

DNA methylation and brain dysmaturation in preterm infants

Emily N. W. Wheeler¹, Paola Galdi¹, Daniel L. McCartney², Manuel Blesa¹, Gemma Sullivan¹, David Q. Stoye¹, Gillian Lamb¹, Sarah Sparrow¹, Lee Murphy², Nicola Wrobel², Alan J. Quigley³, Scott Semple^{4,5}, Michael J. Thrippleton^{4,6}, Joanna M. Wardlaw⁶, Mark E. Bastin⁶, Riccardo E. Marioni², Simon R. Cox⁷, James P. Boardman^{1,6}

Author affiliations:

¹MRC Centre for Reproductive Health, The University of Edinburgh, Queen's Medical Research Institute, Edinburgh EH16 4TJ, UK

²Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK.

³Department of Paediatric Radiology, Royal Hospital for Sick Children, NHS Lothian, UK.

⁴Edinburgh Imaging, University of Edinburgh, EH16 4SB, Edinburgh, UK,

⁵Centre for Cardiovascular Science, The University of Edinburgh, Queen's Medical Research Institute, Edinburgh EH16 4TJ, UK

⁶Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK

⁷Department of Psychology, The University of Edinburgh, Edinburgh, UK

Correspondence to: Professor James P Boardman

MRC Centre for Reproductive Health,

Queen's Medical Research Institute,

47 Little France Crescent,

Edinburgh EH16 4TJ,

United Kingdom.

E-mail: james.boardman@ed.ac.uk

Running title: DNAm and brain development in neonates

Keywords: brain; neonate; MRI; DNA methylation; development

Abbreviations: ADHD = attention deficit hyperactivity disorder; FA = fractional anisotropy; IQ = intelligence quotient; MD = mean diffusivity; NODDI = neurite orientation dispersion and density imaging; NDI = neurite density index; PS = peak width skeletonised; PCA = principal component analysis; SVA = surrogate variable analysis.

Abstract

Preterm birth is associated with dysconnectivity of structural brain networks and is a leading cause of neurocognitive impairment in childhood. Variation in DNA methylation (DNAm) is associated with early exposure to extrauterine life but there has been little research exploring its relationship with brain development.

Using genome-wide DNA methylation data from saliva of 258 neonates, we investigated the impact of gestational age on the methylome and performed functional analysis to identify enriched gene sets from probes that contributed to differentially methylated probes (DMPs) or regions (DMRs). We tested the hypothesis that variation in DNAm could underpin the association between preterm birth and atypical brain development by linking DMPs with measures of white matter connectivity derived from diffusion MRI metrics: peak width of skeletonised mean diffusivity (PSMD), fractional anisotropy (PSFA) and neurite density index (PSNDI).

Gestational age at birth was associated with widespread differential methylation, with genome-wide significant associations observed for 8,870 CpG probes ($p < 3.6 \times 10^{-8}$) and 1,767 differentially methylated regions. Functional analysis identified 14 enriched gene ontology terms pertaining to cell-cell contacts and cell-extracellular matrix contacts. Principal component analysis of probes with genome-wide significance revealed a first principal component (PC1) that explained 23.5% of variance in DNAm, and this was negatively associated with gestational age at birth. PC1 was associated with PSMD ($\beta = 0.349$, $p = 8.37 \times 10^{-10}$) and PSNDI ($\beta = 0.364$, $p = 4.15 \times 10^{-5}$), but not with PSFA ($\beta = -0.035$, $p = 0.510$); these relationships mirrored the imaging metrics' associations with gestational age at birth.

Gestational age at birth has a profound and widely distributed effect on the neonatal saliva methylome. Enriched gene ontology terms related to cell-cell contacts reveal pathways that could mediate the effect of early life environmental exposures on development. Finally, associations between differential DNAm and image markers of white matter tract microstructure suggest that variation in DNAm may provide a link between preterm birth and the dysconnectivity of developing brain networks that characterises atypical brain development in preterm infants.

Introduction

Preterm birth, defined as birth at <37 weeks of gestation, affects around 11% of births worldwide¹ and is a leading cause of neurodevelopmental and cognitive problems that extend across the life course. These include autism spectrum disorder, social difficulties, language impairment, ADHD, reduced IQ, educational underachievement, and psychiatric diagnoses.²⁻⁹

The neural phenotypes that underlie long-term functional impairment include diffuse white matter injury and subsequent dysmaturation of white matter and grey matter neuroaxonal structures, collectively termed the ‘encephalopathy of prematurity’.¹⁰ A consequence of the encephalopathy is generalised dysconnectivity of developing structural networks, which can be inferred from diffusion MRI (dMRI) and neurite orientation dispersion and density imaging (NODDI) during the neonatal period.¹¹⁻¹⁵ Specifically, normal maturation is characterized by a reduction in mean diffusivity (MD) and increases in both fractional anisotropy (FA) and neurite density index (NDI) in white matter; but MD is increased and FA and NDI are decreased in preterm infants at term equivalent age, compared with healthy controls infants born at term.¹⁶ These changes reflect an increase in water content and a decrease in white matter organization in preterm infants. Peak width of skeletonised mean diffusivity (PSMD) is a method for histogram-based calculation of MD distribution across the entire white matter skeleton; it provides a measure of generalized white matter microstructure, is robust to scanner variation, and is predictive of cognition in later life.¹⁷⁻¹⁹ In previous work we extended the histogram model to neonatal data and included other dMRI and NODDI metrics. We found that PSMD and PSNDI are altered in preterm infants at term equivalent age, and that histogram-based measures have utility for investigating upstream determinants of dysmaturation such as systemic inflammation.^{20,21}

The mechanisms that link the environmental stress of preterm birth with atypical brain development are uncertain. Variation in DNA methylation (DNAm) is a possible mechanism; DNAm is involved in the regulation of gene expression and cell fate during fetal brain development.²² Alterations in DNAm contribute to the pathogenesis of neurodevelopmental disorders (Rett syndrome,²³ Immunodeficiency, Centromeric region instability, Facial anomalies [ICE] syndrome,²⁴ and Angelman and Prader-Willi syndromes)²⁵. There is growing evidence that differential DNAm can mediate the effect of environmental pressures

on brain structure and function across the life course.^{26,27} The neonatal methylome is sensitive to prenatal factors such as maternal smoking,²⁸ maternal body mass index,²⁹ as well as birth weight.³⁰ It is altered in association with co-morbidities of preterm birth,^{26,31,32} and there is some evidence for legacy differences in the methylome two decades after preterm birth.³³ A meta-analysis investigating DNAm from umbilical cord blood identified widespread differential methylation associated with gestational age at birth across the gestational age range of 27-42 weeks, as measured on the Illumina 450k array.³⁴ However, due to differences in the cellular composition of samples, epigenetic signatures observed in different tissues are likely to be distinct.³⁵ The main cellular component of saliva, buccal epithelium, may be more representative of the brain than umbilical cord blood because of ectodermal origin.³⁶⁻³⁸

Here, our first aim was to determine whether gestational age at birth was associated with variation in the salivary methylome and to characterise the biological pathways implicated in the DNAm response to preterm birth. Our second aim was to investigate whether the DNAm signal of preterm birth explains variance in measures of white matter microstructure. We tested the hypotheses that gestational age at birth is associated with widespread differential methylation, and that DNAm contributes to variance in peak width skeletonised metrics of white matter connectivity, apparent during the neonatal period.

Materials and methods

Participants

Preterm and term born infants delivered at the Royal Infirmary of Edinburgh, UK were recruited to the Theirworld Edinburgh Birth Cohort, a longitudinal study designed to investigate the effect of preterm birth on brain development.³⁹ Cohort exclusion criteria were major congenital malformations, chromosomal abnormalities, congenital infection, overt parenchymal lesions (cystic periventricular leukomalacia, hemorrhagic parenchymal infarction) or post-hemorrhagic ventricular dilatation. Ethical approval has been obtained from the National Research Ethics Service, South East Scotland Research Ethics Committee (11/55/0061, 13/SS/0143 and 16/SS/0154). Informed consent was obtained from a person with parental responsibility for each participant. The study was conducted according to the

principles of the Declaration of Helsinki. DNAm data were available from 258 neonates, 214 of whom also had successful dMRI acquisition.

