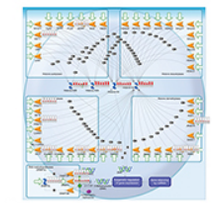
Epigenome-wide association study of global cortical volumes in Generation Scotland

Barbu, Miruna Carmen; Harris, Mat; Shen, Xueyi; Aleks, Stolicyn; Green, Claire; Amador, Carmen

DOI:
10.1080/15592294.2021.1997404

Publication date:
2021

epigenetics



Epigenetics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/kepi20

Epigenome-wide association study of global cortical volumes in Generation Scotland: Scottish Family Health Study

Miruna Carmen Barbu, Mat Harris, Xueyi Shen, Stolicyn Aleks, Claire Green, Carmen Amador, Rosie Walker, Stewart Morris, Mark Adams, Anca Sandu, Christopher McNeil, Gordon Waiter, Kathryn Evans, Archie Campbell, Joanna Wardlaw, Douglas Steele, Alison Murray, David Porteous, Andrew McIntosh & Heather Whalley

To cite this article: Miruna Carmen Barbu, Mat Harris, Xueyi Shen, Stolicyn Aleks, Claire Green, Carmen Amador, Rosie Walker, Stewart Morris, Mark Adams, Anca Sandu, Christopher McNeil, Gordon Waiter, Kathryn Evans, Archie Campbell, Joanna Wardlaw, Douglas Steele, Alison Murray, David Porteous, Andrew McIntosh & Heather Whalley (2021): Epigenome-wide association study of global cortical volumes in Generation Scotland: Scottish Family Health Study, Epigenetics, DOI: 10.1080/15592294.2021.1997404

To link to this article: https://doi.org/10.1080/15592294.2021.1997404

Publisher: Taylor & Francis & Informa UK Limited, trading as Taylor & Francis Group

Journal: *Epigenetics*

DOI: 10.1080/15592294.2021.1997404

Epigenome-wide association study of global cortical volumes in Generation Scotland: Scottish Family Health Study

Miruna Carmen Barbu, PhD[1], Mat Harris, PhD[1], Xueyi Shen, PhD[1], Stolicyn Aleks, PhD[1], Claire Green,MSc[1], Carmen Amador, PhD[2], Rosie Walker, PhD[3,4], Stewart Morris, BSc[3,4], Mark Adams, PhD[1], Anca Sandu, PhD[5], Christopher McNeil,PhD[5], Gordon Waiter,PhD[5], Kathryn Evans, PhD[3,4], Archie Campbell, MA[2], Joanna Wardlaw, MD[4], Douglas Steele, PhD[7], Alison Murray, MD[5], David Porteous, PhD[2,6], * Andrew McIntosh, MD[1,6], * Heather Whalley, PhD[1]

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

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

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

[4]Centre for Clinical Brain Sciences, The University of Edinburgh, UK

[5]Aberdeen Biomedical Imaging Centre, The Institute of Medical Sciences, University of Aberdeen, UK

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

[7]Imaging Science and Technology, School of Medicine, University of Dundee, 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 global cortical volumes Keywords: DNA

methylation, epigenome-wide association study, cortical volumes

Word count (abstract)	216
Word count (main text)	4,903
Figures	3
Tables	2
Supplementary Information	1

Abstract

A complex interplay of genetic and environmental risk factors influence global brain structural alterations associated with brain health and disease. Epigenome-wide association studies (EWAS) of global brain imaging phenotypes have the potential to reveal the mechanisms of brain health and disease and can lead to better predictive analytics through the development of risk scores.

We perform an EWAS of global brain volumes in Generation Scotland using peripherally measured whole blood DNA methylation (DNAm) from two assessments, (i) at baseline recruitment, ~6 years prior to MRI assessment (N=672) and (ii) concurrent with MRI assessment (N=565). Four CpGs at baseline were associated with global cerebral white matter, total grey matter, and whole-brain volume (Bonferroni $p \leq 7.41 \times 10^{-8}$, $\beta_{range} = -1.46 \times 10^{-6}$ to 9.59×10^{-7}). These CpGs were annotated to genes implicated in brain-related traits, including psychiatric disorders, development, and ageing. We did not find significant associations in the meta-analysis of the EWAS of the two sets concurrent with imaging at the corrected level.

These findings reveal global brain structural changes associated with DNAm measured ~6 years previously, indicating a potential role of early DNAm modifications in brain structure. Although concurrent DNAm was not associated with global brain structure, the nominally significant findings identified here present a rationale for future investigation of associations between DNA methylation and structural brain phenotypes in larger population-based samples.

Keywords: DNA methylation; epigenome-wide association study; cortical volumes; Generation Scotland

Introduction

Global brain structure is influenced by genetic and environmental factors, and has previously been associated with health and disorder traits across the lifetime (1–3). For instance, changes in global grey and white matter have been observed in a number of psychiatric and neurological disorders, including schizophrenia (4), major depressive disorder (MDD) (3), bipolar disorder (5), Rett syndrome (6), and Alzheimer's disease (7). Previous studies have also found age-related reductions in both grey and white matter (8,9).

Such global brain structural changes in both health and disease may reflect genetic and environmental factors and their impact. While previous studies have focussed on revealing the genetic architecture of brain structure, there are now opportunities to explore genetic and environmental risk factors through epigenetics, which correlate with changes in gene expression by modulating the genome in different cell types, without altering the underlying genome sequence (10). One such process, DNA methylation (DNAm), implicates the covalent addition of a methyl group to a cytosine nucleotide followed by guanine in DNA, resulting in Cytosine-phosphate-Guanine (CpG) sites (10).

DNAm is modulated by both genetic and environmental factors, and may thus aid in identifying genetic and environmental contributions to health and disease (11). Several brain-related traits and diseases are associated with variation in DNAm. MDD, a moderately heritable disorder, has been associated with differential methylation at several CpG sites, with a methylation risk score explaining 1.75% of the variance in the disorder (12). Further, in an epigenome-wide association study (EWAS) using blood, CpG sites associated with depressive symptoms were annotated to genes involved in axonal guidance (13). Schizophrenia has been associated with epigenetic variation at multiple loci that contribute to the polygenicity of the disorder (14,15). Finally, growing evidence has shown that DNAm can act as a proxy for the biological age of multiple tissues across life (16). These studies indicate that it may be possible, in future, to utilise DNAm modifications as biomarkers for brain-related healthy traits and diseases and to identify novel mechanisms contributing to these traits.

In recent years, increasing efforts have been made to identify epigenetic correlates of brain phenotypes, using both blood and brain tissue (17,18). To maximise statistical

power,

previous studies have focused on candidate genes and candidate epigenetic markers in relation to specific brain regions of interest, such as subcortical volumes in the hippocampus and amygdala, as well as cortical thickness and volume in Freesurfer-derived brain regions (18), although consistency between study findings is modest. Recent advances in high-throughput array technologies that can identify DNAm levels at over 450K and 850K locations along the genome have enabled researchers to identify DNAm-brain associations using a hypothesis-free approach using EWAS (19). DNAm modifications in relation to brain phenotypes have also been identified in patients as opposed to healthy individuals, including in the frontal cortex in schizophrenia (20,21), hippocampal volume in MDD (22), in the cerebral cortex in Alzheimer's disease (23), and in the frontal cortex in Parkinson's disease (24). Structural brain measures may therefore function as endophenotypes that can be used to assess the association between epigenetic modifications and brain health and disease.

The pathogenesis of psychiatric and neurodegenerative disorders has been associated with a multitude of cortical and subcortical brain regions with inconsistent results across studies (3,25–27), potentially indicating a role for whole-brain abnormalities in these disorders. Peripheral DNAm alterations associated with clinically relevant global brain structure may therefore further our mechanistic understanding of brain anatomy in both health and disease, may help to identify modifiable risk factors and may form a basis for the development of more accurate predictive risk scores capturing a wider array of potential influences.

The majority of the studies mentioned above used whole blood as a surrogate tissue for the brain due to inaccessibility of the brain ante-mortem. Although DNAm is reported to be tissue- and cell type-specific, similarities between blood and brain DNAm have also been identified (28). In addition, whole blood has successfully been used in the past to identify meaningful epigenetic differences in brain-related traits, as shown above (18).

Here, we sought to assess DNAm associations with Magnetic Resonance Imaging (MRI) global brain structural phenotypes, including cerebral white matter, total grey matter, and whole-brain volume using the Illumina Infinium MethylationEPIC array, capturing DNAm at approximately 850K CpG sites (29). Using DNAm measured ~6 years prior to MRI data collection, we examined whether CpG sites were associated with global brain structure at a later timepoint in N=672 individuals. We then investigated whether concurrently measured

63 DNAm was associated with global brain structure in N=565 individuals.

Methods

Study population: Generation Scotland: Scottish Family Health Study (GS:SFHS)

GS:SFHS is a large, family-based epidemiological study aiming to investigate the genetics of health and disease in approximately 24,000 individuals aged 18-98 years across Scotland. Data collected between 2006 and 2011 consists of genetic, DNA methylation, and environmental variables (30,31). GS:SFHS received ethical approval from NHS Tayside Research Ethics Committee (REC reference number 05/S1401/89) and has Research Tissue Bank Status (reference: 20/ES/0021). Written informed consent was obtained from all participants.

A total of N=9,618 participants from GS responded when re-contacted at a later timepoint, and further data on mental health, specifically depression, was obtained. N=1,188 were recruited for brain scanning, and approximately N=700 with both DNAm and neuroimaging data were available at the time of the current study. Details of recruitment and study information have been reported previously (32,33). The study was supported by the Wellcome Trust through a Strategic Award (reference 104036/Z/14/Z). Written consent at each stage of the study was obtained from all participants.

