Word count: 5,347
Figure: 6
Table: 3

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

Xueyi Shen[1*+], Miruna Barbu[1*], Doretta Caramaschi[2], Ryan Arathimos[3], Darina Czamara[4], Friederike S. David[5], Anna Dearman[6], Evelyn Dilkes[6], Marisol Herrera-Rivero[7,8], Floris Huider[9], Luise Kühn[10], Kuan-Chen Lu[11], Teemu Palviainen[12], Alicia Marie Schowe[13,14], Gemma Shireby[15,16], Antoine Weihs[10,17], Chloe C. Y. Wong[3], Eleanor Davyson[18], Hannah Casey[1], Mark J Adams[1], Antje-Kathrin Allgaier[19], Michael Barber[20], Joe Burrage[21], Avshalom Caspi[3,22], Ricardo Costeira[23], Erin C. Dunn[24,25], Lisa Feldmann[26], Josef Frank[27], Franz Joseph Freisleder[28], Danni A. Gadd[29], Ellen Greimel[26], Eilis Hannon[21], Sarah E Harris[30], Georg Homuth[31], David M. Howard[32], Stella Iurato[13,33], Tellervo Korhonen[12], Tzu-Pin Lu[34], Nicholas G Martin[35], Jade Martins[36,13], Edel McDermott[37], Susanne Meinert[38,39], Pau Navarro[40,20], Miina Ollikainen[12,41], Verena Pehl[26], Charlotte Piechaczek[26], Aline D. Scherff[26], Frederike Stein[42,43], Fabian Streit[27], Alexander Teumer[10,44], Henry Völzke[45,44], Jenny van Dongen[9], Rosie M. Walker[2], Natan Yusupov[13,46], Louise Arseneault[3], Jordana T. Bell[23], Klaus Berger[47], Elisabeth Binder[48], Dorret I. Boomsma[9], Simon R Cox[30], Udo Dannlowski[38], Kathryn L. Evans[49], Helen L. Fisher[3,50], Andreas J. Forstner[5,51], Hans J. Grabe[10,17], Jaakko Kaprio[12], Tilo Kircher[42,43], Johannes Kopf-Beck[52,53], Meena Kumari[6], Po-Hsiu Kuo[54,55], Qingqin S Li[56], Terrie E. Moffitt[3,22], Hugh Mulcahy[37,57], Therese M. Murphy[58], Gerd Schulte-Körne[26], Jonathan Mill[15,21], Cathryn M. Lewis[3], BeCOME Working Group[48], OPTIMA Working Group[52], PGC MDD Working Group, Naomi R Wray[59,60], Andrew M McIntosh[1+]

1. Division of Psychiatry, University of Edinburgh, Edinburgh, UK
2. Faculty of Health and Life Sciences, Department of Psychology, University of Exeter, Exeter, UK
3. Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK
4. Max-Planck-Institute of Psychiatry, Munich, DE
5. Institute of Human Genetics, University of Bonn, School of Medicine & University Hospital Bonn, Bonn, DE
6. Institute for Social and Economic Research, University of Essex, Colchester, UK
7. Department of Genetic Epidemiology, Institute of Human Genetics, University of Muenster, Münster, DE
8. Department of Psychiatry (current affiliation), University of Muenster, Münster, DE
9. Biological Psychology, Amsterdam Public Health Institute, Vrije Universiteit Amsterdam, Amsterdam, NL
10. Department of Psychiatry and Psychotherapy, University Medicine Greifswald, Greifswald, DE
11. Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, TW
12. Institute for Molecular Medicine Finland - FIMM, University of Helsinki, Helsinki, Finland
13. Department Genes and Environment, Max-Planck-Institute of Psychiatry, Munich, DE
14. Ludwig-Maximilians-Universität, Graduate School of Systemic Neurosciencee, Munich, DE
15. University of Exeter Medical School, University of Exeter, Exeter, UK
16. Institue of Education, University College London, London, UK
17. German Centre for Neurodegenerative Diseases (DZNE), Site Rostock/Greifswald, Greifswald, DE
18. School of Informatics, University of Edinburgh, Edinburgh, UK
19. Department of Human Sciences, University of the Bundeswehr Munich, Munich, DE
20. MRC Human Genetics Unit, University of Edinburgh, Edinburgh, UK
21. Department of Clinical & Biomedical Sciences, University of Exeter, Exeter, UK
22. Departments of Psychology and Neuroscience, Psychiatry and Behavioral Sciences, and Institute for Genome Sciences and Policy, Duke University, North Carolina, USA
23. Department of Twin Research & Genetic Epidemiology, King's College London, London, UK
24. Psychiatric and Neurodevelopmental Genetics Unit, Center for Genomic Medicine, Massachusetts General Hospital, Boston, USA
25. Harvard Medical School, Department of Psychiatry, Boston, USA
26. Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Hospital of the Ludwig-Maximilians-University (LMU) Munich, Munich, DE

27. Department of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, DE
28. kbo-Heckscher-Klinikum gGmbH, Munich, DE
29. Institute of Genetics and Cancer, Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK
30. Lothian Birth Cohorts, Psychology, University of Edinburgh, Edinburgh, UK
31. Interfaculty Institute for Genetics and Functional Genomics, University Medicine Greifswald, Greifswald, DE
32. Social, Genetic & Developmental Psychiatry Centre, King's College London, London, UK
33. Roche Diagnostics GmbH, Munich, DE
34. Department of Public Health & Institute of Health Data Analytics and Statistics, National Taiwan University, Taipei, TW
35. Queensland Institute of Medical Research Berghofer, Brisbane, AU
36. Life Sciences, Capgemini Engineering, Munich, DE
37. Centre for Colorectal Disease, St Vincent's University Hospital, Dublin, IRL
38. Institute for Translational Psychiatry, University of Münster, Münster, DE
39. Institute for Translational Neuroscience, University of Münster, Münster, DE
40. The Roslin Institute, University of Edinburgh, Edinburgh, UK
41. Minerva Foundation Institute for Medical Research, Helsinki, Finland
42. Department of Psychiatry and Psychotherapy, University of Marburg, Marburg, DE
43. Center for Mind, Brain and Behavior (CMBB), University of Marburg, Marburg, DE
44. German Centre for Cardiovascular Research (DZHK), Partner Site Greifswald, Greifswald, DE
45. Institute for Community Medicine, University Medicine Greifswald, Greifswald, DE
46. International Max Planck Research School for Translational Psychiatry (IMPRS-TP), Munich, DE
47. Institute of Epidemiology and Social Medicine, University of Muenster, Münster, DE
48. Department Genes and Environment, Max Planck Institute of Psychiatry, Munich, DE
49. Centre for Genomic and Experimental Medicine, University of Edinburgh, Edinburgh, UK
50. ESRC Centre for Society and Mental Health, King's College London, London, UK
51. Institute of Neuroscience and Medicine (INM-1), Research Center Jülich, Jülich, DE
52. Clinical Department, Max Planck Institute of Psychiatry, Munich, DE
53. Department of Psychology, LMU Munich, Munich, DE
54. Department of Public Health & Institute of Epidemiology and Preventive Medicine, National Taiwan University, Taipei, TW
55. Department of Psychiatry, National Taiwan University Hospital, Taipei, TW
56. Neuroscience, Janssen Research & Development, LLC, Titusville, NJ, USA
57. School of Medicine, University College Dublin, Dublin, IRL
58. School of Biological, Health and Sports Sciences, Technological University Dublin, Dublin, IRL
59. Institute for Molecular Bioscience, University of Queensland, Brisbane, AU
60. Department of Psychiatry, University of Oxford, Oxford, UK