DNA extraction and methylation measurement

Saliva obtained at term equivalent age was collected in Oragene OG-575 Assisted Collection kits, by DNA Genotek, and DNA extracted using prepIT.L2P reagent (DNA Genotek, Ontario, Canada). DNA was bisulfite converted and methylation levels were measured using Illumina HumanMethylationEPIC BeadChip (Illumina, San Diego, CA, USA) at the Edinburgh Clinical Research Facility (Edinburgh, UK). The arrays were imaged on the Illumina iScan or HiScan platform and genotypes were called automatically using GenomeStudio Analysis software version 2011.1 (Illumina). DNAm was analysed in two batches.

DNA Methylation pre-processing

Raw intensity (.idat) files were read into R environment (version 3.4.4) using minfi. `wateRmelon` and `minfi` used for preprocessing, quality control and normalisation.^{40,41} The `pfilter` function in `wateRmelon` was used to exclude: samples with 1 % of sites with a detection p-value greater than 0.05; sites with beadcount <3 in 5% of samples; and sites with 1% of samples with detection p value >0.05. Cross hybridising probes and probes targeting single nucleotide polymorphisms with overall minor allele frequency ≥ 0.05 were also removed.⁴² Control probes were also removed. Samples were removed if there was a mismatch between predicted sex (`minfi`) and recorded sex ($n = 3$). Data was `danet` normalised which includes background correction and dye bias correction.⁴¹ Saliva contains different cells types, including buccal epithelial cells and leukocytes. Epithelial cell proportions were estimated with epigenetic dissection of intra-sample heterogeneity with the reduced partial correlation method implemented in the R package `EpiDISH`.⁴³ Probes located on sex chromosomes were removed prior to analysis. Data from one of each twin pair was removed randomly ($n=20$).

MRI acquisition

This study incorporates data from two phases of MRI acquisition. MRI was obtained at the same appointment as saliva sample collection for DNAm analysis. Structural and diffusion MRI were performed on 93 infants using a MAGNETOM Verio 3T clinical MRI scanner

(Siemens Healthcare GmbH, Erlangen, Germany) and 12-channel phased-array head coil, which were used to acquire: dMRI employed a protocol consisting of 11 T2- and 64 diffusion-weighted ($b = 750 \text{ s/mm}^2$) single-shot spin-echo echo planar imaging (EPI) volumes acquired with 2 mm isotropic voxels ($TE = 106 \text{ ms}$ and $TR = 7300 \text{ ms}$).

For the second phase ($n=121$), structural and diffusion MRI were performed on 121 infants using a MAGNETOM Prisma 3T clinical MRI scanner (Siemens Healthcare GmbH, Erlangen, Germany) and 16-channel phased-array pediatric head and neck coil were used to acquire dMRI in two separate acquisitions: the first consisted of 8 baseline volumes ($b=0 \text{ s/mm}^2$ [b_0]) and 64 volumes with $b=750 \text{ s/mm}^2$; the second consisted of 8 b_0 , 3 volumes with $b=200 \text{ s/mm}^2$, 6 volumes with $b=500 \text{ s/mm}^2$ and 64 volumes with $b = 2500 \text{ s/mm}^2$. An optimal angular coverage for the sampling scheme was applied.⁴⁴ In addition, an acquisition of 3 b_0 volumes with an inverse phase encoding direction was performed. All dMRI volumes were acquired using single-shot spin-echo EPI with 2-fold simultaneous multislice and 2-fold in-plane parallel imaging acceleration and 2 mm isotropic voxels; all three diffusion acquisitions had the same parameters (TR/TE 3500/78.0ms). Images affected by motion artifact were re-acquired multiple times as required; dMRI acquisitions were repeated if signal loss was seen in 3 or more volumes.

Infants were fed and wrapped and allowed to sleep naturally in the scanner without sedation. Pulse oximetry, electrocardiography and temperature were monitored. Flexible earplugs and neonatal earmuffs (MiniMuffs, Natus) were used for acoustic protection. All scans were supervised by a doctor or nurse trained in neonatal resuscitation. Structural images were reported by an experienced pediatric radiologist (A.J.Q.)

dMRI pre-processing

Diffusion images that were acquired on the MAGNETOM Verio scanner were denoised using a Marchenko-Pastur-PCA-based algorithm⁴⁵⁻⁴⁷; eddy current distortion and head movement were corrected using outlier replacement⁴⁸⁻⁵⁰; bias field inhomogeneity correction was performed by calculating the bias field of the mean b_0 volume and applying the correction to all the volumes.⁵¹ FA and MD were calculated from the dMRI data.

The two dMRI acquisitions from the MAGNETOM Prisma scanner were first concatenated and then denoised using a Marchenko-Pastur-PCA-based algorithm⁴⁵⁻⁴⁷; eddy current, head movement and EPI geometric distortions were corrected using outlier replacement and slice-to-volume registration^{48-50,52}; bias field inhomogeneity correction was performed by calculating the bias field of the mean b0 volume and applying the correction to all the other volumes.⁵¹ From the dMRI data we calculated the three eigenvalues and eigenvectors of the water diffusion tensor, and NODDI (Bingham distribution) parametric maps using cuDIMOT (intracellular volume fraction [NDI] and the overall orientation dispersion index [ODI_{TOT}]).^{12,13,53} FA and MD were calculated using single-shell data to match the Verio scanner.

Peak width of skeletonized water diffusion parameters

All the subjects were registered to the Edinburgh Neonatal Atlas (ENA50) using DTI-TK.²⁰ The diffusion tensor derived maps of each subject (FA and MD) were calculated after registration; NDI was then propagated to the template space using the previously calculated transformations. The data was skeletonized using the ENA50 skeleton and then multiplied by a custom mask. Finally, the peak width of the histogram of values computed within the skeletonized maps was calculated as the difference between between the 95th and 5th percentiles.^{17,20}

Statistical analysis

Epigenome-wide association analyses

Unless otherwise stated, analysis was completed in R version 3.4.4.⁵⁴ Surrogate variable analysis of the data matrix was carried out, in order to adjust for potential confounders such as batch, using the *sva* function in the *sva* package in R.^{55,56} A fully adjusted model was specified prior to SVA to retain signal explained by biological variables of interest: CpG ~ Gestational age at birth + Gestational age at scan + Birthweight z score + Maternal smoking + Sex + Epithelial cells. SVA identified 17 significant surrogate variables which were subsequently included in the analysis.

EWAS was performed using the *limma* package in R.⁵⁷ Beta values of each of 776025 CpG sites were regressed (as dependent variables) on gestational age at birth using linear regression. Covariates were added to adjust for sex, birthweight z score, gestational age at sample collection, maternal smoking, estimated epithelial cell proportions and surrogate

variables. A significance threshold of 3.6×10^{-8} was selected, which represents genome-wide significance.⁵⁸

Differentially methylated region analysis

Differentially methylated regions were assessed using the *dmrff* function in the *dmrff* package in R.⁵⁹ Here a differentially methylated region is a region containing two or above sites separated by ≤ 500 bp with EWAS analysis $p \leq 0.05$ and methylation changes in a consistent direction. Following *dmrff*'s subregion selection step, DMRs with Bonferroni-adjusted $p \leq 0.05$ were significant.

Gene set testing

Gene set enrichment analysis was carried out using the GO and KEGG databases, and using the *gometh* function in *missMethyl* package, which controls for multiple probe bias.⁶⁰ Sites that reached genome-wide significance in EWAS, and those that contributed to differentially methylated regions, were included in this analysis.

Principal component analysis

Principal components analysis (PCA) was conducted on CpG probes that reached genome-wide significance, using the *prcomp* function in R. CpGs were pre-corrected for the effects of biological covariates and surrogate variables via linear regression. The scree plot was visually inspected in order to select a principal component (eigenvalue > 1) to be carried forward for subsequent analysis.

Linear regression between DNAm and PS metrics

Pearson's correlation coefficient was used to assess the relationship between the first PC identified from PCA and gestational age at birth. This PC was used in linear regression models, as an independent variable, to test the associations between PSMD, PSNDI and PSFA with DNAm, conducted in R version 4.0.1.⁵⁴ In models testing PSMD and PSFA, MRI scanner was included as a binary covariate as MRI data from both phases of data collection were included. PSNDI was only available for one phase of data collection and so it was not necessary to include scanner as a covariate. We report standardised regression coefficients and p-values.

Data availability

The atlas with templates can be found at <https://git.ecdf.ed.ac.uk/jbrl/ena> and the code necessary to calculate histogram-based metrics is at <https://git.ecdf.ed.ac.uk/jbrl/psmd>. Requests for original image data will be considered through the BRAINS governance process: www.brainsimagebank.ac.uk.⁶¹ DNA methylation data will be deposited in NCBI's Gene Expression Omnibus upon publication.

