

BRAIN COMMUNICATIONS

DNA methylation in relation to gestational age and brain dysmaturation in preterm infants

Emily N. W. Wheeler,¹ Paola Galdi,¹ Daniel L. McCartney,² Manuel Blesa,¹ Gemma Sullivan,¹ David Q. Stoye,¹ Gillian Lamb,¹ Sarah Sparrow,¹ Lee Murphy,² Nicola Wrobel,² Alan J. Quigley,³ Scott Semple,^{4,5} Michael J. Thrippleton,^{4,6} Joanna M. Wardlaw,⁶ Mark E. Bastin,⁶ Riccardo E. Marioni,² Simon R. Cox⁷ and James P. Boardman^{1,6}

Preterm birth is associated with dysconnectivity of structural brain networks and is a leading cause of neurocognitive impairment in childhood. Variation in DNA methylation 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 the 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 or regions. We tested the hypothesis that variation in DNA methylation could underpin the association between low gestational age at birth and atypical brain development by linking differentially methylated probes with measures of white matter connectivity derived from diffusion MRI metrics: peak width skeletonized mean diffusivity, peak width skeletonized fractional anisotropy and peak width skeletonized neurite density index. Gestational age at birth was associated with widespread differential methylation at term equivalent age, with genome-wide significant associations observed for 8870 CpG probes ($P < 3.6 \times 10^{-8}$) and 1767 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 that explained 23.5% of the variance in DNA methylation, and this was negatively associated with gestational age at birth. The first principal component was associated with peak width of skeletonized mean diffusivity ($\beta = 0.349$, $P = 8.37 \times 10^{-10}$) and peak width skeletonized neurite density index ($\beta = 0.364$, $P = 4.15 \times 10^{-5}$), but not with peak width skeletonized fraction anisotropy ($\beta = -0.035$, $P = 0.510$); these relationships mirrored the imaging metrics' associations with gestational age at birth. Low gestational age at birth has a profound and widely distributed effect on the neonatal saliva methylome that is apparent at term equivalent age. 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 DNA methylation and image markers of white matter tract microstructure suggest that variation in DNA methylation may provide a link between preterm birth and the dysconnectivity of developing brain networks that characterizes atypical brain development in preterm infants.

- 1 MRC Centre for Reproductive Health, The University of Edinburgh, Queen's Medical Research Institute, Edinburgh EH16 4TJ, UK
- 2 Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
- 3 Department of Paediatric Radiology, Royal Hospital for Sick Children, NHS Lothian, Edinburgh, UK
- 4 Edinburgh Imaging, University of Edinburgh, EH16 4SB Edinburgh, UK
- 5 Centre for Cardiovascular Science, The University of Edinburgh, Queen's Medical Research Institute, Edinburgh EH16 4TJ, UK
- 6 Centre for Clinical Brain Sciences, The University of Edinburgh, Edinburgh, UK
- 7 Department of Psychology, The University of Edinburgh, Edinburgh, UK

Received April 12, 2021. Revised December 10, 2021. Accepted March 4, 2022. Advance access publication March 8, 2022

© The Author(s) 2022. Published by Oxford University Press on behalf of the Guarantors of Brain.

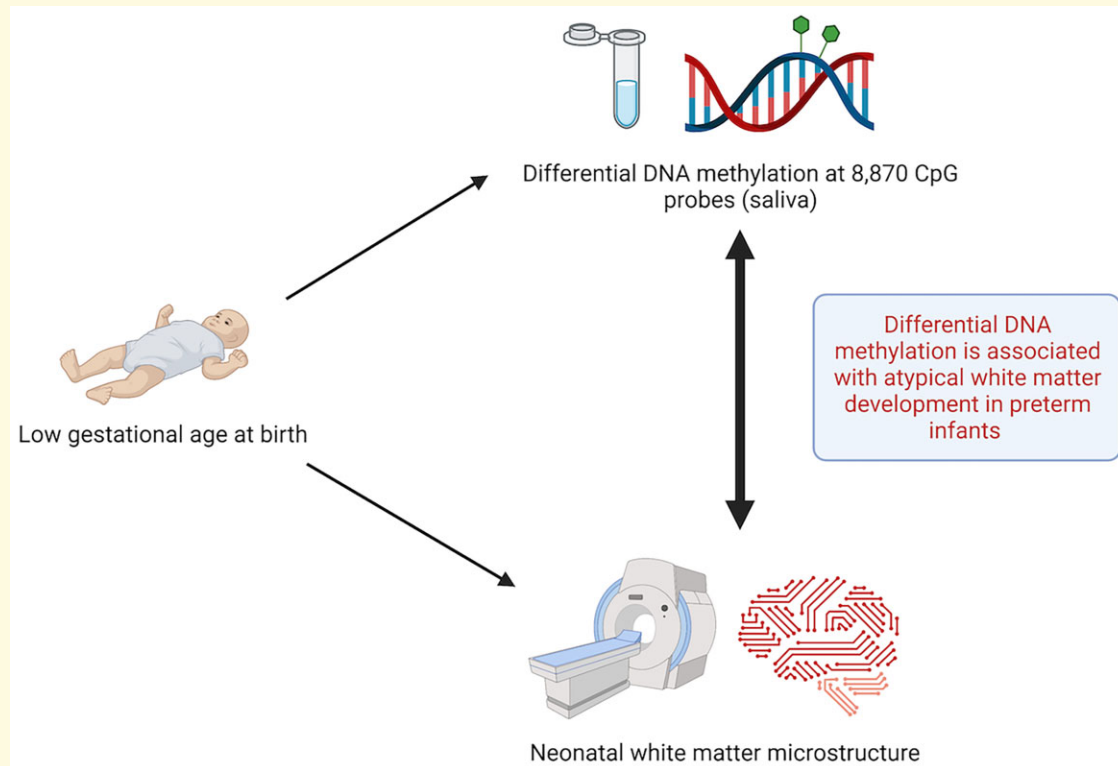
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Correspondence to: Professor James P. Boardman
MRC Centre for Reproductive Health
Queen's Medical Research Institute
47 Little France Crescent, Edinburgh EH16 4TJ, UK
E-mail: james.boardman@ed.ac.uk