* Joint contribution
+ Corresponding authors
Dr Xueyi Shen
E-Mail: xueyi.shen@ed.ac.uk
Address:
University of Edinburgh, Royal Edinburgh Hospital, Morningside Park
Edinburgh EH10 5HF
United Kingdom

Prof Andrew M McIntosh
E-Mail: andrew.mcintosh@ed.ac.uk
Address:
University of Edinburgh, Royal Edinburgh Hospital, Morningside Park
Edinburgh EH10 5HF
United Kingdom

Abstract

Major Depression (MD) is a leading cause of global disease burden, and both experimental and population-based studies suggest that differences in DNA methylation (DNAm) may be associated with the condition. However, previous DNAm studies have not so far been widely replicated, suggesting a need for larger meta-analysis studies. In the present study, the Psychiatric Genomics Consortium Major Depressive Disorder working group conducted a meta-analysis of methylome-wide association analysis (MWAS) for life-time MD across 18 studies of 24,754 European-ancestry participants (5,443 MD cases) and an East Asian sample (243 cases, 1846 controls). We identified fifteen CpG sites associated with lifetime MD with methylome-wide significance ($p < 6.42 \times 10^{-8}$). Top CpG effect sizes in European ancestries were positively correlated with those from an independent East Asian MWAS ($r = 0.482$ and $p = 0.068$ for significant CpG sites, $r = 0.261$ and $p = 0.009$ for the top 100 CpG sites). Methylation score (MS) created using the MWAS summary statistics was significantly associated with MD status in an out-of-sample classification analysis ($\beta = 0.122$, $p = 0.005$, AUC = 0.53). MS was also associated with five inflammatory markers, with the strongest association found with Tumor Necrosis Factor Beta ($\beta = -0.154$, $p = 1.5 \times 10^{-5}$). Mendelian randomisation (MR) analysis demonstrated that 23 CpG sites were potentially causally associated with MD and six of those were replicated in an independent mQTL dataset (Wald's ratio test, absolute β ranged from 0.056 to 0.932, p ranged from 7×10^{-3} to 4.58×10^{-6}). CpG sites located in the Major Histocompatibility complex (MHC) region showed the strongest evidence from MR analysis of being associated with MD. Our study provides evidence that variations in DNA methylation are associated with MD, and further evidence supporting involvement of the immune system. Larger sample sizes in diverse ancestries are likely to reveal replicable associations to improve mechanistic inferences with the potential to inform molecular target identification.

Introduction

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%[4], and polygenic risk scores (PRS) trained on genome-wide-association (GWAS) currently explain 1.5-3.2% of the variance in MD[3].

DNA methylation (DNAm) is one of the most studied epigenetic processes and is influenced by both genetic and environmental factors[5]. DNAm is dynamic and is associated with changes in environmental and lifestyle factors, including smoking[6], alcohol[7], and BMI[8], all factors that are implicated in MD[3]. A recent study[9] identified associations between methylation scores (MS), calculated using methylome-wide association study (MWAS) summary statistics for several relevant lifestyle factors and MD[10]. 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. (2018) 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 guidance[11]. Starnawska et al. (2019) investigated depressive symptomatology in a sample of 724 monozygotic twins, with top findings annotated to genes previously implicated in depression[12]. Finally, Huls et al. (2020) 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 MD[13]. In addition, MS calculated using penalised regression models have previously been utilised to investigate MD[14]. We recently used lasso regression to calculate an MD MS in N=1,780 participants and found that the scores explain approximately 1.75% of the variance in liability to life-time MD, acting additively with PRS[14]. Findings from these studies have been somewhat inconsistent regarding the specific associations. Larger MWAS studies may provide more reliable estimates of the differences in DNAm between MD cases and controls, and in doing so, bring insight to the molecular mechanisms associated with MD.

Given previous evidence highlighting the role of DNAm in MD, we conducted an MD MWAS meta-analysis using DNA extracted from whole blood in 18 cohorts comprising 24,754 individuals (5,443 cases) of European ancestry. We sought to identify whether differences in DNAm were a potentially cause or consequence of MD using a two-sample Mendelian Randomisation 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.

Methods

Cohort information

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 Methods. Written consents were obtained from all participants.

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 normalisation were conducted by each individual study team. Details are provided in the protocol papers for each individual study (Table 1 and Supplementary Methods). In brief, the majority of studies used functional normalisation for methylation data pre-processing, unless stated otherwise in the Supplementary Methods[15]. 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-hybridising probes that map to common genetic variants (at MAF > 0.05) and polymorphic probes[15]. M-values were used for the association analysis[16].

Diagnosis of Major Depression (MD)

Life-time diagnosis for MD were 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)[17]. 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 7 studies defined MD cases using structural clinical interview (N cases = 3,246), 5 studies used DSM-criteria questionnaires (N cases = 1,130), 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 Materials, and Methods sections.

Additional exclusion criteria per study are stated in the Supplementary Methods.

Association 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 Methods. The pipeline uses the R package

'limma' for linear regression on large omic data[18]. Three models representing increasingly rigorous correction for potential confounders were estimated. The simplest model contained age, sex, batch, the first 20 methylation principal components (PCs) or surrogate variables (SVs)[19] 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[19]. 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[20] 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-adjusted model (referred to as the 'main model') are reported as the main findings. Standardised Cohen's d between MD cases versus controls were reported as effect sizes.

Meta analysis

Meta analysis was conducted using METAL (version 2011)[21] 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).
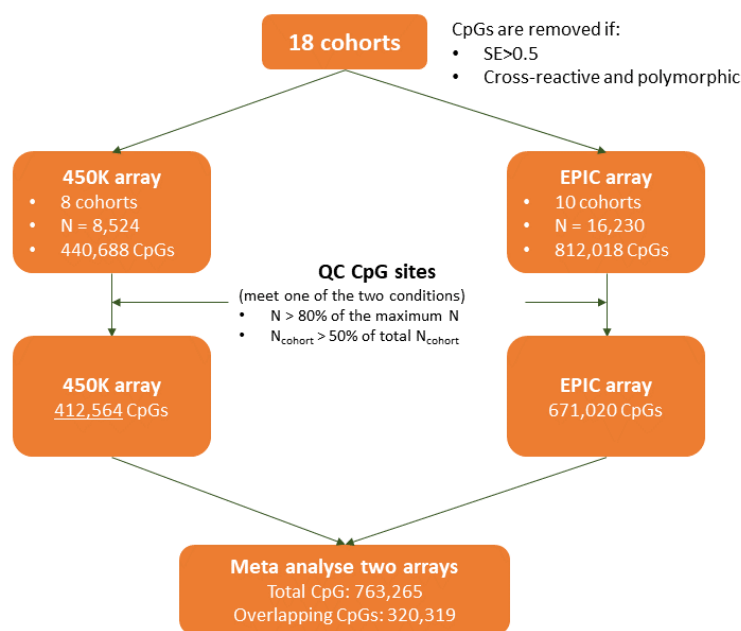
Figure 1. Workflow of meta-analysis.

