

ARTICLE OPEN



Methylome-wide association study of antidepressant use in Generation Scotland and the Netherlands Twin Register implicates the innate immune system

Miruna C. Barbu¹✉, Floris Huider², Archie Campbell³, Carmen Amador⁴, Mark J. Adams¹, Mary-Ellen Lynall⁵, David M. Howard^{1,6}, Rosie M. Walker³, Stewart W. Morris³, Jenny Van Dongen², David J. Porteous³, Kathryn L. Evans³, Edward Bullmore⁵, Gonneke Willemsen², Dorret I. Boomsma², Heather C. Whalley¹ and Andrew M. McIntosh¹

© The Author(s) 2021

Antidepressants are an effective treatment for major depressive disorder (MDD), although individual response is unpredictable and highly variable. Whilst the mode of action of antidepressants is incompletely understood, many medications are associated with changes in DNA methylation in genes that are plausibly linked to their mechanisms. Studies of DNA methylation may therefore reveal the biological processes underpinning the efficacy and side effects of antidepressants. We performed a methylome-wide association study (MWAS) of self-reported antidepressant use accounting for lifestyle factors and MDD in Generation Scotland (GS: SFHS, $N = 6428$, EPIC array) and the Netherlands Twin Register (NTR, $N = 2449$, 450 K array) and ran a meta-analysis of antidepressant use across these two cohorts. We found ten CpG sites significantly associated with self-reported antidepressant use in GS:SFHS, with the top CpG located within a gene previously associated with mental health disorders, *ATP6V1B2* ($\beta = -0.055$, $p_{\text{corrected}} = 0.005$). Other top loci were annotated to genes including *CASP10*, *TMBIM1*, *MAPKAPK3*, and *HEBP2*, which have previously been implicated in the innate immune response. Next, using penalised regression, we trained a methylation-based score of self-reported antidepressant use in a subset of 3799 GS:SFHS individuals that predicted antidepressant use in a second subset of GS:SFHS ($N = 3360$, $\beta = 0.377$, $p = 3.12 \times 10^{-11}$, $R^2 = 2.12\%$). In an MWAS analysis of prescribed selective serotonin reuptake inhibitors, we showed convergent findings with those based on self-report. In NTR, we did not find any CpGs significantly associated with antidepressant use. The meta-analysis identified the two CpGs of the ten above that were common to the two arrays used as being significantly associated with antidepressant use, although the effect was in the opposite direction for one of them. Antidepressants were associated with epigenetic alterations in loci previously associated with mental health disorders and the innate immune system. These changes predicted self-reported antidepressant use in a subset of GS:SFHS and identified processes that may be relevant to our mechanistic understanding of clinically relevant antidepressant drug actions and side effects.

Molecular Psychiatry; <https://doi.org/10.1038/s41380-021-01412-7>

INTRODUCTION

Major Depressive Disorder (MDD) is a leading cause of disability worldwide [1] and is caused by a combination of environmental and genetic factors [2]. MDD has a number of effective treatments, with antidepressant drugs amongst the most commonly prescribed evidence-based therapies worldwide [3, 4]. Response to antidepressants is nevertheless unpredictable and highly variable, with only approximately 50% of individuals achieving remission after completing two treatments [5]. The unmet needs of many individuals with MDD indicate the urgency of understanding the mechanisms of effective antidepressant action [6] so that their efficacy can be improved and their delivery targeted to those most likely to benefit.

The mode of action of antidepressants was originally assumed to be through the inhibition of monoamine reuptake. However, since synaptic monoamine concentration poorly mirrors the trajectory of symptomatic improvement, their mechanism of action remains uncertain [7]. Studies of antidepressants in both animal and human studies have implicated a number of possible modes of action. These include evidence that antidepressants lead to changes in DNA methylation (DNAm) and gene expression in a number of potentially relevant pathways [8]. Francois et al. [9] demonstrated that antidepressants reversed stress-induced changes in DNAm and expression of *5-HT1A* in an animal model, while Zimmermann et al. [10] found that antidepressants affect activity of DNA methyltransferase 1 (*DNMT1*). These studies

¹Division of Psychiatry, The University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK. ²Faculty of Behavioural and Movement Sciences, Biological Psychology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. ³Centre for Genomic and Experimental Medicine, The Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK. ⁴MRC Human Genetics Unit, The Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, UK. ⁵Department of Psychiatry, University of Cambridge, Cambridge, UK. ⁶Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK. ✉email: mbarbu@ed.ac.uk

Table 1. Demographic characteristics for GS individuals with self-reported antidepressant use and SSRI prescribing data in 12 months prior to blood draw date included in MWAS, including lifestyle variables and MDD.

Demographic characteristics	Self-reported AD use sample		SSRI sample from NHS Scotland records	
	AD use (<i>N</i> = 740)	No AD use (<i>N</i> = 5688)	SSRI use in 12-month interval (<i>N</i> = 401)	No SSRI use (<i>N</i> = 6705)
Age				
Mean (SD), range	51.33 (11.07), 18–87	49.5 (13.73), 18–87	49.16 (12.35), 18–85	50.21 (13.65), 18–94
Sex (%)				
Female	564 (76%)	3336 (59%)	364 (75%)	3631 (54%)
Male	176 (24%)	2352 (41%)	123 (25%)	3074 (46%)
Wave (%)				
1	480 (65%)	2918 (51%)	260 (65%)	3389 (51%)
2	260 (35%)	2770 (49%)	141 (35%)	3316 (49%)
BMI				
Mean (SD), range	28.17 (5.58), 16.11–51.29	26.60 (4.95), 14.78–67.62	28.37 (5.97), 17.50–51.29	26.68 (4.96), 15.93–67.62
Alcohol units				
Mean (SD), range	8.99 (11.75), 0–105	10.56 (11.36), 0–146	9.46 (10.59), 0–72	10.91 (12.24), 0–326
Smoking status (%)				
Current smoker	192 (26%)	861 (15%)	104 (26%)	1006 (15%)
Former smokers (quit < 1 year ago)	15 (2%)	146 (2%)	2 (1%)	154 (2%)
Former smokers (quit > 1 year ago)	229 (31%)	1634 (29%)	116 (30%)	1907 (29%)
Never smoked tobacco	304 (41%)	3047 (54%)	179 (43%)	3638 (54%)
Pack years				
Mean (SD), range	11.03 (16.03), 0–116	7.01 (13.51), 0–133	9.59 (14.58), 0–72.85	7.27 (14.07), 0–133
MDD status (%)				
Cases	420 (57%)	740 (13%)	201 (50%)	706 (11%)
Controls	320 (43%)	4948 (87%)	200 (50%)	5999 (89%)

AD antidepressant.

demonstrate that antidepressants may lead to epigenetic changes that reveal their biologically relevant mechanisms of action.

Epigenetic processes are associated with alterations in DNA activity without alterations to the underlying genome sequence [11]. DNAm is perhaps the most commonly investigated epigenetic change, owing to the availability of reliable high-throughput array technologies that can identify changes in DNAm at over 800 K locations throughout the genome. Methylome-wide association studies (MWAS) have begun to identify a number of cytosine-phosphate-guanine (CpG) sites that are associated with MDD [12] and relevant environmental risk factors [13]. DNAm changes may therefore capture an environmental archive of exogenous factors, including drug treatments, relevant to the onset and maintenance of MDD.

Recent studies have shown the potential of a DNAm-based risk score to predict MDD and its associated lifestyle and environmental factors, including antidepressant use [14, 15]. We previously [15] trained a DNAm predictor of MDD in 3047 individuals using a penalised regression model and showed that it was significantly associated with self-reported antidepressant use in an independent cohort. This suggests that CpG sites conferring risk to MDD may also be linked to antidepressant use. A methylation-based predictor of self-reported antidepressant use may therefore be able to predict antidepressant use and its effects in other samples and may signpost the development of clinical biomarkers quantifying drug action.

Here, we sought to identify the DNAm changes associated with self-reported antidepressant use in 6428 individuals

($N_{\text{antidepressant use}} = 740$) from the Generation Scotland study (GS:SFHS) using the Illumina Infinium MethylationEPIC array [16], capturing DNAm at approximately 850 K sites and in 2449 individuals ($N_{\text{antidepressant use}} = 74$) from the Netherlands Twin Register (NTR) using the Infinium HumanMethylation450 Bead-Chip Kit [17], capturing DNAm at 450 K sites. We then conducted a meta-analysis using the two cohorts. Further, by splitting GS:SFHS into two DNAm datasets where the number of participants differed according to data availability, we were able to train a DNAm-based methylation score (MS) of self-reported antidepressant use in one dataset ($N = 3799$; $N_{\text{antidepressant use}} = 585$) and test its ability to predict self-reported antidepressant use in a second ($N = 3360$; $N_{\text{antidepressant use}} = 317$). We also addressed confounding by indication and by smoking in planned sensitivity analyses [18]. Using linked National Health Service (NHS) Scotland medication prescribing data, we were able to identify whether the MWAS findings from self-report were also found when using contemporaneous electronic healthcare information collected around the time of the blood draw.

METHODS

Study populations

Generation Scotland—the Scottish Family Health Study (GS:SFHS). GS:SFHS is a family-based population cohort designed to investigate the genetic and environmental causes of common diseases and well-being in approximately 24,000 participants aged 18–98 years in Scotland. Baseline data was collected between 2006 and 2011 [19, 20] and contains detailed

information on a broad range of variables. DNA is also available from blood samples taken on more than 20,000 consenting participants.