Results

Cohort

DNAm data were collected from 311 neonates. Twenty-nine did not meet DNAm preprocessing QC criteria and were excluded. One participant with a congenital abnormality was removed, as were three participants whose sex predicted from DNAm data did not match their recorded sex. This group of 278 neonates included 20 sets of twins. After random removal of one twin from each set there was no evidence of imbalance for birthweight ($t=0.157$, $p=0.88$), or sex ($\text{Chi}=0.417$, $p=0.52$).

The study group consisted of 258 neonates: 155 participants were preterm and 103 were controls born at full term, see Table 1 for participant characteristics. Among the preterm infants, 38(25%) had bronchopulmonary dysplasia (defined as need for supplementary oxygen ≥ 36 weeks gestational age), 9(6%) developed necrotising enterocolitis requiring medical or surgical treatment, and 31(20%) had an episode of postnatal sepsis defined as either blood culture positivity with a pathogenic organism, or physician decision to treat for ≥ 5 days in the context of growth of coagulase negative staphylococcus from blood or a negative culture. Of the 258 participants with DNAm data, 214 also had MRI data.

Widespread differential saliva DNAm in association with gestational age at birth

We conducted an epigenome-wide association study whereby CpG methylation at 776,025 sites was regressed on gestational age at birth, controlling for birthweight z-score, infant sex, gestational age at sample collection, maternal smoking, estimated epithelial cell proportion and surrogate variables. The genomic inflation factor was 1.72 (Supplementary Fig. 1). Differential methylation in relation to gestational age at birth was identified at 8,870 CpG sites at genome-wide significance ($p < 3.6 \times 10^{-8}$), Fig. 1. Of these, 4,250 (47.9%) sites

demonstrated a positive association with gestational age, while 4,620 (52.1%) were negatively associated. Following Bonferroni adjustment, 1,767 DMRs corresponding to 4,664 CpG sites were significant $p < 0.05$. Of these, 11 had ten or more CpG sites contributing to the DMR. The largest DMR mapped to a genomic region that encodes two genes: *NNAT* and *BLCAP*. The 29 probes mapped to this region were all located within islands and positively associated with gestational age at birth, indicating that longer gestation corresponds to hypermethylation. Of the 10 most significant DMPs, 3 probes were localised to the *IRX4* gene, 1 probe to the *GAL3ST4* gene, and 1 to *LOXL4* (Table 2; Supplementary Fig. 2). The probes with the largest absolute magnitude effect size (top 5 hypermethylated and hypomethylated) mapped to the following genes: *IRX2*, *SMIM2*, *INTS1*, *HEATR2*, *ZBP1* and *UBXN11*, Table 3.

Pathways implicated in functional testing

Based on the 8,870 sites that met the genome-wide significance threshold ($p < 3.6 \times 10^{-8}$), no KEGG terms remained significant following FDR correction for multiple comparisons. Two Gene Ontology terms were enriched following FDR correction: one for anchoring (GO:0070161; $q = 0.0062$) and one for adherens junction (GO:0005912; $q = 0.0062$). In an analysis that incorporated all 11,752 distinct CpGs from both EWAS and DMR analysis, 14 GO terms were enriched, Table 4.

Gestational age at birth is associated with metrics of white matter microstructure in neonates

DNAm and PSMD and PSFA were both available for 214 infants (Supplementary Table 1), and DNAm and PSNDI were available for the 121 infants from phase 2 (Supplementary Table 2). Gestational age at birth was significantly associated with PSMD ($\beta = -0.602$, $p < 2 \times 10^{-16}$) and PSNDI ($\beta = -0.594$, $p = 2.17 \times 10^{-9}$), but not with PSFA ($\beta = -0.005$, $p = 0.933$), Fig. 2A and 2B; Table 5.

Differential DNAm is associated with white matter microstructure

The first unrotated PC derived from the 8,870 genome-wide significant CpGs accounted for 23.5% of the variance, the second PC accounted for 2.5%, Fig. 3A. There was no evidence of batch effects in the PCs (Supplementary Fig. 3). PC1 was significantly correlated with gestational age at birth ($r = -0.622$; $p < 2.2 \times 10^{-16}$), Fig. 3B. PC1 was also positively associated

with PSNDI ($\beta=0.364$, $p=4.15 \times 10^{-5}$), and in models adjusted for scanner it was positively associated with PSMD ($\beta=0.349$, $p=8.37 \times 10^{-10}$) but not PSFA ($\beta=-0.035$, $p=0.510$), Table 5. All models were adjusted for gestational age at scan.

Discussion

By studying a unique database of DNA linked to brain imaging in a population of preterm infants and healthy controls, we have identified extensive differential methylation in association with gestational age at birth and revealed an association between the principal axis of methylation and brain dysmaturation within the same sample. Differentially methylated regions and probes were distributed widely across the genome, indicating that gestation duration has a global effect on DNA methylation. Gene enrichment analysis of changes associated with gestational age identified gene sets pertaining to cell contacts and cytoskeleton. A single principal component that explained 23.5% of the variance in differential DNAm linked to preterm birth was closely associated with markers of generalised dysconnectivity across the white matter skeleton.

The data are consistent with studies that have reported associations between length of gestation and genome-wide variation in DNAm within fetal brain²² and umbilical cord blood^{34,62}; and widely distributed variation is reported in association with postmenstrual age at sampling of preterm infants (a proxy for GA at birth).³¹ The signature we identified in saliva sampled at term equivalent age included 233 probes that were previously shown to be differentially methylated in association with GA in a meta-analysis of umbilical cord blood samples that reported 8,899 gestation-dependent CpGs.³⁴ The limited overlap could be explained by differences in cellular composition of assessed tissues, different array types used to measure DNAm, or due to the time of sampling. Cord blood is collected at birth and so methylation changes at this time reflect fetal maturity and/or prenatal experience, whereas the methylation signature at term equivalent age reflects the allostatic load of early postnatal experiences as well as the prenatal environment. We chose to sample at term equivalent age because postnatal co-morbidities of preterm birth and NICU care practices such as painful stress exposures alter DNAm profiles, and because cumulative DNAm variations over this time period may link exposure to behavioural outcome in preterm infants.^{32,63,64}

Functional analyses of DMPs identified two enriched gene ontology terms, for adherens and anchoring junctions. When distinct probes that contributed to both DMPs and DMRs were combined, gene ontology analysis identified an additional 14 terms related to cell-cell adhesion, cell adhesions with the extracellular matrix, and signalling from the extracellular membrane; 12 of these were also identified in the meta-analysis of gestational age effects on DNAm obtained at birth from umbilical cord blood.³⁴ The most significant DMR mapped to a site encoding two genes including *NNAT*, which encodes the neural fate initiator neuronatin, the expression of which decreases throughout development⁶⁵; there was a positive association with increasing gestational age at birth. Hypomethylation of *NNAT* is associated with a corresponding increase in expression of neuronatin.⁶⁶ Furthermore, the candidacy of *NNAT* as a gene of interest whose expression may be modified by perinatal exposures is supported by previous EWAS of gestational age, maternal body mass index, maternal smoking, and schizophrenia.^{22,29,62,67–70}

Probes that demonstrated the largest magnitude of effect in association with gestational age mapped to genes previously associated with gestational age or maternal risk factors in EWAS.⁷⁰ Hypermethylated genes include: *ZBPI*, which was identified in EWAS investigating gestational age and hypertensive disorders of pregnancy^{62,69,71}; *INTS1*, which has been identified in EWAS of gestational age, hypertensive disorders of pregnancy, maternal body-mass-index, birth weight and breastfeeding duration.^{22,30,62,67,71,72} Hypomethylated genes included: *UBXN11*, which was identified in studies of gestational age^{22,62}; and *IRX4*. Three of the ten most significant DMPs mapped to the *IRX4* gene, all of which displayed a negative association with gestational age at birth. *IRX4* is associated with cardiac development in vertebrates, including humans.⁷³ Its homologues have been implicated in retinal axon guidance in zebrafish, and neural patterning in *Xenopus*^{74,75}, and it has been identified in previous EWAS of hypertensive disorders of pregnancy⁷¹ and prenatal maternal stress.⁷⁶

The novel pathways and genes implicated by EWAS studies of gestational age could provide a strong empirical basis for the selection of genes in targeted analyses in association with neuroimaging.²⁷ For example, one of the genes identified in our EWAS has been implicated in neurodevelopmental disorders; biallelic mutations in *INTS1* have been associated with a rare neurodevelopmental syndrome characterised by intellectual disability.^{77–79}

We used metrics of generalised white matter connectivity to assess relationships between DNAm and brain development because generalised dysconnectivity of structural networks is a hallmark of the preterm brain dysmaturation.^{10,14,80} PSNDI and PSMD were strongly associated with the first principal component of GA-dependent variation in DNAm, but PSFA was not. This suggests that variations in DNAm could contribute to the higher variability in water content and intra-axonal volume that characterise preterm brain dysmaturation.²⁰ We have previously reported that differential DNAm is associated with FA of the genu of the corpus callosum and tract shape of the right corticospinal tract²⁶; it is most likely that we did not observe an association between neither GA nor DNAm with PSFA because this metric is subject to histogram shift²⁰, meaning that although there are groupwise differences in FA values across the skeleton, the spread of values is the same. Associations between GA at birth and both DNAm and image markers of dysconnectivity, and between DNAm and image features, suggest that differential DNAm contributes, in part, to the relationship between GA and brain network dysconnectivity in preterm infants.

