medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Epigenome-wide association study of self-reported antidepressant use in Generation Scotland implicates the innate immune system

Barbu MC, PhD[1], Campbell A, MA[2], Amador, C, PhD[3], Adams, MJ, PhD[1], Lynall, ME, PhD[4], Howard, DM, PhD[1,5], Walker, RM, PhD[2,6], Bretherick, AD, PhD, Morris SW, MSc[2,6], Porteous, DJ, PhD[2,6], Evans, KL, PhD[2,6], Haley, C, PhD[3], Bullmore, E, PhD[4], Whalley, HC, PhD[1], McIntosh, AM, MD[1,6]

[1]Division of Psychiatry, The University of Edinburgh, Royal Edinburgh Hospital, Edinburgh, UK

[2]Centre for Genomic and Experimental Medicine, The Institute of Genetics and Molecular Medicine, The University of Edinburgh, UK

[3]MRC Human Genetics Unit, The Institute of Genetics and Molecular Medicine, The University of Edinburgh, UK

[4]Department of Psychiatry, University of Cambridge, Cambridge, UK

[5]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK

[6]Centre for Cognitive Ageing and Cognitive Epidemiology, School of Philosophy, Psychology and Language Sciences, The University of Edinburgh, UK

Correspondence can be sent to:

Miruna C Barbu, mbarbu@ed.ac.uk, 0131 537 6691

Division of Psychiatry, Kennedy Tower, Royal Edinburgh Hospital, Edinburgh EH10 5HF

Running title: Epigenome-wide association study of self-reported antidepressant use

Keywords: DNA methylation, epigenome-wide association study, methylation-based score, innate immune system, major depressive disorder, antidepressant use

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Abstract

Background: 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 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.

Methods: We performed an epigenome-wide association study (EWAS) of self-reported antidepressant use accounting for lifestyle factors and MDD in the Generation Scotland cohort (N=6,428).

Results: We found 10 CpG sites significantly associated with self-reported antidepressant use, with the top CpG located within a gene previously associated with mental health disorders, *ATP6V1B2* (β=-0.055, $p_{corrected}$=0.005). Other top loci, annotated to genes including *CASP10, TMBIM1, MAPKAPK3,* and *HEBP2*, have previously been implicated in the innate immune response. We also identified a number of disease-associated single nucleotide polymorphisms as *trans* and *cis* methylation quantitative trait loci (mQTLs) for CpG sites identified here. Next, using penalised regression, we trained a methylation-based score (MS) of self-reported antidepressant use in 3,799 individuals that predicted antidepressant use in a second subset of Generation Scotland (N=3,360,β=0.377, $p=3.12\times10^{-11}$,R^2=2.12%). Lastly, in an EWAS analysis of SSRI antidepressant prescribing data from electronic health records, we showed convergent findings with those based on self-report.

Conclusions: Antidepressants may exert their effects through epigenetic alterations in regions previously associated with mental health disorders and those involved in the innate immune system. These changes predicted self-reported antidepressant use in a subset of Generation Scotland and identify processes that may be relevant to our mechanistic understanding of clinically relevant antidepressant drug actions and side effects.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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 approximately 50% of individuals achieving remission after 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 re-uptake. 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. (2015) (9) demonstrated that stress leads to changes in gene expression and DNAm of 5-HT1A that is partly reversed by antidepressants in an animal model. Zimmermann et al. (2012) (10) found that antidepressants affect activity of DNA methyltransferase 1 (*DNMT1*) (10). The studies above demonstrate that antidepressants lead to epigenetic alterations that may help to reveal their biologically relevant properties and modes 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 800K locations throughout the genome. Epigenome-wide association studies (EWAS) have begun to identify a number of cytosine-phosphate-guanine (CpG) sites that are associated with MDD (12) and relevant environmental factors (13). Some DNAm changes may therefore act as an environmental archive capturing the effects of exogenous factors, including drug treatments that may be relevant to the onset and maintenance of MDD.

Recent studies have also shown the potential of a DNAm-based risk score to predict MDD and its associated lifestyle and environmental factors, including antidepressant use (14,15). Barbu et al. (2020) (15) trained a DNAm predictor of MDD in 3,047 individuals using a penalised regression model and showed that it was significantly associated with self-

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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 6,428 individuals ($N_{antidepressant\ use}$=740) from the Generation Scotland study using the Illumina Infinium MethylationEPIC array (16), capturing DNAm at approximately 850K sites. Further, by splitting the Generation Scotland cohort into two DNAm datasets, we were able to train a DNAm-based methylation score (MS) of self-reported antidepressant use in one dataset (N=3,799; $N_{antidepressant\ use}$=585), and test its ability to predict self-reported antidepressant use in a second (N=3,360; $N_{antidepressant\ use}$=317). We also addressed confounding by indication and by smoking in planned sensitivity analyses (17). Using linked National Health Service (NHS) Scotland medication prescribing data, we were able to identify whether the EWAS findings from self-report were also found when using contemporaneous electronic healthcare information collected around the time of the blood draw.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Methods

Study population: *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 (18,19) and contains detailed information on a broad range of variables, including lifestyle and environmental factors, medication, mental health, and linkage to comprehensive routine healthcare records. 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: 15/ES/0040). Written informed consent was obtained from all participants.

Phenotypes

Self-reported antidepressant use

For self-reported antidepressant use, a text-based questionnaire was used for participants recruited between June 2009 and March 2011. 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". 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 medication. There were 7,174 individuals with available self-reported data who also had DNAm data available; this number decreased when including environmental and lifestyle variables in statistical analyses (see Table 1).

Selective serotonin reuptake inhibitors (SSRI) prescription and data linkage

Almost all individuals registered with a General Practitioner (GP) in Scotland are assigned a Community Health Index (CHI) number which acts as a unique identifier. The CHI number was used to record-link GS:SFHS data collected in self-reported questionnaires to the national Prescribing Information System (PIS) administered by NHS National Services

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Scotland Information Services Division (20). PIS provides information on patient-level prescriptions since April 2009 and allows the identification of medications by approved drug name or paragraph code within the British National Formulary (BNF) (21). GS:SFHS obtained PIS-prescription data for April 2009-March 2011.

Here, we selected antidepressants from BNF paragraph code 4.3.3 containing the following SSRIs: Citalopram, Escitalopram, Fluoxetine, Fluvoxamine Maleate, Paroxetine, and Sertraline (21). To investigate DNAm signatures of SSRI use, we restricted analyses to those individuals with SSRI dispensing records within the 12 months prior to blood draw date. Individuals that had any number of dispensed prescriptions within 12 months before the blood draw date were marked as cases (N=487). Individuals with DNAm data that did not have any SSRI dispensed prescription records were marked as controls (N=7,463; this number decreased when including environmental and lifestyle variables in statistical analyses, see Table 1).