Two timepoints were used for the current study: blood samples were collected at baseline measurement (2006-2011), and concurrently with neuroimaging data (2015-2019).

Phenotypes

Global brain volumes

T1 images were processed using standard ENIGMA protocols (34) with FreeSurfer 5.3 and all output was visually quality checked. Manual edits were applied as required to correct for inclusion of skull tissue, exclusion of brain tissue or for errors in parcellation. Global measures were extracted from the final output following all edits. Manual editing, although necessary, did introduce a degree of subjective bias, therefore 'editing' was included as a binary covariate (values: yes/no). Further, as the complete set of T1s was processed, quality checked and edited in two parts, 'batch' was also included as a covariate.

We used 3 global volume measures in the current study. Total cerebral white matter includes hyperintensities and excludes anything that is not white matter. Total grey matter is rendered by the sum of the cortex within the left and right hemispheres, as well as subcortical and cerebellar grey matter. Finally, whole-brain volume includes both grey and white matter, and corresponds to brain volume without the brain stem, ventricles, cerebrospinal fluid, and choroid plexus.

Baseline lifestyle factors and MDD status

Body mass index (BMI) was calculated using height (m) and weight (kg) as measured by clinical staff at baseline recruitment. Participants were asked to report 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 by multiplying the number of cigarette packs (20 cigarettes/pack) smoked per day by the number of years a person has smoked (35). MDD status was assessed at baseline using the Structured Clinical Interview of the Diagnostic and Statistical Manual, version IV (SCID) (36). 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.

Concurrent lifestyle factors and MDD status

At the follow-up assessment, participants were sent study packages that included questionnaires. Here, BMI was calculated using height (m) and weight (kg). Participants also recorded the number of units consumed during the past week, whether they were current, former, or non-smokers, and (if they smoked) the number of cigarettes smoked in an average week. Finally, MDD status was ascertained through the Composite International Diagnostic Interview-Short Form (CIDI-SF) (37), and participants with no MDD were those individuals who did not fulfil criteria for current or previous MDD diagnoses based on responses.

DNA methylation

Baseline DNAm data was pre-processed and quality-checked for all individuals by Amador et al. in 2019 (38). At the concurrent timepoint, samples were placed on the array

at two different time points and were therefore processed separately. The main difference between processing and analysis pipelines related to how key covariates were adjusted for. At baseline these were regressed out during pre-processing, whereas for the concurrent batches they were included as covariates in downstream analyses. However, across all batches, standard quality check (QC) and pre-processing steps with regards to sample and probe exclusions were identical (see below). We note however that differences in the processing resulted in different numbers of final CpG sites included for analysis.

Cross-reactive (N=42,558) and polymorphic (N=10,971) CpGs, obtained from McCartney et al. (2016) were removed from both the baseline and concurrent DNAm datasets (39).

Baseline 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 (29). Samples were obtained and DNA was extracted between 2006-2011. DNAm profiling using the Illumina Human-MethylationEPIC BeadChip (29) was performed in two sets (in 2016, set A_N=5101; in 2019, set B_N=4,450) and pre-processing and QC was conducted once the second set was released, as detailed in Amador et al. (38,40,41). Participants were removed due to a number of reasons, including sex mismatch ($N_{removed}$=24), having more than 1% CpG sites with a detection p-value>0.05 ($N_{removed}$=52), being an outlier for bisulphite conversion control probes ($N_{removed}$=1), having a median methylated signal intensity more than 3 standard deviations lower than expected ($N_{removed}$=74), and other technical and dataset-specific issues ($N_{removed}$=602, see Supplementary Materials). A total of 10,495 CpG sites were removed due to low beadcount, poor detection p-value, and sub-optimal binding.

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. Correction was then 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 type proportions (CD8T, CD4T, NK, Bcell, Mono, Gran) (fixed effects) (42). The final number of CpG sites that converged for these analyses was 674,246 across the 22 autosomes.

Concurrent DNA methylation

Genome-wide DNAm data profiled from whole blood samples was available for a total of 710 individuals using the Illumina Human-MethylationEPIC BeadChip (29). Pre-processing was carried out in two separate sets ($N_{set\ 1}$=404; $N_{set\ 2}$=306) intended as discovery and replication datasets, by Walker et al. (43,44). Meffil (45) was use to remove samples if: there was a mismatch betweenself-reported and methylation-predicted sex and if >0.5% of probes failed the detection p- value threshold (>0.01); probes were removed if >1% samples failed the detection p- value >0.01 and if >5% of samples failed the beadcount threshold (N=3). In addition, samples were removed if they showed evidence of dye bias and they were outliers for the bisulphite conversion control probes. ShinyMethyl (46) was then used to plot the log median intensity of methylated and unmethylated signals per array and inspect the output from the control probes; outlying samples detected by visual inspection were excluded. Meffil (45) was then used again to remove any additional samples who had a sex mismatch. PC plots were made using the first two methylation principal components and any additional outlying samples on the basis of these plots were removed. Finally, data were normalised using the dasen method in wateRmelon, and M-values were generated using the beta2m function in lumi (47). The final number of CpG sites after pre-processing was N=768,068 (set 1) and N=765,695 (set2) across the 22 autosomes.

Statistical methods

Epigenome-wide association

We used the "limma" package (48) in R to run linear regression models for both baseline and concurrent DNAm data, where each CpG was included as an outcome variable. Brain cortical volumes, specifically cerebral white matter, total grey matter, and whole brain volume were included as predictor variables in separate EWAS at each DNAm timepoint.

185 TheR code for these analyses is available in the Supplementary Materials.

186 Covariates for each model using baseline DNAm were MRI site (to account for
187 different data collection sites; see Supplementary Materials), age, age^2, sex, intracranial
188 volume, and set (to account for different DNAm data pre-processing sets). Due to the impact
189 of lifestyle factors on DNAm (49–52), BMI, alcohol units, smoking status, and pack years
190 were also included as covariates. Lastly, due to the increased prevalence of MDD in the
191 dataset, MDD status was included as a covariate in all models. Technical (batch,
192 appointment date) and biological (relatedness, cell type estimations, methylation principal
193 components) variables were regressed out during pre-processing and were not included as
194 covariates in downstream analyses. After QC, there were 674,246 CpGs and epigenome-
195 wide significance was determined by a Bonferroni correction (0.05/674,246, $p \leq 7.41 \times 10^{-8}$).

196 For both sets at the concurrent DNAm timepoint, covariates for each model were
197 DNAm batch, 5 cell type proportion estimations (granulocytes, natural killer cells, B-
198 lymphocytes, CD4+ T-lymphocytes and CD8+ T-lymphocytes), MRI site, age, age^2, sex,
199 intercranial volume, BMI, smoking status, number of cigarettes smoked/week, alcohol units,
200 MDD status, and 20 methylation PCs. Bonferroni correction was applied based on the number
201 of CpGs remaining in each set after QC (set 1: 0.05/768,068 CpGs, $p \leq 6.51 \times 10^{-8}$; set 2:
202 0.05/765,695 CpGs, $p \leq 6.52 \times 10^{-8}$).

203 The Blood Brain DNA Methylation Comparison Tool (53)
204 (http://epigenetics.essex.ac.uk/bloodbrain/) investigates the correlation between DNAm
205 from whole blood and four brain regions (prefrontal cortex, entorhinal cortex, superior
206 temporal gyrus, and cerebellum) for all probes on the Illumina 450K array (54). We used this
207 resource to investigate the strength of correlation between the two tissues for CpGs identified
208 here.

209

210 *Meta-analysis using METAL – concurrent timepoint*

211 At the concurrent timepoint, in set 1, N=331 individuals were available with global
212 volume and methylation data after QC and N=234 were available in set 2. Meta-analysis of
213 these two datasets was performed in METAL (55) using p-value based analysis (N=565). The
214 meta-analysis was based on N=769,263 CpGs across both sets and a Bonferroni correction

(0.05/769,263) was used to define epigenome-wide significance ($p \leq 6.49 \times 10^{-8}$).

Pathway analysis

We annotated CpG sites to genes through the Infinium MethylationEPIC BeadChip database (29). The database provides information about genes, chromosome location, start and end sites, and other features.

We used missMethyl (56), accessed via methylGSA (57), to assess pathway enrichment for differentially-methylated CpG sites. The package allows correction for biases in the representation of genes on the Infinium BeadChip. Gene Ontology (GO) terms were accessed using the msigdbr package (58). Pathways included in the analysis were all GO pathways of size 1-250 genes inclusive. CpG sites included in the analysis were those significant at a threshold of $p<1 \times 10^{-5}$, as used in previous studies (59). Information on GO pathways can be accessed via www.geneontology.org using Gene Ontology identifiers, comprised of "GO" followed by a string of numbers (e.g. GO:0000000).

Power analysis – concurrent timepoint

Since the concurrent data was formed by two smaller samples of pre-processed data, we additionally conducted power analysis to determine whether our concurrent samples had sufficient power to detect a significant effect. This was conducted using effect sizes from the baseline data to inform the power calculations. We used the "pwr.f2.test" function in package "pwr" in R and the set parameters were as follows:

1. Regression coefficients: DNAm batch, 5 cell type estimations (granulocytes, natural killer cells, B-lymphocytes, CD4+ T-lymphocytes and CD8+ T-lymphocytes), MRI site, age, age^2, sex, intercranial volume, BMI, smoking status, number of cigarettes smoked/week, alcohol units, MDD status, 20 methylation principal components.

2. Effect size: we input the largest effect size identified in EWAS at baseline (N=672) for each global volume.

3. Significance level: to adjust for multiple testing correction (FDR), the p-value for a single potential test was set based on the number of CpG sites in each dataset (set 1: $0.05/768,068=6.51 \times 10^{-8}$; set 2: $0.05/765,695=6.53 \times 10^{-8}$).