GS:SFHS received ethical approval from NHS Tayside Research Ethics Committee (REC reference number 05/S1401/89) and has Research Tissue Bank Status (reference: 20/ES/0021). Written informed consent was obtained from all participants.

Netherlands Twin Register (NTR). NTR is a population-based cohort of over 200,000 participants, consisting of twins, their parents, spouses, and siblings. Between 2004 and 2010, biosamples were collected in the NTR-Biobank project. Full details have been reported previously [21, 22]. DNA was collected from peripheral whole blood.

Informed consent was obtained from all participants. The study was approved by the Central Ethics Committee on Research Involving Human Subjects of the VU University Medical Centre, Amsterdam, an Institutional Review Board certified by the U.S. Office of Human Research Protections (IRB number IRB00002991 under Federal-wide Assurance- FWA00017598; IRB/institute codes, NTR 03-180).

Phenotypes

GS:SFHS

Self-reported antidepressant use: For self-reported antidepressant use, a text-based questionnaire was sent to participants 1–2 weeks before attending an in-person appointment where DNAm was collected. The questionnaire recorded medication use through a “yes/no” checkbox with the following accompanying question: “Are you regularly taking any of the following medications?” where available options included “Antidepressants”.

Demographic data on individuals with self-reported antidepressant use is presented in Table 1. Participants with no self-reported antidepressant use were defined as those individuals who answered “No” to the “Antidepressants” sub-section of the questionnaire. There were 6428 individuals with available self-reported data who also had DNAm, lifestyle, and MDD status data available.

Selective serotonin reuptake inhibitors (SSRI) prescription and data linkage: Here, we selected antidepressants from BNF paragraph code 4.3.3 containing the following SSRIs: Citalopram, Escitalopram, Fluoxetine, Fluvoxamine Maleate, Paroxetine, and Sertraline [23] for individuals with medication data in GS:SFHS (see Supplementary Materials for more information). To investigate DNAm signatures of SSRI use, we restricted analyses to those individuals with SSRI dispensing records in the 12 months prior to blood draw date ($N = 401$). Individuals with DNAm data that did not have any SSRI dispensed prescription records were marked as controls ($N = 5688$).

We used dates of dispensing, not prescription, when restricting antidepressant use based on time interval. Demographic data on individuals included here are presented in Table 1. Further details with regards to SSRI prescriptions are presented in Supplementary Table 1.

Lifestyle factors and MDD status: Body mass index (BMI) was computed using height (cm) and weight (kg) as measured by clinical staff at baseline recruitment. Participants reported the number of units of alcohol consumed during the past week and their smoking status (never, former, current); pack years was used to measure heaviness of smoking in current smokers [24].

MDD status was assessed using the Structured Clinical Interview of the Diagnostic and Statistical Manual, version IV (SCID) [25]. Participants with no MDD were defined as those individuals who did not fulfil criteria for a current or previous MDD diagnosis following the SCID interview [26].

Further details about lifestyle factors, other diagnoses, and medication intake are presented in the Supplementary Materials and Supplementary Tables 2 and 3. Participants who suffered from other psychiatric disorders (e.g., bipolar disorder, psychosis) were excluded from analyses ($N = 11$).

NTR. Demographic data on individuals with antidepressant use is presented in Table 2.

Self-reported antidepressant use: At blood draw, participants were asked about all current medication use and for all medicines the dose, brand and chemical names were recorded directly from the medication packaging. Following the Anatomical Therapeutic Chemical (ATC) classification system, subclasses of N06A were considered antidepressants. Antidepressant medication use was coded as 1 (current user)/0 (non-user).

Table 2. Demographic characteristics for NTR individuals with self-reported antidepressant use data within a week of blood draw included in MWAS, including lifestyle variables and MDD.

Demographic characteristics	Self-reported AD use sample	
	AD use ($N = 74$)	No AD use ($N = 2375$)
Age		
Mean (SD), range	37.74 (12.20), 18–72	36.76 (3.97), 18–80
Sex (%)		
Female	52 (70%)	1,618 (68%)
Male	22 (30%)	757 (32%)
BMI		
Mean (SD), range	25.06 (4.25), 18.30–34.60	24.24 (3.97), 14.50–50.70
Smoking status (%)		
Current smokers	13 (17%)	434 (18%)
Former smokers	19 (26%)	506 (21%)
Never smoked	42 (57%)	1435 (61%)
tobacco		
Pack years		
Mean (SD), range	3.56 (5.83), 0–25	3.95 (8.69), 0–105
MDD status (%)		
Cases	10 (14%)	395 (17%)
Controls	64 (86%)	1980 (83%)

AD antidepressant.

Lifestyle factors and MDD status: Body mass index (kg/m^2) was computed based on weight and height obtained at the time of blood sampling. Current and past smoking behaviour was dummy-coded as current/former/never smokers and pack years were recorded as the number of cigarettes smoked per day/20 $\times$ number of years smoked.

Lifetime MDD was derived from the Composite International Diagnostic Interview (CIDI) [27], the Lifetime Depression Assessment Self-report (LIDAS) [28], and DSM-oriented Adult Self Report scales (ASR) of the Achenbach System of Empirically Based Assessment (ASEBA) [29] (Supplementary Materials). Data from the CIDI, LIDAS and ASR scales were combined to form one numeric variable coded as 1 (lifetime MDD case)/0 (lifetime MDD control). When data from multiple sources were available for an individual, the information from CIDI was favoured over that of LIDAS and ASR, and data from LIDAS over that of ASR.

DNA methylation

GS:SFHS. Genome-wide DNAm data profiled from whole blood samples was available for 9873 individuals in GS:SFHS using the Illumina HumanMethylationEPIC BeadChip [16]. DNAm data for individuals was initially released in two waves (wave 1 $_N = 5101$; wave 2 $_N = 4450$). Quality control (QC) and normalisation were conducted using R packages ShinyMethyl [30] and watermelon [31]. Details of the protocol are described in Supplementary Materials.

NTR. DNAm in the NTR-Biobank study [21, 32, 33] was assessed with the Infinium HumanMethylation450 BeadChip Kit ([17], Illumina, San Diego, CA, USA) by the Human Genotyping facility (HugeF) of ErasmusMC, the Netherlands (<http://www.glimdna.org/>) as part of the Biobank-based Integrative Omics Study (BIOS) consortium [34]. DNAm measurements have been described previously [33, 34] (see Supplementary Materials for more information).

Statistical methods

Methylome-wide association

GS:SFHS: We used linear regression models run in the “limma” package [35] in R to analyse the association of each CpG site, included as an

outcome variable, with self-reported antidepressant use included as the predictor variable. The R code for the current analyses is available in the Supplementary Materials. The following covariates were included: age, sex, wave (indicating different DNAm data pre-processing waves), as well as BMI, alcohol units, smoking status, and pack years to observe whether the inclusion of lifestyle factors attenuates the effect of self-reported antidepressant use. MDD status was also included to observe whether associations between CpGs and self-reported antidepressant use were influenced by the presence of MDD and potentially confounded by the indication for prescription, as some individuals not classed as having MDD have indicated antidepressant use ($N = 320$). There were 674,246 CpGs available after QC and a Bonferroni correction ($0.05/674,246$) was used to define methylome-wide significance ($p \leq 7.42 \times 10^{-8}$).

Using the same model as above, we further carried out sensitivity analyses in [1] non-smokers ($N = 3351$, $N_{\text{antidepressant use}} = 304$) [2], individuals with ($N = 5368$, antidepressant users = 420) and without ($N = 5268$, antidepressant users = 320) MDD, and [3] individuals with SSRI dispense dates within 12 months prior to blood draw date ($N = 7106$, $N_{\text{antidepressant use}} = 401$) (see Supplementary Materials for more information).

We compared findings in the main antidepressant use MWAS with summary statistics from a MDD MWAS in the same sample ($N = 8844$, MDD cases = 1461), by looking at effect sizes for significant CpGs in both analyses (Supplementary Figs. 1 and 2).

NTR: The association between antidepressant medication and DNAm was tested under a generalised estimation equation (GEE) model with DNAm M-values as outcome and self-reported antidepressant use as a predictor. The model included the following covariates: MDD status, sex, age at blood sampling, percentages of monocytes, eosinophils, and neutrophils, HM450k array row, 96-wells bisulfite sample plate (dummy coding), BMI, smoking, and smoking pack years. MWAS analyses were performed with the R package 'gee'. The following settings were used: Gaussian link function (for continuous data), 100 iterations, and the 'exchangeable' option to account for the correlation structure within families and within persons.

Meta-analysis: There were $N = 6428$ and $N = 2449$ participants in GS and NTR, respectively, and $N = 318,856$ overlapping CpGs between the two cohorts. Meta-analysis of these two datasets was performed in METAL [36] using p value based analysis ($N = 8877$). The meta-analysis was based on $N = 318,856$ total overlapping CpGs across GS:SFHS and NTR and a Bonferroni correction ($0.05/318,856$) was used to define methylome-wide significance ($p \leq 1.57 \times 10^{-7}$).

DNAm score analysis: To create a training and testing dataset, we separated individuals in GS:SFHS by DNAm pre-processing wave ($N_{\text{wave 1}} = 3799$; $N_{\text{wave 2}} = 3360$). This number differed from the final number included in the MWAS ($N = 6428$) due to data availability for covariates included in different analyses. R package "biglasso" was used to train DNAm predictors (Supplementary Materials).