We used dates of dispensing, not prescription, when restricting antidepressant use based on time interval. The total number of prescriptions dispensed for each of the 6 SSRIs during the 12 months (where one participant may have multiple dispensed prescriptions) is presented in 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 (22).

MDD status was assessed using the Structured Clinical Interview of the Diagnostic and Statistical Manual, version IV (SCID) (23). 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 (24). Further details are presented in the Supplementary Materials.

DNA methylation

Genome-wide DNAm data profiled from whole blood samples was available for 9,873 individuals in GS:SFHS using the Illumina Human-MethylationEPIC BeadChip (16). DNAm data for individuals was initially released in two waves (wave 1_N=5101; wave 2_N=4,450). The raw data for all participants in the present study was pre-processed and

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

quality checked for all individuals, after participant removal due to a number of reasons, including having more than 1% CpG sites with a detection p-value > 0.05, showing evidence of dye bias, being an outlier for bisulphite conversion control probes, and having a median methylated signal intensity more than 3 standard deviations lower than expected. A total of 10,495 CpG sites were removed due to low beadcount, poor detection p-value, and sub-optimal binding.

Firstly, R package "minfi" was used to read in the IDAT files, compute M and beta values, and remove probes with large detection p-values, and to compute principal components (PC) of control probes. Secondly, correction was applied for (1) technical variation, where M values were included as outcome variables in a mixed linear model adjusting for appointment date and Sentrix ID (random effects), jointly with Sentrix position, batch, clinic, year, weekday, and 10 PCs (fixed effects); and (2) biological variation by fitting residuals of (1) as outcome variables in a second mixed linear model adjusting for genetic and common family shared environmental contributions (random effects classed as G: common genetic; K: kinship; F: nuclear family; C: couple; and S: sibling) and sex, age, and estimated cell types proportions (CD8T, CD4T, NK, Bcell, Mono, Gran) (fixed effects) (25). The final number of CpG sites that converged for these analyses was 736,940 across the 22 autosomes.

Statistical methods

Epigenome-wide association

We used linear regression models run in the "limma" package (26) 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, 20 PCs, wave (indicating different 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 CpG sites 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. There were 713,522 CpG sites available after QC and a Bonferroni correction (0.05/713,522) was used to define epigenome-wide significance ($p \leq 7.007 \times 10^{-8}$).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We also performed the above linear regression using SSRI prescribing data as a predictor, to identify whether there are differences in DNAm in relation to antidepressant use when using SSRI PIS data as a phenotype. We restricted SSRI dispense dates to 12 months prior to the blood draw date for each individual. The covariates included here are the same as in the model specified above.

Overlap with disease-associated single nucleotide polymorphisms (SNPs)

To identify SNPs associated with both disease traits and CpG sites that play a role in antidepressant use, we analysed the overlap between disease-associated SNPs (27–31) and SNPs identified as mQTLs (Bretherick et al., in preparation) for the CpG sites we identified in the current EWAS. Candidate diseases were selected based on their association with genes annotated to CpGs identified in the current EWAS.

We used methylation quantitative trait loci (mQTL) analyses carried out in the Generation Scotland wave 1 and GWAS for a number of disease traits, specifically MDD, schizophrenia, cardiovascular disease (CVD), inflammatory bowel syndrome (IBS), and rheumatoid arthritis (RA), to find potential disease-associated mQTLs for CpG sites identified in the current EWAS. The mQTL analysis investigated associations between SNPs and methylation levels at 638,737 CpG sites, and only those associations with $p<1x10^{-3}$ were kept. SNPs associated with MDD (N_{SNPs}=125), schizophrenia (N_{SNPs}=128), CVD (N_{SNPs}=46), IBS (N_{SNPs}=7), and RA (N_{SNPs}=107) at genome-wide significance ($p<5x10^{-8}$) were identified from: Howard et al.'s (2019) (27) MDD GWAS results (excluding Generation Scotland from the meta-analysis); Ripke et al. (2014) (28); Deloukas et al. (2013) (29); Bonfiglio et al. (2018) (30) and Yarwood et al. (2016) (31), respectively. Further details of these analyses are provided in each publication.

DNAm score analysis

We separated individuals in DNAm waves 1 and 2 into training and testing datasets, respectively. In wave 1, self-reported antidepressant use (N=3,799; antidepressant users=585) was first residualised for sex, age, and 10 genetic PCs. The R package "biglasso" was then used to train DNAm predictors, where probes were restricted to only those available on the Illumina 450K array (N=365,837), to maximise the availability and use of our predictor to other cohorts using the 450K array. Penalised regression was applied using the "cv.lasso" function and 10-fold cross validation. Non-zero coefficients from this model, with the lambda value corresponding to the mean square error, were then used to create MS for self-reported antidepressant use in Generation Scotland wave 2.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Due to the known association between smoking and DNAm (17), we also trained DNAm predictors on a set of non-smokers in wave 1, to exclude potentially confounding smoking signals within our predictor. This was achieved by excluding those individuals who had a smoking history (i.e. answered "yes" to the question, "have you ever smoked tobacco?") from all individuals with self-reported antidepressant use data (N excluded=1,847). The training dataset here consisted of 1,952 individuals (226 antidepressant users). Similarly, self-reported antidepressant use was residualised for age, sex, and 10 genetic PCs, and penalised regression was applied as above.

Finally, to exclude MDD-associated effects within our predictor, we trained a further DNAm score on a set of individuals with no MDD diagnosis in wave 1 (MS-control). The training dataset consisted of 2,791 individuals (antidepressant users=195).

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 1, 2, and 3. MS, MS-ns, and MS-control were created for each individual in wave 2 (N=3,360; 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, as well as a number of lifestyle factors and MDD.

Pathway and regulatory element overlap analysis

We used the Infinium MethylationEPIC BeadChip database to annotate significantly-associated CpG sites 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 (32)**,** accessed via methylGSA (33). Gene Ontology (GO) terms were accessed using the msigdbr package (34). 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 $p<1x10^{-5}$, as used in previous studies (35).

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 (36), accessed via the web tool at https://eforge.altiusinstitute.org and using default settings.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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 6,428 individuals in the final EWAS of self-reported antidepressant use ($N_{antidepressant\ use}$: 740). Descriptive and demographic characteristics for these individuals, as well as individuals with prescribing data, are presented in Table 1.

Self-reported antidepressant use

EWAS identified 10 CpG sites that were associated with self-reported antidepressant use ($p \leq 7.007 \times 10^{-8}$). Nine CpG sites were significantly hypomethylated in antidepressant users, while 1 site (cg26277237) was hypermethylated. Information about each CpG site is shown in Table 2 and depicted as a Manhattan plot in Figure 1.

