8-5-2024

Epigenetic associations with neonatal age in infants born very preterm, particularly among genes involved in neurodevelopment.

Kenyaita M. Hodge

Amber A. Burt

Marie Camerota

Brian S. Carter
Children's Mercy Hospital

Jennifer Check

See next page for additional authors

Recommended Citation

Hodge KM, Burt AA, Camerota M, et al. Epigenetic associations with neonatal age in infants born very preterm, particularly among genes involved in neurodevelopment. Sci Rep. 2024;14(1):18147. Published 2024 Aug 5. doi:10.1038/s41598-024-68071-w

scientific reports

Check for updates

OPEN

Epigenetic associations with neonatal age in infants born very preterm, particularly among genes involved in neurodevelopment

Kenyaita M. Hodge[1], Amber A. Burt[1], Marie Camerota[2,3], Brian S. Carter[4], Jennifer Check[5], Karen N. Conneely[6], Jennifer Helderman[5], Julie A. Hofheimer[7], Anke Hüls[1,8,9], Elisabeth C. McGowan[10], Charles R. Neal[11], Steven L. Pastyrnak[12], Lynne M. Smith[13], Sheri A. DellaGrotta[3], Lynne M. Dansereau[3], T. Michael O'Shea[7], Carmen J. Marsit[1,8], Barry M. Lester[2,3,10] & Todd M. Everson[1,8]✉

The time from conception through the first year of life is the most dynamic period in human development. This time period is particularly important for infants born very preterm (< 30 weeks gestation; VPT), as they experience a significant disruption in the normal developmental trajectories and are at heightened risk of experiencing developmental impairments and delays. Variations in the epigenetic landscape during this period may reflect this disruption and shed light on the interrelationships between aging, maturation, and the epigenome. We evaluated how gestational age (GA) and age since conception in neonates [post-menstrual age (PMA)], were related to DNA methylation in buccal cells collected at NICU discharge from VPT infants (n = 538). After adjusting for confounders and applying Bonferroni correction, we identified 2,366 individual CpGs associated with GA and 14,979 individual CpGs associated with PMA, as well as multiple differentially methylated regions. Pathway enrichment analysis identified pathways involved in axonogenesis and regulation of neuron projection development, among many other growth and developmental pathways (FDR q < 0.001). Our findings align with prior work, and also identify numerous novel associations, suggesting that genes important in growth and development, particularly neurodevelopment, are subject to substantial epigenetic changes during early development among children born VPT.

Keywords Preterm, Epigenetics, Gestational age, Neonatal, Perinatal, Methylation

The neonatal and infancy periods are the most dynamic in human growth and development. During these periods, the nervous, respiratory, vascular, and immune systems, among others undergo vital changes. For instance, the brain begins to develop in the third week of gestation with robust changes in cortical grey and white matter

[1]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University, 1518 Clifton Road NE, Atlanta, GA 30322, USA. [2]Department of Psychiatry and Human Behavior, Warren Alpert Medical School of Brown University, Providence, RI, USA. [3]Brown Center for the Study of Children at Risk, Women and Infants Hospital, Providence, RI, USA. [4]Department of Pediatrics-Neonatology, Children's Mercy Hospital, Kansas City, MO, USA. [5]Department of Pediatrics, Wake Forest School of Medicine, Winston-Salem, NC, USA. [6]Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA, USA. [7]Department of Pediatrics, University of North Carolina School of Medicine, Chapel Hill, NC, USA. [8]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [9]Department of Biostatistics and Bioinformatics, Rollins School of Public Health, Emory University, Atlanta, GA, USA. [10]Department of Pediatrics, Warren Alpert Medical School of Brown University and Women and Infants Hospital, Providence, RI, USA. [11]Department of Pediatrics, University of Hawaii John A. Burns School of Medicine, Honolulu, HI, USA. [12]Department of Pediatrics, Spectrum Health-Helen Devos Hospital, Grand Rapids, MI, USA. [13]Department of Pediatrics, Harbor-UCLA Medical Center, Torrance, CA, USA. ✉email: todd.m.everson@emory.edu

beginning at 20 weeks of gestation and the volume of grey matter increasing up until the second year of life[1,2]. Similarly, development of the respiratory system involves a series of intricate stages, from the formation of the airway tree in gestational weeks 6 and 7, to the branching and growth of bronchioles, then production of pulmonary surfactant and the formation of alveoli between gestational weeks 32 and 36[3]. Preterm birth disrupts the normal progression of these processes, and of other organ systems, which can give rise to a range of neonatal medical complications and developmental delays. Clearly, age since conception is a critical indicator of infant health and development and the timing of birth can perturb healthy development. While phenotypic presentations of aging and development are clear, age also plays an important role in molecular regulation as studies have observed widespread and robust changes in the epigenome during early postnatal development, most notably with DNA methylation[4]. However, there are few studies that focus on the relationship between age and epigenetic patterns during the neonatal period[5].

Early life stress, such as that caused by preterm birth, has been linked to changes in DNA methylation (DNAm)[6–8]. DNAm is a biological process in which methyl groups are added to the DNA at cytosine (C) bases that are followed by guanine (G) bases, referred to as CpG sites. This process influences the expression potential of genes but does not change the sequence of DNA. DNAm is an epigenetic mechanism that has been studied extensively due to its ability to capture information about exposures throughout the course of life[9]. Several studies on the development of epigenetic clocks have successfully utilized buccal tissue, which primarily consists of epithelial cells, to accurately predict age in pediatric populations. For example, the development of the pediatric-buccal-epigenetic (PedBE) clock leveraged DNAm profiles from buccal epithelial cells from healthy individuals ranging 0 to 20 years of age[10]. When compared to the pan-tissue Horvath DNAm clock[11], the PedBE clock demonstrated higher accuracy in predicting chronological age in longitudinal sampling during the first year of life and in adolescence. McEwen et al., found that higher gestational age (GA) was associated with a higher PedBE age in typically developing individuals[10]. However, children with a neurodevelopmental disorder, as compared to typically developing children, had higher predicted PedBE age. This finding suggests that deviations in predicted PedBE age are associated with altered development. Our group developed the NEOage clock, an epigenetic clock that predicts chronological age in neonates born preterm[12]. Like the PedBE clock, the NEOage clock leverages DNAm profiles from buccal epithelial cells and has high accuracy in predicting neonatal age in both the training data and an independent saliva data set. These studies provide strong evidence that buccal tissue can be used to investigate the epigenetics of aging in neonatal and pediatric populations[10]. While these prior projects focused on the development of algorithms to estimate age from DNA methylation, our study aims to characterize the genes and genomic regions that are associated age-specific epigenetic changes in infants born preterm.