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

Abbreviations: FA = fractional anisotropy; MD = mean diffusivity; NDI = neurite density index; NODDI = neurite orientation dispersion and density imaging; PCA = principal component analysis; SVA = surrogate variable analysis

Graphical Abstract



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, attention-deficit hyperactivity disorder, reduced intelligence quotient, 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 generalized 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 control infants born at term.¹⁶ These changes reflect an increase in water content and a decrease in white matter organization in preterm infants. The peak width of skeletonized 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 peak width skeletonized neurite density index (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 foetal brain development.²² Alterations in DNAm contribute to the pathogenesis of neurodevelopmental disorders [Rett syndrome,²³ Immunodeficiency, Centromeric region instability, Facial anomalies (ICE) syndrome²⁴ and Angleman 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 comorbidities 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 GA at birth (across the range 27–42 weeks) affecting 2% of sites and involving both hypo- and hypermethylation, 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 low gestational age at birth was associated with variation in the salivary methylome at term equivalent age and to characterize the biological pathways implicated in the DNAm response to preterm birth. Our second aim was to investigate whether the DNAm signal of gestational age explains variance in measures of white matter microstructure at term equivalent age. We tested the hypotheses that low gestational age at birth is associated with widespread differential methylation apparent at the end of neonatal intensive care and that DNAm contributes to variance in peak width skeletonized metrics of white matter connectivity.

Materials and methods

Participants

All participants were born at the Royal Infirmary of Edinburgh, UK. Preterm infants, defined as GA of <33 weeks of gestation and term infants, defined as GA > 37 weeks based on first-trimester ultrasound scan dating, were recruited to the Theirworld Edinburgh Birth Cohort. This is a longitudinal study designed to investigate the effect of preterm birth on brain development.³⁹ Exclusion criteria were major congenital malformations, chromosomal abnormalities, congenital infection, overt parenchymal lesions

(cystic periventricular leukomalacia, haemorrhagic parenchymal infarction) or posthaemorrhagic 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, on the same day as MRI acquisition, was collected in Oragene OG-575 Assisted Collection kits, by DNA Genotek, and DNA extracted using prepIT.L2P reagent (DNA Genotek, ON, Canada). DNA was bisulphite 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 processed in two batches.

DNA methylation preprocessing

Raw intensity (.idat) files were read into the R environment (version 3.4.4) using `minfi`. `wateRmelon` and `minfi` were used for preprocessing, quality control and normalization.^{40,41} The `pfilter` function in `wateRmelon` was used to exclude samples with 1% of sites with a detection P -value > 0.05; sites with beadcount < 3 in 5% of samples and sites with 1% of samples with detection P -value > 0.05. Cross hybridizing 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 were `danet` normalized which includes background correction and dye bias correction.⁴¹ Saliva contains different cell types, including buccal epithelial cells and leucocytes. 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 before analysis. Data from one of each twin pair were removed randomly ($n = 20$).

MRI acquisition

MRI was obtained at the same appointment as saliva sample collection for DNAm analysis. Structural and dMRI 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 using 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 (echo time (TE) = 106 ms and repetition time (TR) = 7300 ms).

One hundred and twenty-one infants were scanned using a MAGNETOM Prisma 3T clinical MRI scanner (Siemens Healthcare GmbH) and 16-channel phased-array paediatric head and neck coil. dMRI was acquired in two separate acquisitions: the first consisted of eight 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 three 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 3500/78.0 ms). Images affected by motion artefact were re-acquired multiple times as required; dMRI acquisitions were repeated if the signal loss was seen in three 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 paediatric radiologist (A.J.Q.)

dMRI preprocessing

Diffusion images that were acquired on the MAGNETOM Verio scanner were denoised using a Marchenko–Pastur-principal component analysis (PCA)-based algorithm^{45–47}; eddy current distortion and head movement were corrected using outlier replacement;^{48–50} 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 algorithm^{45–47}; 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 b_0 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.

The 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 were 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 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 An overview of the analysis pipeline is shown in [Supplementary Fig. 1](#). Surrogate variable analysis (SVA) of the data matrix was carried out, 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 before SVA to retain signal explained by biological variables of interest: CpG \sim gestational age at birth + age at scan + birthweight Z-score + maternal smoking + sex + epithelial cells. SVA identified 17 significant surrogate variables (SVs) which were subsequently included in the analysis.

An epigenome-wide association study (EWAS) was performed using the *limma* package in R.⁵⁷ Beta values of each of 776 025 CpG sites were regressed (as dependent variables) on gestational age (GA) at birth using linear regression. Covariates were added to adjust for sex, birthweight Z-score, age at sample collection, maternal smoking, estimated epithelial cell proportions and 17 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 (DMRs) were assessed using the *dmrff* function in the *dmrff* package in R.⁵⁹ Here, a differentially methylated region is a region containing two or more 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 gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) databases, and using the *gometh* function in *missMethyl* package, which controls for multiple probe bias.⁶⁰ We performed an analysis that included those sites that reached genome-wide significance in EWAS and a

second analysis that also incorporated those sites that contributed to differentially methylated regions.

Principal component analysis

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

Linear regression between DNAm and peak width skeletonized 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 peak width of skeletonised mean diffusivity (PSMD), peak width of skeletonised neurite density index (PSNDI) and peak width of skeletonised fractional anisotropy (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 two scanners were included. PSNDI was only available from data acquired on one scanner, so it was not necessary to include the scanner as a covariate. We report standardized 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 DNAm data available upon request from Theirworld Edinburgh Birth Cohort, University of Edinburgh (james.boardman@ed.ac.uk or <https://www.tebc.ed.ac.uk/2019/12/data-access-and-collaboration/>). The DNAm and

metadata are not publicly available due to them containing information that could compromise participant consent.

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 between the twins that were removed and the twins that remained in the sample ($t = -0.157$, $P = 0.88$) or for sex ($\chi^2 = 0.417$, $P = 0.52$).