The EWAS catalogue (http://www.ewascatalog.org/) includes CpG sites represented on the Illumina 450K 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. Eight of the CpG sites were represented only on the Infinium MethylationEPIC BeadChip (containing approximately 850K CpG sites), and were therefore not included in the catalogue. Searches conducted on other databases, including EWASdb (37) indicate that these 8 CpG sites have not been previously associated with any trait.

cg05603985 was previously found to be associated with smoking ($p=1.8 \times 10^{-43}$; (17)), alcohol consumption ($p=7.9 \times 10^{-13}$ in African ancestry and $p=1.9 \times 10^{-8}$ in European ancestry; (38)) and gestational age ($p=4.8 \times 10^{-5}$; (39)); cg27589594 was found to be associated with smoking ($p=6.9 \times 10^{-5}$; (17)) and gestational age ($p=3.5 \times 10^{-16}$; (39)).

Overlap with disease-associated SNPs

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

We found several disease-associated SNPs to be *cis* and *trans* mQTLs for CpGs identified here; results are indicated in Table 3. Regional visualisation plots for all SNPs are presented in Supplementary Figures 1-12.

For MDD, the association between antidepressant treatment and DNAm was negative for both CpGs, and the same direction of effect was shown in the association between the two CpGs and the risk alleles at the associated SNPs. rs2056445 was not previously associated with any phenotypes. As indicated in the GWAS catalogue (https://www.ebi.ac.uk/gwas/), rs3793577 was previously associated with depressive symptoms ($p=7x10^{-13}$; (40)), neuroticism ($p=7x10^{-13}$ in (40); $p=4x10^{-7}$ in (41)), and well-being ($p=7x10^{-13}$; (40)), in addition to depression (27).

The two schizophrenia-associated SNPs were positively associated with the CpGs identified here, while the CpG-antidepressant use association was negative. rs8044995 and rs7523273 were not found to be previously associated with any traits in addition to schizophrenia (28).

Similarly, for CVD, the direction of effect for the SNP-CpG association was positive, whereas it was negative for the antidepressant use-CpG association. In addition to CVD (29), rs3184504 was previously associated with a host of traits, including leukocyte count ($p=8x10^{-222}$; (42)); hypothyroidism ($p=1x10^{-114}$; (42)); smoking status ($p=6x10^{-67}$; (43)); diastolic blood pressure ($p=2x10^{-64}$; (44)); and coronary artery disease ($p=5x10^{-30}$; (45)).

There were 6 RA-associated SNPs that were found to be mQTLs for CpGs identified here; 3 were in the same direction of effect as the antidepressant use-CpG association (rs6715284, rs2075876, rs6732565), and 4 were opposite (rs13330176, rs6732565, rs10774624, rs2075876). rs6732565 was found to be an mQTL for both cg05186879 and cg20494891. Interestingly, rs6715284, rs2075876, rs6732565, rs13330176, and rs2075876 were not associated with other traits with the exception of RA. rs10774624 was found to be associated with eosinophil count ($p=1x10^{-300}$; (42)); vitiligo ($p=6x10^{-23}$; (46)); systolic blood pressure ($p=3x10^{-19}$; (47)); and intraocular pressure measurement ($p=3x10^{-10}$; (48)), in addition to RA (31).

Finally, IBS-associated SNPs were not identified as mQTLs for any CpGs.

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 Generation Scotland wave 2.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Briefly, the MS was associated with self-reported antidepressant use in a model adjusted for age, sex, and 10 genetic PCs (β=0.377, p=3.12x10^{-11}, R^2=2.12%); when including BMI, alcohol units, smoking status, pack years, and MDD as covariates, this association was still significant, although the variance explained decreased (β=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 (β=0.106, p=0.075, R^2=0.16%) and in the model with the above covariates included (β= 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 (β=0.254, p=1.15x10^{-5}, R^2=0.93%), but not when including covariates (β=0.064, p=0.381, R^2=0.05%). Table 4 below indicates further associations uncovered for all scores.

SSRI

EWAS conducted using individuals with dispensing data identified additional 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 4 and Supplementary Figure 13.

Additionally, all 10 CpG sites that were associated with self-reported antidepressant use ($p \leq 7.007 \times 10^{-8}$) were found to be nominally significant ($p<0.05$) in the EWAS including SSRI dispensing records as a phenotype. The direction of effect was consistent for all CpG sites identified in the self-report EWAS: 9 were hypomethylated and 1 was hypermethylated, presented in Supplementary Figure 13 and Supplementary Table 5 (β and p-value for smallest and largest associations included cg09511513 (β=-0.026, p=0.028) and cg26277237 (β=0.027, p=3.09x10^{-7}). In turn, a majority of CpG sites identified in the EWAS of SSRI dispense records in the 12 months prior to blood draw date were nominally significant in the EWAS where the self-report variable was input as the predictor (Supplementary Table 6).

Pathway and regulatory element overlap analysis

To assess whether differentially methylated regions associated with antidepressant use preferentially affect specific pathways, we used missMethyl (32) to perform an over-representation analysis of GO Biological Pathways for the set of genes annotated to the 144 CpG sites differentially expressed at $p<1 \times 10^{-5}$. Following FDR adjustment for multiple comparisons, the only over-represented pathways were related to regulation of myeloid cell

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

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

We also used eFORGE (36) to test for any tissue-specific signal in the overlap of differentially methylated CpG sites with regulatory elements. Testing for overlap of differentially methylated CpG sites at $p<1x10^{-5}$ 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 (Figure 2), with significant enrichment for monocyte enhancers ($q=4x10^{-11}$) and regions flanking active transcription start sites in monocytes ($q=1x10^{-9}$).

We repeated the tests for enrichment of gene sets and tissue-specific chromatin states using the results from the EWAS of antidepressant prescriptions, including CpGs significant at p<1x10-5 (N=484 CpGs). We did not find enrichment of any specific GO biological processes, but replicated the finding from the self-reported antidepressant use EWAS that monocytes were the peripheral blood cell subset most enriched for the EWAS signal (enhancers in primary monocytes from peripheral blood were enriched at $q=4x10^{-94}$) and also found strong enrichment for enhancers in peripheral blood neutrophils ($q=4x10^{-79}$).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Discussion

Self-reported antidepressant use is associated with differences in DNAm at 10 CpG sites that have previously been associated with psychiatric disorders and the innate immune system. We found that all 10 CpG sites were also nominally associated with antidepressant prescription data obtained from the electronic health record, indicating agreement between self-reported and record linkage data. A DNAm score trained on one sub-sample was associated with self-reported antidepressant use, 4 lifestyle factors, and MDD in a second unrelated sample from the same study. Lastly, we found 2 SNPs previously associated with MDD to be *trans* mQTLs for 2 CpG sites identified here.

The CpG site that had the strongest association with self-reported antidepressant use in terms of effect size was cg09511513 (β=-0.055, $p_{corrected}$=0.005), 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 (49). Importantly, a single nucleotide polymorphism (SNP) within *ATP6V1B2*, rs1106634, has been associated with MDD ($p= 7x10^{-7}$) and bipolar disorder ($p=5.63x10^{-6}$) in previous GWAS (50–52).