In the United States approximately 1 in 10 babies are born prematurely each year[13]. Globally, prematurity is the leading cause of death in children under five years old[14], resulting in a significant public health burden. Infants born very preterm (born < 30 weeks gestation) are at heightened risk of long-term health and development impairments, compared to their term-born counterparts. For example, bronchopulmonary dysplasia (BPD), the most common neonatal morbidity among preterm infants, has been associated with chronic diseases and neurodevelopmental impairments[15,16]. Little is known about how these early life stressors impact the epigenome of infants born preterm. However, studies have shown that preterm infants have different methylation patterns than those born at term, and these differences can be detected in multiple tissues, with special emphasis on minimally invasive tissues such as placenta, cord blood and saliva[17–19]. Variations in the epigenetic landscape during early life may shed light on the interrelationships between aging, maturation, the methylome, and the health of infants born preterm.

We aimed to address this gap in knowledge by conducting an epigenome-wide association study (EWAS) to determine if there are neonatal age associated differences in DNA methylation (DNAm) in buccal tissue of infants born very preterm, and to characterize implicated biological pathways. We also examined whether the buccal DNAm levels were correlated with brain DNAm levels of the identified differentially methylated CpGs, using an independent publicly available data source (GEO Accession: GSE111165)[20].

Results

Characteristics of the study population

There were 538 infants in the neonatal neurobehavior and outcomes in very preterm infants (NOVI) study for whom DNAm data, post-menstrual age (PMA) at NICU discharge, and GA at birth, and covariate data were available. In NOVI, GA ranged from 22.0 to 29.9 weeks (mean = 27.0 weeks) and PMA ranged from 32.1 to 51.4 weeks (mean = 39.2 weeks; Table 1). There were 115 infants (21.4%) born of multiple gestation, and 423 infants were singleton births. Of the 538 infants, 113 were born at a hospital which did not have capacity for neonatal intensive care (hereafter referred to as "outborn") (21.0%) which was associated with increasing PMA at discharge ($p = 0.0001$). The majority of infants were diagnosed with at least one neonatal morbidity including BPD (51.5%), serious brain injury (SBI; 12.8%), infection (19.0%), or severe retinopathy of prematurity (ROP; 6.3%). GA at birth was inversely associated with number of medical morbidities ($r = -0.48$, $p = 7.83 \times 10^{-33}$), which in turn was positively associated with PMA at discharge ($r = 0.59$, $p = 1.23 \times 10^{-52}$), when buccal swabs were collected.

Summary of GA and PMA EWAS

In our EWAS models, we adjusted for the potential confounding effects of neonatal morbidities, as well as the inverse correlation between GA and PMA by including both age variables and a cumulative morbidity risk score (described below) in our models. We also adjusted for sex, study site, batch, outborn status (i.e., requiring transfer into study site hospital), epithelial cell proportion, and the five most influential surrogate variables (SVs; described in the methods section)–SVs represent major sources of variation in the DNA methylation data and are included to account for unmeasured confounding.

Sample characteristics (n = 538)	
Overall	
GA (weeks)	27 (1.92)
PMA (weeks)	39 (3.38)
Infant race (%)	
White	278 (51.7)
Black	122 (22.7)
Other	51 (9.5)
Asian	41 (7.6)
Hawaiian/Pacific Islander	38 (7.0)
Native American	n < 5
Not reported	n < 5
Infant ethnicity (% Hispanic)	115 (21.4)
Outborn (%)	113 (21.0)
Male (%)	299 (55.6)
Maternal age at birth (years)	29.1 (6.36)
Maternal education < HS/GED (%)	72 (134)
Low SES (hollingshead level V; %)	42 (7.8)
Number of neonatal medical morbidities (%)	
0	214 (39.8)
1	196 (36.4)
2	99 (18.4)
3+	29 (5.4)
Proportion of epithelial cells (%)	98.98 (0.03)

Table 1. Distribution of demographic characteristics, neonatal morbidities and maternal/fetal characteristics of the NOVI study population. *GA* gestational age at birth, *PMA* post-menstrual age at discharge.

Overall, from our EWAS we observed 2,366 CpGs associated with GA and 14,979 CpGs associated (Bonferroni correction: $P_{Bonf.} = 0.05/706{,}323$) with PMA while mutually adjusting for each other, as well as other confounders and technical factors. There were 1,151 CpGs significantly associated with both GA and PMA; these coefficients had a strong inversely correlated relationship ($r = -0.95$, $p < 2.2 \times 10^{-16}$). Approximately 68% (n = 1608) of the CpGs significantly associated with GA were hypermethylated with older GA at birth. Of the CpGs significantly associated with PMA, 70% (n = 10,568) were hypermethylated with increasing PMA. We provide the complete set of coefficients, standard errors, and p-values of the GA and PMA associations as supplementary data files (Supplementary Tables S1 and S2).

