Title: The effects of CYP2C19 metaboliser status on DNA methylation

Chen Shen[1,2], Mark Adams[3], Eleanor Davyson[3], Matthew Iveson[3], Andrew M McIntosh[3], Xueyi Shen[3]

Addresses

[1]MRC Centre for Environment and Health, Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, W2 1PG, United Kingdom
[2]National Institute for Health Research Health Protection Research Units in Environmental Exposures and Health & Chemical and Radiation Threats and Hazards, in partnership with UK Health Security Agency, Imperial College London, W2 1PG, United Kingdom
[3]Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, United Kingdom

Corresponding author

Xueyi Shen
Division of Psychiatry, University of Edinburgh, Royal Edinburgh Hospital, Morningside Park, Edinburgh EH10 5HF, UK
Email: xueyi.shen@ed.ac.uk

Abstract

Background

CYP2C19 enzyme is crucial in the metabolism of drugs and wider environmental exposures. CYP2C19 metaboliser status is determined by variation within the *CYP2C19* gene and has been linked to the efficacy of drug treatment and side effects. Studies of the effects of CYP2C19 metaboliser status on DNA methylation could reveal mechanistic pathways underlying the effect of CYP2C19 metaboliser status on drug response and health consequences.

Methods

A methylome-wide association study was conducted in the Generation Scotland cohort of 18,413 individuals to investigate the association between CYP2C19 metaboliser status and genome-wide DNA methylation, including its interaction with CYP2C19-metabolised medication use. Pathway enrichment analysis was conducted for the cytosine-guanine dinucleotide (CpG) probes significantly associated with CYP2C19 metaboliser status. We then conducted a phenome-wide association analysis (PheWAS) in the UK Biobank (n=2,408 to 370,419) to examine the effects of CYP2C19 metaboliser status on behavioural, cognitive and neuroimaging traits.

Results

48 CpG probes were significantly associated with CYP2C19 metaboliser status ($P_{Bonferroni} < 0.05$), annotated to genes involving drug metabolism, liver function, and lipid profile. There was no interaction with CYP2C19-metabolised medication use. Pathway enrichment analysis showed enrichment in biological processes involving metabolic activities and the P450 pathway. PheWAS showed CYP2C19 metaboliser status associated with the structure of brain regions in relation to physical conditions (such as pain) that are comorbid with depression.

Conclusions

This research suggests that CYP2C19 metaboliser status may impact DNA methylation and metabolic pathways that do not depend on the prescription of drugs metabolised by this enzyme. DNA methylation may provide an alternative means of assessing CYP2C19 activity in future studies.

Introduction

The activity of many drug-metabolising enzymes influences an individual's response and side effects to drug treatment. Therefore, drug metabolising status is likely to have an important role in precision medicine by determining the most effective and safest prescription for an individual. The cytochromes P450 (CYP) superfamily comprises multiple genes that code for enzymes involving the metabolism of various exogenous and endogenous compounds such as drugs, environmental pollutants, cholesterol, steroids, and Vitamin D[1]. CYP family 2 subfamily C, polypeptide 19 (CYP2C19) enzyme, encoded by *CYP2C19* gene on chromosome 10 (10q24)[2], plays a key role in the metabolism of a wide range of clinically relevant drugs such as antidepressants, proton pump inhibitors (PPI), and antiplatelet medicine[3-5].

Some *CYP2C19* variant alleles (e.g., *CYP2C19*x2, *CYP2C19*x3) are associated with reduced or loss of CYP2C19-mediated drug metabolising activity[6]. Carriers of these variants have lower active metabolites converted by these drugs, resulting in reduced treatment and/or increased side effects. For instance, evidence has indicated an association between low activity *CYP2C19* genetic variants and intolerance of citalopram, resulting in adverse impacts on the treatment of depression[7,8]. In contrast, carriers of other *CYP2C19* variant alleles (e.g., *CYP2C19*x17) have increased drug metabolism capacity, which may result in a lack of response to standard antidepressant and PPI treatments due to an unusually rapid clearance of drugs[8,9].

Apart from the well-established link between CYP2C19 enzyme and drug metabolism, *CYP2C19* polymorphism is also related to the metabolism of wider environmental exposures such as toxicants (e.g., heavy metals, organic pollutants) and nutrients (e.g.,omega-3 fatty acids), which also have health impact[10]. It has also been shown that individuals with high CYP2C19 enzymatic capacity (i.e., rapid/ultrarapid metaboliser) had more depressive symptoms regardless of antidepressant treatment[11,12], but the biological mechanism is still unknown. Epigenetic change can act as an environmental archive of exogenous exposures (including medication use) by reflecting duration, intensity, and individual susceptibility to exposure[13]. DNA methylation at cytosine-phosphate-guanine (CpG) sites is a commonly investigated epigenetic process that regulates gene expression, which is involved in many physiological and pathological processes[14]. Previous methylome-wide association studies (MWAS) have found DNA methylation at specific CpG sites in relation to major depressive disorder and antidepressant use[15,16]. However, to our knowledge, no studies have investigated the association between metaboliser status determined by *CYP2C19* variants and DNA methylation. Such a study may shed important light on the mechanistic pathway underlying the effect of CYP2C19 metaboliser status on drug response and health consequences.

The present study investigated the association between *CYP2C19-determined* metaboliser status and genome-wide DNA methylation from a large-scale cohort, Generation Scotland (GS), including its interaction with current or historical use of *CYP2C19*-metabolised drugs. We also conducted a phenome-wide association analysis (PheWAS) from the UK Biobank examining the association between CYP2C19 metaboliser status and behavioural, cognitive and neuroimaging traits.