4. Power: to observe different power percentages, we input 60%, 80%, 90%, 95% and 99% power.

Results

Demographic characteristics

There were N=672 individuals in the baseline EWAS, N=331 in the set 1 concurrent EWAS, and N=234 in the set 2 concurrent EWAS. Demographic characteristics for all individuals are presented in Table 1. Further descriptive characteristics regarding global volumes are presented in Supplementary Table 1.

Demographic characteristics	Baseline (N=672)	Concurrent set 1 (N=331)	Concurrent set 2 (N=234)
Age – Mean (SD), range	52.29 (9.93), 18-75	60.45 (8.42), 28-78	59.61 (10.21), 28-81
Sex			
Female	406	193	132
Male	266	138	102
Set			
1	621	-	-
2	51	-	-
BMI – Mean (SD), range	27.13 (4.96), 15.96-56.60	27.48 (5.18), 16.42-51.75	28.23 (5.31), 19-20-52.81
Alcohol units – Mean (SD), range	10.53 (16.44), 0-326	7.12 (8.91), 0-60	7.39 (9.67), 0-60
Smoking status			
Current smoker	83	16	12
Former smokers (quit < 1 year ago)	10	124	92
Former smokers (quit > 1 year ago)	208		
Never smoked tobacco	371	191	130
Pack years – Mean (SD), range	7.59 (14.56), 0-111	-	-
Cigarettes smoked/week			
1-10 cigarettes	-	10	6
11-20 cigarettes	-	10	9
MDD status			
Cases	121	83	83
Controls	551	248	151

Table 1. Demographic characteristics for individuals with global volume data, including lifestyle variables and MDD. "-" indicates that there was no data of the sort for the respective dataset. Former smokers at the baseline measurement were split into those who quit less than a year ago and those who quit more than a year ago; at the concurrent timepoint, this division is not made.

Baseline EWAS

Baseline EWAS identified 1, 3, and 2 CpG sites that were associated with cerebral white matter, total grey matter, and whole-brain volume, respectively ($p \leq 7.41 \times 10^{-8}$). Both CpGs associated with whole brain volume were also associated with total grey matter and were significantly hypermethylated. One CpG site associated with cerebral white matter and one associated with total grey matter were hypomethylated. As shown in Figure 1A-C, CpG associations with grey matter were stronger than with white matter. Information about each CpG site is shown in Table 2.

Phenotype	CpG site	Gene	C	β	P-value	P-corr	CpG – previously associated traits	Gene – previously associated traits
Total grey matter	cg07585845 (EPIC)	-	3	9.59×10^{-7}	1.02×10^{-9}	0.0007	-	-
Whole-brain volume	cg07585845 (EPIC)	-	3	4.47×10^{-7}	1.38×10^{-8}	0.009		
Total grey matter	cg02325951 (450K)	*FOXN3*	14	6.53×10^{-7}	1.31×10^{-9}	0.0009	Sex ($p=2 \times 10^{-54}$;1.8×10^{-42}; (54))	Acute myeloid leukemia ($p=8 \times 10^{-21}$ $p=3 \times 10^{-14}$; (60)) Heel bone mineral density ($p=2 \times 10^{-12}$; (61)) Intelligence ($p=1 \times 10^{-11}$; (62)) Self-reported educational attainment ($p=8 \times 10^{-11}$; (63)) Cognitive function measurement ($p=2 \times 10^{-9}$; (63)) Mathematical ability ($p=3 \times 10^{-9}$; (63)) Smoking status measurement ($p=7 \times 10^{-9}$; (64)) Risk-taking behaviour ($p=8 \times 10^{-9}$; (65))
Whole-brain volume	cg02325951 (450K)	*FOXN3*	14	3.26×10^{-7}	1.45×10^{-9}	0.001		
Cerebral white matter	cg10297662 (EPIC)	*PNKP*	19	-1.46×10^{-6}	4.92×10^{-8}	0.03	-	Involved in DNA repair; mutations at locus associated with microcephaly, seizures, and developmental delay (66)
Total grey matter	cg04190002 (450K)	*SHANK 3*	22	-3.75×10^{-7}	7.31×10^{-9}	0.04	Sex ($p=5.4 \times 10^{-19}$; (62))	Self-reported educational attainment ($p=2 \times 10^{-20}$; (63)) Mathematical ability ($p=1 \times 10^{-17}$; (63)) Cognitive function measurement ($p=3 \times 10^{-12}$; (63)) Schizophrenia ($p=3 \times 10^{-12}$; (67))

Table 2. CpG sites significantly associated with cerebral white matter, total grey matter, and whole-brain volume (N=672), along with gene annotations (Gene), chromosome (C),

standardised effect size (β), nominal (P-value) and multiple comparison-corrected p-values (P-corr). Traits previously associated with each CpG site were extracted from EWAS catalogues (http://www.ewascatalog.org/, association between traits and CpGs on Illumina 450K array at $p \leq 1.0 \times 10^{-4}$; and http://www.bioapp.org/ewasdb/, (68)), association between traits and CpGs on Illumina 450K and EPIC arrays at $p \leq 1.0 \times 10^{-3}$). Gene information was extracted from the GWAS catalogue (https://www.ebi.ac.uk/gwas/; associations between traits and SNPs at $p < 1.0 \times 10^{-5}$). All associations included in the table from these two catalogues are genome-wide significant.

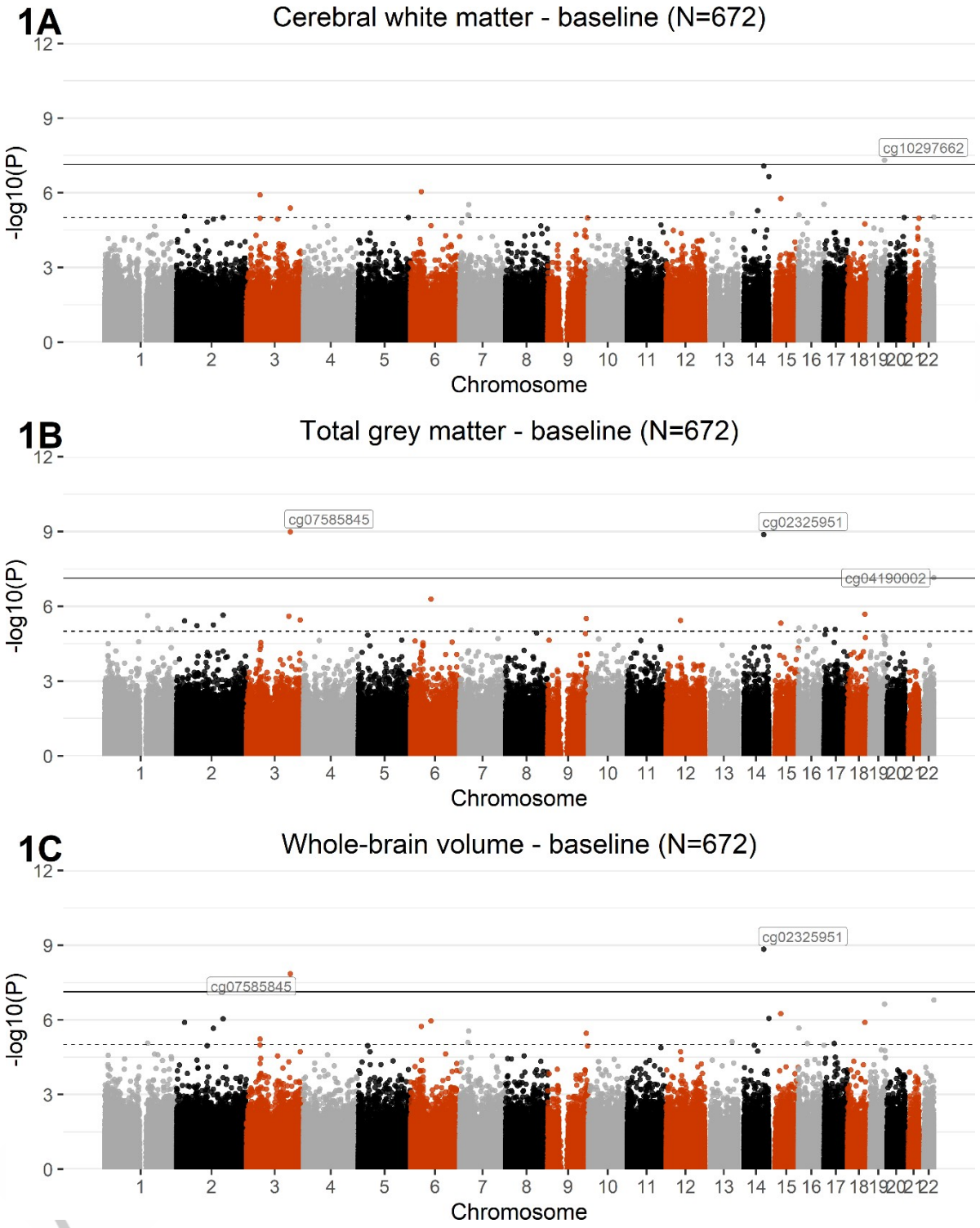


Figure 1A, 1B, 1C. Manhattan plots showing the results from EWASs of cerebral white matter (1A), total grey matter (1B), and whole-brain volume (1C), using baseline DNAm data (N=672). The black line defines the threshold for epigenome-wide significance ($p \leq 7.41 \times 10^{-8}$) and the dotted line defines CpG sites at $p \leq 1 \times 10^{-5}$. Epigenome-wide significant hits for each phenotype are labelled on the graph.