As in the MWAS, we carried out sensitivity analyses by training a score in wave 1 in [1] non-smokers ($N = 1952$, $N_{\text{antidepressant use}} = 226$) and [2] individuals with no MDD diagnosis ($N = 2791$, antidepressant users = 195) (see Supplementary Materials).

There were 76, 20, and 35 CpG sites identified for the full (MS), smoker-excluded (MS-ns), and MDD-excluded (MS-control) datasets described above, respectively. The list of CpG sites and their corresponding weights are presented in Supplementary Tables 4, 5, and 6. MS, MS-ns, and MS-control were created for each individual in wave 2 ($N = 3360$; antidepressant users = 317) by taking the sum of the product of the identified DNAm residualised M-values and supplied model coefficient values. Regression models were then run to identify associations between the scores and self-reported antidepressant use, lifestyle factors and MDD.

Pathway and regulatory element overlap analysis: We used the Infinium MethylationEPIC BeadChip database to annotate significant CpG sites in GS:SFHS to genes [16]. The database provides information with regards to genes, chromosome location, start and end sites, and other features.

To assess pathway enrichment for differentially methylated CpG sites while correcting for biases in the representation of genes on the Infinium BeadChip, we used missMethyl [37], accessed via methylGSA [38]. Gene

Ontology (GO) terms were accessed using the msigdb package [39]. Pathways included in the analysis were all GO Biological Process pathways of size 20–500 genes inclusive. CpG sites included in the analysis were those significant at a threshold of $p < 1 \leq 10^{-5}$ ($N = 144$), as used in previous studies [40].

To assess overlap of differentially methylated CpG sites with the 15 chromatin states across 127 tissues from the Consolidated Roadmap Epigenomics dataset, we used eFORGE 2.0 [41], accessed via the web tool at <https://eforge.altiusinstitute.org> and using default settings.

Identification of differentially methylated regions (DMR): We used the dmrff function implemented in the dmrff R package [42] to identify differentially methylated regions in GS:SFHS, defined as regions containing 2–30 CpG sites separated by ≤ 500 bp. DMR analysis uses CpGs with unadjusted MWAS p values ≤ 0.05 and methylation changes in a consistent direction, and DMRs with a Bonferroni-corrected p value ≤ 0.05 are considered statistically significant.

RESULTS

Demographic characteristics

Individuals with self-reported antidepressant use who had incomplete lifestyle (BMI, alcohol consumption, and smoking) and disorder (MDD status) data were excluded. There were 6428 individuals in the final MWAS of self-reported antidepressant use in GS:SFHS ($N_{\text{antidepressant use}} = 740$) and 2449 ($N_{\text{antidepressant use}} = 74$) in NTR. Descriptive and demographic characteristics for these individuals, as well as individuals with prescribing data, are presented in Table 1 for GS:SFHS and Table 2 for NTR.

Methylome-wide association

GS:SFHS

Self-reported antidepressant use: MWAS identified ten CpG sites that were associated with self-reported antidepressant use ($p \leq 7.42 \times 10^{-8}$). Nine CpG sites were significantly hypomethylated in antidepressant users, while 1 site (cg26277237) was hypermethylated. Information about each CpG site is shown in Table 3 and depicted as a Manhattan plot in Fig. 1A.

The EWAS catalogue (<http://www.ewascatalog.org/>) includes CpG sites represented on the Illumina 450 K methylation array that have shown associations with traits at $p \leq 1 \times 10^{-4}$ in genome-wide analyses. The catalogue was used to cross-reference all CpG sites associated with self-reported antidepressant use with the wider literature. cg05603985 was previously found to be associated with smoking ($p = 1.8 \times 10^{-43}$; [18]) and alcohol consumption ($p = 7.9 \times 10^{-13}$) in African ancestry and $p = 1.9 \times 10^{-8}$ in European ancestry [43]; cg27589594 was found to be associated with gestational age ($p = 3.5 \times 10^{-16}$; [44]). Eight of the CpG sites were represented only on the Infinium MethylationEPIC BeadChip (containing approximately 850 K CpG sites) and were therefore not included in the catalogue. Searches conducted on other databases, including EWASdb [45] indicate that these 8 CpG sites have not been previously associated with any trait.

The Pearson's correlation between the MWAS effect sizes from the full sample MWAS and MWAS in non-smokers across all CpGs was $r = 0.655$ (95% CI = 0.636–0.674). When restricting CpGs to those that reached a p -value $\leq 5 \times 10^{-5}$, the correlation increased further to $r = 0.877$ (95% CI = 0.869–0.884), and Bonferroni-corrected CpGs ($N = 10$) in the full sample and the same CpGs in the non-smoker sample had a high correlation ($r = 0.980$, 95% CI = 0.979–0.981). Supplementary Table 7 indicates the effect size and nominal p -value for all CpGs found to be associated with self-reported antidepressant use in the full sample and in the non-smoker sample.

The Pearson's correlation between effect sizes obtained from the MWAS of antidepressant use in individuals with MDD and MWAS of antidepressant use in those without MDD across all CpGs was $r = 0.094$ (95% CI = 0.069–0.118). Restricting CpGs to a significance cut-off of $p \leq 5 \times 10^{-5}$ increased the correlation in

Table 3. CpGs significantly associated with self-reported antidepressant use in GS ($N = 6428$; antidepressant use = 740; N CpGs = 10) along with gene annotations, chromosome, standardised effect size, nominal and Bonferroni-corrected p values.

CpG site	Gene	Chrom	β	p value	P -corr	CpG site information	Gene information
cg05603985	SKI	1	-0.022	3.92×10^{-10}	0.0002	Smoking [63], alcohol consumption [64]	Platelet count [65], Red cell distribution width [65], body height [66], lung function [66], self-reported educational attainment [66]
cg05273171	PNKD; TMBIM1	2	-0.013	2.35×10^{-8}	0.017	-	PNKD: red cell distribution width [65], body height [66], monocyte, reticulocyte, leukocyte, granulocyte, neutrophil, and eosinophil count [67], waist-hip ratio [66] TMBIM1: red cell distribution width [65], body height [66], inflammatory bowel syndrome [68], Crohn's disease [69], cardiovascular disease [66] Related pathways: innate immune system
cg03864397	CASP10	2	-0.030	1.05×10^{-8}	0.007	-	Mean corpuscular haemoglobin [65], chronic lymphocytic leukaemia or rheumatoid arthritis [70]
cg05186879	MAPKAPK3; CISH	3	-0.023	3.3×10^{-9}	0.002	-	MAPKAPK3: self-reported educational attainment [71], BMI [72], schizophrenia [73] CISH: eosinophil count [66]
cg16315329	CISH; MAPKAPK3	3	-0.039	2.52×10^{-8}	0.018	-	MAPKAPK3: self-reported educational attainment [71], BMI [72], schizophrenia [73] CISH: eosinophil count [66]
cg25753411	HEBP2	6	-0.036	2.89×10^{-10}	0.0002	-	Related pathways: innate immune system
cg09511513	ATP6V1B2	8	-0.055	6.73×10^{-9}	0.005	-	Depression, mental or behavioural disorder [53]
cg26277237	KANK1	9	0.024	2.87×10^{-8}	0.020	-	Uterine fibroid [74], body height [66], systolic blood pressure [75], PHF-tau measurement [76]
cg20494891	MYO1E	15	-0.026	2.32×10^{-8}	0.017	-	Serum metabolite measurement [77], pulse pressure [75], coronary artery calcification [78], BMI [66], diffuse plaque measurement [76]
cg27589594	SLC5A10	17	-0.025	4.91×10^{-9}	0.004	Gestational age [44]	1.5 anhydroglucitol measurement [79], response to paliperidone (schizophrenia) [80]

Background information for each CpG and gene was extracted from EWAS (<http://www.ewascatalog.org>); association between traits and CpGs on Illumina 450 K array at $p \leq 1.0 \times 10^{-4}$; and GWAS (<https://www.ebi.ac.uk/gwas/>); associations between traits and SNPs at $p \leq 1.0 \times 10^{-5}$ catalogue databases. All associations included in the table from these two catalogues are genome-wide significant.

effect sizes to $r = 0.961$ (95% CI = 0.959–0.963). Supplementary Table 8 indicates the effect size and nominal p -value for all CpGs found to be associated with self-reported antidepressant use in the full sample and in the sensitivity analysis sample.

SSRI: MWAS conducted using individuals with dispensing data identified 12 CpG sites, all hypomethylated, associated with SSRI use (the most commonly prescribed treatment for depression) within 12 months prior to blood draw date, which are shown in Supplementary Table 9 and Supplementary Fig. 3.

Additionally, all ten CpG sites that were associated with self-reported antidepressant use ($p \leq 7.42 \times 10^{-8}$) were found to be nominally significant ($p < 0.05$) in the MWAS including SSRI dispensing records as a phenotype. The direction of effect was consistent for all CpG sites identified in the self-report MWAS: 9 were hypomethylated and 1 was hypermethylated, presented in Supplementary Fig. 3 and Supplementary Table 10 (β and p value for smallest and largest associations included cg09511513 ($\beta = -0.026$, $p = 0.028$) and cg26277237 ($\beta = 0.027$, $p = 3.09 \times 10^{-7}$)). The majority (10/12) of CpG sites identified in the MWAS of SSRI also were nominally significant in the MWAS where the self-report variable was input as the predictor (Supplementary Table 11).