Methods

Participants

A total of 18,413 participants with genetic and DNA methylation in Generation Scotland (GS) were included in the study (41.2% male, mean age = 47.5 years, standard deviation (SD) of age = 14.9 years)[17]. GS is family-based, population study mainly consists of people with European ancestry from across Scotland[18]. Written consent was obtained from all participants. Ethical approval for this study was obtained from the NHS Tayside Research Ethics Committee (05/s1401/89).

Genetic data

GS

Genotyping was conducted on blood samples collected at baseline using IlluminaHumanOmniExpressExome-8v1.0 BeadChip (48.8%) or Illumina HumanOmniExpressExome-8 v1.2 BeadChip (51.2%)[19]. Quality check is consistent with previous publications[20], which includes sex mismatch detection, Hardy-Weinberg equilibrium P value < $1x10^{-6}$, and MAF ≤ 1%. Imputation in GS was conducted using the Sanger Imputation Server (URL: https://www.sanger.ac.uk/tool/sanger-imputation-service/) with the HRC v1.1 as the reference sample.[21]

CYP2C19 metaboliser status

CYP2C19 metaboliser status was derived using the PGxPOP pipeline to assign likely metaboliser status based on diplotypes[22]. Five levels were included: poor metaboliser, intermediate metaboliser, normal metaboliser, rapid metaboliser, and ultrarapid metaboliser. The *CYP2C19* gene is highly polymorphic with 38 variant alleles that encode CYP2C19 enzyme with no function (e.g., *CYP2C19**2), decreased function (e.g., *CYP2C19**9), normal function (e.g., *CYP2C19**1), or increased function (e.g., *CYP2C19**17). We ran PGxPOP on phased, imputed genotypes for chromosome 10 with hg19 coordinates and extracted phenotypes for the *CYP2C19* gene. According to the diplotype to phenotype definitions, participants with two no-function alleles were categorised as "poor metabolisers" (e.g., *CYP2C19**2/*2). Participants with two decreased function alleles or one normal/decreased function allele and one no-function allele were categorised as "intermediate metabolisers" (e.g., *CYP2C19**9/*9 or *CYP2C19**1/*2). Participants with two normal function alleles or one decreased function and one increased function alleles were categorised as "normal metabolisers" (e.g., *CYP2C19**1/*1 or CYP2C19*17/*9). Participants with one increase function allele and one no-function allele were categorised as "rapid metabolisers" (e.g., *CYP2C19**1/*2). Participants with one normal and one increased function or two increased function alleles were categorised as "ultrarapid metabolisers" (e.g., CYP2C19*1/*17 or *CYP2C19**17/*17).

DNA methylation data

DNA methylation data was extracted from blood stored at the baseline[17]. The quality check protocol used for a previous release of half of the sample has been described in detail elsewhere[17]. The protocol was applied to the entire GS. In short, CpG sites were removed if they had an outlying beta value (>+/-3 SD, P < 0.001), detection p value <0.005, and low or outlying bead counts (< 3 or >5% of the entire sample). Participants with sex mismatch with self-reported data and an outlier detection P < 0.01 for more than 5% of all CpG sites were removed from the analysis. Cross-hybridising and polymorphic CpG sites mapping to common SNPs (minor allele frequency (MAF) > 0.05) were removed from analysis[23]. M-values were used for further analysis[18]. To account for relatedness, M-values were residualised against the genomic relationship matrix created using GCTA[24]. The residualised M-values were then carried into MWAS[25].

MWAS statistical model

MWAS was conducted using the Omic-data-based complex trait analysis (OSCA) software (version 2.0)[26]. Results from the linear regression model were presented as the main findings. The mixed-linear-model-based method (MOA) was used as a secondary method, accounting for collinearity between CpG sites. For the linear regression model, residualised M-values were set as the independent variable and CYP2C19

metaboliser status as the outcome variable. Age, sex, genotyping array, first ten genetic principal components (PCs), smoking status (current/past/non-smoker), pack-years, first 20 methylation PCs (derived from M-values residualised against genomic relationship matrix, age, sex, and assessment centre), DNA methylation batch, DNA methylation-estimated cell proportions (CD8+T, CD4+T, natural killer cells, B cells and granulocytes). Bonferroni correction was applied across the entire methylome, with a P value 6.64 $x10^{-8}$ (0.05/752,741) set as the significance threshold.

Pathway enrichment analysis

We used the 'gometh' function from the 'missmethyl' R package for Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG)[27]. Gometh analysis accounts for the number of CpG sites per gene to balance the statistical bias introduced by differing number of probes per gene present on the array and CpGs that are annotated to multiple genes[28]. Significant CpG sites found in the linear model MWAS were used as the target list and the entire EPIC array as the background list. GO terms and KEGG pathway enrichment analyses were conducted separately. Default settings were used for the analysis.

Differentially methylated region (DMR) analysis

We conducted a DMR analysis using the 'dmrff' package in R[29]. DNA methylation data from a randomly selected list of participants from GS (N=1,000) was used as the reference panel. Significant DMRs were identified based on the following criteria: (1) the distance between two nearby probes within a DMR was at most 500 base pairs; (2) false discovery rate (FDR)-adjusted P value < 0.05, and (3) the same direction of MWAS effect estimates for the individual probes within a DMR.