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.

Second, the summary statistics for 450K and EPIC array were meta-analysed. 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)[22]. 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

We used the 'gometh' function from the 'missMethyl' R package[23] 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. Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were analysed separately. P-values for both enrichment analyses were FDR corrected.

Identification of differentially methylated regions (DMR)

DMR were analysed using the 'dmrff' R package[24]. A DMR was identified if 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-analysed within the identified region. P-values were Bonferroni-corrected for all regions (>2 CpG sites) and single CpG sites altogether. Significant DMR were identified if the Bonferroni-corrected $p < 0.05$ and the number of CpG sites within the region was ≥ 2.

Analysis of the confounding effect of body mass index (BMI) and alcohol consumption

BMI and alcohol consumption are risk factors for MD and are known to have widespread associations with DNAm[7,25]. To investigate whether the signals found in the main model were due to the effects of BMI and consumption, we conducted an additional meta-analysis of a 'complex model' for the 14 cohorts that had BMI and alcohol consumption data available (N_{total} = 20,196, see Supplementary Figure 2 for sample sizes of individual studies). BMI and alcohol consumption were added as additional covariates for the complex model. Effect sizes and p values for the significant associations between the basic model and complex model were compared.

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: $p \leq 1$, ≤ 0.01, ≤ 0.001, $\leq 1 \times 10^{-6}$, and $\leq 5 \times 10^{-8}$, resulting in five MD-MS.

7

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-MS (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

Generation Scotland (GS) DNAm data Set 3 was used for out-of-sample classification (GS DNAm data Set 1 and 2 were included in the meta-MWAS). GS DNAm data Set 3 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 genomic relationship matrix. Quality check and preprocessing were kept consistent with the GS sample used in the meta-MWAS (Set 1 and 2).

MD diagnosis was derived using electronic health records (EHR) from General Practitioners' (GP) diagnosis[26]. Details for the EHRs were explained in the protocol paper by Kerr *et al*[26]. 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 (https://phenotypes.healthdatagateway.org/phenotypes/PH1021/version/2204/detail/)[27,28] and PH149 (https://phenotypes.healthdatagateway.org/phenotypes/PH149/version/298/detail/)[29] inventories of the HDR UK Phenotype Library for primary care data of depression were used to identify cases of life-time 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 wide-spread association between protein abundance and DNAm[30]. We conducted a proteome-wide association analysis study (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 (Ilumina, San Diego, CA, USA). Proteomic data was profiled using lithium heparin collected plasma samples analysed using a 92-plex proximity extension assay (inflammation panel), Olink® Bioscience, Uppsala Sweden. For 22 proteins over 40% of samples fell below the lowest limit of detection, leaving 70 post-quality-check proteins. Full details on sample preparation and quality check have been reported previously[31].

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

8

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 analyse the potential effect of age across studies. A mixed-effect model with the 'metareg' R package[32] was used for the meta-regression analysis. Mean age of each individual cohort was set as a random effect. Standardised 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 (REML) method. The standardised 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. Generation Scotland (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 1000 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 a MWAS of BMI and conducted a similar analysis, for comparison with MD, to assess whether BMI-DNAm associations were similarly heterogeneous.

Mendelian randomisation (MR)

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 Biobank[3]. A total of 807,553 individuals (246,363 cases and 561,190 controls) of European ancestry were included in the MD meta-GWAS.

Genetics of DNA Methylation Consortium (GoDMC) and GS methylation quantitative trait loci (mQTL)

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 (OSCA) was used for mQTL estimation[33]. Details for the mQTL analysis in GS can be found elsewhere[34]. Covariates were kept consistent with the main model for the MD MWAS, except for using self-reported smoking behaviour (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/), using the same pipeline described in the GoDMC protocol paper by Min et al. (2021)[35]. 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 $< 1\times10^{-5}$ were obtained from participating cohorts. This initial stage created a candidate list of mQTL associations ($n = 120,212,413$). Next, meta-analyses for mQTL were then conducted on these candidate associations. A full description of the GoDMC analysis can be found elsewhere[34,35].

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

<u>Selection of CpG list</u>

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 analysis[35].

In the GoDMC dataset, a total of 156 CpG sites that met the above criteria had at least one *cis* mQTL. *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, $p<5\times10^{-8}$), 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.

<u>MR methods</u>

MR analysis was conducted using the 'TwoSampleMR' R package (version 0.5.6)[36]. To identify causal effects of DNAm on MD, we used the Wald's ratio MR method[37] to analyse causal effects on MD using cis mQTLs (within a 1Mb window in vicinity of the chosen CpG site). The most significant mQTL for each CpG site that reached the threshold of $p<5\times10^{-8}$ 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 $p \leq 5\times10^{-8}$, with a 1Mb window and r=0.001. Causal effects from MD to DNAm were tested using mQTL data from the entire GS sample (Set 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 Materials. As in the main meta-analysis, biological and technical covariates, as well as age, sex, and smoking (indexed by AHRR 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 in 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 $p < 1x10^{-5}$ ($N_{CpG}=24$), as used in previous studies[38].

Results

MD meta-MWAS

A total of 15 CpG sites were significant after Bonferroni correction from the basic model (p < 6.55×10^{-8}, $p_{Bonferroni}$ < 0.05, see Table 1 for details of participating studies, Figure 2 and Table 2 for significant findings).

Five of the significant CpG sites positionally map to genes associated with mental health, neurodegenerative and developmental disorders. The gene mapped from cg17002328 (*CCDC88C*) was associated with schizophrenia[39] and ADHD[40] in epigenetic studies, and brain structural measures in GWAS studies, such as cortical surface area[41] and accumbens volume[42]. CpG site cg17494199 was associated with preterm birth[43], birthweight[44] and foetal brain development[45] 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[46-48].

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), mapping to gene *LZTFL1*. From the EWAS catalog, the *LZTFL1* gene was associated with biomarkers related to pain, such as glucose level[49], and autoimmune disease or markers, such as rheumatoid arthritis[50] and C9 protein levels[30]. Four of the significant CpG sites, cg12187985 (*SS18*), cg10145533, cg09726355 (*GPATCH8*) and cg07063853 (*API5*), map to genes previously found associated with autoimmune markers (e.g. rheumatoid arthritis[50]) 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, BMI[25] and cholesterol to total lipids ratio[51].