Correlation between whole blood DNAm and four brain regions

We used the Blood Brain DNA Methylation Comparison Tool (53) to investigate the correlation between blood and brain methylation measurements for two of the CpGs identified here, located on the 450K array, and four brain regions. cg04190002 was strongly correlated with prefrontal cortex ($r=0.579$, $p=6.55\times10^{-8}$), entorhinal cortex ($r=0.564$, $p=2.94\times10^{-7}$), superior temporal gyrus ($r=0.598$, $p=1.5\times10^{-8}$), and cerebellum ($r=0.663$, $p=3.02\times10^{-10}$), while cg02325951 was strongly correlated with prefrontal cortex ($r=0.858$, $p=1.73\times10^{-22}$), entorhinal cortex ($r=0.868$, $p=1.19\times10^{-22}$), and superior temporal gyrus ($r=0.871$, $p=3.32\times10^{-24}$).

Baseline Pathway Analysis

Enrichment of differentially methylated regions in biological pathways was analysed using missMethyl (56), where an over-representation analysis of GO pathways was performed for sets of genes annotated to CpG sites differentially expressed at $p<1\times10^{-5}$ ($N_{\text{cerebral white matter}}$: 19, $N_{\text{total grey matter}}$: 22, $N_{\text{whole-brain volume}}$: 21).

There were no over-represented pathways after multiple correction. A number of brain-related biological processes, molecular functions, and cellular components were included in the top 10 significant pathways (Supplementary Table 2). For instance, guanylate kinase-associated protein clustering, which facilitates assembly of post-synaptic density of neurons (GO:0097117), was found to be over-represented for all three imaging phenotypes (cerebral white matter nominal p-value=0.0007; total grey matter nominal p-value=0.001; whole-brain volume nominal p-value=0.0009). Positive regulation of synapse structural plasticity (GO:0051835) was over-represented in both cerebral white matter (nominal p-value=0.002) and total grey matter (nominal p-value=0.002). Finally, forebrain generation of neurons (GO:0021872; nominal p-value=0.001) was over-represented for cerebral white matter.

Concurrent EWAS

Meta-analysis of EWAS across the two concurrent sets did not reveal any Bonferroni-corrected CpG sites associated with any of the global volumes (Figure 2A-C). A list of the top

10 CpGs associated with cerebral white matter ($EWAS_{set\ 1}$ $\beta_{range}=4.71x10^{-6}$- $6.53x10^{-6}$; $EWAS_{set\ 2}$ $\beta_{range}=1.02x10^{-5}$-$8.75x10^{-6}$) total grey matter ($EWAS_{set\ 1}$ $\beta_{range}=6.71x10^{-6}$-$8.03x10^{-6}$; $EWAS_{set\ 2}$ $\beta_{range}=1.03x10^{-5}$-$8.84x10^{-6}$), and whole-brain volume ($EWAS_{set\ 1}$ $\beta_{range}=2.69x10^{-6}$-$4.05x10^{-6}$; $EWAS_{set\ 2}$ $\beta_{range}=6.23x10^{-6}$-$6.69x10^{-6}$), is presented in Supplementary Tables 3-5. Genes annotated to these top 10 CpGs have previously been implicated in brain-related phenotypes, including psychiatric disorders (MDD (69–72), schizophrenia (73)), neurodegenerative disorders (neurofibrillary tangles and PHF-tau measurement in Alzheimer's Disease (74)), and cognitive traits (mathematical ability, self- reported educational attainment (75)). Results reported here are nominal and should be supported by further large-scale cohorts.

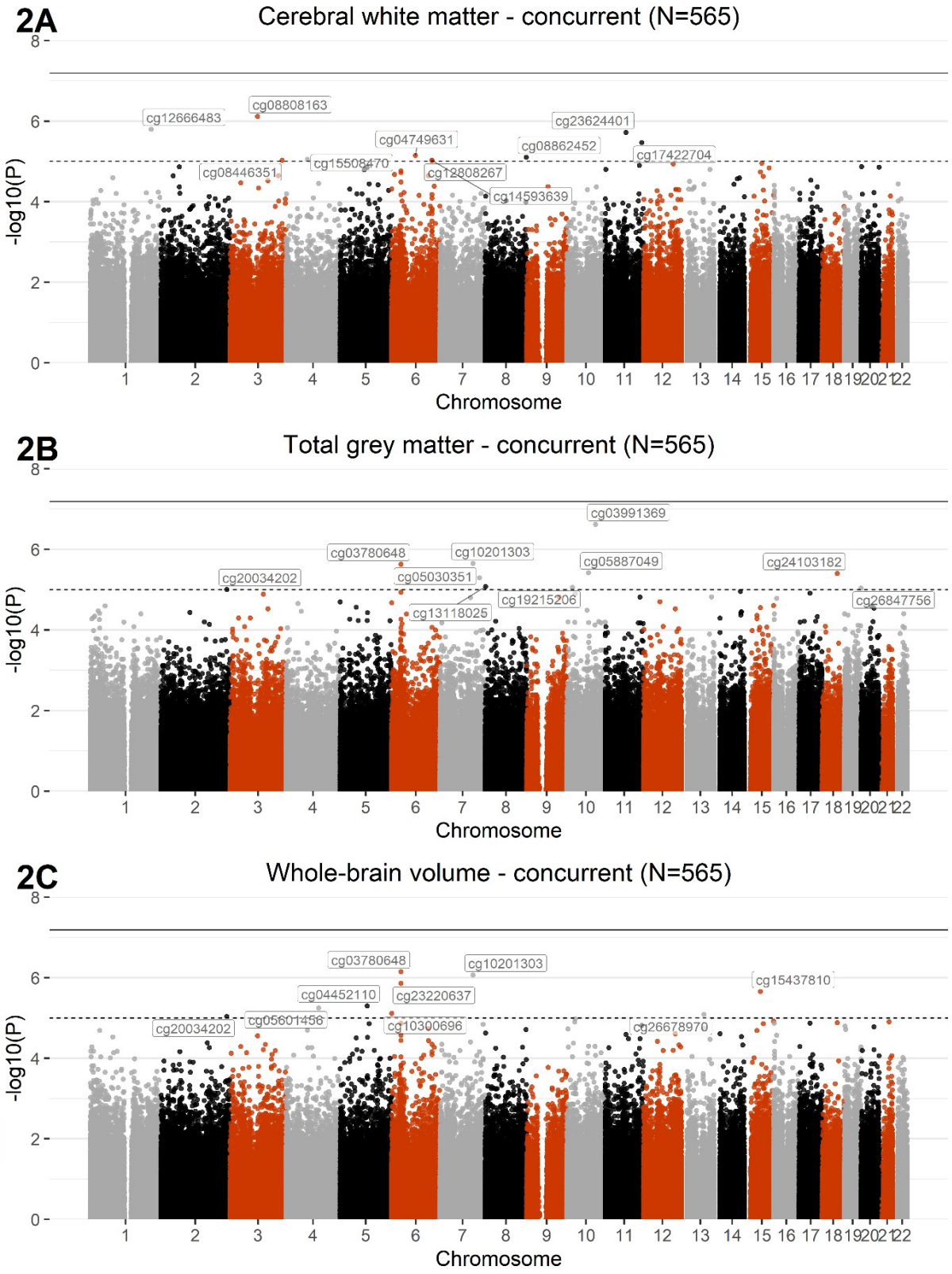


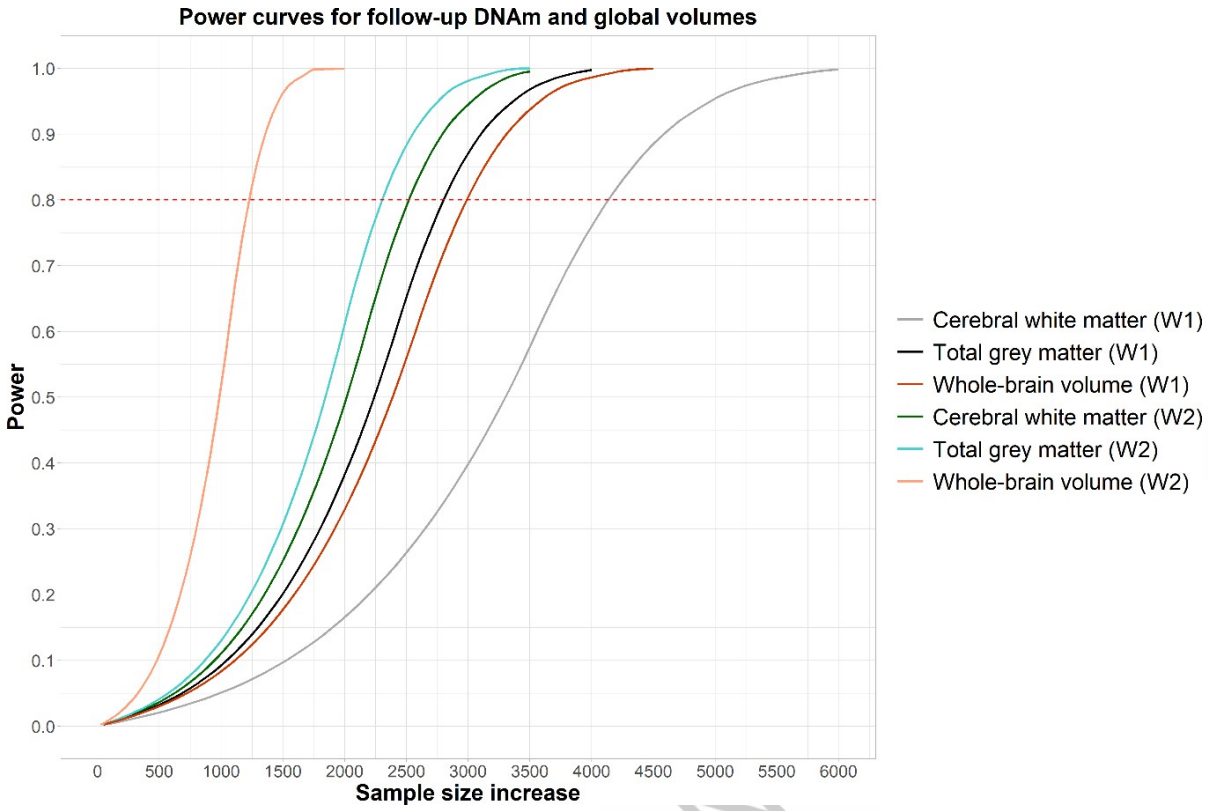
Figure 2A, 2B, 2C. Manhattan plots showing meta-analysis of EWAS of cerebral white matter (2A), total grey matter (2B), and whole-brain volume (2C), across the 2 concurrent sets ($N_{set\ 1}$=331; $N_{set\ 2}$=234; N_{total}=565). The black line defines the threshold for epigenome-wide significance ($p \leq 6.5 \times 10^{-8}$) and the dotted line defines $p \leq 1 \times 10^{-5}$. CpGs that met a significance of $p \leq 1 \times 10^{-5}$ are labelled on the graph.