NTR: There were no methylome-wide significant findings in NTR. The top ten CpG sites were annotated to genes implicated in psychiatric- and cognition-related phenotypes and in disorders of the immune system and are presented Supplementary Table 12. Figure 1B represents a Manhattan plot of NTR results.

Meta-analysis: Meta-analysis of antidepressant use MWAS across GS:SFHS and NTR revealed two Bonferroni-corrected CpG sites, which were on the 450 K array; both were identified in the GS:SFHS MWAS; cg05603985 ($\beta = -0.021$, $p = 3.01 \times 10^{-8}$, $q = 0.009$) had an opposite direction of effect across cohorts (negative in GS:SFHS and positive in NTR), and cg27589594 ($\beta = -0.024$, $p = 6.46 \times 10^{-8}$, $q = 0.021$) had the same direction of effect (negative) in both. Methylome-wide results are depicted in Fig. 1C.

There were $N = 318,856$ overlapping CpGs between GS:SFHS and NTR, and effect sizes were not strongly correlated (Pearson's correlation $r = 0.081$, 95% CI = 0.06–0.102). When restricting CpGs to those that reached a p value ≤ 0.05 across both MWAS ($N = 1714$), the correlation between effect sizes increased ($r = 0.32$, 95% CI = 0.301–0.339). Further information with regards to concordant signals across the two cohorts is in Supplementary Materials.

DNAm score analysis: LASSO regression selected 76, 20, and 35 CpGs to calculate MS, MS-ns, and MS-control for self-reported antidepressant use, respectively, in GS:SFHS wave 2.

Briefly, the MS was associated with self-reported antidepressant use in a model adjusted for age, sex, and 10 genetic PCs ($\beta = 0.377$, $p = 3.12 \times 10^{-11}$, $R^2 = 2.12\%$); After adjustment for BMI, alcohol units, smoking status, pack years, and MDD, this association remained significant, although the variance explained decreased ($\beta = 0.213$, $p = 0.0035$, $R^2 = 0.56\%$). The association between MS-ns and self-reported antidepressant use was in the same direction but became non-significant, both in the covariate-free model ($\beta = 0.106$, $p = 0.075$, $R^2 = 0.16\%$) and in the model with the above covariates included ($\beta = 0.04$, $p = 0.565$, $R^2 = 0.02\%$). Finally, MS-control was significantly associated with self-reported antidepressant use in the covariate-free model ($\beta = 0.254$, $p = 1.15 \times 10^{-5}$, $R^2 = 0.93\%$), but not when including covariates ($\beta = 0.064$, $p = 0.381$, $R^2 = 0.05\%$). Table 4 indicates associations with lifestyle traits and MDD for all scores.

Pathway and regulatory element overlap analysis: Following FDR adjustment for multiple comparisons, the only over-represented pathways were related to regulation of myeloid cell differentiation (GO Negative Regulation of Myeloid Cell Differentiation and

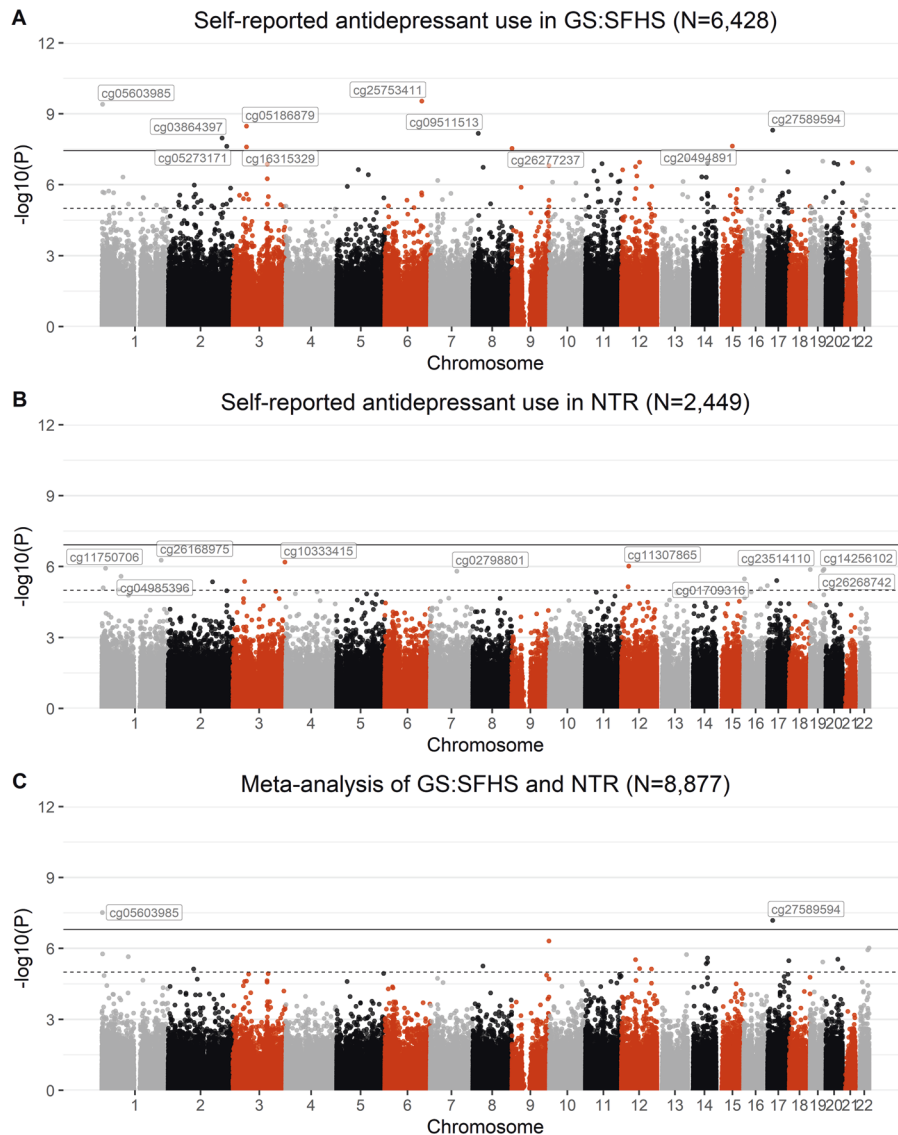


Fig. 1 Manhattan plots for GS:SFHS (A), NTR (B), and MWAS meta-analysis in GS:SFHS and NTR (C). Manhattan plots showing MWAS of self-reported antidepressant use in GS:SFHS (A), MWAS of self-reported antidepressant use in NTR (B), and meta-analysis of MWAS in GS:SFHS and NTR (C). The black line defines methylome-wide significance for each analysis (GS:SFHS: $p \leq 3.6 \times 10^{-8}$; NTR: $p \leq 1.22 \times 10^{-7}$; meta-analysis: $p \leq 1.57 \times 10^{-8}$) and the dotted line defines $p \leq 1 \times 10^{-5}$. Methylome-wide significant hits (GS:SFHS: 10; meta-analysis: 2) and the top 10 CpGs in NTR are labelled on the graph.

GO Negative Regulation of Myeloid Leucocyte Differentiation, $p_{\text{adjusted}} = 0.04$ for both). Enrichment of these pathways was driven by hypomethylation in the antidepressant use group in the body or 5'UTR of the following genes: *GAT2*, *HOXA7*, *INPP5D*, *MEIS1*, *RAR*, and *UBASH3B*.

Testing for overlap of differentially methylated CpG sites at $p < 1 \times 10^{-5}$ with chromatin states in 127 tissues from the Roadmap Epigenomics Consortium, the greatest enrichment was seen in blood cell tissues. Of the peripheral blood cell subsets tested, the cell type showing the greatest enrichment was monocytes (Fig. 2), with significant enrichment for monocyte enhancers ($q = 4 \times 10^{-11}$) and regions flanking active transcription start sites in monocytes ($q = 1 \times 10^{-9}$).

We repeated the tests above using the results from the MEWAS of antidepressant prescriptions (N CpGs at $p < 1 \times 10^{-5} = 484$). We did not find enrichment of any specific GO biological processes but replicated the finding from the self-reported antidepressant use MWAS that monocytes were the peripheral blood cell subset most enriched in the MWAS signal (enhancers in primary

monocytes from peripheral blood were enriched at $q = 4 \times 10^{-94}$) and also found strong enrichment for enhancers in peripheral blood neutrophils ($q = 4 \times 10^{-79}$).

DMR analysis: We identified 159 DMRs, ranging from 2–11 CpGs in length, to be associated with self-reported antidepressant use using the DMR adjusted p-values (Supplementary Excel File 1). The region with the largest significance ($p_{\text{adjust}} = 7.54 \times 10^{-12}$) spanned 4 CpGs annotated to *GATA2* on chromosome 3, a gene involved in stem cell maintenance and in which mutations are associated with wide-ranging immunological disorders [46, 47]. The largest DMR ($p_{\text{adjust}} = 2.04 \times 10^{-8}$) spanned 11 CpGs located on chromosome 2, where most were annotated to *OTX1*, a gene previously associated with risk-taking behaviour [48]. The gene may also play a role in the brain and development of sensory organs [49]. Only one DMR spanning 2 CpGs ($p_{\text{adjust}} = 2.55 \times 10^{-7}$) contained a CpG that was also associated with the trait at the level of individual probes (cg05186879, $p = 4.91 \times 10^{-9}$).

Table 4. Association between MS (methylation score), MS-ns (methylation score trained on non-smokers), and MS-control (methylation score trained on individuals with no MDD diagnosis), and self-reported antidepressant use, MDD, and 4 lifestyle factors (BMI, smoking status, pack years, alcohol units).