Sensitivity analysis

Interaction effect of CYP2C19 metaboliser status and CYP2C19-metabolised medication use on DNA methylation

We examined the interaction effect between CYP2C19 metaboliser status and CYP2C19-metabolised medication use on DNA methylation to investigate whether the effect of CYP2C19 metaboliser status on DNA methylation is driven by medication use. A list of drug names was extracted from PharmGKB (https://www.pharmgkb.org/) (Table S1). Drugs with clinical implications, including efficacy, toxicity, and metabolism determined by *CYP2C19* variants, were selected. This list was used to flag records in community-dispensed prescription data (Prescribing Information System; https://www.ndc.scot.nhs.uk/National-Datasets/data.asp?SubID=9) covering 2009 to 2021, regardless of formulation, dose or prescribing duration. Any participant with a flagged record were treated as users of CYP2C19-metabolised drugs (n=11,721), and those without a flagged record were treated as controls (n=6675). We extracted M-values of significant CpG sites derived from MWAS as the dependent variables and checked the significance of the interaction term of CYP2C19 metaboliser status and CYP2C19-metabolised medication use (yes vs. no) in linear regression models, adjusting for age, sex, genotyping array, first ten genetic PCs, smoking status, pack-years, first 20 methylation PCs, DNA methylation batch, and DNA methylation-estimated cell proportions as covariates.

Conditional analysis

Significant CpG associations may be explained solely from the *cis*-mQTL of the *CYP2C19* gene. To investigate whether there is any secondary signal, we conducted another MWAS, adding the M-value of the top CpG site annotated to *CYP2C19* as an additional covariate.

PheWAS in the UK Biobank (UKB)

The UKB is a large prospective study (n=502,492) of participants recruited between the ages of 40 and 69[30]. Baseline data were collected between 2006-2010, including genotyping and extensive phenotyping for a range of health outcomes. Blood samples from 488,363 participants were genotyped using the UK BiLEVE array (n=49,949) and UKB Axiom array (n=438,417). The quality control procedures are largely consistent with the GS protocol, i.e., participants with non-European-ancestry and/or with mismatching self-reported and genome-inferred sex information were removed from further analysis. Details of genotyping and quality

control are described elsewhere[30]. Imputed data release version 3 was used in the present study. The analysis and data acquisition for the present study were conducted under application #4844. Ethical approval was obtained by the North West National Health Service Research Ethics Committee (11/NW/0382). Written consent was acquired from all participants.

The PheWAS was conducted in the UK Biobank (n = 2,408 to 370,419). A total of 1,382 phenotypes under 19 categories were included in the analysis. Behavioural and blood biomarker categories were mental health, sociodemographic variables, early life factors, physical health, physical measure, physical activity, diet, lifestyle, cognition and blood biomarkers. Categories for MRI measures were: body MRI, bulk brain-tissues, task fMRI, white matter microstructure, subcortical measures, subcortical sub-regions, pial measures, grey-white matter contrast and cortical measures. A detailed description of the phenotype chosen for the present analysis can be found elsewhere[17]. General linear model was used to test linear association between the CYP2C19 metaboliser status and the phenome using the 'glm' function in R. CYP2C19 metaboliser status was the independent variable, and other phenotypes were set as the outcome. Age, sex, assessment centre, first ten genetic PCs and genotyping array were controlled for. For brain imaging phenotypes, scanner position on the X, Y and Z axis and assessment centre of imaging assessment were included as additional covariates. For cortical measures, in particular, Freesurfer-estimated intracranial volume was controlled for in the linear regression model. The models and covariates were kept consistent with a previous PheWAS study[31]. Detailed models for each phenotype category can be found elsewhere[31].

Results

MWAS

The linear model MWAS showed that 48 CpG probes were significantly associated with CYP2C19 metaboliser status with the Bonferroni threshold ($P < 6.64 \times 10^{-8}$). The Manhattan plot and QQ plot are shown in Figure 1 and Figure S1, respectively. The most significant CpG sites associated with CYP2C19 metaboliser status were found annotated to *CYP2C19, NOC3L*, *PDLIM1*, *TBC1D12*, and *PLCE1* genes on Chromosome 10 (Table S2). Specifically, faster metaboliser status was associated with two hypermethylated CpG sites (cg02808805 and cg00051662) and one hypomethylated CpG site (cg20031717), all annotated to the *CYP2C19* gene. The MOA MWAS showed similar results to the linear model (Table S3). All the 48 CpG sites detected in the linear model MWAS remained significant.

Pathway enrichment analysis

Analysis of the GO terms showed enrichment in biological processes involving metabolic activities and the P450 pathway. There were 181 enriched GO terms with a P value < 0.05, but none was significant after FDR correction. The top ten GO terms are listed in Table 1. The test of KEGG pathways showed that pathways such as chemical carcinogenesis and metabolic pathways, including drug metabolism, were enriched. There were 11 nominally significant KEGG pathways with a P value < 0.05, but none was significant after FDR correction (Table 2).

DMR analysis

A total of 5 DMRs annotated to *PDLIM1* (97051104 - 97051225; 97051255 - 97051319), *NOC3L* (96121776 - 96121853), *CYP2C18* (96442621 - 96443071), and an open-sea area (96928199 - 96928657) were identified as associated with CYP2C19 metaboliser status at FDR-adjusted P value < 0.05 (Table 3). Faster metaboliser status was associated with two hypomethylated DMRs annotated to the *PDLIM1* gene, one hypermethylated DMR annotated to the *NOC3L* gene, and one hypomethylated DMR annotated to the *CYP2C18* gene. No significant DMR was identified in the *CYP2C19* gene.