Two SNPs-CpG *trans* associations were identified, one of which was novel (rs2056445–cg03864397). Interestingly, the mQTL for cg27589594, rs3793577, was previously associated with neuroticism, depressive symptoms, and wellbeing, in addition to MDD (27,40,41). In addition to MDD-associated mQTLs, we identified a number of disease-associated SNPs, including schizophrenia (28), CVD (29), and RA (31), to be both *trans* and *cis* mQTLs for CpGs identified here. Most disease-associated variants are non-coding and may exert their effects through gene expression modifications (53). Further, the direction of effect in the current study differed according to disease when investigating the two CpGs in relation to SNPs and antidepressant use. The results here indicate that genetic liability for

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

disease-associated traits may influence DNAm alterations at the same loci associated here with antidepressant use. Further work is required to disentangle the effects of antidepressants on DNAm from the effects of the conditions they are prescribed for.

We were able to show that a MS for self-reported antidepressant use was associated with a number of lifestyle and environmental factors in a subset of Generation Scotland. This included self-reported antidepressant use, MDD status, and 4 lifestyle factors (BMI, alcohol units, smoking status, and pack years). 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 effects may be partly independent from the condition they are prescribed for. Further, only one CpG overlaps between the CpG sites in our MS and CpG sites comprising an MDD DNAm risk score in Barbu et al. (2020) (15). This CpG (cg09935388) has been associated with smoking in previous studies (54,55), and may therefore capture smoking-associated effects with both MDD and antidepressant use. It should also be noted that in the current study, the decreased variance explained may be in part due to differences in the training sample sizes for the two scores (MS_N=3,799, antidepressant users=585; MS-control$_N$=2,791, antidepressant users=195), although variation explained in lifestyle factors is comparable between the two scores (MS R^2: BMI=0.75%; alcohol units=1.93%; smoking status=6.11%; pack years=9.14%; MS-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, being associated with both antidepressant use and DNAm alterations. Similarly to MS-control, these results may be due to differences in the training sample sizes for the two scores (MS_N=3,799; MS-ns$_N$=1,952), and the reduced number of antidepressant users included in the training sets (MS_N=585; MS-ns$_N$=226). Future studies would benefit from training DNAm predictors of self-reported antidepressant use in larger samples of lifelong non-smokers.

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 (56). Several of the genes annotated to CpGs identified at genome-wide significance here

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

were associated with the innate immune response in previous GWAS; this may complement the finding that genes annotated to the 144 CpG sites differentially expressed at $p<1x10^{-5}$ are enriched for myeloid cell differentiation.

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 (57). 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 pro-inflammatory cytokine levels, such as TNF-α, IL-6, and IL-1β, although their effects are complex and incompletely understood (57). The current study may provide novel links between DNAm, antidepressants, and their association with processes involved in the innate immune system. Studies that investigate whether antidepressants exert their effect through the alteration of DNAm 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.

The most commonly dispensed SSRI in the current study was citalopram (Table 1; $N_{prescriptions}$=1,243). Kanherkar et al. (2018) (58) investigated the effect of citalopram (administrated for 30 days) on *in vitro* genome-wide DNAm in HEK-293 cells (human embryonic kidney cells). They found that 626 gene promoters (2.46% of a total of 25,437 genes) showed significant differential methylation. Specifically, of the 626 gene promoters, 272 were hypomethylated and 354 were hypermethylated in treated individuals as compared to controls. Among the top gene networks that were differentially regulated were pathways involved in depression, nervous system development and function, as well as cellular growth and proliferation (58). The above study indicates that citalopram may be one of the SSRIs that may have widespread effects on the methylome. It may be that some of the findings in the current study are driven by citalopram, although replication and longitudinal studies including other SSRIs and other classes of antidepressants would be needed to verify this.

Self-reported antidepressant use was not restricted to specific dates prior to the blood draw date, as this variable was measured using a single question that asked individuals whether they had ever taken antidepressants. This may explain the different top CpG sites based on the predictor input in EWAS. Nonetheless, as indicated in Supplementary Table 6, most of the CpG sites significantly associated with SSRI use within 12 months prior to blood draw date were nominally significant in the EWAS where self-reported antidepressant use was fit as the predictor. This pattern was also observed when investigating the top 10 CpG

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

sites associated with the self-reported measure in the SSRI use EWAS (Supplementary Table 5). 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. (2018) (59), 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 (59).

There are a number of potential limitations to the current study that should be taken into account when considering the main findings. Firstly, there are differences between the antidepressant use phenotypes used in EWAS, specifically a self-reported measure and PIS data. While agreement has previously been shown between these two variables in GS:SFHS (59), the different CpG sites identified in each EWAS may be due in part to the non-overlapping individuals used in each analysis. Nevertheless, since all self-reported EWAS associations were also nominally significant in the SSRI EWAS utilising prospectively collected dispensing data before the blood draw, our findings are unlikely to be false positives or due to recall bias.

Secondly, as we identified two *trans* mQTLs that were previously associated with MDD, it is possible that these changes in DNAm were as a result of liability to MDD rather than antidepressant prescribing. It is however also possible that antidepressant use may lead to epigenetic changes in genes that have been previously implicated in the aetiology of MDD.

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 (60). The current study is also based exclusively on a European ancestry cohort, and the generalisability to diverse ancestries is unknown.

In conclusion, we conducted an EWAS of self-reported antidepressant use and identified 10 novel CpG sites 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. Finally, SSRI use derived from linkage health records showed

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

convergent effects indicating agreement with the findings derived using only self-report. Our findings highlight biological processes that may be relevant to furthering our understanding of antidepressant actions and their side effects.

References

1. World Health Organization. Depression and Other Common Mental Disorders. Who. 2017;24.
2. 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(JUL).
3. I K. Antidepressants and the Placebo Effect. Z Psychol. 2014;222(3).
4. 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 [Internet]. 2018;391(10128):1357–66. Available from: http://dx.doi.org/10.1016/S0140-6736(17)32802-7
5. Thase ME, Entsuah AR, Rudolph RL. Remission rates during treatment with venlafaxine or selective serotonin reuptake inhibitors. Br J Psychiatry. 2001;178(MARCH.):234–41.
6. Keers R, Uher R. Gene-environment interaction in major depression and antidepressant treatment response. Vol. 14, Current Psychiatry Reports. Curr Psychiatry Rep; 2012. p. 129–37.
7. Berton O, Nestler EJ. New approaches to antidepressant drug discovery: Beyond monoamines. Vol. 7, Nature Reviews Neuroscience. Nat Rev Neurosci; 2006. p. 137–51.
8. Menke A, Binder EB. Epigenetic alterations in depression and antidepressant

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

treatment. Dialogues Clin Neurosci. 2014;16(3):395–404.