Strengths of this study are that we studied a population of preterm and control infants across the gestational age range of 23 to 42 weeks, who were uniquely phenotyped with DNAm and dMRI. We sampled after the period of NICU care in order to capture the allostatic load of preterm birth. We measured DNAm from neonatal saliva samples, which has consistency with brain DNAm patterns and is non-invasive. The Illumina EPIC platform provided extensive coverage of the methylome (850,000 sites) and we controlled for cell composition. Finally, we used an image phenotype that is robust to scanner variation.¹⁷ There are some limitations. First, control for cell composition was based on estimation of cell proportions rather than measurement, so we cannot rule out the possibility that some of the signal identified was related to variation in cell composition. Second, mediation analysis to assess causation was not possible because the association between the DNAm PC1 and PS metrics might result from the DNAm PC being derived from CpG sites that are associated with gestational age, so variance attributable to the mediating variable cannot be assumed. This could be addressed by out-of-sample validation, which will require other neonatal cohorts with both methylome and dMRI data. There was also some evidence of inflation of test statistics based on the genomic inflation factor. However, genomic inflation factor is thought to provide an overestimate of inflation and corrections based on it may be overly conservative.⁸¹ The value of the genomic inflation factor was also similar to those previously reported in neonatal cohorts.³⁴ In addition, visual inspection of our results via Manhattan

plots suggests that our finding, of widespread differences in DNAm in relation to gestational age at birth, are in line with previous studies that have investigated this in cord blood and in fetal brain tissue.^{22,34}

In conclusion, gestational age at birth has a profound and widely distributed impact on the neonatal saliva methylome at term equivalent age, which reflects the allostatic load of preterm birth itself and postnatal exposures during neonatal intensive care. Gene ontology terms related to cell-cell contacts were enriched, indicating that cell contacts and organisation are implicated in the phenotype. Associations between DNAm and PSMD and PSNDI suggest that variations in DNAm could contribute to white matter dysconnectivity commonly seen in preterm infants, and this analysis identified several genes and gene regions that could provide further insight into the molecular mechanisms by which early exposure to extrauterine life influences neurodevelopment.

Acknowledgements

Some of the participants were scanned in the University of Edinburgh Imaging Research MRI Facility at the Royal Infirmary of Edinburgh which was established with funding from The Wellcome Trust, Dunhill Medical Trust, Edinburgh and Lothians Research Foundation, Theirworld, The Muir Maxwell Trust, and other sources. We thank Thorsten Feiweier at Siemens Healthcare for collaborating with dMRI acquisitions (Works-in-Progress Package for Advanced EPI Diffusion Imaging).

We are grateful to the families who consented to take part in the study and to all the University's imaging research staff for providing the infant scanning.

Funding

ENWW is supported by the Wellcome Trust Translational Neuroscience PhD fellowship programme at the University of Edinburgh (203769/Z/16/A). This work was supported by Theirworld (www.theirworld.org) and was carried out in the MRC Centre for Reproductive Health, which was funded by MRC Centre Grant (MRC G1002033).

Competing interests

REM has received a speaker fee from Illumina and is an advisor to the Epigenetic Clock Development Foundation. LM has carried out paid presentations and consultancy for Illumina. The remaining authors report no competing interests.

References

1. Chawanpaiboon, S. *et al.* Global, regional, and national estimates of levels of preterm birth in 2014: a systematic review and modelling analysis. *Lancet Glob. Heal.* **7**, e37–e46 (2019).
2. Agrawal, S., Rao, S. C., Bulsara, M. K. & Patole, S. K. Prevalence of autism spectrum disorder in preterm infants: A meta-Analysis. *Pediatrics* **142**, (2018).
3. Franz, A. P. *et al.* Attention-Deficit / Hyperactivity Disorder and Very Preterm / Very Low Birth Weight: A Meta-analysis. *Pediatrics* **141**, (2018).
4. Nosarti, C. *et al.* Preterm birth and psychiatric disorders in young adult life. *Arch. Gen. Psychiatry* **69**, (2012).
5. Van Lieshout, R. J. *et al.* Trajectories of psychopathology in extremely low birth weight survivors from early adolescence to adulthood: a 20-year longitudinal study. *J. Child Psychol. Psychiatry Allied Discip.* **59**, 1192–1200 (2018).
6. Burnett, A. C. *et al.* Prevalence of psychiatric diagnoses in preterm and full-term children, adolescents and young adults: A meta-analysis. *Psychol. Med.* **41**, 2463–2474 (2011).
7. Mackay, D. F., Smith, G. C. S., Dobbie, R. & Pell, J. P. Gestational Age at Delivery and Special Educational Need: Retrospective Cohort Study of 407 , 503 Schoolchildren. **7**, 1–10 (2010).
8. Johnson, S. *et al.* Autism Spectrum Disorders in Extremely Preterm Children. *J. Pediatr.* **156**, 525–531.e2 (2010).
9. Johnson, S. *et al.* Neurodevelopmental outcomes following late and moderate prematurity: A population-based cohort study. *Arch. Dis. Child. Fetal Neonatal Ed.* **100**, F301–F308 (2015).
10. Volpe, J. J. Brain injury in premature infants: a complex amalgam of destructive and developmental disturbances. *Lancet Neurol.* **8**, 110–124 (2009).
11. Pietsch, M. *et al.* A framework for multi-component analysis of diffusion MRI data over the neonatal period. *Neuroimage* **186**, 321–337 (2019).
12. Zhang, H., Schneider, T., Wheeler-Kingshott, C. A. & Alexander, D. C. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage* **61**, 1000–1016 (2012).
13. Tariq, M., Schneider, T., Alexander, D. C., Gandini Wheeler-Kingshott, C. A. & Zhang, H. Bingham-NODDI: Mapping anisotropic orientation dispersion of neurites using diffusion MRI. *Neuroimage* **133**, 207–223 (2016).
14. Telford, E. J. *et al.* A latent measure explains substantial variance in white matter microstructure across the newborn human brain. *Brain Struct. Funct.* **222**, (2017).
15. Ball, G. *et al.* Thalamocortical connectivity predicts cognition in children born preterm. *Cereb. Cortex* **25**, 4310–4318 (2015).
16. Boardman, J. P. & Counsell, S. J. Factors associated with atypical brain development in preterm infants: insights from magnetic resonance imaging. *Neuropathol. Appl. Neurobiol.* **46**, 413–421 (2020).
17. Baykara, E. *et al.* A Novel Imaging Marker for Small Vessel Disease Based on Skeletonization of White Matter Tracts and Diffusion Histograms. *Ann. Neurol.* **80**, 581–592 (2016).