Outcome variable (N)	MS			MS-ns			MS-control		
	β	p value	R ²	β	p value	R ²	β	p value	R ²
AD use (3360)	0.377	3.12×10^{-11}	2.12%	0.106	0.075	0.16%	0.254	1.15×10^{-5}	0.93%
AD use* (3009)	0.213	0.0035	0.56%	0.04	0.565	0.02%	0.064	0.381	0.05%
MDD (3360)	0.28	5.78×10^{-7}	1.14%	0.158	0.007	0.35%	0.202	0.0004	0.57%
BMI (3336)	0.088	2.54×10^{-7}	0.75%	0.021	0.23	0.01%	0.075	1.33×10^{-5}	0.53%
Alcohol units (3099)	0.141	$<2 \times 10^{-16}$	1.93%	0.007	0.665	0%	0.127	5.08×10^{-14}	1.56%
Smoking status (3322)	0.632	$<2 \times 10^{-16}$	6.11%	0.077	0.028	0.11%	0.567	$<2 \times 10^{-16}$	5.08%
Pack years (3310)	0.305	$<2 \times 10^{-16}$	9.14%	0.061	0.0004	0.34%	0.287	$<2 \times 10^{-16}$	8.09%

All regression models include age, sex, and 10 genetic principal components as covariates, except for "AD use*", which also includes BMI, smoking status, pack years, alcohol units, and MDD as covariates. Effect sizes represent standardised betas. R² represents the variance explained in the outcome variables by each score. The number of individuals in each model varies based on different available data for each variable.

DISCUSSION

Self-reported antidepressant use is associated with differences in DNAm at ten CpG sites in GS:SFHS located in genes that have previously been associated with psychiatric disorders and the innate immune system. There were no methylome-wide findings in NTR. We found that all ten CpG sites in GS:SFHS were also nominally associated with antidepressant prescription data obtained from the electronic health record, indicating agreement between self-reported and record linkage data. Sensitivity analyses in GS:SFHS indicated that results were not significantly confounded by smoking. As shown in Supplementary Table 7, the effect for all ten associated CpGs are in the same direction across sensitivity analyses, and all *p* values in the sensitivity analyses are nominally significant. The effect sizes of the top ten CpGs were also highly correlated in the full sample and in the non-smoker sample ($r = 0.980$). A DNAm score trained on one sub-sample was associated with self-reported antidepressant use, four lifestyle factors, and MDD in a second unrelated sample from the same study. Pathway analysis revealed an enrichment of differentially methylated genes in myeloid cell differentiation. However, the effect size correlation between the 10 associated CpGs and the same CpGs in a MDD MWAS ($r = 0.674$) may indicate that DNAm signatures of antidepressants act as a proxy for MDD, although further analyses with much larger samples will be needed to disentangle the effects between antidepressant use and MDD on DNAm. Lastly, meta-analysis of MWAS in GS:SFHS and NTR identified two CpGs that were also found in GS:SFHS.

The CpG site with the largest effect size, both in GS:SFHS and the meta-analysis ($\beta = -0.055$) was cg09511513, located in *ATP6V1B2*. This CpG site has not been associated with any other traits previously, to the best of our knowledge. *ATP6V1B2* encodes a component of vacuolar ATPase, which is a multisubunit enzyme that mediates the acidification of eukaryotic intracellular organelles, including endosomes and lysosomes, and may be involved in neurotransmission [50]. Importantly, a single nucleotide polymorphism (SNP) within *ATP6V1B2*, rs1106634, has exhibited a suggestive *p*-value in a meta-analysis in schizophrenia and bipolar disorder ($p = 3.97 \times 10^{-6}$) and has been associated with lifetime risk of depression ($p < 0.001$) in a recent study [51–53].

In NTR, there were no methylome-wide significant findings. The ten most significant CpGs (Supplementary Table 12) were annotated to genes implicated in similar processes as those found in GS:SFHS, including the innate immune and psychiatric diseases. Although there were no significant findings, the pattern of results in NTR is similar to significant findings in GS:SFHS. The two CpGs identified in the meta-analysis that were present on both array types were also associated in GS:SFHS and were annotated to genes previously implicated in psychiatric disease and the innate immune system (*SKI*, *SLC5A10*) (see Table 3). These were not nominally significant in NTR (Supplementary Table 13). cg05603985 was hypomethylated in GS:SFHS and hypermethylated in NTR, and cg27589594 was hypomethylated in both cohorts. Although the signal from one CpG was the same across the two cohorts, further large-scale analyses on DNAm collected on the same array will be needed to uncover the role of antidepressants in relation to psychiatric and immune processes in DNAm.

We were able to show that a MS for self-reported antidepressant use was associated with self-reported antidepressant use, MDD status, and 4 lifestyle factors (BMI, alcohol units, smoking status, and pack years) in a subset of GS:SFHS. In addition, when excluding MDD-associated signals in a second DNAm score (MS-control), we found that the predictor was still associated with antidepressant use, although with reduced variance explained (MS $R^2 = 2.12\%$; MS-control $R^2 = 0.93\%$). This suggests that, although antidepressant use may in part be a marker of MDD effects on DNAm, antidepressant prescription associations may be partly independent from the condition for which they are prescribed. However, when including lifestyle factors in the association

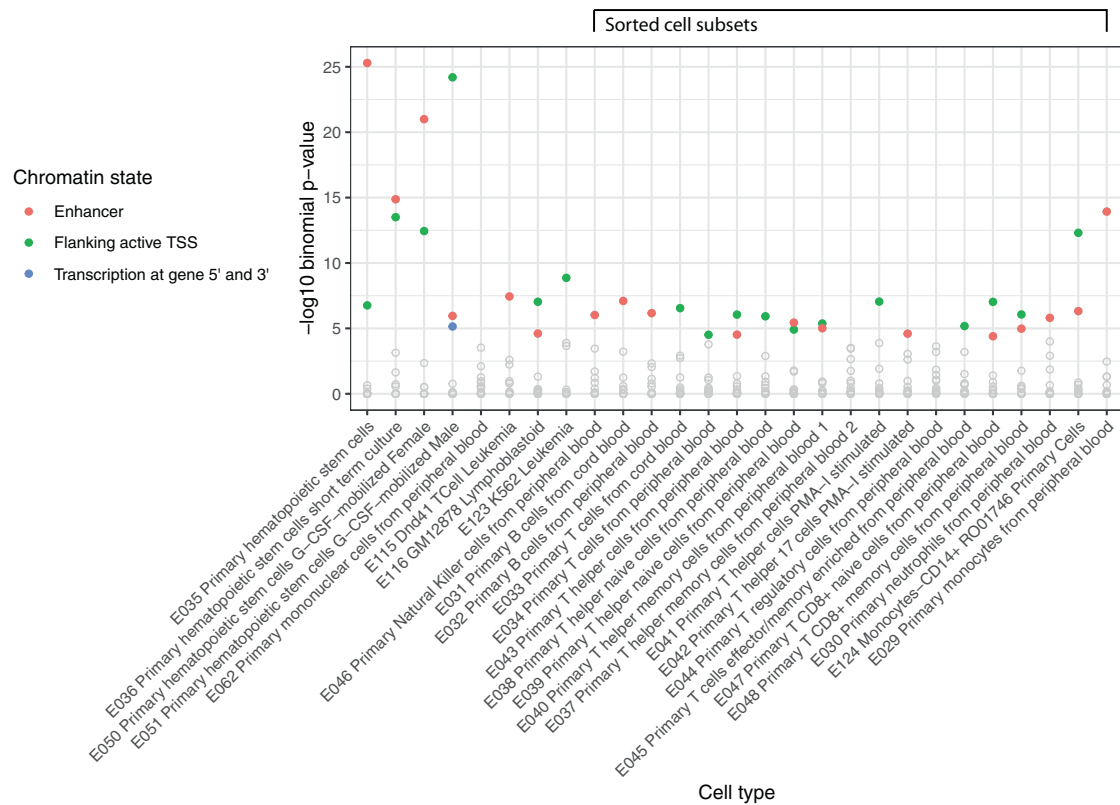


Fig. 2 Differentially methylated CpG sites at $p < 1 \times 10^{-5}$ in GS:SFHS enriched for peripheral blood cell subsets. The x-axis represents peripheral blood cell types, while the y-axis indicates corrected p value.

between MS-control and antidepressant use, the variance explained further decreased ($R^2 = 0.05\%$). This may indicate that lifestyle risk factors for MDD may play an important role in this association, even in the absence of an MDD diagnosis.

The decreased variance explained may be in part due to differences in the training sample sizes for the scores ($MS_N = 3799$, antidepressant users = 585; $MS\text{-}control_N = 2791$, antidepressant users = 195; $MS\text{-}ns_N = 1952$, antidepressant users = 226), although variation explained in lifestyle factors is comparable between the MS and MS-control (MS R^2 : BMI = 0.75%; alcohol units = 1.93%; smoking status = 6.11%; pack years = 9.14%; $MS\text{-}control$ R^2 : BMI = 0.53%; alcohol units = 1.56%; smoking status = 5.08%; pack years = 8.09%).