Associations with GA

The CpG most significantly associated with GA was cg13211971 ($p = 9.79 \times 10^{-46}$) which was annotated to *PDE9A*. We highlight the 10 CpG associations with the smallest p-values, their relationships with GA, and their genomic annotations in (Table 2). We then performed differentially methylated regions (DMR) analysis to identify spatially clustered CpGs that exhibited the same direction of coefficient associated with GA. 417 CpGs identified from the EWAS were within 160 DMRs (Supplementary Tables S3 and S4). The DMRs included *IRX4* ($P_{Bonf} = 7.24 \times 10^{-20}$),

CpG	Coefficient	Standard Error	p-value	P_{Bonf}	Chromosome	Position	Gene	Region
cg13211971	0.2637	0.0186	9.79E-46	6.91E-40	21	44194350	*PDE9A*	Body
cg11424970	0.2264	0.0192	3.76E-32	2.66E-26	2	71969999	*2q13.2*	–
cg10202436	0.2272	0.0196	3.79E-31	2.68E-25	5	138733726	*SPATA24*	Body
cg03558436	0.2009	0.0176	2.69E-30	1.90E-24	5	173060723	*5q35.2*	–
cg18955208	0.1661	0.0146	4.07E-30	2.87E-24	6	21873871	*CASC15*	Body
cg09672187	0.2528	0.0227	9.00E-29	6.35E-23	5	1885367	*IRX4*	Body
cg23720947	−0.1766	0.0162	1.17E-27	8.28E-22	19	5967163	*RANBP3*	Body
cg03173167	0.1818	0.0169	4.28E-27	3.02E-21	4	154598228	*LOC100419170*	–
cg27258182	0.1931	0.0181	1.22E-26	8.61E-21	2	171500500	*MYO3B*	Body
cg14771313	0.1604	0.015	1.32E-26	9.35E-21	3	71161285	*FOXP1*	Body

Table 2. The top 10 most significant ($\alpha = 0.05/706323$) CpGs associated with gestational age (GA) in NOVI.

SLC7A5 ($P_{Bonf} = 7.20 \times 1^{-19}$), *LINC-PINT; FLJ43663* ($P_{Bonf} = 9.74 \times 10^{-19}$), *STAT5A* ($P_{Bonf} = 2.21 \times 10^{-16}$), *PHLDB1* ($P_{Bonf} = 1.14 \times 10^{-15}$), *NXN* ($P_{Bonf} = 7.60 \times 10^{-15}$), *TPM4* ($P_{Bonf} = 8.90 \times 10^{-14}$), *RBFOX2* ($P_{Bonf} = 1.00 \times 10^{-13}$), *PML* ($P_{Bonf} = 1.34 \times 10^{-13}$), and *ANP32A* ($P_{Bonf.} = 1.71 \times 10^{-13}$), as the annotated genes with the smallest p-values associated with GA (Table 3). We highlight the GA-DMRs that were annotated to these genes, along with our overall EWAS results with volcano and Manhattan plots (Fig. 1a,b). Interestingly, some genes not only had multiple CpGs that showed differential methylation, but also several DMRs. *SLC7A5* had the most DMRs (6) associated with GA which contained 22 CpGs in total, followed by *IRX4* with four DMRs containing 20 CpGs. Lastly, the most CpGs within a GA-DMR was eight (cg17650747, cg18246442, cg01875784, cg25378916, cg06654549, cg20426932, cg09155145, and cg14024788) which are located 1 kb away from the transcription start site of *IRX4*. The majority of the CpGs identified in highlighted DMRs are located within the gene body of their respective genes and were hypermethylated with increasing GA.

Associations with PMA

The CpG most significantly associated with PMA was cg06413398 which was annotated to *DDO* ($p = 3.30 \times 10^{-75}$), and we highlight the 10 CpGs with the smallest p-values, their relationships with PMA, and their genomic annotations in (Table 4). From the DMR analysis, we identified 4,498 CpGs within 1,562 Bonferroni significant DMRs (Supplementary Tables S5 and S6). The top DMRs with gene annotations were *DDO* ($P_{Bonf} = 1.54 \times 10^{-73}$), *RBPJ*

Gene	# of CpGs	Estimate	Standard Error	p-value	P_{Bonf}	Chromosome	Position (start)	Position (end)	Region
IRX4	3	1.5214	0.1451	1.02E-25	7.24E-20	5	1885778	1886309	Body
SLC7A5	3	1.2703	0.1237	1.01E-24	7.20E-19	16	87878292	87879698	Body
LINC-PINT; FLJ43663	2	1.2968	0.1267	1.37E-24	9.74E-19	7	130626559	130630462	Body
STAT5A	3	−0.9553	0.0985	3.11E-22	2.21E-16	17	40463425	40463806	3'UTR
6q21.2	3	1.5263	0.1587	6.57E-22	4.67E-16	12	76084308	76084549	–
PHLDB1	7	−1.1414	0.1198	1.61E-21	1.14E-15	11	118502137	118506352	Body;ExonBnd
SLC7A5	3	1.4335	0.1525	5.44E-21	3.86E-15	16	87870597	87873389	Body
NXN	6	−0.6998	0.0750	1.07E-20	7.60E-15	17	867763	872770	Body
4q31.3	2	−1.1678	0.1262	2.13E-20	1.51E-14	4	152720133	152720170	–
7q36.1	2	1.4169	0.1563	1.22E-19	8.69E-14	7	151620691	151620888	–
TPM4	3	1.1060	0.1220	1.25E-19	8.90E-14	19	16178426	16178570	5'UTR;1stExon
RBFOX2	2	−0.9662	0.1067	1.41E-19	1.00E-13	22	36278499	36278607	Body
PML	2	1.6763	0.1859	1.89E-19	1.34E-13	15	74290287	74290466	Body
ANP32A	2	−0.9526	0.1059	2.41E-19	1.71E-13	15	69082449	69084888	Body

Table 3. The most significant differentially methylated regions (DMR) containing 2 or more CpGs within 1500 base pairs of another significant CpG associated with gestational age (GA).