18. Deary, I. J. *et al.* Brain Peak Width of Skeletonized Mean Diffusivity (PSMD) and Cognitive Function in Later Life. *Front. Psychiatry* **10**, 1–10 (2019).
19. Wei, N. *et al.* A Neuroimaging Marker Based on Diffusion Tensor Imaging and Cognitive Impairment Due to Cerebral White Matter Lesions. **10**, 1–7 (2019).
20. Blesa, M., Galdi, P., Sullivan, G. & Wheeler, E. N. Peak width of skeletonized water diffusion MRI in the neonatal brain. *Front. Neurol.* (2020)
doi:10.3389/fneur.2020.00235/abstract.
21. Sullivan, G. *et al.* Interleukin-8 dysregulation is implicated in brain dysmaturation following preterm birth. *Brain. Behav. Immun.* **90**, 311–318 (2020).
22. Spiers, H. *et al.* Methylomic trajectories across human fetal brain development. *Genome Res.* 1–16 (2015) doi:10.1101/gr.180273.114.Freely.
23. Lyst, M. J. & Bird, A. Rett syndrome: A complex disorder with simple roots. *Nat. Rev. Genet.* **16**, 261–274 (2015).
24. Weemaes, C. *et al.* Heterogeneous clinical presentation in ICF syndrome: correlation with underlying gene defects. 1219–1225 (2013) doi:10.1038/ejhg.2013.40.
25. Rangasamy, S., D’Mello, S. R. & Narayanan, V. Epigenetics, Autism Spectrum, and Neurodevelopmental Disorders. *Neurotherapeutics* **10**, 742–756 (2013).
26. Sparrow, S. *et al.* Epigenomic profiling of preterm infants reveals DNA methylation differences at sites associated with neural function. *Transl. Psychiatry* **6**, e716 (2016).
27. Wheeler, E. N. W. *et al.* DNA methylation and brain structure and function across the life course: A systematic review. *Neurosci. Biobehav. Rev.* **113**, 133–156 (2020).
28. Joubert, B. R. *et al.* DNA Methylation in Newborns and Maternal Smoking in Pregnancy: Genome-wide Consortium Meta-analysis. *Am. J. Hum. Genet.* **98**, 680–696 (2016).
29. Sharp, G. C. *et al.* Maternal BMI at the start of pregnancy and offspring epigenome-wide DNA methylation: Findings from the pregnancy and childhood epigenetics (PACE) consortium. *Hum. Mol. Genet.* **26**, 4067–4085 (2017).
30. Küpers, L. K. *et al.* Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nat. Commun.* **10**, 1–11 (2019).
31. Everson, T. M. *et al.* Serious neonatal morbidities are associated with differences in DNA methylation among very preterm infants. *Clin. Epigenetics* **12**, 1–15 (2020).
32. Everson, T. M. *et al.* Epigenome-wide Analysis Identifies Genes and Pathways Linked to Neurobehavioral Variation in Preterm Infants. 1–13 (2019) doi:10.1038/s41598-019-42654-4.
33. Cruickshank, M. N. *et al.* Analysis of epigenetic changes in survivors of preterm birth reveals the effect of gestational age and evidence for a long term legacy. *Genome Med.* **5**, (2013).
34. Merid, S. K., Novoloaca, A., Sharp, G. C., Küpers, L. K. & Kho, A. T. Epigenome-wide meta-analysis of blood DNA methylation in newborns and children identifies numerous loci related to gestational age. *Genome Med.* **12**, 1–17 (2020).
35. Armstrong, D. A., Lesueur, C., Conradt, E., Lester, B. M. & Marsit, C. J. Global and gene-specific DNA methylation across multiple tissues in early infancy: Implications for children’s health research. *FASEB J.* **28**, 2088–2097 (2014).
36. Lowe, R. *et al.* Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics* **8**, 445–454 (2013).
37. Braun, P. R. *et al.* Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Transl. Psychiatry* **9**, (2019).
38. Smith, A. K. *et al.* DNA extracted from saliva for methylation studies of psychiatric traits: Evidence tissue specificity and relatedness to brain. *Am. J. Med. Genet. Part B*

- Neuropsychiatr. Genet.* **168**, 36–44 (2015).
39. Boardman, J. P. *et al.* Impact of preterm birth on brain development and long-term outcome: Protocol for a cohort study in Scotland. *BMJ Open* **10**, 1–11 (2020).
 40. Aryee, M. J. *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
 41. Pidsley, R. *et al.* A data-driven approach to preprocessing Illumina 450 K methylation array data. *BMC Genomics* **14**, 293 (2013).
 42. McCartney, D. L. *et al.* Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data* **9**, 22–24 (2016).
 43. Zheng, S. C. *et al.* A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. *Epigenomics* **10**, 925–940 (2018).
 44. Caruyer, E., Lenglet, C., Sapiro, G. & Deriche, R. Design of multishell sampling schemes with uniform coverage in diffusion MRI. *Magn. Reson. Med.* **69**, 1534–1540 (2013).
 45. Veraart, J., Fieremans, E. & Novikov, D. S. Diffusion MRI noise mapping using random matrix theory. *Magn. Reson. Med.* **76**, 1582–1593 (2016).
 46. Veraart, J. *et al.* Denoising of diffusion MRI using random matrix theory. *Neuroimage* **142**, 394–406 (2016).
 47. Tournier, J. D. *et al.* MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage* **202**, 116137 (2019).
 48. Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage* **23**, 208–219 (2004).
 49. Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage* **125**, 1063–1078 (2016).
 50. Andersson, J. L. R., Graham, M. S., Zsoldos, E. & Sotiropoulos, S. N. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *Neuroimage* **141**, 556–572 (2016).
 51. Tustison, N. J. *et al.* N4ITK: Improved N3 bias correction. *IEEE Trans. Med. Imaging* **29**, 1310–1320 (2010).
 52. Andersson, J. L. R. *et al.* Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. *Neuroimage* **152**, 450–466 (2017).
 53. Hernandez-Fernandez, M. *et al.* Using GPUs to accelerate computational diffusion MRI: From microstructure estimation to tractography and connectomes. *Neuroimage* **188**, 598–615 (2019).
 54. RCoreTeam. R: A language and environment for statistical computing. *R Found. Stat. Comput. Vienna, Austria.* (2020).
 55. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724–1735 (2007).
 56. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
 57. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
 58. Saffari, A. *et al.* Estimation of a significance threshold for epigenome-wide association studies. *Genet. Epidemiol.* **42**, 20–33 (2018).

59. Suderman, M. *et al.* Dmrff: Identifying differentially methylated regions efficiently with power and control. *bioRxiv* 1–26 (2018) doi:10.1101/508556.
60. Maksimovic, J., Oshlack, A. & Phipson, B. Gene set enrichment analysis for genome-wide DNA methylation data. *bioRxiv* 2020.08.24.265702 (2020) doi:10.1101/2020.08.24.265702.
61. Job, D. E. *et al.* A brain imaging repository of normal structural MRI across the life course: Brain Images of Normal Subjects (BRAINS). *Neuroimage* **144**, 299–304 (2017).
62. Bohlin, J. *et al.* Prediction of gestational age based on genome-wide differentially methylated regions. *Genome Biol.* **17**, 1–9 (2016).
63. Provenzi, L., Guida, E. & Montirosso, R. Preterm behavioral epigenetics: A systematic review. *Neurosci. Biobehav. Rev.* 0–3 (2017) doi:10.1016/j.neubiorev.2017.08.020.
64. Fontana, C. *et al.* Early maternal care restores LINE-1 methylation and enhances neurodevelopment in preterm infants. 1–16 (2021).
65. Lin, H. H. *et al.* Neuronatin promotes neural lineage in ESCs via Ca²⁺ signaling. *Stem Cells* **28**, 1950–1960 (2010).
66. Numata, S. *et al.* DNA methylation signatures in development and aging of the human prefrontal cortex. *Am. J. Hum. Genet.* **90**, 260–272 (2012).
67. Sharp, G. C. *et al.* Maternal pre-pregnancy BMI and gestational weight gain, offspring DNA methylation and later offspring adiposity: Findings from the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **44**, 1288–1304 (2015).
68. Hannon, E. *et al.* An integrated genetic-epigenetic analysis of schizophrenia: Evidence for co-localization of genetic associations and differential DNA methylation. *Genome Biol.* **17**, 1–16 (2016).
69. Hannon, E. *et al.* Variable DNA methylation in neonates mediates the association between prenatal smoking and birth weight. *Philos. Trans. R. Soc. B Biol. Sci.* **374**, (2019).
70. Battram, T. *et al.* The EWAS Catalog: a database of epigenome-wide association studies. *OSF Prepr.* 2–5 (2021).
71. Kazmi, N. *et al.* Hypertensive Disorders of Pregnancy and DNA Methylation in Newborns: Findings from the Pregnancy and Childhood Epigenetics Consortium. *Hypertension* **74**, 375–383 (2019).
72. Odintsova, V. V *et al.* DNA Methylation Signatures of Breastfeeding in Buccal Cells Collected in Mid-Childhood. *Nutrients* 1–26 (2019).
73. Bruneau, B. G. *et al.* Cardiac expression of the ventricle-specific homeobox gene *Irx4* is modulated by *Nkx2-5* and *dHand*. *Dev. Biol.* **217**, 266–277 (2000).
74. Jin, Z. *et al.* *Irx4*-mediated regulation of *Slit1* expression contributes to the definition of early axonal paths inside the retina. *Development* **130**, 1037–1048 (2003).
75. Garriock, R. J., Vokes, S. A., Small, E. M., Larson, R. & Krieg, P. A. Developmental expression of the *Xenopus* Iroquois-family homeobox genes, *Irx4* and *Irx5*. *Dev. Genes Evol.* **211**, 257–260 (2001).
76. Rijlaarsdam, J. *et al.* An epigenome-wide association meta-analysis of prenatal maternal stress in neonates: A model approach for replication. *Epigenetics* **11**, 140–149 (2016).
77. Oegema, R. *et al.* Human mutations in integrator complex subunits link transcriptome integrity to brain development. *PLoS Genet.* **13**, e1006923 (2017).
78. Krall, M. *et al.* Biallelic sequence variants in *INTS1* in patients with developmental delays, cataracts, and craniofacial anomalies. *Eur. J. Hum. Genet.* **27**, 582–593 (2019).
79. Zhang, X. *et al.* Biallelic *INTS1* Mutations Cause a Rare Neurodevelopmental Disorder in Two Chinese Siblings. *J. Mol. Neurosci.* **70**, 1–8 (2020).