9. 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 Oct 1;82:332–41.
10. Zimmermann N, Zschocke J, Perisic T, Yu S, Holsboer F, Rein T. Antidepressants inhibit DNA methyltransferase 1 through reducing G9a levels. Biochem J. 2012 Nov 15;448(1):93–102.
11. Bird A. Perceptions of epigenetics. Vol. 447, Nature. Nature Publishing Group; 2007. p. 396–8.
12. 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(9):949–59.
13. 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(1):136.
14. 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 Jun 1;25(6):1334–43.
15. Barbu MC, Shen X, Walker RM, Howard DM, Evans KL, Whalley HC, et al. Epigenetic prediction of major depressive disorder. Mol Psychiatry [Internet]. 2020 Jun 10 [cited 2020 Jun 11];1–12. Available from: http://www.nature.com/articles/s41380-020-0808-3
16. Hansen K. IlluminaHumanMethylationEPICanno.ilm10b2.hg19: Annotation for Illumina's EPIC methylation arrays. R Packag version 060 [Internet]. 2016; Available from: https://bitbucket.com/kasperdanielhansen/Illumina_EPIC
17. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet. 2016;436–47.
18. 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.
19. 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

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

J Epidemiol. 2013;(July 2012):689–700.

20. Alvarez-Madrazo S, McTaggart S, Nangle C, Nicholson E BM. Data Resource Profile: The Scottish National Prescribing Information System (PIS) [Internet]. International journal of epidemiology. 2016 [cited 2020 Jun 11]. p. 714. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5005947/
21. Britain RPS of G. British National Formulary 58. 2009.
22. Karen Leffondré, Michal Abrahamowicz, Jack Siemiatycki BR. Modeling Smoking History: A Comparison of Different Approaches. Am J Epidemiol. 2002;156(9):813–23.
23. 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 reliablity study. J Pers Disord. 1995 Feb 8;9(2):92–104.
24. 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 Nov 1;10(11).
25. Xia C, Amador C, Huffman J, Trochet H, Campbell A, Porteous D, et al. Pedigree- and SNP-Associated Genetics and Recent Environment are the Major Contributors to Anthropometric and Cardiometabolic Trait Variation. PLoS Genet. 2016 Feb 1;12(2).
26. Matthew E Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi GKS. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43(7):47.
27. Howard DM, Adams MJ, Clarke TK, Hafferty JD, Gibson J, Shirali M, et al. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat Neurosci [Internet]. 2019;22(3):343–52. Available from: http://dx.doi.org/10.1038/s41593-018-0326-7
28. Ripke S, Neale BM, Corvin A, Walters JTR, Farh KH, Holmans PA, et al. Biological insights from 108 schizophrenia-associated genetic loci. Nature [Internet]. 2014 [cited 2020 Sep 17];511(7510):421–7. Available from: https://pubmed.ncbi.nlm.nih.gov/25056061/
29. Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, Thompson JR, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet [Internet]. 2013 Jan 2 [cited 2020 Sep 17];45(1):25–33. Available from: https://www.nature.com/articles/ng.2480

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

30. Bonfiglio F, Zheng T, Garcia-Etxebarria K, Hadizadeh F, Bujanda L, Bresso F, et al. Female-Specific Association Between Variants on Chromosome 9 and Self-Reported Diagnosis of Irritable Bowel Syndrome. Gastroenterology. 2018 Jul 1;155(1):168–79.
31. Yarwood A, Huizinga TWJ, Worthington J. The genetics of rheumatoid arthritis: Risk and protection in different stages of the evolution of RA. Rheumatol (United Kingdom) [Internet]. 2015 Oct 5 [cited 2020 Sep 17];55(2):199–209. Available from: https://academic.oup.com/rheumatology/article/55/2/199/1821726
32. Phipson B, Maksimovic J, Oshlack A. MissMethyl: An R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics. 2016;32(2):286–8.
33. 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(11):1958–9.
34. Dolgalev I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. R package version 7.1.1. [Internet]. 2020 [cited 2020 Jul 29]. Available from: https://cran.r-project.org/web/packages/msigdbr/index.html
35. 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 [Internet]. 2017 Dec 1;8(1):1–12. Available from: www.nature.com/naturecommunications
36. 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(22):4767–9.
37. Liu D, Zhao L, Wang Z, Zhou X, Fan X, Li Y, Xu J, Hu S, Niu M, Song X LY. EWASdb: epigenome-wide association study database. Nucleic Acids Res. 2019;47:989–93.
38. Liu C, Marioni RE, Hedman ÅK, Pfeiffer L, Tsai P, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. Mol Psychiatry. 2018;(February 2016):422–33.
39. Bohlin J, Håberg 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 Oct 7;17(1).
40. Baselmans BML, Jansen R, Ip HF, van Dongen J, Abdellaoui A, van de Weijer MP, et al. Multivariate genome-wide analyses of the well-being spectrum. Nat Genet. 2019 Mar 1;51(3):445–51.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

41. Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke TK, Shirali M, et al. Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nat Genet. 2018 Jan 1;50(1):6–11.

42. 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 [Internet]. 2019 Jan 3 [cited 2020 Sep 17];104(1):65–75. Available from: https://pubmed.ncbi.nlm.nih.gov/30595370/

43. Sung YJ, Winkler TW, de las Fuentes L, Bentley AR, Brown MR, Kraja AT, et al. A Large-Scale Multi-ancestry Genome-wide Study Accounting for Smoking Behavior Identifies Multiple Significant Loci for Blood Pressure. Am J Hum Genet [Internet]. 2018 Mar 1 [cited 2020 Sep 17];102(3):375–400. Available from: https://pubmed.ncbi.nlm.nih.gov/29455858/

44. Hoffmann TJ, Ehret GB, Nandakumar P, Ranatunga D, Schaefer C, Kwok PY, et al. Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation. Nat Genet [Internet]. 2017 Jan 1 [cited 2020 Sep 17];49(1):54–64. Available from: https://pubmed.ncbi.nlm.nih.gov/27841878/

45. Van Der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. Circ Res [Internet]. 2018 Feb 2 [cited 2020 Sep 17];122(3):433–43. Available from: https://pubmed.ncbi.nlm.nih.gov/29212778/

46. Jin Y, Andersen G, Yorgov D, Ferrara TM, Ben S, Brownson KM, et al. Genome-wide association studies of autoimmune vitiligo identify 23 new risk loci and highlight key pathways and regulatory variants. Nat Genet [Internet]. 2016 Nov 1 [cited 2020 Sep 17];48(11):1418–24. Available from: https://pubmed.ncbi.nlm.nih.gov/27723757/

47. Giri A, Hellwege JN, Keaton JM, Park J, Qiu C, Warren HR, et al. Trans-ethnic association study of blood pressure determinants in over 750,000 individuals. Nat Genet [Internet]. 2019 Jan 1 [cited 2020 Sep 17];51(1):51–62. Available from: https://pubmed.ncbi.nlm.nih.gov/30578418/

48. Khawaja AP, Cooke Bailey JN, Wareham NJ, Scott RA, Simcoe M, Igo RP, et al. Genome-wide analyses identify 68 new loci associated with intraocular pressure and improve risk prediction for primary open-angle glaucoma. Nat Genet [Internet]. 2018 Jun 1 [cited 2020 Sep 17];50(6):778–82. Available from: https://pubmed.ncbi.nlm.nih.gov/29785010/

49. Petschner P, Gonda X, Baksa D, Eszlari N, Trivaks M, Juhasz G, et al. Genes Linking

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Mitochondrial Function, Cognitive Impairment and Depression are Associated with Endophenotypes Serving Precision Medicine. Vol. 370, Neuroscience. Elsevier Ltd; 2018. p. 207–17.