When smokers were excluded from the training sample, MS-ns associations with all variables remained in the same direction but became non-significant. This suggests that smoking may partially confound the associations between antidepressant use and DNAm. However, when excluding smokers in our MWAS analyses, correlations between the effect sizes in the full-sample and non-smoker MWAS indicated that results were not likely confounded by smoking ($r = 0.980$). Further, effects for all significant CpGs were in the same direction in the reduced sample as in the full sample, and were all nominally significant in the non-smoker sample (Supplementary Table 7). Future studies would benefit from training DNAm predictors of self-reported antidepressant use in larger samples of lifelong non-smokers. Only one CpG overlaps between the CpG sites in our MS and CpG sites comprising an MDD DNAm risk score (MS) in Barbu et al. [15]. This CpG (cg09935388) has been associated with smoking in previous studies [54, 55], and may therefore capture smoking-associated effects related to both MDD and antidepressant use. Further, in both studies, MDD MS and antidepressant use MS were both more strongly associated with lifestyle factors than with the trait of interest (MDD MS R^2 in Barbu et al.

[15]: MDD = 1.75%, BMI = 0.097%, alcohol units = 0.7%, smoking status = 3.2%, pack years = 6.5%; Antidepressant use MS R^2 : antidepressant use = 2.12%, BMI = 0.75%, alcohol units = 1.93%, smoking status = 6.11%, pack years = 9.14%). The stronger associations for both MS with lifestyle factors may indicate that the scores capture an archive of relevant current and past lifestyle factors that are also associated with depression and the prescription of antidepressants. Previous large-scale studies have found profound effects of each of the lifestyle factors on the methylome [56–58]. However, the MS was still associated with antidepressant use after accounting for the lifestyle factors mentioned above ($R^2 = 0.56\%$), indicating a possible independent effect of antidepressants. Due to the complex associations between DNAm, MDD, and lifestyle factors, future large-scale studies, including longitudinal studies, are warranted to investigate independent effects of each variable in relation to MDD.

To investigate potential relationships between DNAm links to antidepressant use and MDD, we conducted a MDD MWAS in GS ($N = 8844$, cases = 1641). While none of the top ten CpGs in the antidepressant use MWAS were significantly associated with MDD, the effect sizes showed a moderate correlation ($r = 0.674$; See Supplementary Figs. 2 and 3). This suggests that DNAm signatures of antidepressant use may partially reflect confounding by indication and much larger samples will be needed to disentangle this relationship.

Using missMethyl, we identified altered DNAm nearby several genes involved in myeloid cell differentiation. Myeloid cells play an important role in the innate immune response, and their activation and differentiation depend in part on epigenetic mechanisms [59]. Several of the genes annotated to CpGs identified at genome-wide significance here were associated with the innate immune response in previous GWAS; this may complement the finding that genes annotated to the 144 CpG

sites differentially methylated at $p < 1 \times 10^{-5}$ are enriched for myeloid cell differentiation.

Further, the differentially methylated regions were annotated to genes that play a role in cell maintenance, as well as brain and sensory organ development [46, 49]. Importantly, mutations in *GATA2* have previously been related to wide-ranging immunological disorders, providing further evidence linking antidepressant use and the immune system [47].

The role of the innate immune system has previously been investigated in relation to psychiatric disorders. In MDD, studies have found a number of pro-inflammatory cytokines, such as TNF- α and IL-6, to have higher concentrations in depressed individuals, as compared to healthy controls [60]. Importantly, antidepressants, particularly SSRIs, have been shown to exert effects on the immune system by causing a reduction in inflammatory markers. Generally, antidepressants have been found to reduce levels of pro-inflammatory cytokine, such as TNF- α , IL-6, and IL-1 β , although their effects are complex and incompletely understood [60]. The current study suggests novel links between DNAm, antidepressants, and processes involved in the innate immune system. Studies that investigate links between antidepressants and DNAm alterations at CpG sites that are known to be associated with the expression of innate immune system, myeloid cell differentiation, and MDD-specific genes may be of interest.

There are a number of potential limitations to the current study. First, although self-reported antidepressant use could not be restricted to specific dates prior to the blood draw date in GS:SFHS, the variable is considered contemporaneous to blood draw date as self-report questionnaires were sent 1–2 weeks before the clinic appointment. However, the questionnaire did not record types of antidepressants, which may explain the different top CpG sites based on the predictor input in MWAS. The different CpG sites identified in each MWAS may also be due in part to the proportion of overlapping individuals used in each analysis ($N_{\text{overlap in antidepressant users}} = 289$; $N_{\text{overlap in controls}} = 4811$). Nonetheless, as indicated in Supplementary Table 11, most of the CpG sites significantly associated with SSRI use within 12 months prior to blood draw date were nominally significant in the MWAS where self-reported antidepressant use was fit as the predictor. This pattern was also observed when investigating the top 10 CpG sites associated with the self-reported measure in the SSRI use MWAS (Supplementary Table 10). The findings above indicate there is agreement between different forms of data collection, specifically self-reported medication use and record linkage to PIS. The variables have previously been investigated in a study by Hafferty et al. [61], who showed that self-reported antidepressant use in GS:SFHS showed very good agreement with record linkage data at 3- and 6-month fixed time windows [61].

Secondly, since we chose to restrict PIS data linkage to 12 months before the blood draw date, it seems likely that the DNAm changes reported may occur in response to antidepressant treatment, although this requires further replication. Nevertheless, the DNAm data provided here is cross-sectional, and it is not possible to show how the trajectory of DNAm at each CpG site evolved over time. Future studies should therefore collect DNAm at multiple time points and further address the possibility of confounding by antidepressant indication.

It should also be noted that SSRI prescribing data is unable to confirm whether individuals took the medication as prescribed [62]. In addition, the current study is based exclusively on European ancestry cohorts, and the generalisability to diverse ancestries is unknown.

Finally, correlations of effect sizes between the two cohorts are of low magnitude, confidence intervals are large, and only one of the two CpGs uncovered in the meta-analysis was consistent in terms of direction of effect. However, the top 10 CpGs identified in NTR, although not significant, were implicated in similar processes

as those in GS:SFHS, and the meta-analysis uncovered two CpGs that were also uncovered in GS:SFHS. These novel findings should serve as a stimulus for larger, international collaborations and meta-analyses in future.

In conclusion, we conducted an MWAS of self-reported antidepressant use across two large, population-based cohorts and identified 10 novel CpG sites in GS:SFHS located in genes previously associated with mental health disorders, implicating the innate immune system. Further, MS predictors were associated with self-reported antidepressant use, MDD, and a number of lifestyle factors in a second sample from the same study. SSRI use derived from linkage health records showed convergent effects indicating agreement with the findings derived using only self-report. Finally, a meta-analysis revealed similar findings to the ones in GS:SFHS and indicated that further analyses should be conducted to confirm findings here. Our findings highlight biological processes that may be relevant to furthering our understanding of antidepressant actions and their side effects.

REFERENCES

- World Health Organization. Depression and other common mental disorders. Geneva: World Health Organization; 2017; pp. 1–24.
- Wray NR, Gottesman II. Using summary data from the Danish National Registers to estimate heritabilities for schizophrenia, bipolar disorder, and major depressive disorder. *Front Genet.* 2012;3:118.
- Kirsch I. Antidepressants and the placebo effect. *Z Psychol.* 2014;222:128–134.
- Cipriani A, Furukawa TA, Salanti G, Chaimani A, Atkinson LZ, Ogawa Y, et al. Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *Lancet.* 2018;391:1357–66. [https://doi.org/10.1016/S0140-6736\(17\)32802-7](https://doi.org/10.1016/S0140-6736(17)32802-7).
- Rush AJ, Trivedi MH, Wisniewski SR, Nierenberg AA, Stewart JW, Warden D, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: A STAR*D report. *Am J Psychiatry.* 2006;163:1905–17. <https://pubmed.ncbi.nlm.nih.gov/17074942/>.
- Keers R, Uher R. Gene-environment interaction in major depression and antidepressant treatment response. *Curr Psychiatry Rep.* 2012;14:129–37.
- Berton O, Nestler EJ. New approaches to antidepressant drug discovery: beyond monoamines. *Nat Rev Neuroscience.* 2006;7:137–51.
- Menke A, Binder EB. Epigenetic alterations in depression and antidepressant treatment. *Dialogues Clin Neurosci.* 2014;16:395–404.
- Le François B, Soo J, Millar AM, Daigle M, Le Guisquet AM, Leman S, et al. Chronic mild stress and antidepressant treatment alter 5-HT1A receptor expression by modifying DNA methylation of a conserved Sp4 site. *Neurobiol Dis.* 2015;82:332–41.
- Zimmermann N, Zschocke J, Perisic T, Yu S, Holsboer F, Rein T. Antidepressants inhibit DNA methyltransferase 1 through reducing G9a levels. *Biochem J.* 2012;448:93–102.
- Bird A. Perceptions of epigenetics. *Nature.* 2007;447:396–8.
- Jovanova OS, Nedeljkovic I, Spieler D, Walker RM, Liu C, Luciano M, et al. DNA methylation signatures of depressive symptoms in middle-aged and elderly persons: meta-analysis of multiethnic epigenome-wide studies. *JAMA Psychiatry.* 2018;75:949–59.
- McCartney DL, Hillary RF, Stevenson AJ, Ritchie SJ, Walker RM, Zhang Q, et al. Epigenetic prediction of complex traits and death. *Genome Biol.* 2018;19:136.
- Clark SL, Hattab MW, Chan RF, Shabalin AA, Han LKM, Zhao M, et al. A methylation study of long-term depression risk. *Mol Psychiatry.* 2019;25:1334–43.
- Barbu MC, Shen X, Walker RM, Howard DM, Evans KL, Whalley HC, et al. Epigenetic prediction of major depressive disorder. *Mol Psychiatry.* 2020;1–12. <http://www.nature.com/articles/s41380-020-0808-3>.
- Hansen K. IlluminaHumanMethylationEPICanno.ilm10b2.hg19: annotation for illumina's EPIC methylation arrays. R Package version 060. 2016. https://bitbucket.com/kasperdanielhansen/Illumina_EPIC.
- Hansen KD. IlluminaHumanMethylation450kanno.ilmn12.hg19: Annotation for Illumina's 450k methylation arrays. 2016; R package version 0.6.0.
- Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet.* 2016;9:436–47.
- Smith BH, Campbell H, Blackwood D, Connell J, Connor M, Deary IJ, et al. Generation Scotland: the Scottish Family Health Study; a new resource for researching genes and heritability. *BMC Med Genomics.* 2006;9:1–9.

20. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort profile: generation Scotland: Scottish Family Health Study (GS: SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol*. 2013;42:689–700.
21. Willemsen G, de Geus EJ, Bartels M, van Beijsterveldt CE, Brooks AL, Estourgie-van Burk GF, et al. The Netherlands Twin Register Biobank: a resource for genetic epidemiological studies. *Twin Res Hum Genet*. 2010;13:231–45. <https://www.cambridge.org/core/journals/twin-research-and-human-genetics/article/netherlands-twin-register-biobank-a-resource-for-genetic-epidemiological-studies/94AA0448D70BFEA780DC0D01EB7DF05>.
22. Ligthart L, van Beijsterveldt CEM, Kevenaar ST, de Zeeuw E, van Bergen E, Bruins S, et al. The Netherlands Twin Register: longitudinal research based on twin and twin-family designs. *Twin Res Hum Genet*. 2019;22:623–36. <https://www.cambridge.org/core/journals/twin-research-and-human-genetics/article/netherlands-twin-register-longitudinal-research-based-on-twin-and-twinfamily-designs/C68488FDD92716CB9ED7C04089776E2F>.
23. British Medical Association, Royal Pharmaceutical Society of Great Britain. British national formulary. 58 edn. UK: BMJ Publishing Group. 2009.
24. Leffondré K, Abrahamowicz M, Siemiatycki J, Rachet B. Modeling smoking history: a comparison of different approaches. *Am J Epidemiol*. 2002;156:813–23.
25. First MB, Spitzer RL, Gibbon M, Williams JBW, Davies M, Borus J, et al. The structured clinical interview for DSM-III-R personality disorders (SCID-II). Part II: multi-site test-retest reliability study. *J Pers Disord*. 1995;9:92–104.
26. Fernandez-Pujals AM, Adams MJ, Thomson P, McKechanie AG, Blackwood DHR, Smith BH, et al. Epidemiology and heritability of major depressive disorder, stratified by age of onset, sex, and illness course in generation Scotland: Scottish family health study (GS: SFHS). *PLoS ONE*. 2015;10:e0142197.
27. Kessler RC, Wittchen H-U, Abelson JM, McGonagle K, Schwarz N, Kendler KS, et al. Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US national comorbidity survey (NCS). *Int J Methods Psychiatr Res*. 1998;7:33–55. <https://onlinelibrary.wiley.com/doi/full/10.1002/mpr.33>.
28. Bot M, Middeldorp CM, de Geus EJ, Lau HM, Sinke M, van Nieuwenhuizen B, et al. Validity of LIDAS (Lifetime Depression Assessment Self-report): a self-report online assessment of lifetime major depressive disorder. *Psychol Med*. 2017;47:279–89. <https://www.cambridge.org/core/journals/psychological-medicine/article/validity-of-lidas-lifetime-depression-assessment-selfreport-a-selfreport-online-assessment-of-lifetime-major-depressive-disorder/97602A258E2DABA0745F8039C0F20907>.
29. The Achenbach System of Empirically Based Assessment (ASEBA) for Ages 18 to 90 Years. - PyscNET. 2021. <https://pyscnet.apa.org/record/2004-14941-004>.
30. Fortin J-P, Fertig E, Hansen K. shinyMethyl: interactive quality control of Illumina 450K DNA methylation arrays in R. *F1000Research*. 2014;3:175 <https://doi.org/10.12688/f1000research.4680.1>.
31. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 2013;14:1–10. <https://bmcbgenomics.biomedcentral.com/articles/10.1186/1471-2164-14-293>.
32. Boomsma DI, Willemsen G, Sullivan PF, Heutink P, Meijer P, Sondervan D, et al. Genome-wide association of major depression: description of samples for the GAIN Major Depressive Disorder Study: NTR and NESDA biobank projects. *Eur J Hum Genet*. 2008;16:335–42. <https://www.nature.com/articles/5201979>.
33. van Dongen J, Nivard MG, Willemsen G, Hottenga J-J, Helmer Q, Dolan CV, et al. Genetic and environmental influences interact with age and sex in shaping the human methylome. *Nat Commun*. 2016;7:1–13. <https://www.nature.com/articles/ncomms11115>.
34. Bonder MJ, Luijk R, Zhernakova DV, Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. *Nat Genet*. 2017;49:131–8.
35. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43:47.
36. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genome-wide association scans. *Bioinformatics*. 2010;26:2190–1. <https://pubmed.ncbi.nlm.nih.gov/20616382/>.
37. Phipson B, Maksimovic J, Oshlack A. MissMethyl: An R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*. 2016;32:286–8.
38. Ren X, Kuan PF. methylGSA: a Bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. *Bioinformatics*. 2019;35:1958–9.
39. Dolgalev I. msigdb: MSigDB gene sets for multiple organisms in a tidy data format. R package version 7.1.1. 2020. <https://cran.r-project.org/web/packages/msigdb/index.html>.
40. Chu AY, Tin A, Schlosser P, Ko YA, Qiu C, Yao C, et al. Epigenome-wide association studies identify DNA methylation associated with kidney function. *Nat Commun*. 2017;8:1–12. www.nature.com/naturecommunications.
41. Breeze CE, Reynolds AP, Van Dongen J, Dunham I, Lazar J, Neph S, et al. EFORGE v2.0: updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics*. 2019;35:4767–9.
42. Suderman M, Staley J, French R, Arathimos R, Simpkin A, Tilling K. dmrff: identifying differentially methylated regions efficiently with power and control. *bioRxiv*. 2018. <https://doi.org/10.1101/508556>.
43. Liu C, Marioni RE, Hedman AK, Pfeiffer L, Tsai P, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. *Mol Psychiatry*. 2018;23:422–33.
44. Bohlin J, Häberg SE, Magnus P, Reese SE, Gjessing HK, Magnus MC, et al. Prediction of gestational age based on genome-wide differentially methylated regions. *Genome Biol*. 2016;17:1–9.
45. Liu D, Zhao L, Wang Z, Zhou X, Fan X, Li Y, et al. EWASdb: epigenome-wide association study database. *Nucleic Acids Res*. 2019;47:989–93.
46. Li Y, Qi X, Liu B, Huang H. The STAT5–GATA2 pathway is critical in basophil and mast cell differentiation and maintenance. *J. Immunol*. 2015;194:4328–38. <http://www.jimmunol.org/content/194/9/4328> <http://www.jimmunol.org/content/194/9/4328.full#ref-list-1>.
47. Collin M, Dickinson R, Bigley V. Haematopoietic and immune defects associated with GATA2 mutation. *Br J Haematol*. 2015;169:173–87. <https://onlinelibrary.wiley.com/doi/full/10.1111/bjh.13317>.
48. Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nat Genet*. 2019;51:245–57. <https://pmc/articles/PMC6713272/?report=abstract>.
49. Frantz GD, Weimann JM, Levin ME, McConnell SK. Otx1 and Otx2 define layers and regions in developing cerebral cortex and cerebellum. *J Neurosci*. 1994;14:5725–40. <https://www.jneurosci.org/content/14/10/5725>.
50. Petschner P, Gonda X, Baksa D, Eszlari N, Trivaks M, Juhasz G, et al. Genes linking mitochondrial function, cognitive impairment and depression are associated with endophenotypes serving precision medicine. *Neuroscience*. 2018;370:207–17.
51. Wang KS, Liu XF, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. *Schizophr Res*. 2010;124:192–9.
52. Shyn SI, Shi J, Kraft JB, Potash JB, Knowles JA, Weissman MM, et al. Novel loci for major depression identified by genome-wide association study of sequenced treatment alternatives to relieve depression and meta-analysis of three studies. *Mol Psychiatry*. 2011;16:202–15.
53. Gonda X, Eszlari N, Anderson IM, Deakin JFW, Bagdy G, Juhasz G. Association of ATP6V1B2 rs1106634 with lifetime risk of depression and hippocampal neuro-cognitive deficits: possible novel mechanisms in the etiopathology of depression. *Transl Psychiatry*. 2016;6:e945.
54. Joubert BR, Häberg SE, Nilsen RM, Wang X, Vollset SE, Murphy SK, et al. 450K epigenome-wide scan identifies differential DNA methylation in newborns related to maternal smoking during pregnancy. *Environ Health Perspect*. 2012;120:1425–31.
55. Küpers LK, Xu X, Jankipersadsing SA, Vaez A, La Bastide-van Gemert S, Scholtens S, et al. DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. *Int J Epidemiol*. 2015;44:1224–37.
56. Wahl S, Drong A, Lehne B, Loh M, Scott WR, Kunze S, et al. Epigenome-wide association study of body mass index, and the adverse outcomes of adiposity. *Nature*. 2017;541:81–6. <https://www.nature.com/articles/nature20784>.
57. Liu C, Marioni RE, Hedman AK, Pfeiffer L, Tsai PC, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. *Mol Psychiatry*. 2018;23:422–33.
58. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic signatures of cigarette smoking. *Circ Cardiovasc Genet*. 2016;9:436–47. <http://circgenetics.ahajournals.org/lookup/suppl/>.
59. Álvarez-Erro D, Vento-Tormo R, Sieweke M, Ballestar E. Epigenetic control of myeloid cell differentiation, identity and function. *Nat Rev Immunol*. 2015;15:7–17. <https://pubmed.ncbi.nlm.nih.gov/25534619/>.
60. Eyre HA, Lavretsky H, Kartika J, Qassim A, Baune BT. Modulatory effects of antidepressant classes on the innate and adaptive immune system in depression. *Pharmacopsychiatry*. 2016;49:85–96.
61. Hafferty JD, Campbell AI, Navrady LB, Adams MJ, MacIntyre D, Lawrie SM, et al. Self-reported medication use validated through record linkage to national prescribing data. *J Clin Epidemiol*. 2018;94:132–42.
62. Arnet I, Abraham I, Messerli M, Hersberger KE. A method for calculating adherence to polypharmacy from dispensing data records. *Int J Clin Pharm*. 2014;36:192–201.
63. Batram T, Richmond RC, Baglietto L, Haycock PC, Perduca V, Bojesen SE, et al. Appraising the causal relevance of DNA methylation for risk of lung cancer. *Int J Epidemiol*. 2019;48:1493–504. <https://academic.oup.com/ije/article/48/5/1493/5573019>.
64. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and alcohol use. *Nat Genet*. 2020;51:237–44. <https://pmc/articles/PMC6358542/?report=abstract>.