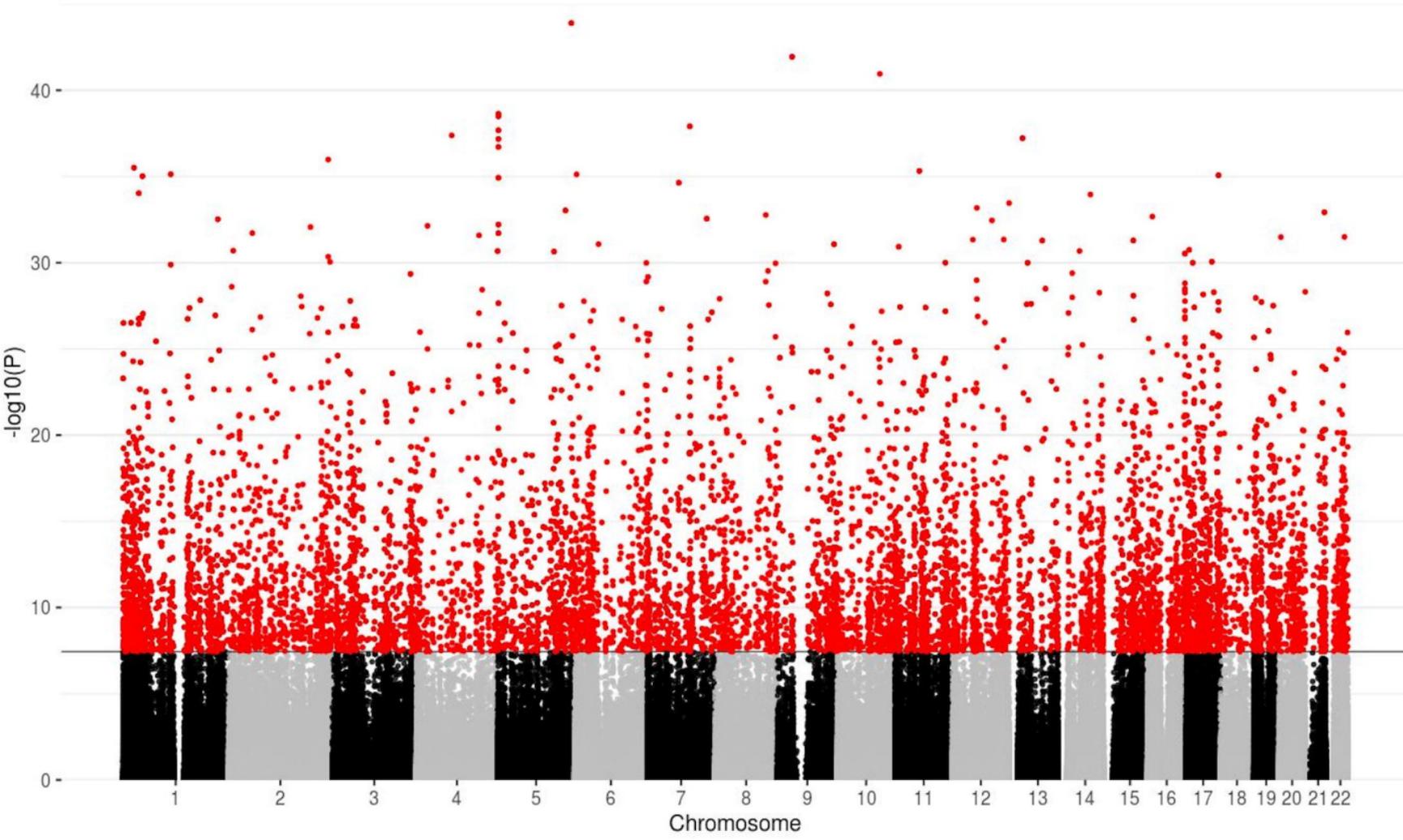
80. Galdi, P. *et al.* Neonatal morphometric similarity mapping for predicting brain age and characterizing neuroanatomic variation associated with preterm birth. *NeuroImage Clin.* **25**, 102195 (2020).
81. van Iterson, M. *et al.* Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol.* **18**, 1–13 (2017).

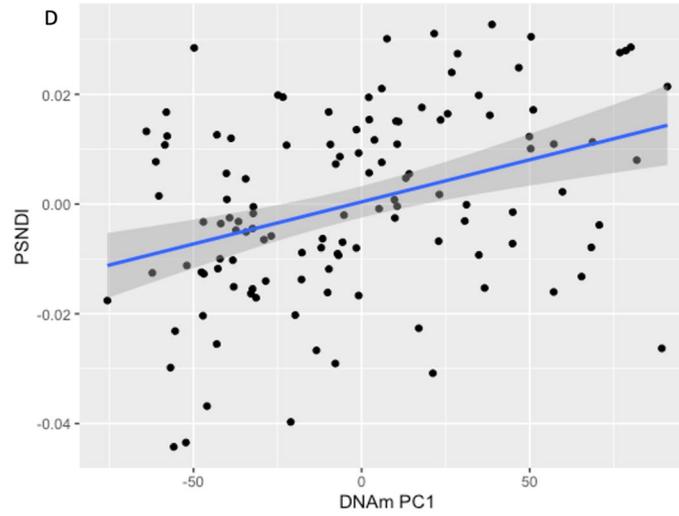
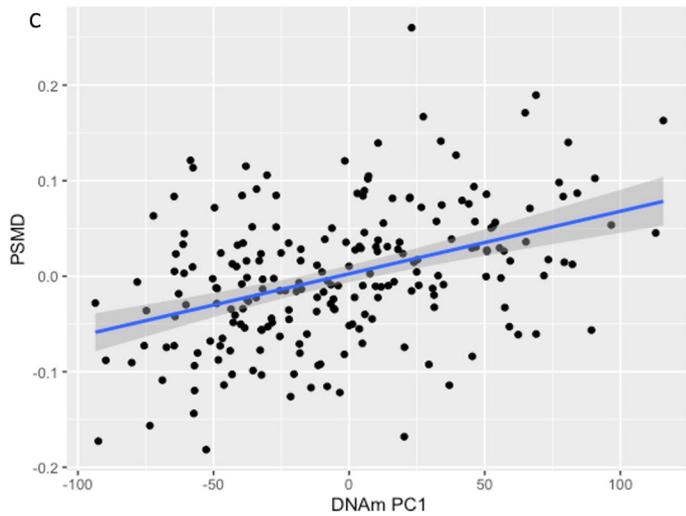
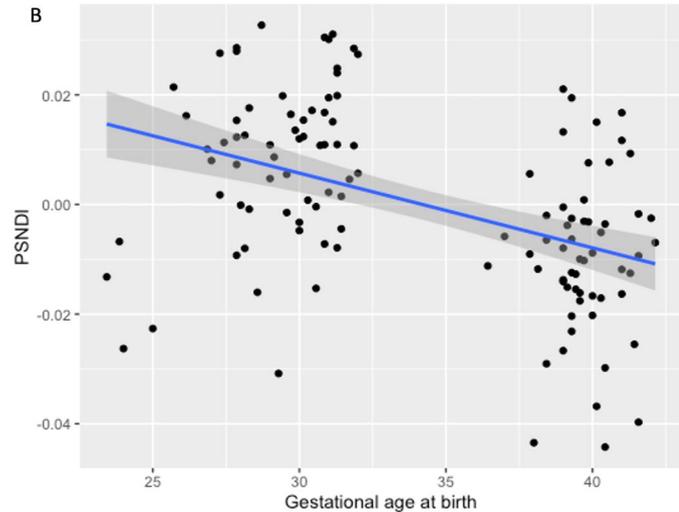
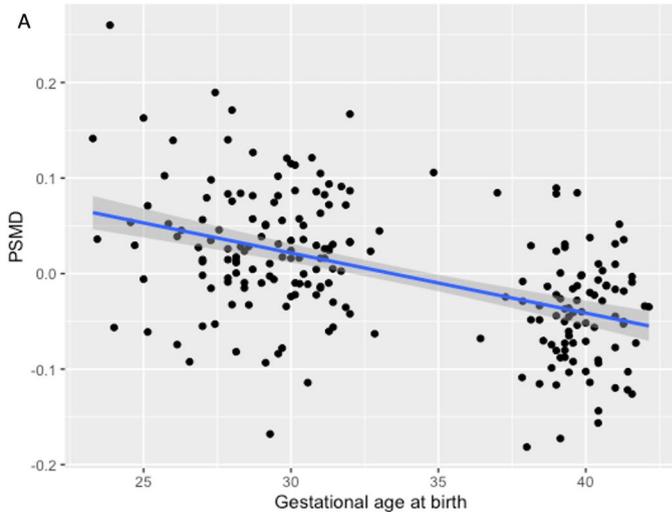
Figure legends