50. Wang KS, Liu XF, Aragam N. A genome-wide meta-analysis identifies novel loci associated with schizophrenia and bipolar disorder. Schizophr Res. 2010 Dec;124(1–3):192–9.

51. 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 Feb;16(2):202–15.

52. Gonda X, Eszlari N, Anderson IM, Deakin JFW, Bagdy G, Juhasz G. Association of ATP6V1B2 rs1106634 with lifetime risk of depression and hippocampal neurocognitive deficits: possible novel mechanisms in the etiopathology of depression. Transl Psychiatry. 2016 Nov 8;6(11):e945.

53. Bonder MJ, Luijk R, Zhernakova D V., Moed M, Deelen P, Vermaat M, et al. Disease variants alter transcription factor levels and methylation of their binding sites. Nat Genet. 2017 Jan 1;49(1):131–8.

54. Joubert BR, Håberg 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 Oct;120(10):1425–31.

55. Küpers LK, Xu X, Jankipersadsing SA, Vaez A, La Bastide-van Gemert S, Scholtens S, Nolte IM, Richmond RC, Relton CL, Felix JF DL. DNA methylation mediates the effect of maternal smoking during pregnancy on birthweight of the offspring. Int J Epidemiol. 2015;44(4):1224–37.

56. Álvarez-Errico D, Vento-Tormo R, Sieweke M, Ballestar E. Epigenetic control of myeloid cell differentiation, identity and function [Internet]. Vol. 15, Nature Reviews Immunology. Nature Publishing Group; 2015 [cited 2020 Jul 29]. p. 7–17. Available from: https://pubmed.ncbi.nlm.nih.gov/25534619/

57. Eyre HA, Lavretsky H, Kartika J, Qassim A, Baune BT. Modulatory Effects of Antidepressant Classes on the Innate and Adaptive Immune System in Depression. Vol. 49, Pharmacopsychiatry. Georg Thieme Verlag; 2016. p. 85–96.

58. RR K, B G, J B-S, S V, T H, Y T, et al. The Effect of Citalopram on Genome-Wide DNA Methylation of Human Cells. Int J Genomics. 2018;2018.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

59. 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 Feb 1;94:132–42.

60. Arnet I, Abraham I, Messerli M, Hersberger KE. A method for calculating adherence to polypharmacy from dispensing data records. Int J Clin Pharm. 2014 Nov 29;36(1):192–201.

Acknowledgements

ADB would like to acknowledge funding from the Wellcome PhD training fellowship for clinicians (204979/Z/16/Z), the Edinburgh Clinical Academic Track (ECAT) programme. 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). DMH is supported by a Sir Henry Wellcome Postdoctoral Fellowship (Reference 213674/Z/18/Z) and a 2018 NARSAD Young Investigator Grant from the Brain & Behavior Research Foundation (Ref: 27404).

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	Self-reported AD use sample		**SSRI sample from NHS Scotland records**	
Demographic characteristics	**AD use** (N=740)	**No AD use** (N=5,688)	**SSRI use in 12-month interval** (N=487)	**No SSRI use** (N=6,705)
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%)	3,336 (59%)	364 (75%)	3,631 (54%)
Male	176 (24%)	2,352 (41%)	123 (25%)	3,074 (46%)
Wave (%)				
1	480 (65%)	2,918 (51%)	321 (66%)	3,389 (51%)
2	260 (35%)	2,770 (49%)	166 (34%)	3,316 (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%)	125 (27%)	1,006 (15%)
Former smokers (quit < 1 year ago)	15 (2%)	146 (2%)	14 (3%)	154 (2%)
Former smokers (quit > 1 year ago)	229 (31%)	1,634 (29%)	138 (28%)	1,907 (29%)
Never smoked tobacco	304 (41%)	3,047 (54%)	202 (42%)	3,638 (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 (%)				

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Cases	420 (57%)	740 (13%)	250 (51%)	706 (11%)
Controls	320 (43%)	4,948 (87%)	237 (49%)	5,999 (89%)
SSRI dispensing records			N=2,476	
Citalopram Escitalopram Fluoxetine Fluvoxamine maleate Paroxetine Sertraline	N/A	N/A	1,243 162 733 0 135 203	N/A

Table 1. Demographic characteristics for individuals with self-reported antidepressant use and SSRI prescribing data in EWAS, including lifestyle variables and MDD. The time interval indicated in the SSRI sample from NHS Scotland records represents dispensed SSRIs within 12 months prior to the blood draw date. SSRI dispensing records refer to the total number of dispensed SSRIs by any individuals within those 12 months; AD=antidepressant.

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, alcohol consumption, gestational age	Red cell distribution width; lung function; body height; self-reported educational attainment
cg05273171	*PNKD; TMBIM1*	2	-0.013	2.35×10^{-8}	0.017	-	PNKD: systolic blood pressure; body height; waist-hip ratio; leukocyte count; neutrophil and eosinophil count TMBIM1: cardiovascular disease; inflammatory bowel syndrome; height; Crohn's disease; related pathways: innate immune system
cg03864397	*CASP10*	2	-0.030	1.05×10^{-8}	0.007	-	Mean corpuscular haemoglobin; trans fatty acid levels; chronic lymphocytic leukaemia, rheumatoid arthritis
cg05186879	*MAPKAPK3; CISH*	3	-0.023	3.3×10^{-9}	0.002	-	MAPKAPK3: schizophrenia, self-reported educational attainment CISH: eosinophil count

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

cg16315329	*CISH; MAPKAPK3*	3	-0.039	2.52×10^{-8}	0.018	-	MAPKAPK3: schizophrenia, self-reported educational attainment CISH: eosinophil count
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; myeloperoxidase measurement; erythrocyte cadmium measurement
cg26277237	*KANK1*	9	0.024	2.87×10^{-8}	0.020	-	Pulmonary function decline; height; energy intake
cg20494891	*MYO1E*	15	-0.026	2.32×10^{-8}	0.017	-	Serum metabolite measurement, social communication impairment, BMI, coronary artery calcification
cg27589594	*SLC5A10*	17	-0.025	4.91×10^{-9}	0.004	Gestational age, smoking	1.5 anhydroglucitol measurement; schizophrenia; response to paliperidone; schizophrenia symptom severity measurement