A complete list of CpG sites and related genes and traits can be found in Table 2. The inflation factors of the association statistics for each individual cohort can be found in Supplementary Figure 3. See Supplementary Figure 4 for effect sizes of each study for the significant CpG sites.
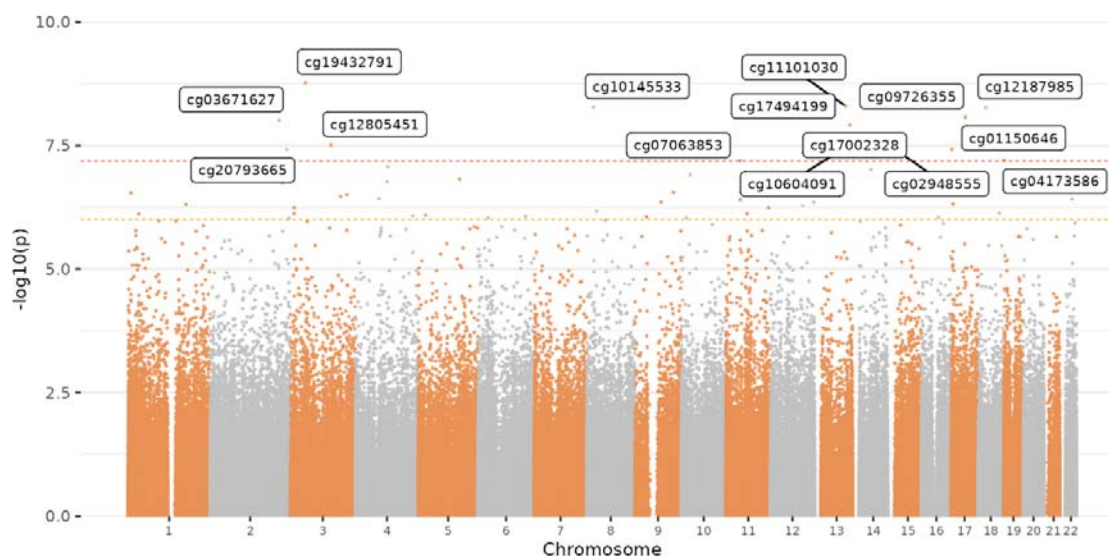
Figure 2. Manhattan plot for the meta-analysis of MWAS for MD. Each dot represents a CpG probe. X-axis represents the relative position of the probes in the genome. Y-axis represents –log10-transformed p-values. The red and yellow dashed line represents the significance threshold for Bonferroni and FDR correction, respectively.

Identification of DMR

A total of 37 DMRs were identified after Bonferroni correction. The most strongly associated region locates within the major histocompatibility complex (MHC) region, discoidin domain receptor tyrosine kinase 1 (*DDR1*) gene (CHR6:30,853,258-30,854,233), previously implicated in major psychiatric disorders (e.g. schizophrenia, MD and bipolar disorder).

The top ten most significant DMRs are listed in Table 3, and the full list of significant DMRs can be found in Supplementary Table 1.

Pathway enrichment analysis

No GO term or KEGG pathway was significantly enriched after FDR correction (Supplementary Table 2). The top 10 most significantly enriched GO terms included pathways relevant to protein and metabolic processes (e.g. negative regulation of protein localisation 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 ($p \le 0.007$).

The basic model versus complex model

Results for the basic model (18 cohorts, N_{total} = 24,754) and complex model (15 cohorts, N_{total} = 20,196, see Supplementary Figure 2) were compared to evaluate the confounding effect of BMI and alcohol consumption (Figure 3). 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 ($p < 2.51 \times 10^{-4}$, $p_{Bonferroni} < 0.004$ corrected across the significant CpG sites in discovery analysis).
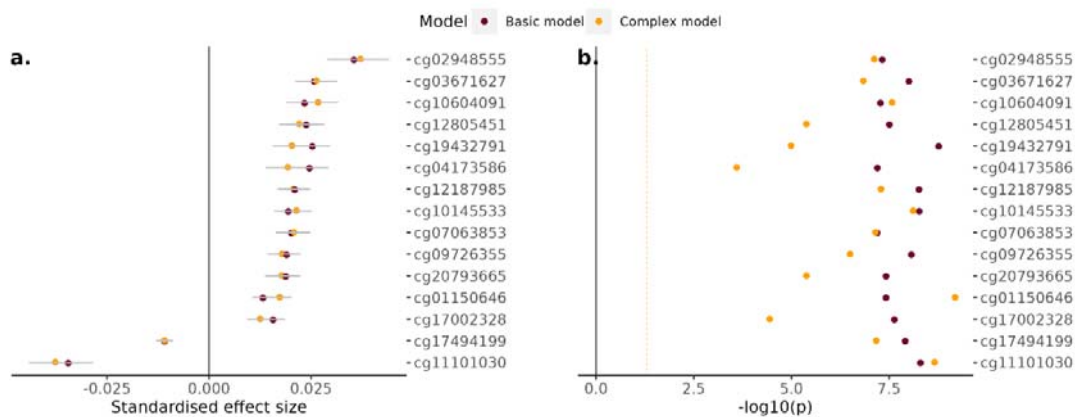
Figure 3. Comparison between basic and complex models. In both (a) comparison of effect sizes and (b) comparison of p values, each dot represents a CpG probe. In panel (a), x-axis represents effect size, and in panel (b), x-axis represents -log10-transformed p value. Differences in p-values reflect both the model used and the sample sizes. Y-axis represents individual CpG site. The yellow dashed line in panel (b) represents the significance threshold for nominal significance (p < 0.05).

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

All five MD-MS 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 becomes increasingly stringent (see Figure 4). Out of the five scores tested, only MS at p-value threshold of $\leq 5 \times 10^{-8}$ was found associated with MD diagnosis ($\beta = 0.13$, p = 0.003, see Figure 4). P-values for other MS ranged from 0.069 to 0.314.
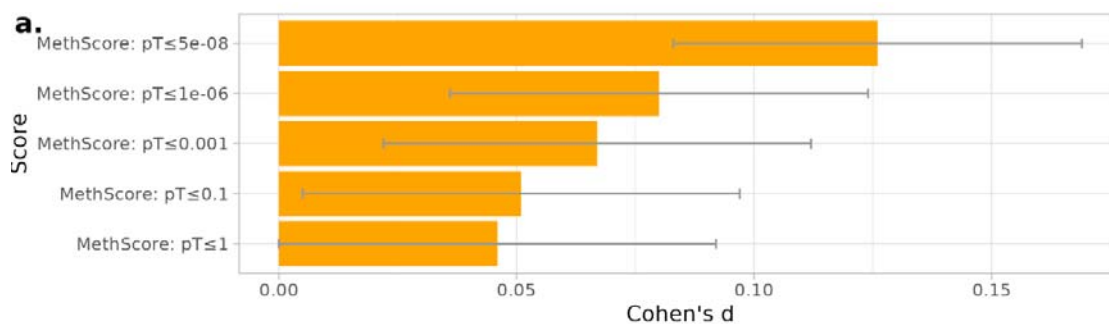
Score	Cohen's d	SE	p
MethScore: pT≤5e-08	0.13	0.04	0.003
MethScore: pT≤1e-06	0.08	0.04	0.069
MethScore: pT≤0.001	0.07	0.04	0.139
MethScore: pT≤0.1	0.05	0.05	0.269
MethScore: pT≤1	0.05	0.05	0.314

13

Figure 4. Out-of-sample classification of MD using MS. (a) Bar-plot for effect sizes of associations. Length of bar represents standardised log-transformed odds ratio and standard error (error bar). Y-axis represents MS. (b) Statistics for MS association tests.