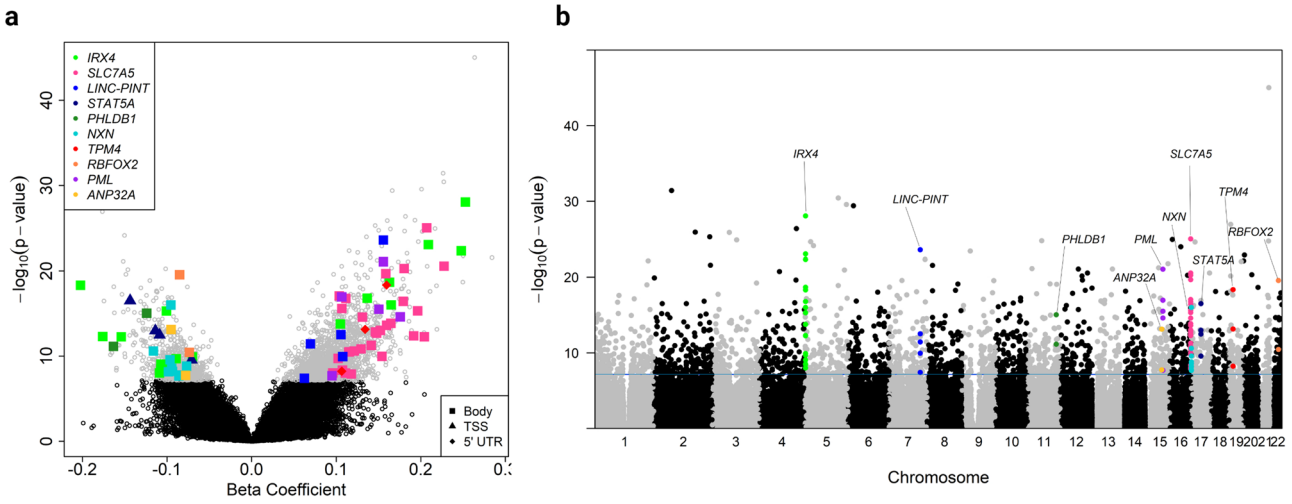

Figure 1. Volcano plot (**a**) of the beta coefficients and −log10(p values) from the epigenome-wide association study (EWAS) for GA (gray = Bonferroni, color = CpGs within the top 10 differentially methylated regions (DMR)), and Manhattan plot (**b**) of the genomic distribution of these results with gene names annotated to those CpGs within top 10 DMRs. Bonferroni significance threshold of $\alpha = 0.05/706323$ (blue horizontal line).

CpG	Coefficient	Standard Error	p-value	P_{Bonf}	Chromosome	Position	Gene	Region
cg06413398	−0.1806	0.0098	3.30E-75	2.33E-69	6	110736865	*DDO*	TSS200
cg10167094	0.1008	0.0056	2.83E-73	2.00E-67	12	53567373	*CSAD*	5'UTR;Body
cg07164639	−0.1772	0.01	2.37E-70	1.67E-64	6	110736958	*DDO*	TSS1500
cg02603733	0.1712	0.0097	5.66E-70	4.00E-64	1	151447362	*1q21.3*	
cg20919287	0.089	0.0051	1.34E-68	9.44E-63	12	53567340	*CSAD*	5'UTR;Body
cg21424090	−0.1978	0.0117	1.00E-63	7.08E-58	11	118550443	*TREH*	TSS200
cg01772842	0.1599	0.0095	1.35E-63	9.54E-58	2	220265394	*2q35*	
cg04453552	−0.2411	0.0145	3.33E-62	2.35E-56	16	57394387	*CCL22*	Body
cg13861278	−0.1426	0.0087	5.15E-61	3.64E-55	1	167868783	*ADCY10*	Body
cg22786472	−0.19	0.0116	2.51E-60	1.77E-54	4	26198975	*RBPJ*	Body

Table 4. The top 10 most significant (α = 0.05/706,323) CpGs associated with post-menstrual age (PMA) in NOVI.

($P_{Bonf} = 1.09 \times 10^{-52}$), *SGCE* ($P_{Bonf} = 3.29 \times 10^{-49}$), *MN1* ($P_{Bonf} = 3.64 \times 10^{-48}$), *DGKZ* ($P_{Bonf} = 6.38 \times 10^{-48}$), *LINC01491* ($P_{Bonf} = 1.83 \times 10^{-47}$), *LRRC17* ($P_{Bonf} = 1.84 \times 10^{-43}$), *IKZF4* ($P_{Bonf} = 1.28 \times 10^{-42}$), *CUX1* ($P_{Bonf} = 2.35 \times 10^{-42}$), and *GGACT* ($P_{Bonf} = 5.93 \times 10^{-42}$) (Table 5). We highlight the DMRs annotated to these genes, along with our overall EWAS results with volcano and Manhattan plots (Fig. 2a,b). Of these top DMRs, *MN1* had the most CpGs (16) within 5 DMRs. Most of the DMRs had only two CpGs within a region however, the annotated DMR with the most CpGs was *C1QTNF1* ($P_{Bonf} = 2.03 \times 10^{-8}$) with 24 CpGs within 1500 bp. All of the significant CpGs within the DMRs of *DDO* and *GGACT* were hypomethylated with increasing PMA, whereas the CpGs within *RBPJ*, *SGCE*, *MN1*, *DGKZ*, *LRRC17*, and *IKZF4* were all hypermethylated with increasing PMA.

Overrepresentation of genes in biological pathways

Enrichment analysis of the genes annotated to GA- and PMA-associated CpGs were used to identify gene ontology (GO) of biological pathways (BP), cellular composition (CC), and molecular functions (MF), as well as KEGG[21–23] and Reactome pathways that were associated with the epigenetic signatures of these age metrics (Supplementary Figures S1-S4). These figures report the pathways that had the most genes within the pathway that were significantly enriched. Interestingly, GO analysis identified two MF pathways ("the regulation of GTPase activity" and "cadherin binding"; Supplementary Figures Sb1 and S2b), two BP pathways ("axonogenesis" and "regulation of neuron projection development"; Supplementary Figures S1c and S2c), and one KEGG pathway ("PI3K-Akt signaling"; Supplementary Figures S3a and S4a) as being significantly over-represented amongst genes differentially methylated with both GA and PMA. The top CC and MF pathways over-represented amongst differentially methylated genes associated with PMA were cell-cell junction ($p = 3.70 \times 10^{-18}$) and actin binding ($p = 4.74 \times 10^{-17}$) (Supplementary Figure S2a & S2b). More specifically, pathway enrichment analysis identified many neurodevelopmental related pathways over-represented amongst genes demonstrating differential methylation associated with both age metrics. For example, neuronal cell body and neuron to neuron synapse were top CC pathways over-represented in genes demonstrating differential methylation with GA (Supplementary Figure S1a) and PMA (Supplementary Figure S2a). Axon guidance ($p = 7.34 \times 10^{-11}$) was also among the most enriched KEGG pathways over-represented in genes demonstrating differential methylation with PMA (Supplementary Figure S4a) and the BP "neuron projection extension" ($p = 4.58 \times 10^{-8}$) was over-represented in genes