Table 2. CpG sites significantly associated with self-reported antidepressant use (N=6,428; antidepressant use=740) along with gene annotations, chromosome, standardised effect size, nominal and multiple comparison-corrected p-values. Background information for each CpG site (CpG site information) and gene (Gene information) was extracted from EWAS (http://www.ewascatalog.org/; association between traits and CpGs on Illumina 450K array at $p \leq 1.0\text{x}10^{-4}$); and GWAS (https://www.ebi.ac.uk/gwas/; associations between traits and SNPs at $p < 1.0\text{x}10^{-5}$) catalogue databases.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

CpG site (chr)	MDD SNPs (chr); β, p	Schizophrenia SNPs (chr); β, p	Cardiovascular disease SNPs (chr); β, p	Rheumatoid arthritis SNPs (chr); β, p
cg05603985 (1)				
cg05273171 (2)				rs13330176 (16); β=0.078, p=0.0009
cg03864397 (2)	rs2056445 (13); β= -0.075, p=0.0009			rs6715284 (2); β= -0.176, p=3.39x10^{-7}
cg05186879 (3)				rs6732565 (2); β= 0.086, p=2.31x10^{-5}
cg16315329 (3)			rs3184504 (12); β= 0.108, p=6.43x10^{-8}	rs10774624 (12); β=0.102, p=4.89x10^{-7}
cg25753411 (6)		*rs8044995 (16); β= 0.230, p=5.2x10^{-17}		rs2075876 (21); β=-0.109, p=0.0009
cg09511513 (8)				
cg26277237 (9)				rs10175798 (2); β= -0.075, p=0.0002
cg20494891 (15)				rs6732565 (2); β= 0.073, p=0.0003
cg27589594 (17)	rs3793577 (9); β=-0.076, p=0.0001	rs7523273 (1); β= 0.086, p=6.62x10^{-5}		

Table 3. Disease-associated SNPs found to be mQTLs for CpGs associated with self-reported antidepressant disease here. Chromosome number for each CpG site and each SNP is indicated in brackets. β and p-values represent the association between each SNP and each CpG reported. Inflammatory bowel syndrome SNPs were not found to be mQTLs for any of the 10 CpGs identified here and is therefore not represented in the table above.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

	MS			**MS-ns**			**MS-control**		
Outcome variable (N)	**β**	**p-value**	**R^2**	**β**	**p-value**	**R^2**	**β**	**p-value**	**R^2**
AD use (3,360)	0.377	3.12×10^{-11}	2.12%	0.106	0.075	0.16%	0.254	1.15×10^{-5}	0.93%
AD use* (3,009)	0.213	0.0035	0.56%	0.04	0.565	0.02%	0.064	0.381	0.05%
MDD (3,360)	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 (3,322)	0.632	$< 2 \times 10^{-16}$	6.11%	0.077	0.028	0.11%	0.567	$< 2 \times 10^{-16}$	5.08%
Pack years (3,310)	0.305	$< 2 \times 10^{-16}$	9.14%	0.061	0.0004	0.34%	0.287	$< 2 \times 10^{-16}$	8.09%

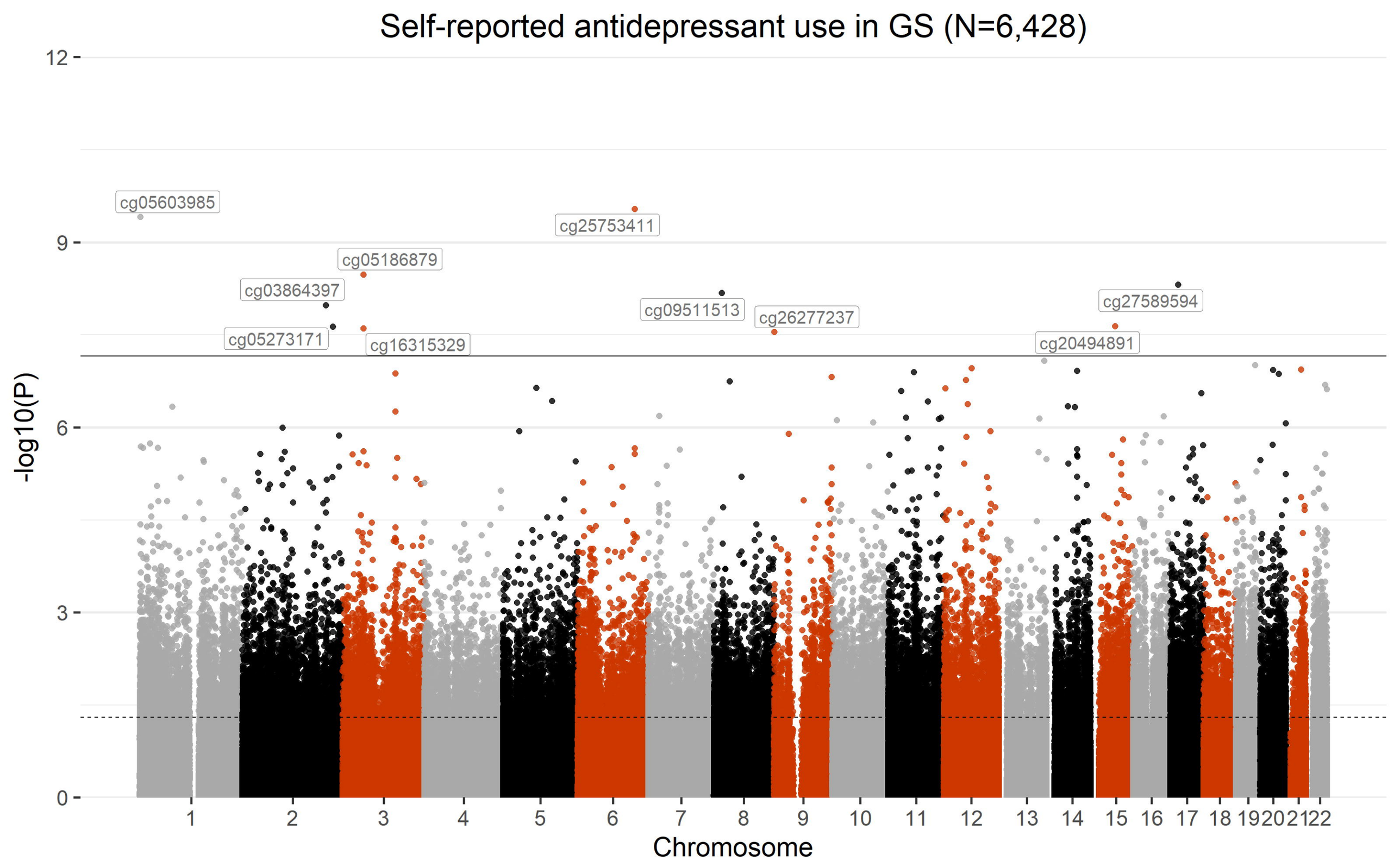
Table 4. Association between MS, MS-ns, and MS-control, and self-reported antidepressant use, MDD, and 4 lifestyle factors (BMI, smoking status, pack years, alcohol units). All regression models include age, sex, and 10 genetic principal components as covariates, except for the row marked *, which also includes BMI, smoking status, pack years, alcohol units, and MDD as covariates. Effect sizes represent standardized betas. MS=Methylation score, MS-ns=Methylation score trained on non-smokers, MS-control=Methylation score trained on individuals with no MDD diagnosis. R^2 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.