Sensitivity analysis

No interaction effect was detected between CYP2C19 metaboliser status and CYP2C19-metabolised medication use on M-values of significant CpG sites (P values for interaction term: all > 0.05), indicating that the association between CYP2C19 metaboliser status and DNA methylation did not vary by CYP2C19-metabolised medication use.

When additionally adjusting for the M-value of the CpG that was most robustly associated with CYP2C19 metaboliser status (cg20031717), the associations between other significant CpG probes (e.g., cg02808805, cg00051662) annotated to the *CYP2C19* gene and CYP2C19 metaboliser status were attenuated but still significant (Table S4), indicating that associations between CYP2C19 metaboliser status and DNA methylation were not explained solely from mQTL of the *CYP2C19* gene. The associations between significant CpG probes (e.g., cg07889765, cg11776334) annotated to other genes and CYP2C19 metaboliser status remained similar.

PheWAS shows that a total of 30 traits were associated with CYP2C19 metaboliser status after FDR correction. Among these significant traits, 16 were brain cortical measures with brain regions defined by the Destrieux atlas[32]. Within these brain cortical structural measures, 12 were cortical surface area, and the remaining four were cortical volumes. The most significant MRI measure was found in the surface area of the anterior transverse temporal gyrus ($\beta = -0.024$, $P = 1.45 \times 10^{-5}$, Figure 3). Six blood biomarkers were associated with CYP2C19 metaboliser status (P ranged from 2.48×10^{-4} to 3.82×10^{-6}). The strongest association of biomarkers was found with higher total bilirubin ($\beta=0.008$, $P =3.82 \times 10^{-6}$). A complete list of significant associations is shown in Table S5.

Table 1 Results for gene ontology (GO) analysis for the MWAS on CYP2C19 metaboliser status at P threshold 6.64 x10^{-8}. BP: Biological process; MF: molecular function; DE: number of genes that are differentially methylated.

GO ID	Ontology	Term	N	DE	P_{DE}	P_{FDR}
GO:0008392	MF	Arachidonic acid epoxygenase activity	17	2	1.98x10^{-5}	0.249536
GO:0019373	BP	Epoxygenase P450 pathway	19	2	2.90x10^{-5}	0.249536
GO:0008391	MF	Arachidonic acid monooxygenase activity	21	2	3.26x10^{-5}	0.249536
GO:0070330	MF	Aromatase activity	25	2	5.21x10^{-5}	0.298608
GO:0008395	MF	Steroid hydroxylase activity	38	2	0.000129	0.451396
GO:0016712	MF	Oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, reduced flavin or flavoprotein as one donor, and incorporation of one atom of oxygen	40	2	0.000137	0.451396
GO:0019825	MF	Oxygen binding	38	2	0.000138	0.451396
GO:0019369	BP	Arachidonic acid metabolic process	58	2	0.000342	0.980497
GO:0098632	MF	Cell-cell adhesion mediator activity	54	2	0.000415	1
GO:0098631	MF	Cell adhesion mediator activity	64	2	0.000566	1

The top ten GO terms are listed (out of 181 GO terms with $P_{DE} < 0.05$)

Table 2 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis for the MWAS on CYP2C19 metaboliser status at P threshold 6.64 x10^{-8}. DE: number of genes that are differentially methylated.

KEGG pathway ID	Description	N	DE	P_{DE}	P_{FDR}
hsa05204	Chemical carcinogenesis - DNA adducts	68	2	0.0002	0.070515
hsa04726	Serotonergic synapse	113	2	0.000523	0.092375
hsa01100	Metabolic pathways	1501	3	0.005525	0.650054
hsa00591	Linoleic acid metabolism	30	1	0.01018	0.898369
hsa00562	Inositol phosphate metabolism	73	1	0.018692	0.98415
hsa00590	Arachidonic acid metabolism	61	1	0.019488	0.98415
hsa00982	Drug metabolism - cytochrome P450	70	1	0.022019	0.98415
hsa00830	Retinol metabolism	67	1	0.022304	0.98415
hsa04070	Phosphatidylinositol signaling system	97	1	0.027180	1
hsa04933	AGE-RAGE signaling pathway in diabetic complications	100	1	0.032271	1
hsa04919	Thyroid hormone signaling pathway	121	1	0.036322	1

KEGG pathways with $P_{DE} < 0.05$ are listed

Table 3 Significant DMRs identified from the MWAS analysis.

CHR	Start	End	Gene	Effect	SE	Adjusted P values	CpGs
10	97051104	97051225	*PDLIM1*	-1.2338614	0.14066702	1.39×10^{-12}	cg11911874, cg03178678
10	97051255	97051319	*PDLIM1*	-0.8881719	0.13296932	1.89×10^{-05}	cg06542614, cg05599883
10	96121776	96121853	*NOC3L*	0.26081666	0.03640475	6.16×10^{-07}	cg15209556, cg17162216, cg10164249
10	96442621	96443071	*CYP2C18*	-0.6596835	0.11595304	0.010056	cg17014018, cg04501839, cg19852607
10	96928199	96928657	Open sea	1.22355361	0.18163946	1.28×10^{-05}	cg25102879, cg24087710

SE: standard error

Figure 1 Manhattan plot for the linear model MWAS of CYP2C19 metaboliser status.