Gene	# of CpGs	Estimate	Standard Error	p-value	P_{Bonf}	Chromosome	Position (start)	Position (end)	Region
DDO	2	−1.4664	0.0777	2.08E-79	1.54E-73	6	110736772	110736865	TSS200
2q35	2	1.2173	0.0736	1.96E-61	1.45E-55	2	220265305	220265394	–
11p15.5	3	1.1819	0.0720	1.72E-60	1.27E-54	11	334561	336031	–
RBPJ	2	−1.1250	0.0697	1.47E-58	1.09E-52	4	26198975	26199269	Body
SGCE;PEG10	4	1.0177	0.0651	4.44E-55	3.29E-49	7	94279660	94284274	Body;TSS1500
MN1	3	1.0009	0.0647	4.92E-54	3.64E-48	22	28191222	28191914	Body
DGKZ	2	1.1833	0.0766	8.62E-54	6.38E-48	11	46383051	46383141	TSS200; Body
LINC01491	3	0.6394	0.0416	2.47E-53	1.83E-47	15	48137313	48137817	Body
LRRC17; FBXL13	4	1.0983	0.0744	2.49E-49	1.84E-43	7	102553437	102557016	5'UTR;;1stExon
16p13.3	2	1.1127	0.0754	3.05E-49	2.26E-43	16	3045060	3048362	–
IKZF4	3	2.1913	0.1497	1.72E-48	1.28E-42	12	56414508	56414698	TSS200;5'UTR;1stExon
CUX1	3	0.8773	0.0601	3.17E-48	2.35E-42	7	101740892	101741221	Body
GGACT	2	−1.2237	0.0842	8.00E-48	5.93E-42	13	101197327	101197428	5'UTR

Table 5. The most significant differentially methylated regions (DMR) containing 2 or more CpGs within 1500 base pairs of another significant CpG associated with post-menstrual age (PMA).

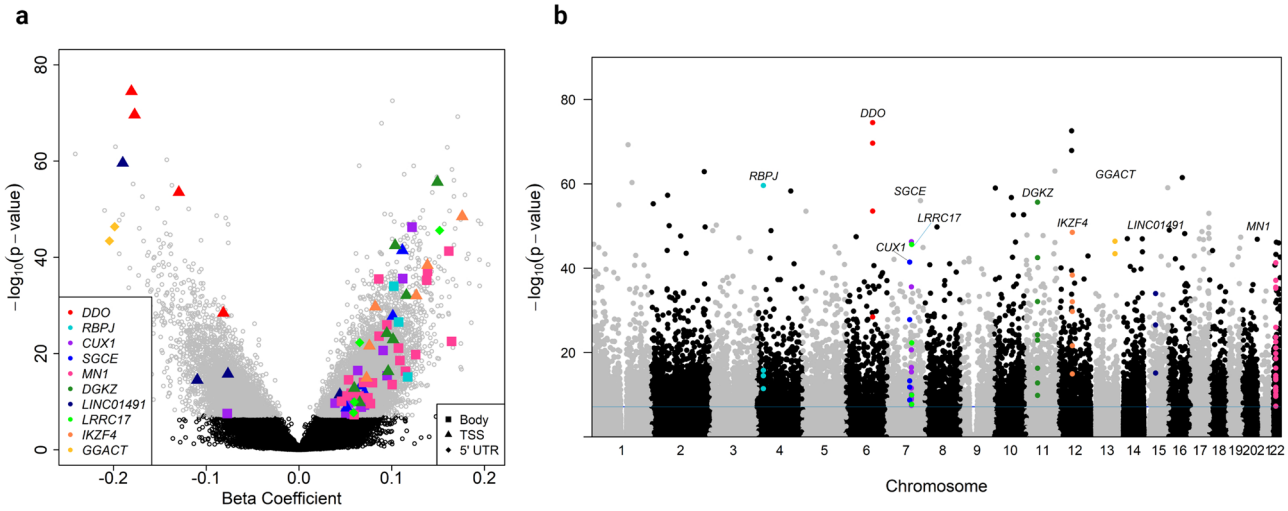

Figure 2. Volcano plot (**a**) of the beta coefficients and −log10(p values) from the epigenome-wide association study (EWAS) for PMA (gray = Bonferroni, color = CpGs within the top 10 differentially methylated regions (DMR)), and Manhattan plot (**b**) of the genomic distribution of these results with gene names annotated to those CpGs within top 10 DMRs. Bonferroni significance threshold of α = 0.05/706323 (blue horizontal line).

demonstrating differential methylation with GA (Supplementary Figure S1c). Pathways associated with growth and development were also significantly over-represented amongst genes with differential methylation associated with neonatal age metrics. Growth factor binding ($p = 2.24 \times 10^{-7}$) was among the top MF pathways over-represented by genes with differential methylation associated with PMA (Supplementary Figure S2a), while cell growth ($p = 4.24 \times 10^{-9}$), the regulation of cell growth ($p = 2.07 \times 10^{-7}$), developmental cell growth ($p = 2.06 \times 10^{-8}$), and the regulation of developmental growth ($p = 8.84 \times 10^{-8}$) were among the top BP pathways enriched with GA associated differential methylation (Supplementary Figure S1c). In addition, several pathways related to both the development and function of muscle tissue and the heart were found to be associated with GA (Supplementary Table S7). Consistent with the GO analysis, the most significant Reactome pathway over-represented amongst genes with GA-associated differential methylation was neuronal systems ($p = 9.63 \times 10^{-4}$) (Supplementary Figure S3b). The most enriched Reactome pathways amongst genes differentially methylated with PMA were RHO GTPase cycle ($p = 7.29 \times 10^{-13}$) and diseases of signal transduction by growth factor receptors and second messengers ($p = 1.63 \times 10^{-7}$) (Supplementary Figure S4b). All results can be found in Supplementary Tables S7-S16.