Association between MS for MD and inflammatory protein abundance

Five proteins were significantly ($p < 3.5\times10^{-3}$) associated with the MD-MS created at the p-threshold of $\leq 5\times10^{-8}$. The strongest association was found in *TNFB* ($\beta = -0.15$, $p = 2.37\times10^{-5}$). Other proteins were: *IL6* ($\beta = 0.12$, $p = 5.18\times10^{-4}$), *TFG alpha* ($\beta = 0.11$, $p = 2.8\times10^{-3}$), CD5 ($\beta = -0.11$, $p = 2.1\times10^{-3}$) and *EN-RAGE* ($\beta = 0.11$, $p = 3.5\times10^{-3}$). See Supplementary Figure 5.

Heterogeneity analysis

Leave-one-out analysis

For every CpG site, all effect sizes for leave-one-out analyses remained in the same direction as the meta-analysis (Supplementary Figure 6). All tested CpG sites remained significant after leaving individual studies out except when the largest study, Generation Scotland, was left out ($p < 1.03\times10^{-4}$). When Generation Scotland was left out, 11 out of the 15 tested CpG sites remained nominally significant ($p < 8.91\times10^{-3}$). The four sites that became non-significant when Generation Scotland 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

Out 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 Figure 7). One site, cg04173586, showed a significant effect of age ($\beta = -9.25\times10^{-4}$, $p = 0.006$). However, leave-one-out analysis both showed highly consistent findings for cg04173586 across studies. There were 11 out of 15 participating cohorts that showed effect sizes consistent with the meta-MWAS (Supplementary Figure 7) and leave-one-out analyses were significant for all iterations (p ranged from 1.03×10^{-4} to 2.38×10^{-8}, Supplementary Figure 4 and Supplementary Table 3).

Correlation matrix for effect sizes

Heterogeneity between studies was analysed by looking at the between-study correlation of effect sizes estimated using the basic model (see Figure 5). 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 Netherland Twin Register and Janssen (r=0.31). Out of the 153 pairwise correlations, 96 were positive (62.7%). Compared to 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).
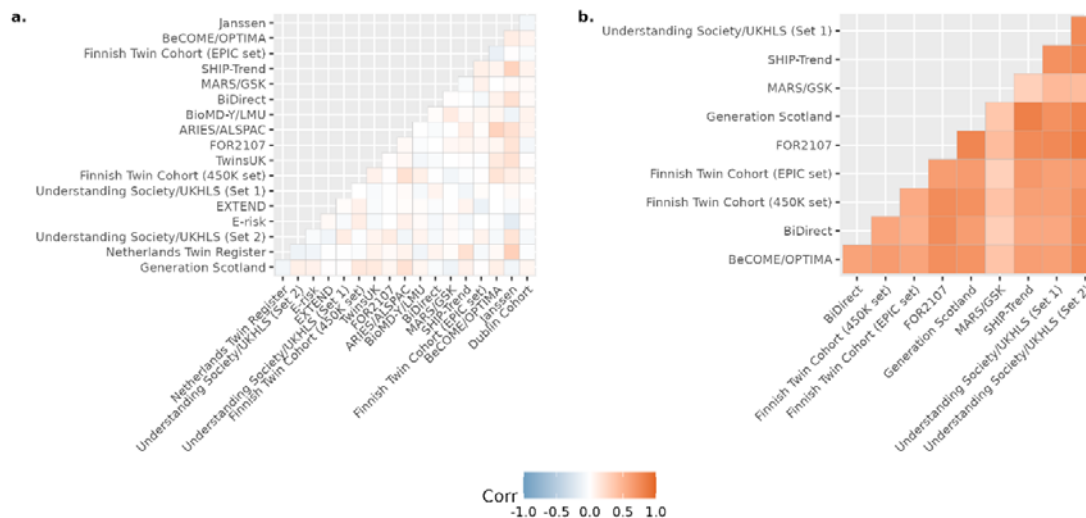
Figure 5. Correlation matrices for MD and BMI MWAS, respectively. Panel (a) shows the correlation matrix for the top 1000 CpG sites in GS MD MWAS. Pair-wise correlation of effect sizes are shown in the figure. Panel (b) shows the correlation matrix for the top 1000 CpG sites in GS BMI MWAS. Pair-wise correlation of effect sizes are shown in the figure.

Mendelian Randomisation (MR)

Using *cis* mQTL data from GS and Wald's ratio 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}, see Figure 6a and Supplementary Data 1). There were 17 CpG sites located in the MHC 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.

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. Seven CpG sites tested were replicated (p<0.003, Figure 6b). Six of the replicated CpG sites were in the MHC region (β: 0.047 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*, β = -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^{-5} to 0.22, $p_{uncorrected}$ > 0.001, p_{FDR} > 0.4).
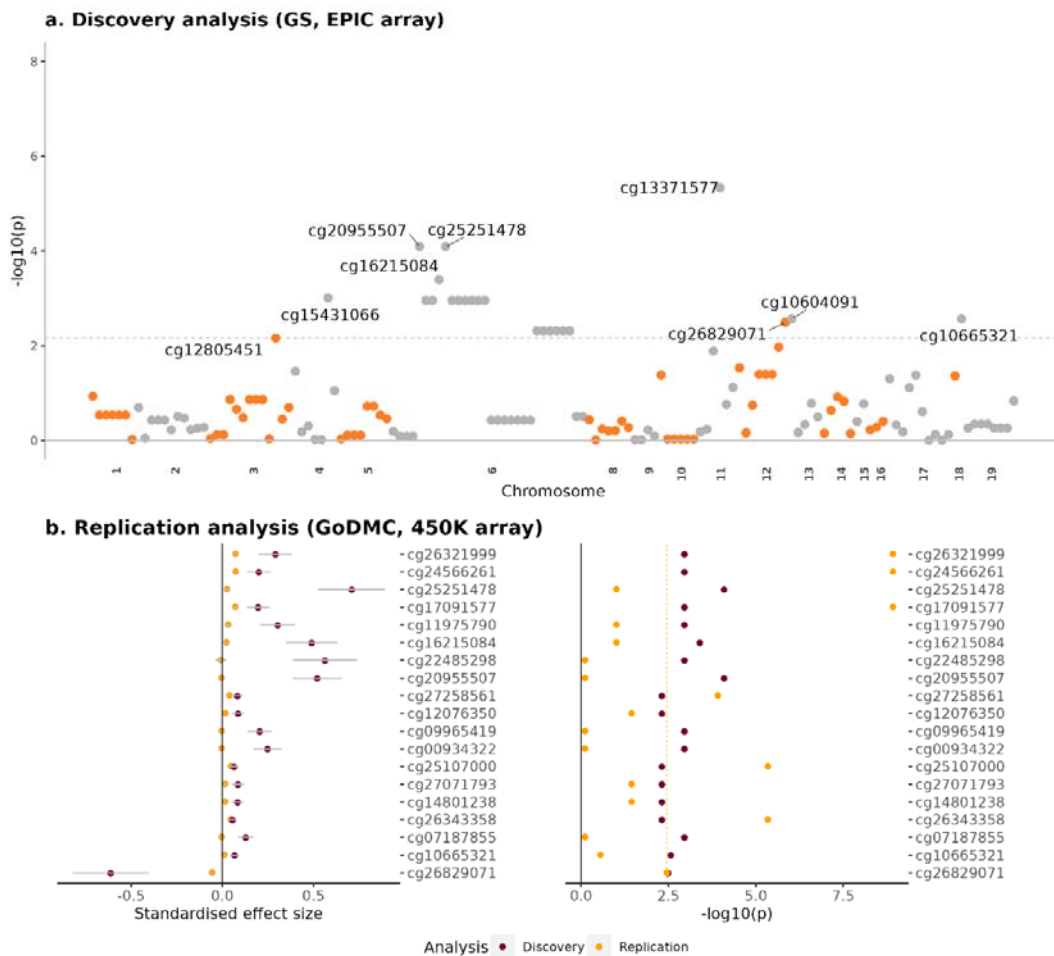
Figure 6. Significance of Mendelian Randomisation analysis of the causal effect of DNAm on MD. (a) P-plot for discovery MR analysis in GS. Each dot represents a CpG site. X-axis represents chromosomes and base pair position. Y-axis represent -log10-tranformed p-value of Wald's ratio MR analysis. The grey 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. Out 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. X-axes represent effect size and -log10-transformed p value on the left and right panels. Y-axis represents individual CpG site. The yellow, dashed line in the right panel represents FDR-significance threshold.