Concurrent Pathway Analysis

As above, enrichment of differentially methylated regions in specific pathways was assessed using missMethyl (50) for sets of genes annotated to CpG sites differentially expressed at $p<1x10^{-5}$ ($N_{cerebral\ white\ matter}$: 10, $N_{total\ grey\ matter}$: 10, $N_{whole\text{-}brain\ volume}$: 9). There were no over-represented pathways following FDR adjustment for multiple comparisons. The top 10 most significant pathways for each phenotype indicated a pattern of phenotype-specific biological processes, molecular functions, and cellular components (Supplementary Table 6). For instance, over-represented pathways in cerebral white matter included myelination (GO:0042552; nominal p-value=0.002), ensheathment of neurons (GO:0007272; nominal p-value=0.002), axon ensheathment (GO:0008366; nominal p-value=0.001), glial cell development (GO:0021782; nominal p-value=0.001) and glial cell differentiation (GO:0010001; nominal p-value=0.004). Total grey matter over-represented pathways included glutamate catabolic process to aspartate (GO:0019550; nominal p-value=0.0009) and to 2-oxoglutarate (GO:0019551; nominal p-value=0.0009). Finally, over-represented pathways in whole-brain volume included several MHC-related biological processes, including regulation (GO:0002586; nominal p-value=0.001) and negative regulation (GO:0002587; nominal p-value=0.0009) of antigen processing and presentation of peptide antigen via MHC class II, negative regulation of antigen processing and presentation of peptide or polysaccharide antigen via MHC class II (GO:0002581; nominal p-value=0.001), as well as N-acetyllactosaminide beta-1,3-N-acetylglucosaminyltransferase (GO:0008532, molecular function, nominal p-value=0.001), an enzyme encoded by the gene *B3GNT2*, which is highly expressed in whole-brain, hippocampus, amygdala, cerebellum, and caudate nucleus (https://www.uniprot.org/uniprot/Q9Z222).

Power curves for concurrent data

Power curves for the three imaging phenotypes are presented in Figure 3. Further details, including effect size for each phenotype, are included in Supplementary Tables 7 and 8. These indicate that approximately 1,000-6,000 individuals (depending on phenotype) would be needed to detect an effect after multiple correction.



362

363

364

365 **Figure 3.** Power curves for cerebral white matter, total grey matter, and whole-brain volume
366 calculated separately for set 1 and set 2. The x-axis indicates how many participants would
367 be needed to detect an effect with 60%, 80%, 90%, 95% or 99% power at $p<6.51x10^{-8}$ (set 1
368 (W1)) and $p<6.53x10^{-8}$ (set 2 (W2)) with 36 regression coefficients included in the linear
369 model. Effect sizes were calculated based on the largest effect size obtained in EWAS for each
370 phenotype at baseline.

Discussion

We report a number of significant associations between DNAm measured ~6 years prior to MRI data collection and cerebral white matter ($N_{significant\ CpGs}$=1), total grey matter ($N_{significant\ CpGs}$=3), and whole-brain volume ($N_{significant\ CpGs}$=2) (N=672), annotated to genes involved in brain-related traits. There were no significant associations between DNAm collected concurrently with MRI data (N=565). In addition, pathway analysis did not uncover any significant findings for either the baseline or concurrent analyses. Power analysis of the concurrent data using baseline data for effect size confirmed that approximately 1,000-6,000 individuals (depending on phenotype) would be needed to detect a statistically significant effect.

For the analysis of associations between DNAm measured at baseline and cortical volumes ~6 years later, one CpG associated with cerebral white matter, cg10297662, was annotated to *PNKP*. This CpG site has not previously been associated with any other traits, to the best of our knowledge. *PNKP* is involved in DNA repair following ionizing radiation or oxidative damage (76) and is expressed in a number of tissues, including the brain. Mutations in this gene have been associated with a number of neural conditions, including microcephaly, developmental delay, seizures, and cerebellar ataxia (66,77). These mutations have been shown to lead to white matter defects, which is the phenotype investigated here (78). Previous evidence also indicates that loss of *PNKP* strongly impacts oligodendrocytes, leading to white matter abnormalities (79). Efforts should be made to identify whether the relationship between *PNKP* mutations and defects in white matter is mediated by differential DNAm at specific sites.

Two CpGs, cg07585845 and cg02325951, were associated with both total grey matter and whole-brain volume. cg07585845 has not been previously associated with any traits nor annotated to any genes. cg02325951 was previously associated with sex in a study investigating methylation trajectories across human foetal brain development ($p=2x10^{-54}$; (80)). The gene to which cg02325951 is annotated, *FOXN3*, is involved in several physiological processes, such as development, ageing, obesity, and cancer and is expressed in multiple tissues, including the forebrain and midbrain. Further, animal studies show that mutations within the gene have been associated with craniofacial defects (81). In addition, *FOXN3* has previously been associated with several brain-related phenotypes in previous GWAS,

including intelligence ($p=1x10^{-11}$; (62)), self-reported educational attainment ($p=8x10^{-11}$), cognitive function measurement ($p=2x10^{-9}$), and mathematical ability ($p=3x10^{-9}$) (63). These cognition-related phenotypes have previously been associated with whole brain volume, where higher cognition was associated with a larger brain size (76). Future studies should investigate whether DNAm localized to *FOXN3* plays a role in cognition development through modifications in whole-brain volume.

Finally, in addition to the two CpGs above, total grey matter was also associated with cg04190002, a CpG previously associated with sex in newborns ($p=5.4x10^{-19}$; (82)). The CpG is annotated to *SHANK3*, which encodes multidomain scaffold proteins of the postsynaptic density connecting neurotransmitter receptors, among other membrane proteins and is expressed in the cerebral cortex and the cerebellum. The gene has previously been associated with a host of brain disorders and traits, including self- reported educational attainment ($p=2x10^{-20}$), mathematical ability ($p=1x10^{-17}$), cognitive function measurement ($p=3x10^{-12}$) (63) and schizophrenia ($p=3x10^{-9}$; (67)), and mutations have previously been associated with autism spectrum disorder (83). These disorders in turn have been associated with changes in grey matter (84), and future studies should investigate whether these psychiatric disorders are also associated with differential DNAm at cg04190002, and other probes localized to *SHANK3*, as well as explore whether associations are mediated by global brain phenotypes.

Blood and brain methylation measures for both cg02325951 and cg04190002 (both CpGs on the 450K array) were strongly correlated, indicating that whole blood is a suitable proxy tissue for investigating associations with brain phenotypes, at least for these probes. Future studies exploring DNAm in relation to global brain phenotypes and associated traits may therefore benefit from whole blood DNAm measurements.

DNAm profiled at a different timepoint to phenotype measurement has previously yielded interesting results. Barbu et al. (2020) found that a methylation risk score calculated from DNAm profiled 4-11 years prior to MDD diagnosis was significantly associated with incident cases who were well at DNAm measurement but went on to develop MDD (12). Clark et al. (2020) similarly associated DNAm profiled in MDD patients at baseline with MDD status 6 years later (85). These previous findings indicate that DNAm measured prior to phenotype measurement may provide meaningful insight into phenotype development and change

across time. The findings above relating DNAm measured previously to MRI scans may therefore aid in the investigation of epigenetic differences in brain-related disease and health at a later timepoint, although further longitudinal replication is needed to verify these associations.

Associations between DNAm measured concurrently to MRI scans did not yield any significant findings. Power calculations using the baseline data to derive effect size showed that approximately 1,000-6,000 participants (depending on phenotype) would be needed to uncover a significant effect at epigenome-wide level. This number is supported by previous studies, such as Jia et al. (2019), who analysed 3,337 individuals across 11 cohorts as part of ENIGMA to find 2 CpGs significantly associated with hippocampal volume (19). This may indicate that null findings were due to lack of power at the concurrent timepoint. Null findings here should serve as a stimulus for larger collaborations and meta-analyses in future.

Further, effect sizes for both timepoints were much smaller than those identified in previous studies that analysed larger sample sizes in specific brain regions (19) (largest baseline effect size: 1.46×10^{-6}; largest concurrent effect size: 1.06×10^{-6}), which suggests that findings here should be interpreted with caution. The results here indicate that global associations with DNAm may be weaker than those at a regional level. Future studies may therefore benefit from focussing on lobe- and region-specific correlates of DNAm.