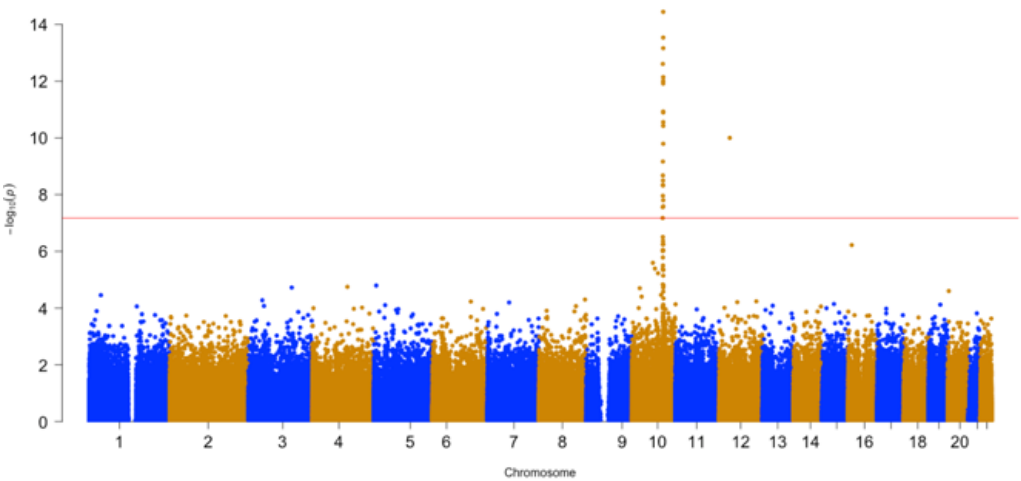


Figure 2 P-plot for phenome-wide association analysis. Each dot represents a trait. Categories are labelled on the x-axis. Y-axis represents -log20-trasformed p-values. The dotted pink line represents FDR significance threshold. Names for the significant phenotypes (FDR corrected) are annotated in the plot.

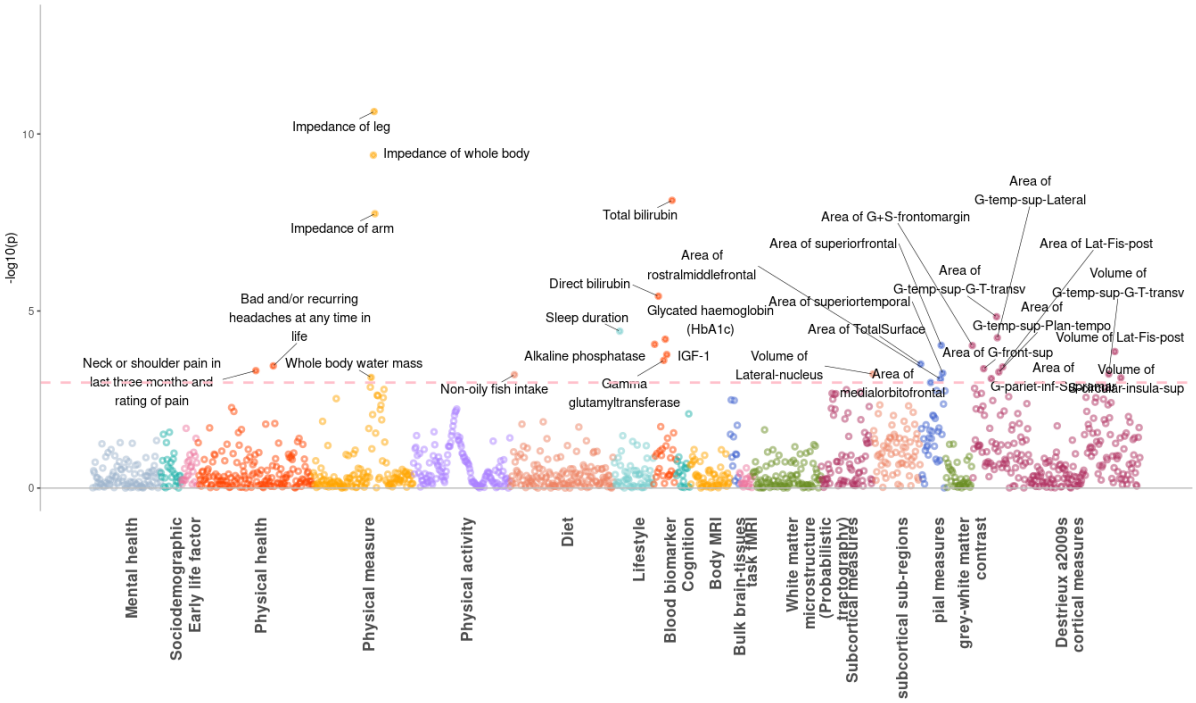
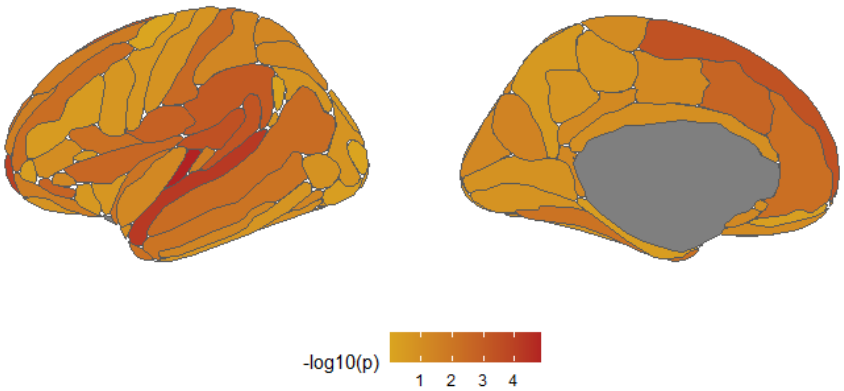


Figure 3 Results for association analysis between CYP2C19 metaboliser status in cortical surface area. Cortical regions were defined by the Destrieux Atlas.