MWAS in East Asian ancestry

We sought to identify DNAm associations with MD in East Asian ancestry using data from Taiwan Biobank (TBB). There were no methylome-wide significant findings identified. Correlation between the methylome-wide significant CpGs identified in the main results (N=15) and the same CpGs in TBB was r = 0.48. Effect direction across the two cohorts was the same for 11 of the 15 CpGs (73.3%

in the same direction, see Supplementary Table 4-6). Positive correlations of effect sizes 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 associations[11]. 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 heterogeneous and low for MD, in contrast to BMI which showed uniformly positive, significant 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-19[52]. In addition to its association with inflammation, *LZTFL1* regulates ciliary localisation of the Bardet–Biedl syndrome (BBSome) protein complex. The BBSome complex is a well-replicated causative protein marker for obesity[53] and highly relevant to the HPA-axis activity by being involved in transducing leptin signals in hypothalamic neurons[54,55]. Consistent with the top CpG site, other sites, such as cg02948555 and cg10604091, mapped to genes associated with C-reactive protein (CRP)[8,56]. CRP is one of the most well-studied predictors of vulnerability to depression[57] and persistence of depressive symptoms[58]. As a pro-inflammatory cytokine, CRP has been found as an indicator of hyper-activity in the HPA-axis associated with stress[59]. The PWAS findings provided additional evidence for the association between the top CpG sites and inflammation, for example, in tumor necrosis factor-beta (*TNFB*) 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, likely due to the limited sample sizes of MWAS conducted on mental health-related traits.

Three CpG sites were associated with traits (e.g 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[3,60]. Within the genes in vicinity of the four CpG sites, *ZNF106* is involved in the regulation of calcium homeostasis that is crucial for cell survival and death[61]. Other genes, including *MYO1C* and *FNDC3B*, are understood to be involved in energy metabolism and homeostasis and highly sensitive to stress coping and inflammation[62,63]. 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 samples[13]. Taking into account 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 observed varying results from different cohorts in our study, with little evidence of systematic differences in age differences between studies 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 > 1000) 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 in size. The considerable costs of DNAm profiling, compared to genotyping, unfortunately act as a barrier towards 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 (e.g. 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 groups[64]. Developing larger non-European DNA methylation samples will be crucial to give statistical balanced comparison between ancestry groups and to identify ancestry-group-specific DNA methylation sites for MD[65].

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. Additionally, there is the 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 types[66]. 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.

Funding and acknowledgement

Cohort name	PMIDs	N cases	Sample size	MD prevalence	Case phenotype	Mean age	Gender (% male)	Genetic ethnicity	DNAm array	Tissue type
ES/ALSPAC	25991711	204	657	31.1%	Questionnaire	29.4/~45(mums)	All women	Europeans	450K	Whole blood
ME/OPTIMA	32393358; 33054737	222	305	72.8%	DSM-criteria questionnaire	37.5	38.4%	European	EPIC	whole blood
BiDirect	24924233	287	561	51.5%	Questionnaire	51.6	49.7%	Europeans	EPIC	Whole Blood
blin Cohort	26419460	37	186	19.9%	Questionnaire	38.0	52.7%	Europeans	450K	PBMCs
E-risk	12236608 ; 12537874	322	1625	19.8%	DSM-criteria questionnaire	18.5	51.0%	Predominantly European	450K	Whole blood
EXTEND	29790996	190	1181	16.1%	Self-reported diagnosis/GP visit	56.3	48.0%	Predominantly European	450K	Whole blood
in Cohort (450K set)	31796134; 31640839	163	985	16.5%	Clinical interview	23.8	45.9%	Finnish, Europeans	450K	Whole blood
in Cohort (EPIC set)	31796134; 31640839	93	366	25.4%	Clinical interview	25.4	49.5%	Finnish, Europeans	EPIC	Whole blood
FOR2107	30267149	340	667	51.0%	Clinical interview	35.3	36.7%	Europeans	EPIC	Whole blood
cotland (Set 1 and 2)	30918249	1,632	9,502	17.2%	Clinical interview	49.8	41.0%	Europeans	EPIC	Whole blood
Janssen	36319817	191	222	86.0%	DSM-criteria questionnaire	42.8	19.7%	Predominantly European	EPIC	whole blood
MD-Y/LMU	30947532	381	633	60.2%	Clinical interview	15.1	33.3%	European	EPIC	whole blood
ARS/GSK	18586274; 16822851	311	497	62.6%	DSM-criteria questionnaire	48.1	42.5%	European	450K	whole blood
nds Twin Register	31666148	436	2701	16.1%	Clinical interview/DSM-criteria questionnaire	36.5	33.0%	Europeans	450K	Whole blood
HIP-Trend	35348705	84	492	17.1%	DSM-criteria questionnaire	51.1	46.3%	Europeans	EPIC	Whole blood
TwinsUK	30760334	201	692	29.0%	Clinical interview	59.0	All women	Europeans	450K	Whole blood
ding Society / UK Longitudinal Study (HLS) Set 1	30401456	72	1121	6.4%	Self-reported diagnosis/GP visit	58.4	41.9%	Europeans	EPIC	Whole blood
ding Society / UK Longitudinal Study (HLS) Set 2	30401456	277	2361	11.7%	Self-reported diagnosis/GP visit	51.1	45.9%	Europeans	EPIC	Whole blood

ormation for cohorts that participated in the methylome-wide meta-analysis.