Figure 1. Manhattan plot for the association between gestational age at birth and DNA methylation, following adjustment for covariates and surrogate variables. The solid horizontal line shows the genome-wide significance level and red dots above this line represent probes that are significant at this threshold ($p < 3.6 \times 10^{-8}$).

Figure 2. Scatter plots with 95% confidence intervals, showing the relationships between gestational age at birth and DNAm with peak width skeletonised mean diffusivity and neurite density index. The associations between gestational age and PSMD and PSNDI are shown in A and B, respectively. The relationships between DNAm PC1 and PSMD and PSNDI are shown in C and D, respectively. PS metrics are residualised for gestational age at scan; PSMD is additionally residualised for scanner variable.

Figure 3. Variation in DNAm probes selected by EWAS captured by principal component, and the relationship with gestational age. Fig 3A, scree plot showing sharp elbow after the first PC, accounting for 23.5% of variance. Fig 3B, scatter plot showing relationship between gestational age at birth and PC1 ($r = -0.622$) with 95% confidence intervals.





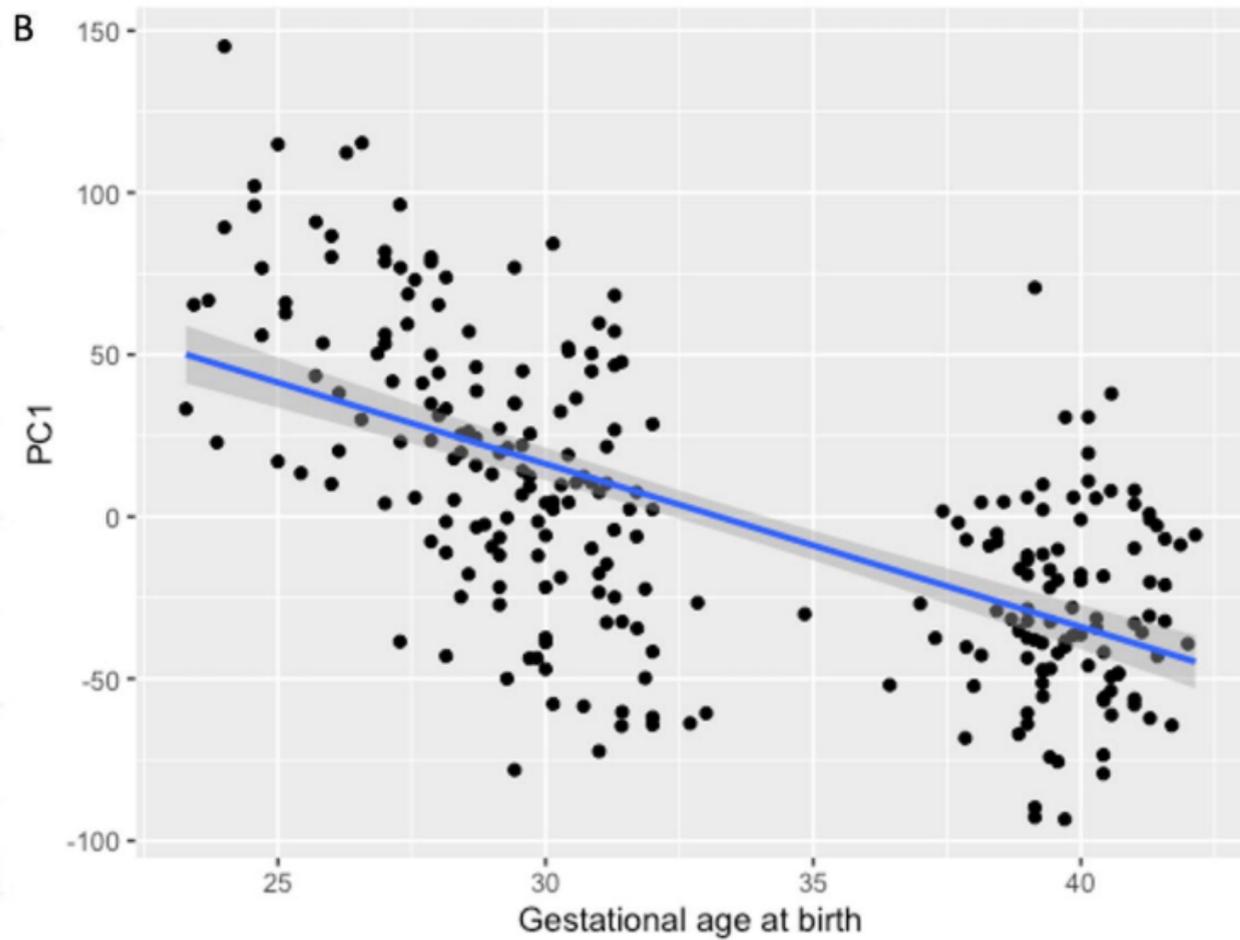
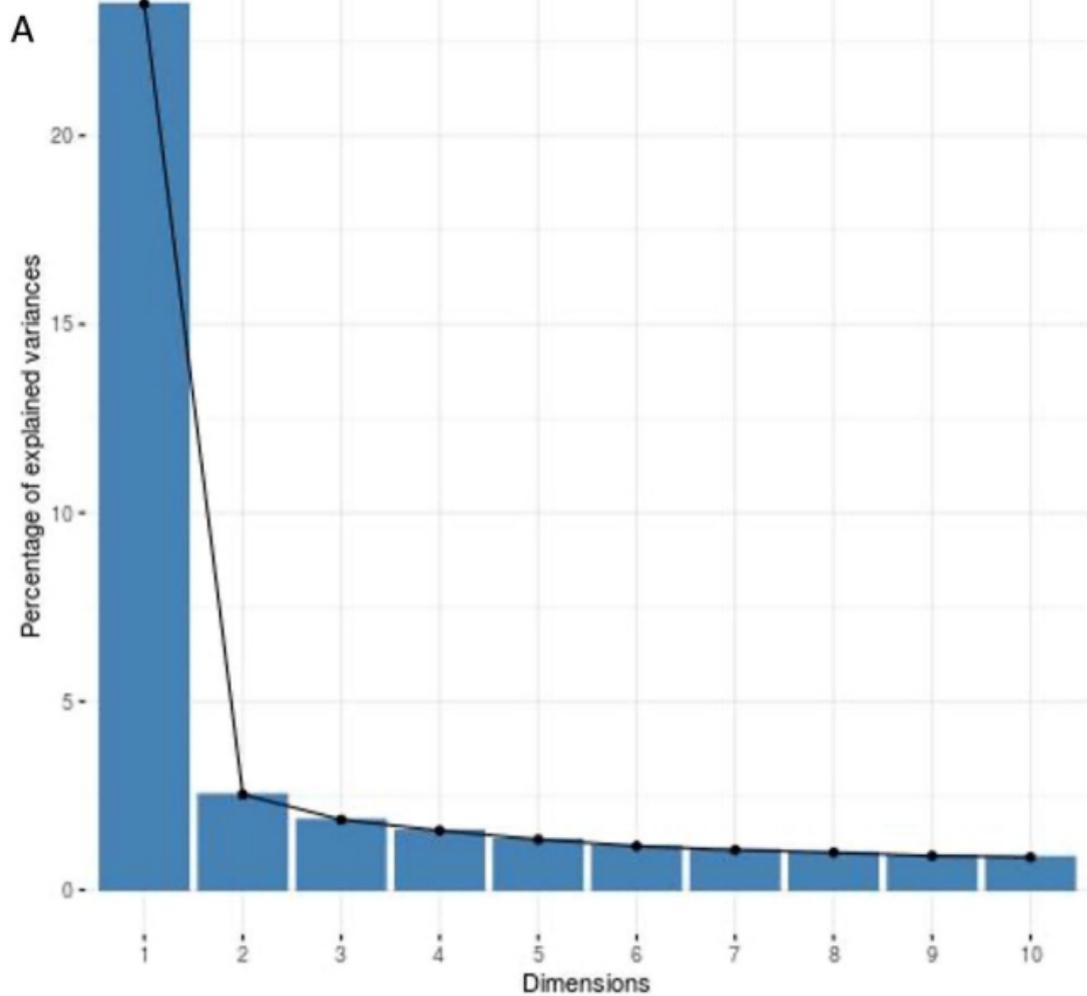