Lastly, we utilized the gene enrichment R package *methylGSA*[24] that accounts for probe bias (Supplemental Tables S17-S22). Positive regulation of GTPase activity ($p = 1.56 \times 10^{-5}$) remained one of the most significant PMA-associated pathways identified by GO analysis. Immune related pathways, such as natural killer cell differentiation ($p = 2.70 \times 10^{-3}$), immunological synapse ($p = 2.74 \times 10^{-3}$) and T cell activation ($p = 2.87 \times 10^{-3}$) were also enriched pathways. Neurodevelopmental pathways including negative regulation of long-term synaptic potentiation ($p = 6.95 \times 10^{-4}$), regulation of modifications to synaptic structure (1.48×10^{-3}) and axon guidance receptor activity ($p = 2.96 \times 10^{-3}$) were also identified as enriched pathways associated with PMA.

Brain-buccal correlations

We tested whether methylation levels at our GA- and PMA-associated CpGs exhibited positive correlations between brain and buccal tissues in an independent external dataset (GEO Accession: GSE111165). Briefly, this external dataset consisted of paired blood, saliva, buccal, and brain tissue samples from living donors. Samples were drawn during surgery, from patients with medically intractable epilepsy between the ages of 5 and 61. DNAm was assessed with the Infinium HumanMethylationEPIC BeadChip arrays (n = 21)[20]. We observed 85 of the 2,366 CpGs with differential methylation associated with GA also had positive and significant brain-buccal correlations (Supplementary Table S23), and of them, 12 CpGs were within GA-DMRs. The CpG with the strongest positive brain-buccal correlation for GA was cg22757362 which was annotated to *RASGEF1A* ($p = 3.70 \times 10^{-7}$, cor = 0.870) (Table 6). *IRX4* (cg17774559, $p = 5.49 \times 10^{-4}$, cor = 0.702) was also among the top genes with positive brain-buccal correlated CpGs. The top CpGs with the most significant brain-buccal correlation also within a GA-associated DMR were cg23605961 ($p = 2.32 \times 10^{-5}$, cor = 0.793) and cg19087971 (p = 0.001, cor = 0.676) annotated to *PRKAR1B*. We also observed that 802 of the 14,979 CpGs associated with PMA also had positive and significant ($p < 0.05$) brain-buccal correlations (Supplementary Table S24). Of the significant and positively correlated CpGs we observed that 267 of them were also within our identified PMA-DMRs. Overall, the most significant brain-buccal correlation that was within a CpG with the strongest differential methylation with PMA was cg26330304, which was annotated to *ZNF710* (p = 0, cor = 0.838) (Table 6). The most significantly correlated CpG within a DMR associated with PMA was cg11193064 which was annotated to *SMAD6* ($p = 8.94 \times 10^{-7}$, cor = 0.830). Of the highlighted genes with differential methylation associated with PMA in (Fig. 2 and Table 5), *MN1* and *LRRC17* were annotated to CpGs with significant positive brain-buccal correlations. Last, we performed the GO-term and pathway enrichment analyses, restricting the analysis to age-associated CpGs with significant,

CpG	Correlation	p-value	Chromosome	Position	Gene	Regulatory Region
GA						
cg22757362	0.8701	3.70E-07	10	43721445	*RASGEF1A*	5'UTR
cg23605961	0.7935	2.32E-05	7	751331	*PRKAR1B*	5'UTR
cg22158475	0.7416	1.82E-04	1	157096908	*CYCSP52; ETV3*	TSS1500; Body
cg17774559	0.7026	5.49E-04	5	1879698	*IRX4*	Body
cg19087971	0.6766	1.03E-03	7	751233	*PRKAR1B*	5'UTR
cg12227401	0.674	1.09E-03	4	185189393		
cg20544356	0.6468	1.97E-03	17	29890935		
cg18552453	0.6234	3.12E-03	5	151903736		
cg16248376	0.6169	3.53E-03	2	175472517	*WIPF1*	5'UTR
cg03688854	0.6169	3.53E-03	20	59767609		
PMA						
cg26330304	0.8377	0.00E+00	15	90569017	*ZNF710*	5'UTR
cg11193064	0.8299	8.94E-07	15	67008444	*SMAD6*	Body
cg02003272	0.7974	1.88E-05	13	50702719		
cg23605961	0.7935	2.32E-05	7	751331	*PRKAR1B*	5'UTR
cg21371176	0.7896	2.83E-05	1	25407112		
cg17250863	0.7649	8.13E-05	20	33451272	*GGT7*	Body
cg26576041	0.7636	8.53E-05	1	1247929	*CPSF3L*	Body
cg22817719	0.7597	9.84E-05	11	130068300	*ST14*	Body
cg13990107	0.7519	1.29E-04	15	63663946	*CA12*	Body
cg09299774	0.7429	1.74E-04	1	1251017	*CPSF3L*	Body

Table 6. The most significant CpGs associated with gestational age (GA) and post-menstrual age (PMA) were also found to have a positive and significant Brain-Buccal correlation.

positive brain-buccal correlations. Several pathways involving the response to and signaling of growth factor pathways were found to be associated with brain-buccal associated GA CpGs ($q < 0.05$). Additionally, we found axonogenesis, regulation of neuron projection development, and cellular adhesion pathways among the top significant GO pathways ($q < 0.05$) associated with the brain-buccal positive PMA CpGs. Finally, the significant GA and PMA brain-buccal CpGs were then run through the online tool eFORGE[25–27] to determine if any sets of these CpGs are overlapping chromatin accessibility in other tissue types. Positive brain-buccal correlated and GA-associated CpGs were found to be accessible in skin ($p = 6.36 \times 10^{-5}$, $q = 0.01$) and breast tissue ($p = 0.002$, $q = 0.202$) (Supplementary Figure S5). Fetal tissue had the most significant associations with positive brain-buccal correlated and PMA-associated CpGs: fetal lung ($p = 2.13 \times 10^{-8}$, $q = 3 \times 10^{-6}$), fetal kidney ($p = 4 \times 10^{-4}$, $q = 0.039$), fetal stomach ($p = 7.41 \times 10^{-4}$, $q = 0.041$), and fetal muscle leg ($p = 0.002$, $q = 0.085$) (Supplementary Figure S6).