Discussion

We analysed the association between CYP2C19 metaboliser status and DNA methylation in a large population health dataset. We found that CYP2C19 metaboliser status was associated with DNA methylation at 48 CpG sites. Significant CpGs were enriched in metabolic pathways involving P450 epoxygenase and drug metabolism. No interactive effect was observed between CYP2C19-metabolised medication use and CYP2C19 metaboliser status on DNA methylation. Additionally, we found significant associations between CYP2C19 metaboliser status and brain imaging traits in UK Biobank.

cg20031717 annotated to the *CYP2C19* gene was most robustly associated with CYP2C19 metaboliser status. Other significant CpG sites annotated to the *CYP2C19* gene were cg02808805 and cg00051662, with hypermethylation within these two CpG sites found in faster metaboliser status. DNA methylation at both cg20031717 and cg00051662 has been reported to decrease with age but is not associated with any other traits[33]. Hypermethylation within cg02808805 has been found associated with the prevalence of prostate cancer[34]. This is corroborated by the KEGG analysis indicating enriched pathways involving chemical carcinogenesis.

Apart from the *CYP2C19* gene, significant CpG sites are also annotated to other genes on Chromosome 10, such as *NOC3L*, *PDLIM1*, *TBC1D12*, and *PLCE1*. Significant DMRs were annotated to *NOC3L*, *PDLIM1, and CYP2C18. NOC3L* regulates DNA replication, affecting the serum concentration of liver enzymes, aspartate aminotransferase and gamma-glutamyl transpeptidase, implicating metabolism and liver pathology[35,36]. The liver is one of the organs which maximally contributes to the overall metabolism and elimination of drugs.[37] *NOC3L* is also associated with low-density lipoprotein (LDL) cholesterol levels in the blood.[38] Previous research highlights the important role CYP enzymes play in the metabolism of cholesterols[39]. This explains that lipid profile, including LDL subfractions, predicts response to antidepressants, as found by another study.[40] *PDLIM1* encodes a cytoplasmic protein with PDZ and LIM domains[41]. This gene has been associated with metabolite levels of warfarin (an anticoagulant drug) and steroid hormone levels (biosynthesis promoted by CYP)[42,43]. *CYP2C18* is another CYP superfamily member associated with warfarin metabolite levels and clopidogrel and serum metabolites[42,44,45].

Our sensitivity analysis did not detect the interaction effect between CYP2C19-related medication use and CYP2C19 metaboliser status on DNA methylation. This indicates that metabolites of these drugs and general metabolic status determined by *CYP2C19* variants may affect DNA methylation independently. This is partly supported by a previous MWAS meta-analysis in European ancestries (GS and Netherlands Twin Register) on antidepressant use[16]. The two CpG signals, cg05603985 (annotated to the *SKI* gene on Chromosome 1) and cg27589594 (annotated to the *SLC5A10* gene on Chromosome 17) associated with antidepressant use are not overlapped with any significant CpG sites or DMRs in our study, indicating that CYP2C19 metaboliser status and antidepressant use may not share biological pathways. However, this has not formally been tested with other CYP2C19-metabolised drugs, e.g., warfarin, clopidogrel, to our knowledge.

The associations between CYP2C19 metaboliser status and brain imaging traits derived from PheWAS in UK Biobank indicate that the structure of brain regions such as insula, medial frontal gyrus, and temporal lobe may be affected by CYP2C19 metaboliser status. These regions are associated with persistent physical conditions that are comorbid with depression, such as chronic pain. This may reflect the physical conditions that may be alleviated by CYP2C19-related medications.[46]

Strengths of our study include that GS is the largest single sample MWAS to date, which allows adequate power to detect individual CpG sites and DMRs associated with CYP2C19 metaboliser status. Additionally, as CYP2C19-determined metaboliser status is genetically determined, reverse causality is not a concern. However, our study has some limitations. First, our analyses only included participants predominantly with European ancestry, so our findings may not be generalisable to individuals of other ancestries. Second, DNA methylation from the whole blood may not be representative of different phenotypes associated with

CYP2C19 metaboliser status, such as brain structure. Future studies should examine DNA methylation in the brain tissue in post-mortem samples to examine the inter-relationship between CYP2C19 metaboliser status, brain structure, and DNA methylation. Third, we only considered the current or historical use of CYP2C19-metabolised medications as a binary variable when examining its interaction with CYP2C19 metaboliser status affecting DNA methylation. More detailed information on medication use (e.g., duration, side effect) would enable a better understanding of the biological effect of CYP2C19 metaboliser status together with CYP2C19-metabolised medication use on DNA methylation. Finally, we did not replicate our analyses in a validation sample, which is required in future research to confirm our findings.

In conclusion, the MWAS of CYP2C19 metaboliser status in a large-scale cohort suggests that CYP2C19-determined metaboliser status may have biological implications relevant to the response of certain drugs, metabolites relevant to CYP, and liver diseases. The use of CYP2C19-metabolised medications did not interact with CYP2C19 metaboliser status to affect DNA methylation, indicating that CYP2C19 metaboliser status may impact DNA methylation and metabolic pathways that do not depend on the use of medications metabolised by this enzyme. Future studies should confirm our findings in other populations and investigate how CYP2C19 metaboliser status interacts with detailed medication records on DNA methylation and health consequences.