The study group consisted of 258 neonates: 155 participants were preterm (gestational age range 23.28–34.84) and 103 were infants born at full term (gestational age range 36.42–42.14), see [Table 1](#) for participant characteristics and [Supplementary Fig. 2](#) for participant flow. Among the preterm infants, 38 (25%) had bronchopulmonary dysplasia (defined as need for supplementary oxygen ≥ 36 weeks gestational age), 9 (6%) developed necrotizing 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 the 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

Table 1 Participant characteristics

	Preterm infants (n = 155)	Term infants (n = 103)	<i>P</i> -value ^a
Gestational age at birth/weeks (range)	28.84 (23.28–34.84)	39.7 (36.42–42.14)	<0.05
Gestational age at scan/weeks (range)	40.56 (37.70–45.14)	42.27 (39.84–47.14)	<0.05
Birth weight/g (range)	1177 (500–2100)	3482 (2346–4670)	<0.05
Birth weight Z-score (range)	-0.19 (range -3.13–1.58)	0.43 (range -2.30–2.96)	<0.05
Sex: female (%)	75 (48)	44 (43)	0.37
Maternal folate supplementation in pregnancy (%)	136 (88)	86 (83)	0.33
Maternal age (years)	31.1 (17–44)	33.7 (19–48)	<0.05
Maternal tobacco smoker in pregnancy (%)	29 (19)	2 (2)	<0.05
Maternal diabetes (%)	10 (6)	6 (6)	0.84
Pregnancy-induced hypertension (%)	22 (14)	7 (7)	0.07
Highest maternal qualification			<0.05
None	4 (3)	0 (0)	
High school	43 (28)	7 (7)	
College/university	106 (68)	95 (92)	

^aStudent's *t*-test was used to analyse continuous variables and χ^2 to analyse proportions. There were three missing for maternal qualifications (two preterm, one term).

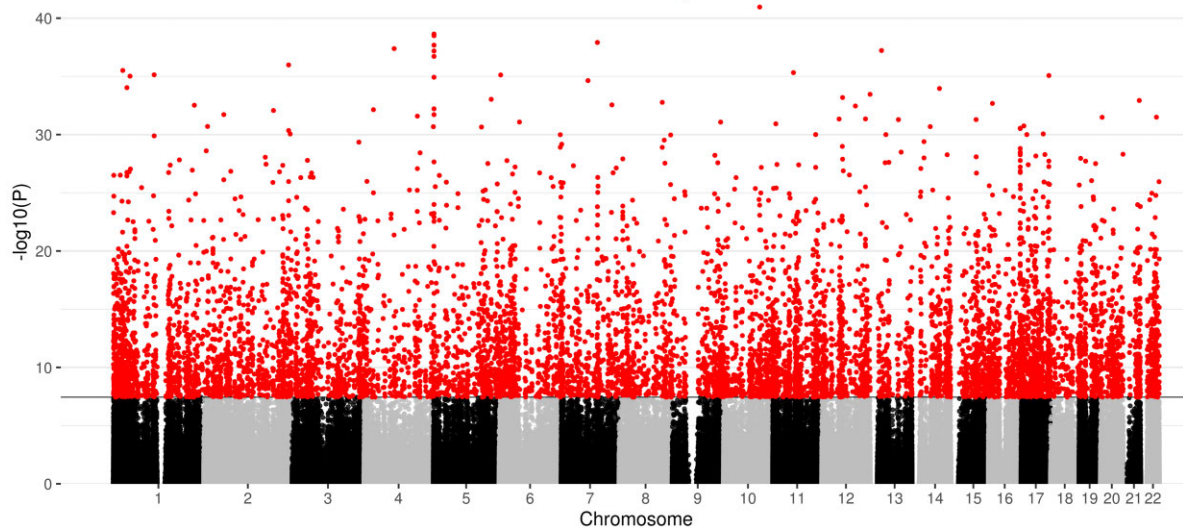


Figure 1 Manhattan plot for the significance [$-\log_{10}(P\text{-value})$ used for visualization purposes] of the association between gestational age at birth (weeks) and DNA methylation, following adjustment for covariates and surrogate variables. The solid horizontal line shows the genome-wide significance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

on gestational age at birth, controlling for birthweight Z-score, infant sex, age at sample collection, maternal smoking, estimated epithelial cell proportion and surrogate variables. The genomic inflation factor was 1.72 (Supplementary Fig. 3). Differential methylation in relation to gestational age at birth was identified at 8870 CpG sites at genome-wide significance ($P < 3.6 \times 10^{-8}$), Fig. 1. Of these, 4250 (47.9%) sites demonstrated a positive association with gestational age, while 4620 (52.1%) were negatively associated. Following Bonferroni adjustment, 1767 DMRs corresponding to 4664 CpG sites were significant at $P < 0.05$. Of these, 11 had 10 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 differentially methylated probes (DMPs), three probes were localized to the *IRX4* gene, one probe to the *GAL3ST4* gene, and one to *LOXL4* (Table 2; Supplementary Fig. 4). The probes with the largest absolute magnitude effect size (top five hypermethylated and hypomethylated) were mapped to the following genes: *IRX2*, *SMIM2*, *INTS1*, *HEATR2*, *ZBP1* and *UBXN11* (Table 3).

Pathways implicated in functional testing

Based on the 8870 sites that met the genome-wide significance threshold ($P < 3.6 \times 10^{-8}$), no KEGG terms remained significant following false discovery rate (FDR) correction for multiple comparisons. Two GO 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

DNA_m and PSMD and PSFA were both available for 214 infants (Supplementary Table 1), and DNA_m 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 B; Table 5).

Differential DNA_m is associated with white matter microstructure

The first unrotated PC (PC1) derived from the 8870 genome-wide significant CpGs accounted for 23.5% of the variance, and the second PC accounted for 2.5% (Fig. 3A). There was no evidence of batch effects in the PCs (Supplementary Fig. 5). 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.