At the concurrent timepoint, DNAm data was pre-processed and quality-checked in 2 sets, resulting in a different number of final CpGs ($N_{CpG\ set\ 1}$=768,068; $N_{CpG\ set\ 2}$=765,695). Pearson's correlations between the EWAS betas from set 1 and set 2 across all CpGs were *r*=0.02 (95% C.I.=0-0.102), *r*=0.04 (95% C.I.=0-122), and *r*=0.03 (95% C.I.=0-0.112) for cerebral white matter, total grey matter, and whole brain volume, respectively. When restricting CpGs to those with a nominal p-value (≤0.05), the beta correlations were slightly higher, although not strong: *r*=0.17 (95% C.I.=0.089-0.249), *r*=0.18 (95% C.I.=0.099-0.259), and *r*=0.22 (95% C.I.=0.14-0.297) for cerebral white matter, total grey matter, and whole-brain volume, respectively. The low effect size correlations may be a further reflection of the small sample investigated here.

There are limitations to the current study. Firstly, we report DNAm changes in whole

blood, which may not be representative of brain phenotypes. However, two of the CpGs identified here, located on the 450K array, were strongly correlated with DNAm in four brain regions (53). Although previous studies have shown that there is considerable agreement between blood and brain (28), future studies should explore DNAm changes in the brain in post-mortem samples where possible to uncover biological mechanisms underpinning brain structure within the same tissue. Further, findings at baseline may indicate that some DNAm changes lie upstream of brain structural changes, although effect sizes for each CpG were small compared to previous concurrent EWAS of brain regions (18,19). In addition, we cannot test the direction of association between brain structural changes and DNAm. In future, studies may apply Mendelian Randomization to investigate whether DNAm may be on the causal path to brain structure alterations in brain health and disease. Finally, in the current study we focussed on global brain phenotypes to explore whether global brain-related changes, previously associated with psychiatric and neurological disorders, are associated with DNAm alterations. Previous evidence includes DNAm associations at both global and regional level (18), and it may be that DNAm may provide more insight into region-specific alterations in relation to brain health and disease.

In conclusion, we report an EWAS of global cortical brain volumes using DNAm data collected ~6 years prior to MRI data collection in 672 individuals and an EWAS meta-analysis of cortical brain volumes using DNAm measured concurrently to MRI data in 565 individuals, both part of a large, population-based cohort. Using baseline DNAm data, we find four CpGs significantly associated with cortical brain volumes ~6 years later, all of which are annotated to genes implicated in brain-related phenotypes. We did not find significant associations at the concurrent timepoint. Findings here should be interpreted with caution, and future studies should aim to determine further links between DNAm changes and brain structure and function, to highlight our understanding of this relationship in health and disease.

Acknowledgements

Generation Scotland received core support from the Chief Scientist Office of the Scottish Government Health Directorates [CZD/16/6] and the Scottish Funding Council [HR03006] and is currently supported by the Wellcome Trust [216767/Z/19/Z]. Genotyping of the GS:SFHS samples was carried out by the Genetics Core Laboratory at the Edinburgh Clinical Research Facility, University of Edinburgh, Scotland and was funded by the Medical Research Council UK and the Wellcome Trust (Wellcome Trust Strategic Award "STratifying Resilience and Depression Longitudinally" (STRADL) Reference 104036/Z/14/Z). MCB is supported by a Guarantors of Brain Non-clinical Post-Doctoral Fellowship. AMM is supported by the Wellcome Trust (104036/Z/14/Z, 216767/Z/19/Z, 220857/Z/20/Z) and UKRI MRC (MC_PC_17209, MR/S035818/1). KLE is supported by the NARSAD Independent Investigator Award (Grant ID: 21956). JMW is supported by UK Dementia Research Institute which is funded by the MRC, Alzheimer's Research UK and Alzheimer's Society, by the Fondation Leducq (16 CVD 05), and the Row Fogo Centre for Research Into Ageing and the Brain (BRO-D.FID3668413). This work is part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 847776.

Financial disclosures

AMM has previously received grant support from Pfizer, Lilly and Janssen. These studies are not connected to the current investigation. Remaining authors report no conflicts of interest. JDS has received funding via an honorarium associated with a lecture for Wyeth and funding from Indivior for a study on opioid dependency, both of which are unconnected with the current investigation.

References

1. DeYoung CG, Hirsh JB, Shane MS, Papademetris X, Rajeevan N, Gray JR. Testing predictions from personality neuroscience. Brain structure and the big five. Psychol Sci a J Am Psychol Soc / APS [Internet]. 2010 [cited 2020 Nov 9];21(6):820–8. Available from: /pmc/articles/PMC3049165/?report=abstract

2. Zatorre RJ, Fields RD, Johansen-Berg H. Plasticity in gray and white: Neuroimaging changes in brain structure during learning [Internet]. Vol. 15, Nature Neuroscience. Europe PMC Funders; 2012 [cited 2020 Nov 9]. p. 528–36. Available from: /pmc/articles/PMC3660656/?report=abstract

3. Shen X, Reus LM, Cox SR, Adams MJ, Liewald DC, Bastin ME, et al. Subcortical volume and white matter integrity abnormalities in major depressive disorder: Findings from UK Biobank imaging data. Sci Rep [Internet]. 2017 Dec 1 [cited 2020 Nov 9];7(1). Available from: /pmc/articles/PMC5514104/?report=abstract

4. Höistad M, Segal D, Takahashi N, Sakurai T, Buxbaum JD, Hof PR. Linking white and grey matter in schizophrenia: Oligodendrocyte and neuron pathology in the prefrontal cortex [Internet]. Vol. 3, Frontiers in Neuroanatomy. Frontiers Media SA; 2009 [cited 2020 Nov 9]. Available from: /pmc/articles/PMC2713751/?report=abstract

5. Favre P, Pauling M, Stout J, Hozer F, Sarrazin S, Abé C, et al. Widespread white matter microstructural abnormalities in bipolar disorder: evidence from mega- and meta-analyses across 3033 individuals. Neuropsychopharmacology [Internet]. 2019 Dec 21 [cited 2020 Nov 9];44(13):2285–93. Available from: http://www.nature.com/articles/s41386-019-0485-6

6. Mahmood A, Bibat G, Zhan AL, Izbudak I, Farage L, Horska A, et al. White matter impairment in rett syndrome: Diffusion tensor imaging study with clinical correlations. Am J Neuroradiol [Internet]. 2010 Feb [cited 2020 Nov 9];31(2):295–9. Available from: /pmc/articles/PMC2909097/?report=abstract

7. Serra L, Cercignani M, Lenzi D, Perri R, Fadda L, Caltagirone C, et al. Grey and white matter changes at different stages of Alzheimer's disease. J Alzheimer's Dis. 2010 Jan 1;19(1):147–59.

8. Salat DH, Buckner RL, Snyder AZ, Greve DN, Desikan RSR, Busa E, et al. Thinning of the cerebral cortex in aging. Cereb Cortex [Internet]. 2004 Jul [cited 2020 Nov 9];14(7):721–30. Available from: https://pubmed.ncbi.nlm.nih.gov/15054051/

9. Cox SR, Ritchie SJ, Tucker-Drob EM, Liewald DC, Hagenaars SP, Davies G, et al. Ageing and brain white matter structure in 3,513 UK Biobank participants. Nat Commun [Internet]. 2016 Dec 15 [cited 2020 Nov 9];7. Available from: /pmc/articles/PMC5172385/?report=abstract

10. Bird A. Perceptions of epigenetics. Vol. 447, Nature. Nature Publishing Group; 2007. p. 396–8.

11. C Greenberg M V, Bourchis D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol [Internet]. [cited 2020 Nov 9];

554 Available from: www.nature.com/nrm

555 12. Barbu MC, Shen X, Walker RM, Howard DM, Evans KL, Whalley HC, et al. Epigenetic
556 prediction of major depressive disorder. Mol Psychiatry [Internet]. 2020 Jun 10 [cited
557 2020 Jun 11];1–12. Available from: http://www.nature.com/articles/s41380-020-
558 0808-3

559 13. Jovanova OS, Nedeljkovic I, Spieler D, Walker RM, Liu C, Luciano M, et al. DNA
560 methylation signatures of depressive symptoms in middle-aged and elderly persons:
561 Meta-analysis of multiethnic epigenome-wide studies. JAMA Psychiatry.
562 2018;75(9):949–59.

563 14. Hannon E, Dempster E, Viana J, Burrage J, Smith AR, Macdonald R, et al. An integrated
564 genetic-epigenetic analysis of schizophrenia: Evidence for co-localization of genetic
565 associations and differential DNA methylation. Genome Biol [Internet]. 2016 Aug 30
566 [cited 2020 Nov 9];17(1):176. Available from:
567 https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-1041-x

568 15. Montano C, Taub MA, Jaffe A, Briem E, Feinberg JI, Trygvadottir R, et al. Association
569 of DNA methylation differences with schizophrenia in an epigenome-wide association
570 study. JAMA Psychiatry [Internet]. 2016 May 1 [cited 2020 Nov 9];73(5):506–14.
571 Available from: https://jamanetwork.com/

572 16. Horvath S, Raj K. DNA methylation-based biomarkers and the epigenetic clock theory
573 of ageing. Nat Rev Genet [Internet]. 2018;19(6):371–84. Available from:
574 http://dx.doi.org/10.1038/s41576-018-0004-3

575 17. Ladd-Acosta C, Pevsner J, Sabunciyan S, Yolken RH, Webster MJ, Dinkins T, et al. DNA
576 methylation signatures within the human brain. Am J Hum Genet [Internet]. 2007 Dec
577 1 [cited 2020 Nov 9];81(6):1304–15. Available from: www.ajhg.org

578 18. Wheater ENW, Stoye DQ, Cox SR, Wardlaw JM, Drake AJ, Bastin ME, et al. DNA
579 methylation and brain structure and function across the life course: A systematic
580 review. Vol. 113, Neuroscience and Biobehavioral Reviews. Elsevier Ltd; 2020. p. 133–
581 56.