Table 1. Participant characteristics

	Preterm infants (n=155)	Term infants (n=103)
Gestational age at birth/weeks (range)	28.84 (23.28 – 34.84)	39.7 (36.42 – 42.14)
Gestational age at scan/weeks (range)	40.56 (37.70-45.14)	42.27 (39.84 – 47.14)
Birth weight/g (range)	1177 (500 – 2100)	3482 (2346 – 4670)
Birth weight z-score (range)	-0.19 (range -3.13 – 1.58)	0.43 (range -2.30 – 2.96)
Sex: Female (%)	75 (48)	44 (43)
Maternal folate supplementation in pregnancy (%)	136 (88)	86 (83)
Maternal age (years)	31.1 (17-44)	33.7 (19-48)
Maternal tobacco smoker in pregnancy (%)	29 (19)	2 (2)

Table 2. Most significant probes associated with gestational age at birth.

Probe	Chromosome	p value	Gene	Direction of effect	Coefficient*	Relation to island
cg03558436	5	1.26x10 ⁻⁴⁴	-	+	0.01016	Open Sea
cg04466438	9	1.13x10 ⁻⁴²	-	+	0.00755	Open Sea
cg23701943	10	1.11x10 ⁻⁴¹	<i>LOXL4</i>	+	0.01042	Open Sea
cg18172877	5	2.31x10 ⁻³⁹	<i>IRX4</i>	-	-0.00612	Island
cg04180086	5	3.22x10 ⁻³⁹	<i>IRX4</i>	-	-0.00736	Island
cg22645539	7	1.22x10 ⁻³⁸	<i>GAL3ST4</i>	-	-0.00815	North Shelf
cg17774559	5	2.09x10 ⁻³⁸	<i>IRX4</i>	-	-0.00635	Island
cg17582074	4	4.12x10 ⁻³⁸	-	+	0.00525	Open Sea
cg08915267	13	6.01x10 ⁻³⁸	-	-	-0.00515	North Shelf
cg04441405	5	6.60x10 ⁻³⁸	-	-	-0.01102	Island

*Coefficient corresponds to methylation change per week of gestation.

Table 3. Probes with the largest absolute magnitude of association with gestational age at birth.

Probe	Chromosome	P value	Gene	Direction of effect	Coefficient ^a	Relation to island
cg10402321	1	3.11×10^{-36}	<i>UBXN11</i>	-	-0.01136	Open Sea
cg04441405	5	6.60×10^{-38}	-	-	-0.01102	Island
cg07167946	5	1.94×10^{-32}	<i>IRX4</i>	-	-0.00985	Island
cg07803375	7	3.6×10^{-22}	<i>HEATR2</i>	-	-0.00908	North Shelf
cg14670058	13	9.24×10^{-23}	<i>SMIM2</i>	-	-0.00907	Open Sea
cg16051275	6	7.53×10^{-36}	-	+	0.01231	Open Sea
cg11460314	20	4.21×10^{-20}	<i>ZBP1</i>	+	0.01242	Open Sea
cg04118102	17	1.00×10^{-30}	-	+	0.01312	South Shelf
cg17368297	16	1.55×10^{-25}	-	+	0.01403	Open Sea
cg14576951	7	6.73×10^{-30}	<i>INTS1</i>	+	0.01438	Island

^aCoefficient corresponds to methylation change per week of gestation.

Gene ontology	Term	FDR Q value	Type	Description
GO:0005912	adherens junction	0.00002317	Cellular component	A cell-cell junction composed of the epithelial cadherin-catenin complex.
GO:0005925	focal adhesion	0.00907565	Cellular component	A cell-substrate junction that anchors the cell to the extracellular matrix and that forms a point of termination of actin filaments.
GO:0007155	cell adhesion	0.01728820	Biological	The attachment of a cell, to another cell or to the extracellular matrix, via cell adhesion

Table 4. Gene ontology (GO) terms that were significantly enriched in an analysis of all probes that contributed to DMPs and DMRs.

			process	molecules.
GO:0007167	enzyme linked receptor protein signalling pathway	0.01243116	Biological process	A series of molecular signals initiated by the binding of an extracellular ligand to a receptor on the target cell plasma membrane, where the receptor possesses catalytic activity or is closely associated with an enzyme such as a protein kinase, and ending with regulation of a downstream cellular process, e.g. transcription
GO:0007169	transmembrane receptor protein tyrosine kinase signalling pathway	0.00907565	Biological process	A series of molecular signals, initiated by the binding of an extracellular ligand to a tyrosine kinase receptor on the target cell plasma membrane, ending with regulation of a downstream cellular process.
GO:0022610	biological adhesion	0.01266030	Biological process	The attachment of a cell to a substrate, another cell, including intracellular attachment between membrane regions.
GO:0030029	actin filament-based process	0.03635020	Biological process	Any cellular process that depends upon, or alters, the actin cytoskeleton (comprising actin filaments and their associated proteins).
GO:0030036	actin cytoskeleton organization	0.04472556	Biological process	The assembly, arrangement of constituent parts, or disassembly of cytoskeletal structures comprising actin filaments and their associated proteins.
GO:0030054	cell junction	0.00907565	Cellular component	Forms a specialized region of connection between two or more cells, or between a cell and the extracellular matrix, or between two membrane-bound components of a cell, such as flagella.
GO:0030055	cell-substrate junction	0.00907565	Cellular component	A cell junction between a cell and the extracellular matrix.
GO:0034330	cell junction organization	0.04159681	Biological process	The assembly, arrangement of constituent parts, or disassembly of a cell junction. A cell junction is a specialized region of connection between two cells or between a cell and the extracellular matrix
GO:0045296	cadherin binding	0.00207297	Molecular function	Interacting selectively and non-covalently with cadherin, a type I membrane protein involved in cell adhesion.
GO:0050839	cell adhesion molecule binding	0.00207297	Molecular function	Interacting selectively and non-covalently with a cell adhesion molecule.
GO:0070161	anchoring junction	0.00002317	Cellular component	A cell junction that mechanically attaches a cell, and its cytoskeleton, to neighbouring cells or the extracellular matrix.

Terms in bold were enriched in an analysis of 8870 genome-wide significant DMPs

Table 5. Associations between global metrics of white matter microstructure, DNAm first principal component (left) and gestational age (right)

PS metric	Metric ~ PC1 DNAm + Gestational age at scan + scanner variable ^a		Metric ~ Gestational age at birth + Gestational age at scan + scanner variable ^a	
	β	<i>p</i>	β	<i>p</i>
PSFA	-0.035	0.510	-0.005	0.933
PSMD	0.349	8.37x10⁻¹⁰	-0.602	< 2x10⁻¹⁶
PSNDI	0.364	4.15x10⁻⁵	-0.594	2.17x10⁻⁹

^ascanner variable was included in the model for PSFA and PSMD but not PSNDI because NODDI imaging was carried out for a subset using a single MRI scanner (n=121). Bold print signifies significant associations.