Table 2 Most significant probes associated with gestational age at birth

Probe	Chromosome	P-value	Gene	Direction of effect	Coefficient ^a	Standard error	Relation to island
cg03558436	5	1.26×10^{-44}	—	+	1.02×10^{-2}	5.82×10^{-4}	Open Sea
cg04466438	9	1.13×10^{-42}	—	+	7.55×10^{-3}	4.47×10^{-4}	Open Sea
cg23701943	10	1.11×10^{-41}	<i>LOXL4</i>	+	1.04×10^{-2}	6.29×10^{-4}	Open Sea
cg18172877	5	2.31×10^{-39}	<i>IRX4</i>	—	-6.12×10^{-3}	3.85×10^{-4}	Island
cg04180086	5	3.22×10^{-39}	<i>IRX4</i>	—	-7.36×10^{-3}	4.64×10^{-4}	Island
cg22645539	7	1.22×10^{-38}	<i>GAL3ST4</i>	—	-8.15×10^{-3}	5.20×10^{-4}	North Shelf
cg17774559	5	2.09×10^{-38}	<i>IRX4</i>	—	-6.35×10^{-3}	4.06×10^{-4}	Island
cg17582074	4	4.12×10^{-38}	—	+	5.25×10^{-3}	3.38×10^{-4}	Open Sea
cg08915267	13	6.01×10^{-38}	—	—	-5.15×10^{-3}	3.32×10^{-4}	North Shelf
cg04441405	5	6.60×10^{-38}	—	—	-1.10×10^{-2}	7.13×10^{-4}	Island

^aCoefficient 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	Standard error	Relation to island
cg10402321	1	3.11×10^{-36}	<i>UBXN11</i>	—	-1.14×10^{-2}	7.60×10^{-4}	Open Sea
cg04441405	5	6.60×10^{-38}	—	—	-1.10×10^{-2}	7.13×10^{-4}	Island
cg07167946	5	1.94×10^{-32}	<i>IRX4</i>	—	-9.85×10^{-3}	7.13×10^{-4}	Island
cg07803375	7	3.6×10^{-22}	<i>HEATR2</i>	—	-9.08×10^{-3}	8.50×10^{-4}	North Shelf
cg14670058	13	9.24×10^{-23}	<i>SMIM2</i>	—	-9.07×10^{-3}	8.35×10^{-4}	Open Sea
cg16051275	6	7.53×10^{-36}	—	+	1.23×10^{-2}	8.30×10^{-4}	Open Sea
cg11460314	20	4.21×10^{-20}	<i>ZBPI</i>	+	1.24×10^{-2}	1.24×10^{-3}	Open Sea
cg04118102	17	1.00×10^{-30}	—	+	1.31×10^{-2}	9.86×10^{-4}	South Shelf
cg17368297	16	1.55×10^{-25}	—	+	1.40×10^{-2}	1.20×10^{-3}	Open Sea
cg14576951	7	6.73×10^{-30}	<i>INTS1</i>	+	1.44×10^{-2}	1.10×10^{-3}	Island

^aCoefficient corresponds to methylation change per week of gestation.

Discussion

By studying a unique database of DNA linked to brain imaging in a population of preterm and term infants, 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 DNAm. 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 generalized 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 foetal brain²² and umbilical cord blood;^{34,62} and widely distributed variation is reported in association with postmenstrual age at the sampling of preterm infants (a proxy for gestational age 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 gestational age in a meta-analysis of umbilical cord blood samples that reported 8899 gestation-dependent CpGs.³⁴ The limited overlap

could be explained by differences in the 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 foetal 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 GO terms, for adherens and anchoring junctions. When distinct probes that contributed to both DMPs and DMRs were combined, GO 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: *NNAT* and *BLCAP*. *NNAT* encodes the neural fate initiator neuronatin, the expression of which decreases throughout development;⁶⁵ there was a positive association with increasing gestational age at birth. Hypomethylation of *NNAT* is associated with a corresponding increase in the expression of neuronatin.⁶⁶ *BLCAP*

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

Gene ontology	Term	FDR Q value	Number of probes associated with the gene ontology/total number of probes in the ontology	Type	Description
GO:0005912	Adherens junction	0.00002317	217/545	Cellular component	A cell–cell junction composed of the epithelial cadherin–catenin complex.
GO:0005925	Focal adhesion	0.00907565	154/404	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	428/1413	Biological process	The attachment of a cell, to another cell or to the extracellular matrix, via cell adhesion molecules.
GO:0007167	Enzyme-linked receptor protein signalling pathway	0.01243116	322/1043	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	237/719	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	431/1420	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	257/756	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	229/664	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	422/1296	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	155/411	Cellular component	A cell junction between a cell and the extracellular matrix.
GO:0034330	Cell junction organization	0.04159681	119/290	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	130/331	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	186/499	Molecular function	Interacting selectively and non-covalently with a cell adhesion molecule.
GO:0070161	Anchoring junction	0.00002317	222/561	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.