Discussion

Our EWAS study identified age-associated differences in the neonatal buccal tissue epigenome of infants born very preterm. Both age metrics were observed to have widespread associations with differential methylation throughout the genome. GA-associated DNAm may represent how timing of birth leaves a lasting and detectable signature in the epigenome when measured at the end of the NICU stay, while PMA-associated changes in methylation reveals how the methylome varies with aging and development between conception and discharge from the NICU. Increasing GA and PMA were both predominately associated with increasing methylation levels; however, a few of the most significant CpGs for GA were observed to have decreased methylation levels with increasing age at birth.

A number of our findings align with those of a recent study of GA in both term-born and preterm-born infants[28]. Multiple CpGs within *IRX4* (cg18172877, cg07167946, and cg04441405) and one CpG in *ZBP1* (cg11460314), exhibited hypomethylation with increasing GA in both of our independent analyses. Consistent with our findings, Wheater et al[28] also found *IRX4* to be the most significant DMR associated with GA. Our DMR analysis identified *IRX4* and *SLC7A5* as having the strongest association with GA as well as having the most CpGs within their respective DMRs. Thus, DNA methylation patterns at these genes are undergoing dynamic changes in early development. *IRX4* is associated with several pathways related to cardiac development, and prematurity is a known risk factor for long-term issues with cardiac function[29]. *SLC7A5* has previously been observed to have multiple differentially methylated regions associated with neural function in preterm infants[30]. Interestingly, both cardiac- and neurodevelopmental processes were also over-represented in our pathway enrichment analyses.

Importantly, we identified novel neonatal age-associated changes in DNAm. For example, our overall EWAS found that cg13211971, annotated to *PDE9A*, had the strongest association with GA. This CpG and gene was not identified as age-associated by Wikenius et al[5], but this gene is involved in cGMP regulation, is highly expressed in neurons, and is being investigated as a promising target for treating neurodevelopmental[31] and neurodegenerative

disorder[32]. Additionally, of the 26 CpGs within *SLC7A5* highlighted by Sparrow et al., we observed 12 with differential methylation as well as additional novel *SLC7A5* CpG sites associated with GA.

PMA-associated CpGs were annotated to genes associated with neurodevelopment. We found the CpG most strongly associated with PMA to be cg06413398 which is annotated to *DDO*. *DDO*, a known age-associated gene[33,34], becomes increasingly hypomethylated as humans age and appears to be a marker of aging. Interestingly, we also observed an overlap between the GA-associated CpGs from Wheater et al[28] and our PMA-associated CpGs where cg04441405, cg17774559 within *IRX4*, and cg11460314 within *ZPB1* had significant opposite associations between our studies. Lastly, we observed two of the GA-associated genes identified by Wheater et al. also being associated with PMA in our study: *HEATR2* and *SMIM2*; however differential methylation was observed at different CpGs between these two studies. Overall, increasing DNA methylation was associated with increasing PMA at the majority of CpGs and this observation aligns with the findings of Wikenius et al[5] who described changes in DNAm associated with PMA between 6 and 52 postnatal weeks. Of the 42 genes identified by Wikenius et al[5] as undergoing changes in DNA methylation in early development, we observed that 29 were associated with PMA in our study. In both studies, *IKZF4* and *MN1* were among the genes that were most hyper-methylated with increasing age, however, we observed a larger number of differentially methylated CpGs within these genes in our DMR analysis. The novel findings that we observe could be due to differences in study population as we focus only on very preterm infants and thus, our range of gestational age at birth differs substantially from that of Wheater et al.[28], Sparrow et al.[30], and Wikenius et al.[5]. Additionally, our study used the EPIC array which provides more epigenomic coverage and our larger sample size increases the power of our study. Despite these differences we captured genes and CpGs that were first identified in these prior studies, and discovered novel neonatal age-associated genes involved in cardiac development and neurodevelopment which is consistent with the developmental processes that are occurring during this critical period. Biological processes identified in our pathway enrichment analysis are fundamental to the development of neonates, specifically those born prematurely.

We also explored whether our neonatal age associated CpGs were captured by the PedBE Clock, which can be used to predict age in epithelial cells sampled from persons age 0–20 years[10]. We found that CpGs within *BMP4*, *UBA7*, and *AMBRA1* were captured in our data as well as the PedBE Clock. *BMP4* encodes a ligand of TGF-beta and has been found to also regulate heart development and adipogenesis. *AMBRA1* has been associated with neural tube defects and is a biomarker of multiple system atrophy[35]. The age-associated methylation changes in these genes may represent developmental processes that are dynamic during the neonatal period but remain dynamic through the pediatric years. We also screened whether our strongest neonatal age-associated CpGs are associated with aging throughout the life course. *DDO* was observed to be differentially hypomethylated as GA increases. Two of our top CpGs within *DDO*, cg02872426 and cg06413398, have been associated with aging in adult populations[33]. Additionally, Cameron et al. (2023), examined whether there were differences in methylation when comparing infants with very low birth weight (VLBW), a condition that overlaps highly with preterm birth, to healthy term birth infants, as well as their methylation status as adults[36]. Interestingly, 12 of the 17 top genes associated with VLBW birth were also found to be significantly associated with GA or PMA in our study. In adults born with VLBW, cg24263062 annotated to *EBF4* was also differentially methylated in our study and associated with PMA (coef = 0.046, $p = 3.81 \times 10^{-9}$). These findings suggest that age at birth has a lasting impact on the methylome throughout the life course.

We acknowledge that there were some limitations to this study. We utilized non-invasive cheek swabs to collect newborns' buccal tissue to study age associated perturbations in DNAm. As such, the epigenetic changes in this tissue may not be reflective of what is occurring in different developing organ systems. Despite this limitation, we identified pathways and CpGs involved in neurodevelopment, heart development, as well as cellular processes that were also found in studies that utilize blood samples. Another potential limitation is that we are studying DNAm at NICU discharge in infants born preterm which may be capturing the changes that are occurring as these children become healthier, as well as epigenetic changes influenced by biological and environmental experiences in the NICU. We attempted to control for this by adjusting for number of neonatal morbidities (BPD, ROP, SBI, and infection) which are associated with longer NICU stays. We recognize that other unmeasured exposures and experiences in the NICU may have influenced our observed associations. Importantly however, many of our findings align with prior evidence which has shown that preterm birth and gestational age are associated with DNA methylation difference and has a lasting effect on health throughout the life course[5,36,37]. Additionally, our study provided a unique opportunity to study changes in DNAm solely among very preterm infants, rather than comparing preterm to term born infants. Another strength of our study is our large sample size compared to prior studies that investigated similar research questions, and thus we were well powered to address the association between neonatal age and DNAm in infants born very prematurely. Additionally, the NOVI study is an ongoing cohort study with longitudinal follow-up up to seven years. Thus, our future research will allow us to the opportunity to study how methylation of these genes and CpGs change throughout childhood.