Acknowledgements
This work is supported by two Wellcome Trust grants to AMM (220857/Z/20/Z and 104036/Z/14/Z). 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. In addition, DNA methylation profiling was supported by funding from NARSAD (Ref 27404) and the Royal College of Physicians of Edinburgh (SIM Fellowship). Genotyping of the GS samples was funded by the MRC and Wellcome Trust [104036/Z/14/Z]. GS also receives support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006]. Data acquisition and analyses were conducted using the UK Biobank Resource under approved project #4844. This work has used the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/). ED was supported by the UK Research and Innovation (Grant No. EP/S02431X/1), UK Research and Innovation Centre for Doctoral Training in Biomedical AI at the University of Edinburgh, School of Informatics. We thank the participants and team members for their ongoing contribution to the recruitment, data management and technical and legal support for GS. Participation of the UK Biobank subjects is gratefully appreciated.

References

1 Nebert, D. W. & Russell, D. W. Clinical importance of the cytochromes P450. *Lancet* **360**, 1155-1162 (2002). https://doi.org:10.1016/s0140-6736(02)11203-7

2 Shuldiner, A. R. *et al.* Association of cytochrome P450 2C19 genotype with the antiplatelet effect and clinical efficacy of clopidogrel therapy. *JAMA* **302**, 849-857 (2009). https://doi.org:10.1001/jama.2009.1232

3 Brøsen, K. Some aspects of genetic polymorphism in the biotransformation of antidepressants. *Therapie* **59**, 5-12 (2004). https://doi.org:10.2515/therapie:2004003

4 Li, X. Q., Andersson, T. B., Ahlström, M. & Weidolf, L. Comparison of inhibitory effects of the proton pump-inhibiting drugs omeprazole, esomeprazole, lansoprazole, pantoprazole, and rabeprazole on human cytochrome P450 activities. *Drug Metab Dispos* **32**, 821-827 (2004). https://doi.org:10.1124/dmd.32.8.821

5 Hulot, J. S. *et al.* Cytochrome P450 2C19 loss-of-function polymorphism is a major determinant of clopidogrel responsiveness in healthy subjects. *Blood* **108**, 2244-2247 (2006). https://doi.org:10.1182/blood-2006-04-013052

6 Scott, S. A. *et al.* Clinical Pharmacogenetics Implementation Consortium guidelines for cytochrome P450-2C19 (CYP2C19) genotype and clopidogrel therapy. *Clin Pharmacol Ther* **90**, 328-332 (2011). https://doi.org:10.1038/clpt.2011.132

7 Mrazek, D. A. *et al.* CYP2C19 variation and citalopram response. *Pharmacogenet Genomics* **21**, 1-9 (2011). https://doi.org:10.1097/fpc.0b013e328340bc5a

8 Campos, A. I. *et al.* Impact of CYP2C19 metaboliser status on SSRI response: a retrospective study of 9500 participants of the Australian Genetics of Depression Study. *Pharmacogenomics J* **22**, 130-135 (2022). https://doi.org:10.1038/s41397-022-00267-7

9 Sim, S. C. *et al.* A common novel CYP2C19 gene variant causes ultrarapid drug metabolism relevant for the drug response to proton pump inhibitors and antidepressants. *Clin Pharmacol Ther* **79**, 103-113 (2006). https://doi.org:10.1016/j.clpt.2005.10.002

10 Parajuli, R. P. *et al.* Variation in biomarker levels of metals, persistent organic pollutants, and omega-3 fatty acids in association with genetic polymorphisms among Inuit in Nunavik, Canada. *Environ Res* **200**, 111393 (2021). https://doi.org:10.1016/j.envres.2021.111393

11 Jukić, M. M. *et al.* Elevated CYP2C19 expression is associated with depressive symptoms and hippocampal homeostasis impairment. *Mol Psychiatry* **22**, 1155-1163 (2017). https://doi.org:10.1038/mp.2016.204

12 Sim, S. C. *et al.* Association between CYP2C19 polymorphism and depressive symptoms. *Am J Med Genet B Neuropsychiatr Genet* **153b**, 1160-1166 (2010). https://doi.org:10.1002/ajmg.b.31081

13 Vineis, P. *et al.* What is new in the exposome? *Environ Int* **143**, 105887 (2020). https://doi.org:10.1016/j.envint.2020.105887

14 Habano, W. *et al.* Analysis of DNA methylation landscape reveals the roles of DNA methylation in the regulation of drug metabolizing enzymes. *Clin Epigenetics* **7**, 105 (2015). https://doi.org:10.1186/s13148-015-0136-7

15 Clark, S. L. *et al.* A methylation study of long-term depression risk. *Mol Psychiatry* **25**, 1334-1343 (2020). https://doi.org:10.1038/s41380-019-0516-z

16 Barbu, M. C. *et al.* Methylome-wide association study of antidepressant use in Generation Scotland and the Netherlands Twin Register implicates the innate immune system. *Mol Psychiatry* **27**, 1647-1657 (2022). https://doi.org:10.1038/s41380-021-01412-7

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

18 Smith, B. H. *et al.* Cohort Profile: Generation Scotland: Scottish Family Health Study (GS:SFHS). The study, its participants and their potential for genetic research on health and illness. *Int J Epidemiol* **42**, 689-700 (2013). https://doi.org:10.1093/ije/dys084