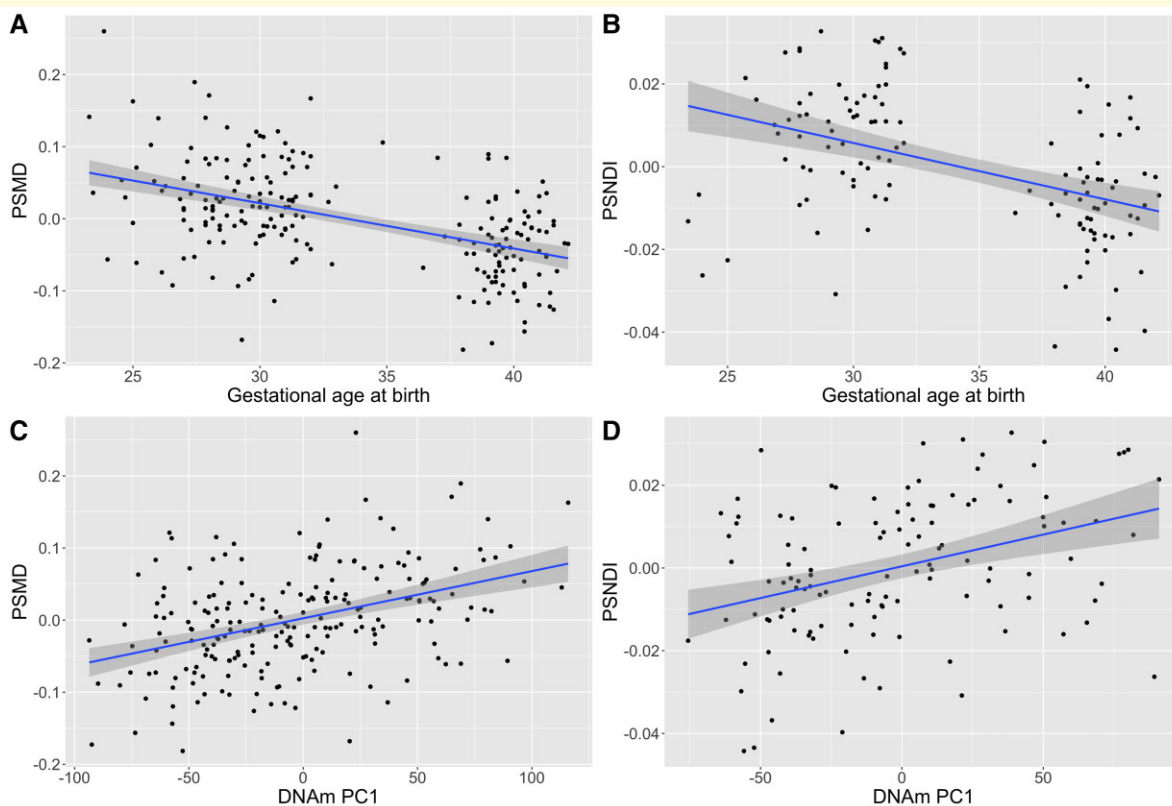


Figure 2 Scatter plots with regression lines and 95% confidence intervals showing the relationships between gestational age at birth (weeks) and DNAm with PSMD and PSNDI, where peak width is the difference between the 95th and 5th centile of histogram values across the white matter skeleton. The associations between gestational age (weeks) 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 residualized for gestational age at scan; PSMD is additionally residualized for scanner variable.

encodes the bladder cancer-associated protein. This is a tumour suppressor that induces apoptosis, with high expression in brain and B lymphocyte. The candidacy of this locus, encoding *NNAT* and *BLCAP*, as a region 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} In addition, *BLCAP* was found to have reduced methylation in placental samples from women who had preeclampsia.⁷¹

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 probes were found in genes including *ZBP1*, which was identified in EWAS investigating gestational age and hypertensive disorders of pregnancy;^{62,69,72} *INTS1*, which has been identified in EWAS of gestational age, hypertensive disorders of pregnancy, maternal body mass index, birthweight and breastfeeding duration.^{22,30,62,67,72,73} Hypomethylated probes were found in genes including *UBXN11*, which was identified in studies of gestational age;^{22,62} and *IRX4*. Three of the 10 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.⁷⁴ Its homologues have been implicated in retinal axon guidance in zebrafish, and neural patterning in *Xenopus*,^{75,76} and it has been identified in previous EWAS of hypertensive disorders of pregnancy⁷² and prenatal maternal stress.⁷⁷

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 characterized by intellectual disability.^{78–80}

We used metrics of generalized white matter connectivity to assess relationships between DNAm and brain development because generalized dysconnectivity of structural networks is a hallmark of preterm brain dysmaturation.^{10,14,81} PSNDI and PSMD were strongly associated with the first principal component of gestational age-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 characterize 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 tract;²⁶ it is most likely that we did not observe an association between gestational age or DNAm with PSFA because this metric is subject to histogram shift,²⁰ meaning that although there are groupwise differences in FA values across the skeleton, the spread of values is the same. Associations between gestational age 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 gestational age and brain network dysconnectivity in preterm infants. However, some of the DNAm signatures may be related to underlying causes of prematurity that operate in foetal life, such as infection or

preeclampsia. We cannot rule out that such changes are involved in the aetiology of preterm birth, which might preclude them from mediating specific associations with brain development.

The strengths of this study are that we studied a population of preterm and term infants across the gestational age range of 23–42 weeks, who were uniquely phenotyped with DNAm and dMRI. We sampled after the period of NICU care 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 peak width skeletonized 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 saliva methylome and dMRI data. Cohorts, with such data, that recruit neonates from across the gestational age continuum would also provide valuable replication of analyses described here. There was also some evidence of inflation of test statistics based on the genomic inflation factor. However, the genomic inflation factor is thought to provide an overestimate of inflation and corrections based on it may

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

PS metric	Metric ~ PCI DNAm + age at scan + scanner variable ^a		Metric ~ gestational age at birth + 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	8.37×10^{-10}	−0.602	$<2 \times 10^{-16}$
PSNDI	0.364	4.15×10^{-5}	−0.594	2.17×10^{-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.