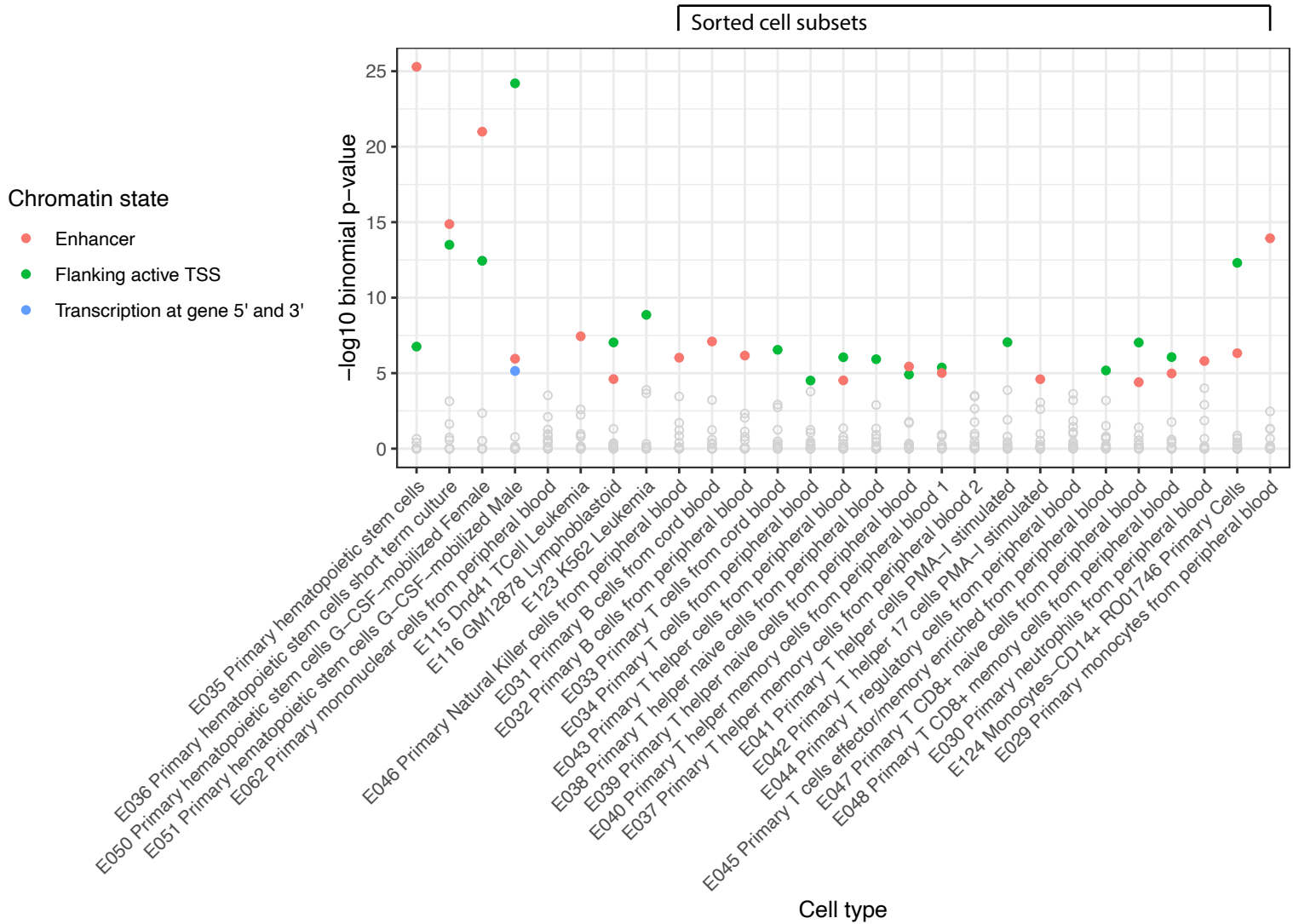
medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 1. Manhattan plot showing EWAS of self-reported antidepressant use in GS:SFHS. The black line defines the threshold for epigenome-wide significance ($p \leq 7.007 \times 10^{-8}$) and the dotted line defines the threshold for nominal significance ($p \leq 0.05$). Epigenome-wide significant hits are labelled on the graph.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.06.20207621; this version posted October 6, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license .

Figure 2. Differentially methylated CpG sites at $p<1x10^{-5}$ enriched for peripheral blood cell subsets. The x-axis represents peripheral blood cell types, while the y-axis indicates corrected p-value.


Self-reported antidepressant use in GS (N=6,428)
-log10(P)
12
9
6
3
0
cg05603985
cg05186879
cg03864397
cg05273171
cg16315329
cg25753411
cg09511513
cg26277237
cg27589594
cg20494891
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
Chromosome


Sorted cell subsets
Chromatin state
Enhancer
Flanking active TSS
Transcription at gene 5' and 3'
−log10 binomial p−value
25
20
15
10
5
0
E035 Primary hematopoietic stem cells
E036 Primary hematopoietic stem cells short term culture
E050 Primary hematopoietic stem cells G−CSF−mobilized Female
E051 Primary hematopoietic stem cells G−CSF−mobilized Male
E062 Primary mononuclear cells from peripheral blood
E115 Dnd41 TCell Leukemia
E116 GM12878 Lymphoblastoid
E123 K562 Leukemia
E046 Primary Natural Killer cells from peripheral blood
E031 Primary B cells from cord blood
E032 Primary B cells from peripheral blood
E033 Primary T cells from cord blood
E034 Primary T cells from peripheral blood
E043 Primary T helper cells from peripheral blood
E038 Primary T helper naive cells from peripheral blood
E039 Primary T helper memory cells from peripheral blood
E040 Primary T helper memory cells from peripheral blood
E037 Primary T helper cells PMA−I stimulated
E041 Primary T helper 17 cells PMA−I stimulated
E042 Primary T regulatory cells from peripheral blood
E044 Primary T cells effector/memory enriched from peripheral blood
E045 Primary T CD8+ naive cells from peripheral blood
E047 Primary T CD8+ memory cells from peripheral blood
E048 Primary neutrophils from peripheral blood
E030 Primary monocytes−CD14+ RO01746 Primary Cells
E124 Monocytes−CD14+ from peripheral blood
E029 Primary monocytes from peripheral blood
Cell type