582 19. Jia T, Chu C, Liu Y, van Dongen J, Papastergios E, Armstrong NJ, et al. Epigenome-wide
583 meta-analysis of blood DNA methylation and its association with subcortical volumes:
584 findings from the ENIGMA Epigenetics Working Group. Mol Psychiatry. 2019;

585 20. Wockner LF, Noble EP, Lawford BR, Young RMD, Morris CP, Whitehall VLJ, et al.
586 Genome-wide DNA methylation analysis of human brain tissue from schizophrenia
587 patients. Transl Psychiatry [Internet]. 2014 Jan 7 [cited 2020 Nov 9];4(1):339.
588 Available from: www.nature.com/tp

589 21. Liu J, Siyahhan Julnes P, Chen J, Ehrlich S, Walton E, Calhoun VD. The association of
590 DNA methylation and brain volume in healthy individuals and schizophrenia patients.
591 Schizophr Res [Internet]. 2015 Dec 1 [cited 2020 Nov 9];169(1–3):447–52. Available
592 from: /pmc/articles/PMC4681600/?report=abstract

593 22. Davis EG, Humphreys KL, McEwen LM, Sacchet MD, Camacho MC, MacIsaac JL, et al.
594 Accelerated DNA methylation age in adolescent girls: associations with elevated

diurnal cortisol and reduced hippocampal volume. Transl Psychiatry [Internet]. 2017 Aug 29 [cited 2020 Nov 9];7(8):e1223. Available from: www.nature.com/tp

23. De Jager PL, Srivastava G, Lunnon K, Burgess J, Schalkwyk LC, Yu L, et al. Alzheimer's disease: Early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat Neurosci [Internet]. 2014 Sep 1 [cited 2020 Nov 9];17(9):1156–63. Available from: /pmc/articles/PMC4292795/?report=abstract

24. Masliah E, Dumaop W, Galasko D, Desplats P. Distinctive patterns of DNA methylation associated with Parkinson disease: Identification of concordant epigenetic changes in brain and peripheral blood leukocytes. Epigenetics [Internet]. 2013 [cited 2020 Nov 9];8(10):1030–8. Available from: /pmc/articles/PMC3891683/?report=abstract

25. van Erp TGM, Walton E, Hibar DP, Schmaal L, Jiang W, Glahn DC, et al. Cortical Brain Abnormalities in 4474 Individuals With Schizophrenia and 5098 Control Subjects via the Enhancing Neuro Imaging Genetics Through Meta Analysis (ENIGMA) Consortium. Biol Psychiatry [Internet]. 2018 Nov 1 [cited 2020 Nov 26];84(9):644–54. Available from: /pmc/articles/PMC6177304/?report=abstract

26. Schmaal L, Pozzi E, C. Ho T, van Velzen LS, Veer IM, Opel N, et al. ENIGMA MDD: seven years of global neuroimaging studies of major depression through worldwide data sharing. Vol. 10, Translational Psychiatry. Springer Nature; 2020.

27. Wright IC, Rabe-Hesketh S, Woodruff PWR, David AS, Murray RM, Bullmore ET. Meta-analysis of regional brain volumes in schizophrenia. Am J Psychiatry [Internet]. 2000 Jan 1 [cited 2020 Nov 26];157(1):16–25. Available from: https://ajp.psychiatryonline.org/doi/abs/10.1176/ajp.157.1.16

28. Walton E, Hass J, Liu J, Roffman JL, Bernardoni F, Roessner V, et al. Correspondence of DNA methylation between blood and brain tissue and its application to schizophrenia research. Schizophr Bull [Internet]. 2016 Mar 1 [cited 2020 Nov 9];42(2):406–14. Available from: https://academic.oup.com/schizophreniabulletin/article/42/2/406/2518914

29. 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

30. 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 Genet [Internet]. 2006 Oct 2 [cited 2020 Nov 9];7(1):74. Available from: http://bmcmedgenet.biomedcentral.com/articles/10.1186/1471-2350-7-74

31. Smith BH, Campbell A, Linksted P, Fitzpatrick B, Jackson C, Kerr SM, et al. Cohort Profile : Generation Scotland : Scottish Family Health Study (GS : SFHS). The study , its participants and their potential for genetic research on health and illness. Int J Epidemiol. 2013;(July 2012):689–700.

32. Navrady LB, Wolters MK, MacIntyre DJ, Clarke TK, Campbell AI, Murray AD, et al. Cohort profile: Stratifying Resilience and Depression Longitudinally (STRADL): A questionnaire follow-up of Generation Scotland: Scottish Family Health Study (GS:

SFHS). Int J Epidemiol [Internet]. 2018 Feb 1 [cited 2020 Nov 9];47(1):13-14g. Available from: https://academic.oup.com/ije/article/47/1/13/3978732

33. Habota T, Sandu A-L, Waiter GD, McNeil CJ, Steele JD, Macfarlane JA, et al. Cohort profile for the STratifying Resilience and Depression Longitudinally (STRADL) study: A depression-focused investigation of Generation Scotland, using detailed clinical, cognitive, and neuroimaging assessments. Wellcome Open Res [Internet]. 2019 Nov 25 [cited 2020 Nov 9];4:185. Available from: https://doi.org/10.12688/wellcomeopenres.15538.1https://doi.org/10.12688/wellcomeopenres.15538.1

34. Zugman A, Harrewijn A, Cardinale EM, Zwiebel H, Freitag GF, Werwath KE, et al. Mega-analysis methods in ENIGMA: The experience of the generalized anxiety disorder working group [Internet]. Human Brain Mapping. John Wiley and Sons Inc.; 2020 [cited 2020 Nov 26]. Available from: https://onlinelibrary.wiley.com/doi/full/10.1002/hbm.25096

35. Leffondré K, Abrahamowicz M, Siemiatycki J, Rachet B. Modeling smoking history: A comparison of different approaches. Am J Epidemiol [Internet]. 2002 Nov 1 [cited 2020 Nov 9];156(9):813–23. Available from: https://pubmed.ncbi.nlm.nih.gov/12396999/

36. 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.

37. Gigantesco A, Morosini P. Development, reliability and factor analysis of a self-administered questionnaire which originates from the World Health Organization's Composite International Diagnostic Interview - Short Form (CIDI-SF) for assessing mental disorders. Clin Pract Epidemiol Ment Heal [Internet]. 2008 Apr 10 [cited 2020 Nov 9];4:8. Available from: /pmc/articles/PMC2329624/?report=abstract

38. Amador C, Zeng Y, Barber M, Walker R, Campbell A, McIntosh AM, et al. Genome-wide methylation data improves dissection of the effect of smoking on body mass index. bioRxiv [Internet]. 2021 Apr 21 [cited 2021 Oct 12];2020.10.08.329672. Available from: https://www.biorxiv.org/content/10.1101/2020.10.08.329672v2

39. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genomics Data [Internet]. 2016 Sep 1 [cited 2020 Nov 9];9:22–4. Available from: /pmc/articles/PMC4909830/?report=abstract

40. Zeng Y, Amador C, Gao C, Walker RM, Morris SW, Campbell A, et al. Lifestyle and Genetic Factors Modify Parent-of-Origin Effects on the Human Methylome. bioRxiv [Internet]. 2021 Jun 30 [cited 2021 Oct 12];2021.06.28.450122. Available from: https://www.biorxiv.org/content/10.1101/2021.06.28.450122v1

41. Barbu M, Huider F, Campbell A, Amador C, Adams M, Lynall M, et al. Methylome-wide association study of antidepressant use in Generation Scotland and the Netherlands Twin Register implicates the innate immune system. medRxiv [Internet]. 2021 Oct 15 [cited 2021 Oct 18];2020.10.06.20207621. Available from:

https://www.medrxiv.org/content/10.1101/2020.10.06.20207621v2

42. 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 [Internet]. 2016 Feb 1 [cited 2020 Nov 9];12(2):e1005804. Available from: www.ed.ac.uk/

43. Gadd DA, Hillary RF, McCartney DL, Zaghlool SB, Stevenson AJ, Nangle C, et al. Epigenetic scores for the circulating proteome replicate protein-disease predictions as tools for biomarker discovery. bioRxiv [Internet]. 2021 Apr 28 [cited 2021 Oct 12];8:2020.12.01.404681. Available from: https://www.biorxiv.org/content/10.1101/2020.12.01.404681v2

44. Gadd DA, Hillary RF, McCartney DL, Shi L, McGeachan RI, Stolicyn A, et al. Epigenome and phenome study reveals circulating markers pertinent to brain health. medRxiv [Internet]. 2021 Sep 6 [cited 2021 Oct 12];2021.09.03.21263066. Available from: https://www.medrxiv.org/content/10.1101/2021.09.03.21263066v1

45. Min JL, Hemani G, Smith GD, Relton C, Suderman M. Meffil: Efficient normalization and analysis of very large DNA methylation datasets. Bioinformatics [Internet]. 2018 Dec 1 [cited 2021 May 31];34(23):3983–9. Available from: https://academic.oup.com/bioinformatics/article/34/23/3983/5042224

46. Fortin J-P, Fertig E, Hansen K. shinyMethyl: interactive quality control of Illumina 450k DNA methylation arrays in R. F1000Research [Internet]. 2014 Jul 30 [cited 2020 Nov 9];3:175. Available from: https://doi.org/10.12688/f1000research.4680.1

47. Du P, Kibbe WA, Lin SM. lumi: A pipeline for processing Illumina microarray. Bioinformatics [Internet]. 2008 Jul [cited 2020 Nov 9];24(13):1547–8. Available from: https://pubmed.ncbi.nlm.nih.gov/18467348/

48. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res [Internet]. 2015 Jan 6 [cited 2020 Nov 9];43(7):e47. Available from: /pmc/articles/PMC4402510/?report=abstract

49. 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.