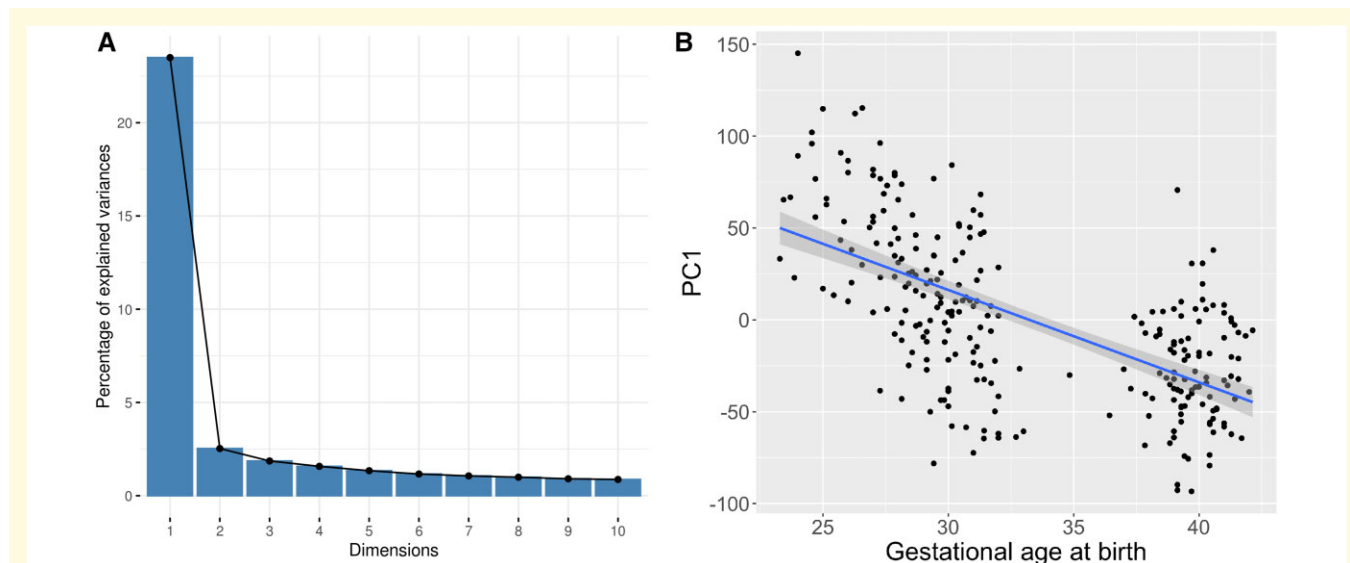


Figure 3 Variation in DNAm probes selected by EWAS captured by principal component analysis, and the relationship with gestational age (weeks). A. A scree plot showing percentage of variance accounted for by the first 10 components, with a sharp elbow after the first PC, accounting for 23.5% of variance. B. A scatter plot, with regression line, showing the relationship between gestational age at birth (weeks) and PC1 ($r = -0.622$) with 95% confidence intervals.

be overly conservative.⁸² 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, is in line with previous studies that have investigated this in cord blood and foetal brain tissue.^{22,34} Finally, preterm birth itself can reflect maternal health or circumstance; in this cohort, mothers of preterm infants were more likely to have pregnancy-induced hypertension and the control mothers were more likely to have college or university qualifications. It is possible that these, or other maternal factors associated with preterm birth such as infection, chronic disease, poor nutrition and alcohol, tobacco or drug use, could influence the neonatal methylome at term equivalent age. A large prospective study with longitudinal sampling from birth would be required to test the hypothesis that maternal factors contribute to changes in the methylome apparent after neonatal intensive care.

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. GO terms related to cell–cell contacts were enriched, indicating that cell contacts and organization 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). The authors 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. The graphical abstract was created with [BioRender.com](https://www.biorender.com).

Funding

E.N.W.W. is supported by the Wellcome Trust funded 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 Medical Research Council Centre for Reproductive Health, which was funded by Medical Research Council Centre Grant (MRC G1002033).

Competing interests

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

Supplementary material

Supplementary material is available at *Brain Communications* online.

References

1. Chawanpaiboon S, Vogel JP, Moller AB, *et al.* Global, regional, and national estimates of levels of preterm birth in 2014: A systematic review and modelling analysis. *Lancet Glob Health*. 2019;7(1): e37–e46.
2. Agrawal S, Rao SC, Bulsara MK, Patole SK. Prevalence of autism spectrum disorder in preterm infants: A meta-analysis. *Pediatrics*. 2018;142(3):e20180134.
3. Franz AP, Bolat GU, Bolat H, *et al.* Attention-deficit/hyperactivity disorder and very preterm/very low birth weight: A meta-analysis. *Pediatrics*. 2018;141(1):e20171645.
4. Nosarti C, Reichenberg A, Murray RM, *et al.* Preterm birth and psychiatric disorders in young adult life. *Arch Gen Psychiatry*. 2012; 69(6):E1–E8.
5. Van Lieshout RJ, Ferro MA, Schmidt LA, *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*. 2018;59(11):1192–1200.
6. Burnett AC, Anderson PJ, Cheong J, Doyle LW, Davey CG, Wood SJ. Prevalence of psychiatric diagnoses in preterm and full-term children, adolescents and young adults: A meta-analysis. *Psychol Med*. 2011;41(12):2463–2474.
7. Mackay DF, Smith GCS, Dobbie R, Pell JP. Gestational age at delivery and special educational need: Retrospective cohort study of 407, 503 schoolchildren. *PLoS Med*. 2010;7(6):e1000289.
8. Johnson S, Hollis C, Kochhar P, Hennessy E, Wolke D, Marlow N. Autism spectrum disorders in extremely preterm children. *J Pediatr*. 2010;15(4):525–531.e2.
9. Johnson S, Evans TA, Draper ES, *et al.* Neurodevelopmental outcomes following late and moderate prematurity: A population-based cohort study. *Arch Dis Child Fetal Neonatal Ed*. 2015; 100(4):F301–F308.
10. Volpe JJ. Brain injury in premature infants: A complex amalgam of destructive and developmental disturbances. *Lancet Neurol*. 2009; 8(1):110–124.
11. Pietsch M, Christiaens D, Hutter J, *et al.* A framework for multi-component analysis of diffusion MRI data over the neonatal period. *Neuroimage*. 2019;186:321–337.
12. Zhang H, Schneider T, Wheeler-Kingshott CA, Alexander DC. NODDI: Practical in vivo neurite orientation dispersion and density imaging of the human brain. *Neuroimage*. 2012;61(4):1000–1016.
13. Tariq M, Schneider T, Alexander DC, Gandini Wheeler-Kingshott CA, Zhang H. Bingham-NODDI: Mapping anisotropic orientation