CHR	BP	beta	se	p	p-adj	N	UCSC gene	Relation to Island	Array	EWAS catalog / EWAS atlas (PMID)	GWAS catalog gene look-up (PMID)
3	45894556	0.025	0.004	1.70E-09	0.001	16545	LZTFL1	OpenSea	EPIC	C9 protein levels (35945220); Glucose (33151971); Schizophrenia (26619358)	Multiple sclerosis (31604244); COVID19 (35910207); CD3 protein levels (32929287)
13	93744566	-0.035	0.006	4.99E-09	0.004	15901	--	OpenSea	EPIC	--	--
8	20350779	0.019	0.003	5.34E-09	0.004	24191	--	OpenSea	Both	Crohn's disease* (27279921); Preterm birth* (28428831); Autoimmune disease* (31024609)	--
18	23654410	0.021	0.004	5.43E-09	0.004	16850	SS18	OpenSea	EPIC	Preterm birth (28428831); Rheumatoid arthritis (23334450); LRIG1 protein (35945220)	Neuroticism (30595370); Autoimmune thyroid disease (32581359)
17	42544621	0.019	0.003	8.53E-09	0.007	16628	GPATCH8	OpenSea	EPIC	Rheumatoid arthritis (23334450); HIV infection (27105112); Alcohol consumption (27843151)	Cerebrospinal fluid progranulin levels (33087363); Brain morphology (32665545); Educational attainment (35361970)
2	209185667	0.026	0.005	9.85E-09	0.008	16850	PIKFYVE	OpenSea	EPIC	Kidney disease* (33933144); Alcohol consumption (31789449); Childhood obesity (28614626)	BMI (36581621); Asthma (27611488)
13	106834942	-0.011	0.002	1.22E-08	0.009	24748	--	OpenSea	Both	Very early preterm birth* (30833390), birthweight* (31015461), fetal brain development* (25650246)	--
14	91751773	0.016	0.003	2.32E-08	0.018	24750	CCDC88C	N_Shelf	Both	Multiple sclerosis* (30479356); Chronic fatigue syndrome (30036399) Schizophrenia (33279932); Autism (27404287); BMI (33517419); ADHD (28785368)	White matter hyperintensity volume (33293549); Subcortical volume (32665545); Autism (35215271); Alzheimer disease (26830138)
3	123672734	0.024	0.004	3.11E-08	0.024	16850	CCDC14	OpenSea	EPIC	--	Intelligence (29844566); General cognitive ability (29326435)
2	232549224	0.019	0.003	3.79E-08	0.029	24752	--	S_Shelf	Both	--	--
17	1377304	0.013	0.002	3.80E-08	0.029	16850	MYO1C	S_Shore	EPIC	Insulin resistence (30792424); BMI (29278407); Alzheimers disease (33257653); Alcohol consumption (27843151); Preterm birth (28428831)	Alzheimer disease (35770850); Educational attainment (30038396)
15	42749848	0.036	0.007	4.72E-08	0.036	15911	ZNF106	OpenSea	EPIC	Downs syndrome (33547282); Early onset intracranial atherosclerotic stenosis (31142690)	CRP (31900758); Cholesterol to total lipid ratio in LDL (35213538)
13	50650086	0.023	0.004	5.27E-08	0.040	16850	DLEU2	OpenSea	EPIC	BMI (29278407); Inflammatory bowel disease (32281463); Alzheimers disease (25129075); Birthweight (31015461); Insufficient sleep (30718923)	BMI (31669095); Creatinine (35710981, 34594039); Non-HDL cholesterol levels (34887591)
19	2167496	0.025	0.005	6.30E-08	0.048	24043	DOT1L	S_Shore	Both	Schizophrenia* (33646943); Alcohol consumption* (31789449)	BMI-adjusted waist-hip ratio (30575882); gamma-linolenic acid (26584805)
11	43333512	0.020	0.004	6.42E-08	0.049	23395	API5	Island	Both	HIV infection* (27105112); Obesity (29692867); Rheumatoid arthritis (27585642); Alzheimer's disease (33257653)	--

op ten significant differentially methylated regions (DMRs) identified in the basic model. Start/End = start/end base pair position of the DMR.

CHR	Start	End	Gene(s)	N_{CpG}	Beta	se	p	p.bonferroni
6	30853258	30854233	DDR1	12	0.036	0.005	9.89E-13	6.75E-08
19	37825211	37825455	HKR1	8	0.036	0.005	5.66E-12	3.87E-07
10	135342560	135343248	CYP2E1	5	0.031	0.005	1.42E-10	9.68E-06
3	18480242	18481064	SATB1	8	0.048	0.009	1.16E-07	0.008
6	30698784	30698936	FLOT1	4	0.046	0.009	1.45E-07	0.010
11	43333495	43333512	API5	3	0.055	0.009	2.59E-09	1.77E-04
17	7832769	7832932	KCNAB3	3	0.036	0.005	1.14E-13	7.78E-09
9	139557250	139557920	EGFL7	4	-0.059	0.011	2.20E-08	0.002
2	219738714	219738732	WNT6	2	0.074	0.014	5.44E-08	0.004
3	11643427	11643630	VGLL4	3	0.083	0.015	4.45E-08	0.003

References

1   Sullivan P.F., Neile M.C. & K.S, K. Genetic Epidemiology of Major Depression: Review and Meta-Analysis. *American Journal of Psychiatry* **157**, 1552-1562 (2000).

2   McIntosh, A. M., Sullivan, P. F. & Lewis, C. M. Uncovering the Genetic Architecture of Major Depression. *Neuron* **102**, 91-103, doi:10.1016/j.neuron.2019.03.022 (2019).

3   Howard, D. M. *et al.* Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. *Nature Neuroscience* **22**, 343-352, doi:10.1038/s41593-018-0326-7 (2019).

4   Pettersson, 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. *Psychological Medicine* **49**, 1166-1173, doi:10.1017/s0033291718002039 (2019).

5   Zeng, Y. *et al.* Parent of origin genetic effects on methylation in humans are common and influence complex trait variation. *Nature Communications* **10**, doi:10.1038/s41467-019-09301-y (2019).

6   Joehanes, R. *et al.* Epigenetic Signatures of Cigarette Smoking. *Circ Cardiovasc Genet* **9**, 436-447, doi:10.1161/CIRCGENETICS.116.001506 (2016).

7   Liu, C. *et al.* A DNA methylation biomarker of alcohol consumption. *Mol Psychiatry* **23**, 422-433, doi:10.1038/mp.2016.192 (2018).

8   Wahl, S. *et al.* Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature* **541**, 81-86, doi:10.1038/nature20784 (2017).

9   Barbu, M. C. *et al.* Complex trait methylation scores in the prediction of major depressive disorder. *eBioMedicine* **79**, 104000, doi:10.1016/j.ebiom.2022.104000 (2022).

10   Nabais, M. F. *et al.* An overview of DNA methylation-derived trait score methods and applications. *Genome Biol* **24**, 28, doi:10.1186/s13059-023-02855-7 (2023).

11   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, doi:10.1001/jamapsychiatry.2018.1725 (2018).

12   Starnawska, A. *et al.* Epigenome-wide association study of depression symptomatology in elderly monozygotic twins. *Translational Psychiatry* **9**, doi:10.1038/s41398-019-0548-9 (2019).

13   Huls, A. *et al.* Association between DNA methylation levels in brain tissue and late-life depression in community-based participants. *Transl Psychiatry* **10**, 262, doi:10.1038/s41398-020-00948-6 (2020).

14   Barbu, M. C. *et al.* Epigenetic prediction of major depressive disorder. *Mol Psychiatry* **26**, 5112-5123, doi:10.1038/s41380-020-0808-3 (2021).

15   Walker, R. M. *et al.* Assessment of dried blood spots for DNA methylation profiling. *Wellcome Open Res* **4**, 44, doi:10.12688/wellcomeopenres.15136.1 (2019).