50. Mendelson MM, Marioni RE, Joehanes R, Liu C, Hedman ÅK, Aslibekyan S, et al. Association of Body Mass Index with DNA Methylation and Gene Expression in Blood Cells and Relations to Cardiometabolic Disease : A Mendelian Randomization Approach. PLOS Med. 2017;1–30.

51. Liu C, Marioni RE, Hedman AK, Pfeiffer L, Tsai PC, Reynolds LM, et al. A DNA methylation biomarker of alcohol consumption. Mol Psychiatry [Internet]. 2018 Feb 1 [cited 2021 Mar 29];23(2):422–33. Available from: www.nature.com/mp

52. Joehanes R, Just AC, Marioni RE, Pilling LC, Reynolds LM, Mandaviya PR, et al. Epigenetic Signatures of Cigarette Smoking. Circ Cardiovasc Genet [Internet]. 2016 Oct 1 [cited 2021 Mar 29];9(5):436–47. Available from: http://circgenetics.ahajournals.org/lookup/suppl/doi:10.1161/CIRCGENETICS.116.001

506/-/DC1.

53. Hannon E, Lunnon K, Schalkwyk L, Mill J. Interindividual methylomic variation across blood, cortex, and cerebellum: Implications for epigenetic studies of neurological and neuropsychiatric phenotypes. Epigenetics [Internet]. 2015 [cited 2021 Jan 22];10(11):1024–32. Available from: /pmc/articles/PMC4844197/?report=abstract

54. Hansen K. IlluminaHumanMethylation450kanno. ilmn12. hg19: annotation for Illumina's 450k methylation arrays. 2016;

55. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics [Internet]. 2010 Jul 8 [cited 2020 Nov 9];26(17):2190–1. Available from: https://pubmed.ncbi.nlm.nih.gov/20616382/

56. Phipson B, Maksimovic J, Oshlack A. MissMethyl: An R package for analyzing data from Illumina's HumanMethylation450 platform. Bioinformatics [Internet]. 2016 Jan 15 [cited 2020 Nov 9];32(2):286–8. Available from: https://pubmed.ncbi.nlm.nih.gov/26424855/

57. 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.

58. 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

59. 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 [cited 2020 Jul 29];8(1):1–12. Available from: www.nature.com/naturecommunications

60. H L, M Z, Z S, J L, S Z, D L, et al. Genome-wide haplotype association study identify the FGFR2 gene as a risk gene for acute myeloid leukemia. Oncotarget [Internet]. 2017 [cited 2021 Sep 16];8(5):7891–9. Available from: https://pubmed.ncbi.nlm.nih.gov/27903959/

61. G K, G B, PR L, S G, K B, MK F, et al. Leveraging Polygenic Functional Enrichment to Improve GWAS Power. Am J Hum Genet [Internet]. 2019 Jan 3 [cited 2021 Sep 16];104(1):65–75. Available from: https://pubmed.ncbi.nlm.nih.gov/30595370/

62. Hill WD, Marioni RE, Maghzian O, Ritchie SJ, Hagenaars SP, McIntosh AM, et al. A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. Mol Psychiatry [Internet]. 2019;24(2):169–81. Available from: http://dx.doi.org/10.1038/s41380-017-0001-5

63. JJ L, R W, A O, E K, O M, M Z, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet [Internet]. 2018 Aug 1 [cited 2021 Sep 16];50(8):1112–21. Available from: https://pubmed.ncbi.nlm.nih.gov/30038396/

64. Liu M, Jiang Y, Wedow R, Li Y, Brazel DM, Chen F, et al. Association studies of up to 1.2 million individuals yield new insights into the genetic etiology of tobacco and

alcohol use [Internet]. Vol. 51, Nature Genetics. Nature Publishing Group; 2019 [cited 2020 Nov 9]. p. 237–44. Available from: /pmc/articles/PMC6358542/?report=abstract

65. Karlsson Linnér R, Biroli P, Kong E, Meddens SFW, Wedow R, Fontana MA, et al. Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. Nat Genet [Internet]. 2019 Feb 1 [cited 2020 Nov 9];51(2):245–57. Available from: /pmc/articles/PMC6713272/?report=abstract

66. Shen J, Gilmore EC, Marshall CA, Haddadin M, Reynolds JJ, Eyaid W, et al. Mutations in PNKP cause microcephaly, seizures and defects in DNA repair. Nat Genet [Internet]. 2010 Mar [cited 2020 Nov 9];42(3):245–9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2835984/

67. Lam M, Hill WD, Trampush JW, Yu J, Knowles E, Davies G, et al. Pleiotropic Meta-Analysis of Cognition, Education, and Schizophrenia Differentiates Roles of Early Neurodevelopmental and Adult Synaptic Pathways. Am J Hum Genet [Internet]. 2019 Aug 1 [cited 2020 Nov 9];105(2):334–50. Available from: /pmc/articles/PMC6699140/?report=abstract

68. 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.

69. Howard DM, Adams MJ, Clarke T-K, 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. bioRxiv. 2018.

70. Day FR, Ong KK, Perry JRB. Elucidating the genetic basis of social interaction and isolation. Nat Commun [Internet]. 2018 Dec 1 [cited 2020 Nov 10];9(1). Available from: https://pubmed.ncbi.nlm.nih.gov/29970889/

71. Nagel M, Jansen PR, Stringer S, Watanabe K, De Leeuw CA, Bryois J, et al. Meta-analysis of genome-wide association studies for neuroticism in 449,484 individuals identifies novel genetic loci and pathways. Nat Genet [Internet]. 2018 Jul 1 [cited 2020 Nov 10];50(7):920–7. Available from: https://pubmed.ncbi.nlm.nih.gov/29942085/

72. BML B, R J, HF I, J van D, A A, MP van de W, et al. Multivariate genome-wide analyses of the well-being spectrum. Nat Genet [Internet]. 2019 Mar 1 [cited 2021 Sep 16];51(3):445–51. Available from: https://pubmed.ncbi.nlm.nih.gov/30643256/

73. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. Mol Autism [Internet]. 2017 [cited 2020 Nov 10];8:21. Available from: https://pubmed.ncbi.nlm.nih.gov/28540026/

74. Wang H, Yang J, Schneider JA, De Jager PL, Bennett DA, Zhang HY. Genome-wide interaction analysis of pathological hallmarks in Alzheimer's disease. Neurobiol Aging [Internet]. 2020 Sep 1 [cited 2020 Nov 19];93:61–8. Available from: https://pubmed.ncbi.nlm.nih.gov/32450446/

75. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and

polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet. 2018;50(8):1112–21.

76. Jilani A, Ramotar D, Slack C, Ong C, Yang XM, Scherer SW, et al. Molecular cloning of the human gene, PNKP, encoding a polynucleotide kinase 3'-phosphatase and evidence for its role in repair of DNA strand breaks caused by oxidative damage. J Biol Chem [Internet]. 1999 Aug 20 [cited 2020 Nov 9];274(34):24176–86. Available from: http://www.jbc.org/

77. Gatti M, Magri S, Nanetti L, Sarto E, Di Bella D, Salsano E, et al. From congenital microcephaly to adult onset cerebellar ataxia: Distinct and overlapping phenotypes in patients with *PNKP* gene mutations. Am J Med Genet Part A [Internet]. 2019 Nov 22 [cited 2020 Nov 9];179(11):2277–83. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/ajmg.a.61339

78. Dumitrache LC, McKinnon PJ. Polynucleotide kinase-phosphatase (PNKP) mutations and neurologic disease. Mech Ageing Dev [Internet]. 2017 Jan 1 [cited 2021 Jul 2];161(Pt A):121–9. Available from: /pmc/articles/PMC5161711/

79. Shimada M, Dumitrache LC, Russell HR, McKinnon PJ. Polynucleotide kinase–phosphatase enables neurogenesis via multiple DNA repair pathways to maintain genome stability . EMBO J [Internet]. 2015 Oct [cited 2021 Jul 2];34(19):2465–80. Available from: /pmc/articles/PMC4601665/

80. Spiers H, Hannon E, Schalkwyk LC, Smith R, Wong CCY, O'Donovan MC, et al. Methylomic trajectories across human fetal brain development. Genome Res [Internet]. 2015 Mar 1 [cited 2020 Nov 9];25(3):338–52. Available from: http://www.genome.org/cgi/doi/10.1101/gr.180273.114.

81. Samaan G, Yugo D, Rajagopalan S, Wall J, Donnell R, Goldowitz D, et al. Foxn3 is essential for craniofacial development in mice and a putative candidate involved in human congenital craniofacial defects. Biochem Biophys Res Commun. 2010;400:60–5.

82. Yousefi P, Huen K, Davé V, Barcellos L, Eskenazi B, Holland N. Sex differences in DNA methylation assessed by 450K BeadChip in newborns. BMC Genomics [Internet]. 2015 Nov 9 [cited 2020 Nov 9];16(1):911. Available from: https://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-015-2034-y

83. Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, et al. Contribution of SHANK3 mutations to autism spectrum disorder. Am J Hum Genet [Internet]. 2007 Dec 1 [cited 2020 Nov 9];81(6):1289–97. Available from: www.ajhg.org

84. Mancuso L, Fornito A, Costa T, Ficco L, Liloia D, Manuello J, et al. A meta-analytic approach to mapping co-occurrent grey matter volume increases and decreases in psychiatric disorders. Neuroimage. 2020 Nov 15;222:117220.

85. Clark SL, Hattab MW, Chan RF, Shabalin AA, Han LKM, Zhao M, et al. A methylation study of long-term depression risk. Mol Psychiatry [Internet]. 2020 Jun 1 [cited 2020 Nov 9];25(6):1334–43. Available from: /pmc/articles/PMC7061076/?report=abstract