- dispersion of neurites using diffusion MRI. *Neuroimage*. 2016;133:207–223.
14. Telford EJ, Cox SR, Fletcher-Watson S, *et al.* A latent measure explains substantial variance in white matter microstructure across the newborn human brain. *Brain Struct Funct*. 2017;222(6):4023–4033.
 15. Ball G, Pazderova L, Chew A, *et al.* Thalamocortical connectivity predicts cognition in children born preterm. *Cereb Cortex*. 2015;25(11):4310–4318.
 16. Boardman JP, Counsell SJ. Factors associated with atypical brain development in preterm infants: Insights from magnetic resonance imaging. *Neuropathol Appl Neurobiol*. 2020;46(5):413–421.
 17. Baykara E, Gesierich B, Adam R, *et al.* A novel imaging marker for small vessel disease based on skeletonization of white matter tracts and diffusion histograms. *Ann Neurol*. 2016;80(4):581–592.
 18. Deary IJ, Ritchie SJ, Muñoz Maniega S, *et al.* Brain peak width of skeletonized mean diffusivity (PSMD) and cognitive function in later life. *Front Psychiatry*. 2019;10:524.
 19. Wei N, Deng Y, Yao L, *et al.* A neuroimaging marker based on diffusion tensor imaging and cognitive impairment due to cerebral white matter lesions. *Front Neurol*. 2019;10:81.
 20. Blesa M, Galdi P, Sullivan G, *et al.* Peak width of skeletonized water diffusion MRI in the neonatal brain. *Front Neurol*. 2020;11:235.
 21. Sullivan, G, Galdi P, Cabez MB, *et al.* Interleukin-8 dysregulation is implicated in brain dysmaturation following preterm birth. *Brain Behav Immun*. 2020;90:311–318.
 22. Spiers H, Hannon E, Schalkwyk LC, *et al.* Methylomic trajectories across human fetal brain development. *Genome Res*. 2015;25(3):338–352.
 23. Lyst MJ, Bird A. Rett syndrome: A complex disorder with simple roots. *Nat Rev Genet*. 2015;16(5):261–274.
 24. Weemaes, CM, van Tol MJ, Wang J, *et al.* Heterogeneous clinical presentation in ICF syndrome: Correlation with underlying gene defects. *Eur J Hum Genet*. 2013;21(11):1219–1225.
 25. Rangasamy S, D’Mello SR, Narayanan V. Epigenetics, autism spectrum, and neurodevelopmental disorders. *Neurotherapeutics*. 2013;10(4):742–756.
 26. Sparrow S, Manning JR, Cartier J, *et al.* Epigenomic profiling of preterm infants reveals DNA methylation differences at sites associated with neural function. *Transl Psychiatry*. 2016;6(1):e716.
 27. Wheeler ENW, Stoye DQ, Cox SR, *et al.* DNA methylation and brain structure and function across the life course: A systematic review. *Neurosci Biobehav Rev*. 2020;113:133–156.
 28. Joubert BR, Felix JF, Yousefi P, *et al.* DNA methylation in newborns and maternal smoking in pregnancy: Genome-wide consortium meta-analysis. *Am J Hum Genet*. 2016;98(4):680–696.
 29. Sharp GC, Salas LA, Monnereau 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*. 2016;26(20):4067–4085.
 30. Küpers LK, Monnereau C, Sharp GC, *et al.* Meta-analysis of epigenome-wide association studies in neonates reveals widespread differential DNA methylation associated with birthweight. *Nat Commun*. 2017;10(1):1893.
 31. Everson TM, O’Shea TM, Burt A, *et al.* Serious neonatal morbidities are associated with differences in DNA methylation among very preterm infants. *Clin Epigenetics*. 2017;12(1):151.
 32. Everson TM, Marsit CJ, Michael O’Shea T, *et al.* Epigenome-wide analysis identifies genes and pathways linked to neurobehavioral variation in preterm infants. *Sci Rep*. 2019;9(1):6322.
 33. Cruickshank MN, Oshlack A, Theda C, *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*. 2013;5(10):96.
 34. Merid SK, Novoloaca A, Sharp GC, *et al.* Epigenome-wide meta-analysis of blood DNA methylation in newborns and children identifies numerous loci related to gestational age. *Genome Med*. 2020;12(1):25.
 35. Armstrong DA, Lesueur C, Conradt E, Lester BM, Marsit CJ. Global and gene-specific DNA methylation across multiple tissues in early infancy: Implications for children’s health research. *FASEB J*. 2014;28(5):2088–2097.
 36. Lowe R, Gemma C, Beyan H, *et al.* Buccals are likely to be a more informative surrogate tissue than blood for epigenome-wide association studies. *Epigenetics*. 2014;8(4):445–454.
 37. Braun PR, Han S, Hing B, *et al.* Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Transl Psychiatry*. 2019;9(4):47.
 38. Smith AK, Kilaru V, Klengel T, *et al.* DNA extracted from saliva for methylation studies of psychiatric traits: Evidence tissue specificity and relatedness to brain. *Am J Med Genet B Neuropsychiatr Genet*. 2015;168B(1):36–44.
 39. Boardman JP, Hall J, Thrippleton MJ, *et al.* Impact of preterm birth on brain development and long-term outcome: Protocol for a cohort study in Scotland. *BMJ Open*. 2020;10(3):e035854.
 40. Aryee MJ, Jaffe AE, Corrada-Bravo H, *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–1369.
 41. Pidsley R, Wong CCY, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450 K methylation array data. *BMC Genomics*. 2013;14:293.
 42. 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. *Genom Data*. 2013;9:22–24.
 43. Zheng SC, Webster AP, Dong D, *et al.* A novel cell-type deconvolution algorithm reveals substantial contamination by immune cells in saliva, buccal and cervix. *Epigenomics*. 2018;10(7):925–940.
 44. Caruyer E, Lenglet C, Sapiro G, Deriche R. Design of multishell sampling schemes with uniform coverage in diffusion MRI. *Magn Reson Med*. 2013;69(6):1534–1540.
 45. Veraart J, Fieremans E, Noviko DS. Diffusion MRI noise mapping using random matrix theory. *Magn Reson Med*. 2016;76(5):1582–1593.
 46. Veraart J, Novikov DS, Christiaens D, Ades-aron B, Sijbers J, Fieremans E. Denoising of diffusion MRI using random matrix theory. *Neuroimage*. 2016;142:394–406.
 47. Tournier JD, Smith R, Raffelt D, *et al.* MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *Neuroimage*. 2019;202:116137.
 48. Smith SM, Jenkinson M, Woolrich MW, *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. 2004;23(Suppl 1):S208–S219.
 49. Andersson JLR, Sotiropoulos SN. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *Neuroimage*. 2016;125:1063–1078.
 50. Andersson JLR, Graham MS, Zsoldos E, Sotiropoulos SN. Incorporating outlier detection and replacement into a non-parametric framework for movement and distortion correction of diffusion MR images. *Neuroimage*. 2016;141:556–572.
 51. Tustison NJ, Avants BB, Cook PA, *et al.* N4ITK: Improved N3 bias correction. *IEEE Trans Med Imaging*. 2010;29(6):1310–1320.
 52. Andersson JLR, Graham MS, Drobniak I, Zhang H, Filippini N, Bastiani M. Towards a comprehensive framework for movement and distortion correction of diffusion MR images: Within volume movement. *Neuroimage*. 2017;152:450–466.
 53. Hernandez-Fernandez M, Reguly I, Jbabdi S, Giles M, Smith S, Sotiropoulos SN. Using GPUs to accelerate computational diffusion MRI: From microstructure estimation to tractography and connectomes. *Neuroimage*. 2019;188:598–615.
 54. RCoreTeam. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, 2020.
 55. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet*. 2007;3(6):1724–1735.

56. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics*. 2012;28(6):882–883.
57. Ritchie ME, Phipson B, Wu D, *et al*. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7):e47.
58. Saffari A, Silver MJ, Zavattari P, *et al*. Estimation of a significance threshold for epigenome-wide association studies. *Genet Epidemiol*. 2018;42(1):20–33.
59. Suderman M, Staley JR, French R, Arathimos R, Simpkin A, Tilling. Dmrrf: Identifying differentially methylated regions efficiently with power and control. *BioRxiv*, doi:10.1101/508556, 31 December 2018, preprint: not peer reviewed.
60. Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide DNA methylation data. *Genome Biol*. 2021;22:173.
61. Job DE, Dickie DA, Rodriguez D, *et al*. A brain imaging repository of normal structural MRI across the life course: Brain images of normal subjects (BRAINS). *Neuroimage*. 2017;144:299–304.
62. Bohlin J, Häberg SE, Magnus P, *et al*. Prediction of gestational age based on genome-wide differentially methylated regions. *Genome Biol*. 2016;17(1):207.
63. Provenzi L, Guida E, Montirosso R. Preterm behavioral epigenetics: A systematic review. *Neurosci Biobehav Rev*. 2017;84:262–271.
64. Fontana C, Marasca F, Provitera L, *et al*. Early maternal care restores LINE-1 methylation and enhances neurodevelopment in pre-term infants. *BMC Med*. 2021;19(1):42.
65. Lin HH, Bell E, Uwanogho D, *et al*. Neuronatin promotes neural lineage in ESCs via Ca²⁺ signaling. *Stem Cells*. 2010;28(11):1950–1960.
66. Numata S, Ye T, Hyde TM, *et al*. DNA methylation signatures in development and aging of the human prefrontal cortex. *Am J Hum Genet*. 2012;90(2):260–272.
67. Sharp GC, Lawlor DA, Richmond RC, *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*. 2015;44(4):1288–1304.
68. Hannon E, Dempster E, Viana J, *et al*. An integrated genetic-epigenetic analysis of schizophrenia: Evidence for colocalization of genetic associations and differential DNA methylation. *Genome Biol*. 2016;17(1):176.
69. Hannon E, Schendel D, Ladd-Acosta C, *et al*. Variable DNA methylation in neonates mediates the association between prenatal smoking and birth weight. *Philos Trans R Soc Lond B Biol Sci*. 2019;374(1770):20180120.
70. Battram, T., Yousefi P, Crawford G, *et al*. The EWAS Catalog: A database of epigenome-wide association studies. OSF Preprint, doi:10.31219/osf.io/837wn, 4 February 2021.
71. Li Y, Cui S, Shi W, *et al*. Differential placental methylation in pre-eclampsia, preterm and term pregnancies. *Placenta*. 2020;93:56–63.
72. Kazmi N, Sharp GC, Reese SE, *et al*. Hypertensive disorders of pregnancy and DNA methylation in newborns: Findings from the pregnancy and childhood epigenetics consortium. *Hypertension*. 2015;74(2):375–383.
73. Odintsova VV, Hagenbeek FA, Suderman M, *et al*. DNA methylation signatures of breastfeeding in buccal cells collected in mid-childhood. *Nutrients*. 2019;11(11):2804.
74. Bruneau BG, Bao ZZ, Tanaka M, *et al*. Cardiac expression of the ventricle-specific homeobox gene *Irx4* is modulated by *Nkx2-5* and *dHand*. *Dev Biol*. 2000;217(2):266–277.
75. Jin Z, Zhang J, Klar A, *et al*. *Irx4*-mediated regulation of *Slit1* expression contributes to the definition of early axonal paths inside the retina. *Development*. 2003;130(6):1037–1048.
76. Garriock RJ, Vokes SA, Small EM, Larson R, Krieg PA. Developmental expression of the *Xenopus* Iroquois-family homeobox genes, *Irx4* and *Irx5*. *Dev Genes Evol*. 2001;211(5):257–260.
77. Rijlaarsdam J, Pappa I, Walton E, *et al*. An epigenome-wide association meta-analysis of prenatal maternal stress in neonates: A model approach for replication. *Epigenetics*. 2016;11(2):140–149.
78. Oegema R, Baillat D, Schot R, *et al*. Human mutations in integrator complex subunits link transcriptome integrity to brain development. *PLoS Genet*. 2016;13(8):e1006923.
79. Krall M, Htun S, Schnur RE, *et al*. Biallelic sequence variants in *INTS1* in patients with developmental delays, cataracts, and craniofacial anomalies. *Eur J Hum Genet*. 2019;27(4):582–593.
80. Zhang, X, Wang Y, Yang F, *et al*. Biallelic *INTS1* mutations cause a rare neurodevelopmental disorder in two Chinese siblings. *J Mol Neurosci*. 2020;70(1):1–8.
81. Galdi P, Blesa M, Stoye DQ, *et al*. Neonatal morphometric similarity mapping for predicting brain age and characterizing neuroanatomic variation associated with preterm birth. *Neuroimage Clin*. 2020;25:102195.
82. van Iterson M, van Zwet EW; BIOS Consortium, Heijmans BT. Controlling bias and inflation in epigenome- and transcriptome-wide association studies using the empirical null distribution. *Genome Biol*. 2017;18(1):19.