16   Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587, doi:10.1186/1471-2105-11-587 (2010).

17   Association, A. P. *Diagnostic And Statistical Manual of Mental Disorders, Fifth Edition*. (2013).

18   Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research* **43**, e47-e47, doi:10.1093/nar/gkv007 (2015).

19   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 Biology* **18**, doi:10.1186/s13059-017-1148-8 (2017).

20   Philibert, R., Dogan, M., Beach, S. R. H., Mills, J. A. & Long, J. D. AHRR methylation predicts smoking status and smoking intensity in both saliva and blood DNA. *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **183**, 51-60, doi:10.1002/ajmg.b.32760 (2020).

21   Willer, C. J., Li, Y. & Abecasis, G. R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191, doi:10.1093/bioinformatics/btq340 (2010).

22   IlluminaHumanMethylationEPICanno.ilm10b4.hg19: Annotation for Illumina's EPIC methylation arrays v. R package version 0.6.0 (2017).

23   Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* **32**, 286-288, doi:10.1093/bioinformatics/btv560 (2016).

24   Suderman, M. *et al. dmrff: identifying differentially methylated regions efficiently with power and control* (Cold Spring Harbor Laboratory, bioRxiv, 2018).

25   Geurts, Y. M. *et al.* Novel associations between blood DNA methylation and body mass index in middle-aged and older adults. *Int J Obes (Lond)* **42**, 887-896, doi:10.1038/ijo.2017.269 (2018).

26   Kerr, S. M. *et al.* Electronic health record and genome-wide genetic data in Generation Scotland participants. *Wellcome Open Res* **2**, 85, doi:10.12688/wellcomeopenres.12600.1 (2017).

27   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, doi:10.1136/bmjopen-2016-013167 (2016).

28   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, doi:10.1017/S0033291716002099 (2016).

29   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, doi:10.1016/S2589-7500(19)30012-3 (2019).

30   Gadd, D. A. *et al.* Integrated methylome and phenome study of the circulating proteome reveals markers pertinent to brain health. *Nature Communications* **13**, doi:10.1038/s41467-022-32319-8 (2022).

31   Hillary, R. F. *et al.* Multi-method genome- and epigenome-wide studies of inflammatory protein levels in healthy older adults. *Genome Med* **12**, 60, doi:10.1186/s13073-020-00754-1 (2020).

32   Schwarzer G., C. J. R., Gucker G. *Meta-Analysis with R*. (Springer Cham, 2015).

33   Zhang, F. *et al.* OSCA: a tool for omic-data-based complex trait analysis. *Genome Biology* **20**, doi:10.1186/s13059-019-1718-z (2019).

34   Shen, X. *et al.* DNA methylome-wide association study of genetic risk for depression implicates antigen processing and immune responses. *Genome Medicine* **14**, doi:10.1186/s13073-022-01039-5 (2022).

35   Min, J. L. *et al.* Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation. *Nature Genetics* **53**, 1311-1321, doi:10.1038/s41588-021-00923-x (2021).

36   Hemani, G. *et al.* The MR-Base platform supports systematic causal inference across the human phenome. *eLife* **7**, doi:10.7554/elife.34408 (2018).

37   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. *Statistics in Medicine* **27**, 1133-1163, doi:10.1002/sim.3034 (2008).

38   Chu, A. Y. *et al.* Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat Commun* **8**, 1286, doi:10.1038/s41467-017-01297-7 (2017).

39   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, doi:10.1038/s41380-020-00968-0 (2021).

40   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, doi:10.1186/s13148-017-0376-9 (2017).

41   Shadrin, A. A. *et al.* Vertex-wise multivariate genome-wide association study identifies 780 unique genetic loci associated with cortical morphology. *Neuroimage* **244**, 118603, doi:10.1016/j.neuroimage.2021.118603 (2021).

42   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, doi:10.1038/s41588-019-0516-6 (2019).

43   Knijnenburg, T. A. *et al.* Genomic and molecular characterization of preterm birth. *Proc Natl Acad Sci U S A* **116**, 5819-5827, doi:10.1073/pnas.1716314116 (2019).

44   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, doi:10.1038/s41467-019-09671-3 (2019).

45   Spiers, H. *et al.* Methylomic trajectories across human fetal brain development. *Genome Res* **25**, 338-352, doi:10.1101/gr.180273.114 (2015).

46   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, doi:10.1038/nn.3786 (2014).

47   Lunnon, K. *et al.* Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat Neurosci* **17**, 1164-1170, doi:10.1038/nn.3782 (2014).

48   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, doi:10.1038/s41467-020-19791-w (2020).

49   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, doi:10.1371/journal.pmed.1003229 (2020).

50   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, doi:10.1038/nbt.2487 (2013).

51   Richardson, T. G. *et al.* Characterising metabolomic signatures of lipid-modifying therapies through drug target mendelian randomisation. *PLoS Biol* **20**, e3001547, doi:10.1371/journal.pbio.3001547 (2022).

52   Downes, D. J. *et al.* Identification of LZTFL1 as a candidate effector gene at a COVID-19 risk locus. *Nat Genet* **53**, 1606-1615, doi:10.1038/s41588-021-00955-3 (2021).

53   Guo, D. F. & Rahmouni, K. Molecular basis of the obesity associated with Bardet-Biedl syndrome. *Trends Endocrinol Metab* **22**, 286-293, doi:10.1016/j.tem.2011.02.009 (2011).

54   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, doi:10.1371/journal.pgen.1005890 (2016).

55   Lopez, M. Hypothalamic Leptin Resistance: From BBB to BBSome. *PLoS Genet* **12**, e1005980, doi:10.1371/journal.pgen.1005980 (2016).

56   Koskeridis, F. *et al.* Pleiotropic genetic architecture and novel loci for C-reactive protein levels. *Nat Commun* **13**, 6939, doi:10.1038/s41467-022-34688-6 (2022).

57   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, doi:10.1097/PSY.0000000000000072 (2014).

58   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, doi:10.1038/s41380-019-0501-6 (2020).

59   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, doi:10.1016/j.bbi.2015.06.001 (2015).

60   Wray, N. R. *et al.* Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature Genetics* **50**, 668-681, doi:10.1038/s41588-018-0090-3 (2018).

61   Joyce, P. I. *et al.* Deficiency of the zinc finger protein ZFP106 causes motor and sensory neurodegeneration. *Hum Mol Genet* **25**, 291-307, doi:10.1093/hmg/ddv471 (2016).

62   Arif, E. *et al.* The motor protein Myo1c regulates transforming growth factor-β–signaling and fibrosis in podocytes. *Kidney International* **96**, 139-158, doi:10.1016/j.kint.2019.02.014 (2019).

63   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, doi:10.3389/fimmu.2022.1027154 (2022).

64   Lam, M. *et al.* Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nature Genetics* **51**, 1670-1678, doi:10.1038/s41588-019-0512-x (2019).

65   Nievergelt, C. M. *et al.* International meta-analysis of PTSD genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nature Communications* **10**, doi:10.1038/s41467-019-12576-w (2019).

66   Braun, P. R. *et al.* Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Translational Psychiatry* **9**, doi:10.1038/s41398-019-0376-y (2019).