19 Gunderson, K. L. Whole-genome genotyping on bead arrays. *DNA Microarrays for Biomedical Research: Methods and Protocols*, 197-213 (2009).
20 Hall, L. S. *et al.* Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. *Transl Psychiatry* **8**, 9 (2018). https://doi.org:10.1038/s41398-017-0034-1
21 Nagy, R. *et al.* Exploration of haplotype research consortium imputation for genome-wide association studies in 20,032 Generation Scotland participants. *Genome Med* **9**, 23 (2017). https://doi.org:10.1186/s13073-017-0414-4
22 Hicks, J. K. *et al.* Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for CYP2D6 and CYP2C19 Genotypes and Dosing of Selective Serotonin Reuptake Inhibitors. *Clin Pharmacol Ther* **98**, 127-134 (2015). https://doi.org:10.1002/cpt.147
23 Shen, X. *et al.* DNA methylome-wide association study of genetic risk for depression implicates antigen processing and immune responses. *Genome Med* **14**, 36 (2022). https://doi.org:10.1186/s13073-022-01039-5
24 Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet* **88**, 76-82 (2011). https://doi.org:10.1016/j.ajhg.2010.11.011
25 Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, 587 (2010). https://doi.org:10.1186/1471-2105-11-587
26 Zhang, F. *et al.* OSCA: a tool for omic-data-based complex trait analysis. *Genome Biol* **20**, 107 (2019). https://doi.org:10.1186/s13059-019-1718-z
27 Phipson, B., Maksimovic, J. & Oshlack, A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics* **32**, 286-288 (2016). https://doi.org:10.1093/bioinformatics/btv560
28 Maksimovic, J., Oshlack, A. & Phipson, B. Gene set enrichment analysis for genome-wide DNA methylation data. *Genome Biol* **22**, 173 (2021). https://doi.org:10.1186/s13059-021-02388-x
29 Suderman, M. *et al. dmrff: identifying differentially methylated regions efficiently with power and control* (Cold Spring Harbor Laboratory, 2018).
30 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203-209 (2018). https://doi.org:10.1038/s41586-018-0579-z
31 Shen, X. *et al.* A phenome-wide association and Mendelian Randomisation study of polygenic risk for depression in UK Biobank. *Nat Commun* **11**, 2301 (2020). https://doi.org:10.1038/s41467-020-16022-0
32 Destrieux, C., Fischl, B., Dale, A. & Halgren, E. Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage* **53**, 1-15 (2010). https://doi.org:10.1016/j.neuroimage.2010.06.010
33 Mulder, R. H. *et al.* Epigenome-wide change and variation in DNA methylation in childhood: trajectories from birth to late adolescence. *Hum Mol Genet* **30**, 119-134 (2021). https://doi.org:10.1093/hmg/ddaa280
34 Hillary, R. F. *et al.* Blood-based epigenome-wide analyses of 19 common disease states: A longitudinal, population-based linked cohort study of 18,413 Scottish individuals. *PLoS Med* **20**, e1004247 (2023). https://doi.org:10.1371/journal.pmed.1004247
35 Chen, V. L. *et al.* Genome-wide association study of serum liver enzymes implicates diverse metabolic and liver pathology. *Nat Commun* **12**, 816 (2021). https://doi.org:10.1038/s41467-020-20870-1
36 Sakaue, S. *et al.* A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* **53**, 1415-1424 (2021). https://doi.org:10.1038/s41588-021-00931-x
37 Almazroo, O. A., Miah, M. K. & Venkataramanan, R. Drug metabolism in the liver. *Clinics in liver disease* **21**, 1-20 (2017).
38 Graham, S. E. *et al.* The power of genetic diversity in genome-wide association studies of lipids. *Nature* **600**, 675-679 (2021). https://doi.org:10.1038/s41586-021-04064-3
39 Pikuleva, I. A. Cholesterol-metabolizing cytochromes P450: implications for cholesterol lowering. *Expert Opin Drug Metab Toxicol* **4**, 1403-1414 (2008). https://doi.org:10.1517/17425255.4.11.1403

40 Caspani, G. *et al.* Metabolomic signatures associated with depression and predictors of antidepressant response in humans: A CAN-BIND-1 report. *Commun Biol* **4**, 903 (2021). https://doi.org:10.1038/s42003-021-02421-6

41 Ono, R., Kaisho, T. & Tanaka, T. PDLIM1 inhibits NF-κB-mediated inflammatory signaling by sequestering the p65 subunit of NF-κB in the cytoplasm. *Sci Rep* **5**, 18327 (2015). https://doi.org:10.1038/srep18327

42 Asiimwe, I. G. *et al.* A genome-wide association study of plasma concentrations of warfarin enantiomers and metabolites in sub-Saharan black-African patients. *Front Pharmacol* **13**, 967082 (2022). https://doi.org:10.3389/fphar.2022.967082

43 Hysi, P. G. *et al.* Metabolome Genome-Wide Association Study Identifies 74 Novel Genomic Regions Influencing Plasma Metabolites Levels. *Metabolites* **12** (2022). https://doi.org:10.3390/metabo12010061

44 Backman, J. D. *et al.* Genome-wide analysis of clopidogrel active metabolite levels identifies novel variants that influence antiplatelet response. *Pharmacogenet Genomics* **27**, 159-163 (2017). https://doi.org:10.1097/fpc.0000000000000272

45 Krumsiek, J. *et al.* Mining the unknown: a systems approach to metabolite identification combining genetic and metabolic information. *PLoS Genet* **8**, e1003005 (2012). https://doi.org:10.1371/journal.pgen.1003005

46 Ferreira, G. E. *et al.* Efficacy, safety, and tolerability of antidepressants for pain in adults: overview of systematic reviews. *BMJ* **380**, e072415 (2023). https://doi.org:10.1136/bmj-2022-072415