65. Chen MH, Raffield LM, Mousas A, Sakaue S, Huffman JE, Moscati A, et al. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell*. 2020;182:1198–1213.e14. <https://pubmed.ncbi.nlm.nih.gov/32888493/>.
66. Kichaev G, Bhatia G, Loh PR, Gazal S, Burch K, Freund MK, et al. Leveraging polygenic functional enrichment to improve GWAS power. *Am J Hum Genet*. 2019;104:65–75. <https://pubmed.ncbi.nlm.nih.gov/30595370/>.
67. Astle WJ, Elding H, Jiang T, Allen D, Ruklisa D, Mann AL, et al. The allelic landscape of human blood cell trait variation and links to common complex disease. *Cell*. 2016;167:1415–1429.e19. <https://pubmed.ncbi.nlm.nih.gov/27863252/>.
68. Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491:119–24. <https://pubmed.ncbi.nlm.nih.gov/23128233/>.
69. Liu JZ, Van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47:979–86. <https://pubmed.ncbi.nlm.nih.gov/26192919/>.
70. Din L, Sheikh M, Kosaraju N, Smedby KE, Bernatsky S, Berndt SI, et al. Genetic overlap between autoimmune diseases and non-Hodgkin lymphoma subtypes. *Genet Epidemiol*. 2019;43:844–63. <https://pubmed.ncbi.nlm.nih.gov/31407831/>.
71. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*. 2018;50:1112–21.
72. Zhu Z, Guo Y, Shi H, Liu CL, Panganiban RA, Chung W, et al. Shared genetic and experimental links between obesity-related traits and asthma subtypes in UK Biobank. *J Allergy Clin Immunol*. 2020;145:537–49. <https://pubmed.ncbi.nlm.nih.gov/31669095/>.
73. Lam M, Chen CY, Li Z, Martin AR, Bryois J, Ma X, et al. Comparative genetic architectures of schizophrenia in East Asian and European populations. *Nat Genet*. 2019;51:1670–8. <https://pubmed.ncbi.nlm.nih.gov/31740837/>.
74. Edwards TL, Giri A, Hellwege JN, Hartmann KE, Stewart EA, Jeff JM, et al. A trans-ethnic genome-wide association study of uterine fibroids. *Front Genet*. 2019;10. <https://pubmed.ncbi.nlm.nih.gov/31249589/>.
75. Evangelou E, Warren HR, Mosen-Ansorena D, Mifsud B, Pazoki R, Gao H, et al. Genetic analysis of over 1 million people identifies 535 new loci associated with blood pressure traits. *Nat Genet*. 2018;50:1412–25. <https://pubmed.ncbi.nlm.nih.gov/30224653/>.
76. Wang H, Yang J, Schneider JA, De Jager PL, Bennett DA, Zhang HY. Genome-wide interaction analysis of pathological hallmarks in Alzheimer's disease. *Neurobiol Aging*. 2020;93:61–8. <https://pubmed.ncbi.nlm.nih.gov/32450446/>.
77. Gallois A, Mefford J, Ko A, Vaysse A, Julienne H, Ala-Korpela M, et al. A comprehensive study of metabolite genetics reveals strong pleiotropy and heterogeneity across time and context. *Nat Commun*. 2019;10. <https://pubmed.ncbi.nlm.nih.gov/31636271/>.
78. Inouye M, Ripatti S, Kettunen J, Lyytikäinen LP, Oksala N, Laurila PP, et al. Novel loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet*. 2012;8. <https://pubmed.ncbi.nlm.nih.gov/22916037/>.
79. Li M, Maruthur NM, Loomis SJ, Pietzner M, North KE, Mei H, et al. Genome-wide association study of 1,5-anhydroglucitol identifies novel genetic loci linked to glucose metabolism. *Sci Rep*. 2017;7. <https://pubmed.ncbi.nlm.nih.gov/28588231/>.
80. Li Q, Wineinger NE, Fu DJ, Libiger O, Alphas L, Savitz A, et al. Genome-wide association study of paliperidone efficacy. *Pharmacogenet Genomics*. 2017;27:7–18. <https://pubmed.ncbi.nlm.nih.gov/27846195/>.

ACKNOWLEDGEMENTS

Generation Scotland is currently supported by the Wellcome Trust Investigator Award in Science 01/06/2021 to 31/05/26 'Exploiting genomic approaches to identify the environmental basis of depression'. (Reference: 220857/Z/20/Z) to AMM. For the purpose of open access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission. The DNA methylation profiling and data preparation was supported by Wellcome Investigator Award 220857/Z/20/Z and Grant 104036/Z/14/Z (PI for both grants: AMM) and through funding from NARSAD (Ref: 27404; PI: DMH and Ref: 21956; PI: KLE) and the Royal College of Physicians of Edinburgh (Sim Fellowship; PI: HCW). CA was supported by an MRC University Unit award to the MRC Human Genetics Unit, University of Edinburgh, MC_UU_00007/10. Genotyping of the GS:SFHS samples was funded by the MRC and Wellcome Trust [104036/Z/14/Z]. Generation Scotland also receives support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. DMH is supported by a Sir Henry Wellcome Postdoctoral Fellowship (Reference 213674/Z/18/Z). MCB is supported by a Guarantors of Brain Non-clinical Post-Doctoral Fellowship.

EB is an NIHR Senior Investigator and is a member of the scientific advisory board of Sosei Heptares. ML is supported by a fellowship from the Medical Research Council (MR/S006257/1).

AUTHOR CONTRIBUTIONS

MCB ran the main analyses included in the manuscript, including the MWAS of antidepressant use and SSRI in Generation Scotland, the meta-analysis of the two cohorts investigated (Generation Scotland and the Netherlands Twin Register), DNAm score analysis, and differential methylation region analysis; and wrote the manuscript. FH ran the MWAS of antidepressant use in the Netherlands Twin register and wrote the sections of the manuscript relating to the Netherlands Twin Register. AC is the Generation Scotland project manager and provided information about the variables used in the manuscript. CA pre-processed and quality-checked the DNA methylation data in Generation Scotland. MJA created the MWAS pipeline which was used to run the MWAS in Generation Scotland. MEL ran the pathway and regulatory element overlap analysis in Generation Scotland. DMH provided information with regards to the variables used in the manuscript as well as valuable comments in the write-up of the manuscript. RMW pre-processed and quality-checked the DNA methylation data in Generation Scotland. SWM pre-processed and quality-checked the DNA methylation data in Generation Scotland. JvD provided useful information about the Netherlands Twin Register data, including information about DNA methylation, lifestyle factors, and major depressive disorder. DP founded Generation Scotland and therefore provided valuable information for the manuscript. KLE coordinated the pre-processing and quality-checking the DNA methylation data in Generation Scotland and provided useful comments with regards to the DNA methylation data investigated here. EB provided useful information with regards to inflammatory processes related to both major depressive disorder and DNA methylation and provided valuable comments to the manuscript. GW supervises the Netherlands Twin Register and therefore provided valuable information to the manuscript. DIB provided useful information about the Netherlands Twin Register and coordinated to ensure that all analyses in the cohort matched the Generation Scotland analyses. HCW helped with analysis and result interpretation and provided useful comments throughout the manuscript. AMM helped with analysis and result interpretation and provided useful comments throughout the manuscript.

COMPETING INTERESTS

AMM has previously received grant support from Pfizer, Lilly and Janssen. These studies are not connected to the current investigation. Remaining authors report no conflicts of interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41380-021-01412-7>.

Correspondence and requests for materials should be addressed to Miruna C. Barbu.

Reprints and permission information is available at <http://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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021