using two independent methQTL datasets (GoDMC for discovery analysis and Generation Scotland for replication). Causal effects of DNAm to Major Depression were tested using these two independent datasets. Out of the 23 potentially causal effects found in the discovery analysis, four CpG sites were specific to the EPIC array, while 19 were available on both array types and could, therefore, be included in the replication analysis.
Randomization	Participants were allocated to case and control groups based on their diagnosis for Major Depression (MD). Life-time diagnosis of MD was derived based on structured clinical interview or self-reported symptoms. Those studies that derived diagnoses of MD based on structural clinical interviews used criteria from The Diagnostic and Statistical Manual of Mental Disorders, Fifth/Fourth Edition (DSM-5/DSM-4)55. Self-declared MD was based on depressive symptoms lasting for more than two weeks and help-seeking due to depression. Studies that derived MD diagnosis based on multiple time points defined cases as those who experienced any depressive episodes during their lifetime, and controls were those who did not declare MD throughout. A total of 9 studies defined MD cases using structural clinical interview (N cases = 3,790), 4 studies used DSM-criteria questionnaires (N cases = 586), 3 studies used self-administered questionnaires for depressive symptoms (N cases = 528), and 3 studies defined MD cases based on self-declared visits to general practitioner (N cases = 539). Details for MD diagnosis for each cohort can be found in Table 1 and the Supplementary Information, Cohort details.
Blinding	Individual-level DNAm and diagnoses for Major Depression were anonymized, and association analysis for each study was conducted using a pipeline that takes individual-level DNAm and diagnosis as inputs to minimize potential bias. Individual study sites then shared summary statistics for inclusion in the meta-MWAS. The meta-analysis was conducted in a double-blind manner as it was independent from the association analyses within individual studies.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Plants

Seed stocks	Not applicable
Novel plant genotypes	Not applicable
Authentication	Not applicable