In conclusion, we identified CpGs and associated genes that undergo dynamic changes in methylation during early development including previously identified developmental and age associated genes like *DDO*, *IRX4* and *SLC7A5*. We found that the majority of the age-associated differentially methylated CpGs were hyper-methylated with increasing GA and PMA which is consistent with other studies that show increased methylation with aging. Among infants born very preterm, the genes with differential methylation associated with PMA and GA were involved in neurodevelopment, growth processes, cardiac development, and cellular function. Additionally, a small subset of these CpGs also exhibit significant correlations in methylation levels between brain and buccal samples, and these were enriched for pathways associated with axonogenesis, response to growth factors, and cell development.

Methods

Study population

The NOVI Study was conducted at 9 university affiliated NICUs in Providence, RI, Grand Rapids, MI, Kansas City, MO, Honolulu, HI, Winston Salem, NC, and Torrance and Long Beach, CA from April 2014 through June 2016. These NICUs were also Vermont Oxford Network (VON) participants. Eligibility was determined based on the following inclusion criteria: (1) birth at < 30 weeks post menstrual age; (2) parental ability to read and speak English or Spanish and (3) residence within 3 h of the NICU and follow-up clinic. Exclusion criteria included maternal age < 18 years, maternal cognitive impairment, maternal death, infants with major congenital anomalies, including central nervous system, cardiovascular, gastrointestinal, genitourinary, chromosomal, and nonspecific anomalies, and NICU death. Parents of eligible infants were invited to participate in the study when survival to discharge was determined to be likely by the attending neonatologist. Written informed consent was obtained from all participants and all study procedures and protocols were carried out with approval from each of the following center's Institutional Review Board *[Children's Mercy Hospital IRB in Kansas City, MO (IRB00004750), Western Institutional Review Board in Puyallup, WA (WIRB20131387), John F. Wolf Human Subjects Committee in Los Angeles, CA (IRB00000389), Spectrum Health Systems, Inc. in Grand Rapids, MI (IRB00009435), Women & Infants Hospital in Providence, RI (IRB00000746), and Wake Forest University Health Sciences in Winston-Salem, NC (IRB00000212)].* All study activities were performed in accordance with relevant guidelines and regulations. Overall, 704 eligible infants were enrolled and written informed consent was obtained for 93% of infants for epigenomic screening via buccal cells swabs that were collected on 624 infants. Ultimately, 538 samples were included in the study from infants for whom DNAm data, PMA, and GA, and covariate data were available.

Maternal interviews were performed to collect demographic information such as age, race and ethnicity, and educational attainment, while the Hollingshead Index was used to assess socioeconomic status (SES; collected by maternal report on education and occupation) with a Hollingshead level V indicating low SES. Infant medical records were reviewed to collect birthweight, gestational age, length of NICU stay, whether the newborn was outborn, and diagnoses of neonatal morbidities described in detail below. Outborn refers to infants that were born in a hospital without subspecialty providers of neonatal intensive care and were transferred, almost always on the day of birth, to a tertiary center for subspecialty care. Gestational age was estimated using the highest quality of available information: first using the dates of embryo retrieval or intrauterine insemination, then using fetal ultrasound, then using date of last menstrual period, and finally assigned by attending neonatologist in the absence of the above information.

Age variables

In this study we investigated PMA, an estimate of the period between conception to buccal tissue collection at NICU discharge, and GA, an estimate of the period between conception and birth. PMA in NOVI was calculated by adding postnatal age at buccal collection to the estimated GA at birth which is an established process[38,39]. GA was estimated by using the date of embryo retrieval, intrauterine insemination, fetal ultrasound, or the date of last menstrual period as described elsewhere by Everson et al[40]. Buccal cells were collected for epigenomic analyses during the week of discharge from the NICU (± 3 days); thus, PMA at buccal swab collection represents the combination of GA at birth plus the length of NICU stay (± 3 days).

Description of confounding variables

As described in Everson et al[40], we adjusted for potential confounders, batch (EPIC array) effects and cellular heterogeneity[41] and we also adjusted for other potential confounders, such as outborn status[42] given that this variable increases the risk of neonatal morbidities and neurobehavioral deficits and is associated with longer stays in the NICU in males[40,43,44]. We adjusted for sex[45] given its association with DNAm. Cell heterogeneity, the proportion of epithelial, fibroblast, and immune cells, were estimated in our buccal samples using reference methylomes[41]. As in our previous studies, we observe that epithelial cells make up 95.7% of cells in 95% of our buccal samples, with the remaining majority are immune cells[40]. Therefore, we adjusted for cell heterogeneity by using the proportions of epithelial cells as a covariate in our statistical models. Lastly, we adjusted for potential unmeasured confounding such as batch effect and unwanted sources of variation with surrogate variables[46]. We included both PMA and GA in our models (as independent variables), so we could characterize the relationships of both age metrics with DNAm, independent of each other. In all models, we adjusted for sex, whether the newborn was outborn, the proportion of estimated epithelial cells, recruitment site (6-level factor), batch (7-level), cumulative morbidity risk score (4-level factor) ranging from 0 to 3 + serious morbidities (BPD, ROP, SBI, and infection) experienced during NICU stay, and five SVs.

DNA methylation (DNAm) measurement, quality control, and preprocessing

To assess the methylation profiles of the NOVI study cohort we analyzed the DNAm data from buccal tissue at NICU discharge measured using the Illumina MethylationEPIC. These data have undergone established quality control and processing pipelines, including the exclusion of samples with > 5% of probes yielding detection p-values > 1.0E-5, with mismatch between reported and predicted sex, or incomplete phenotype data[40]. Functional normalization and beta-mixture quantile (BMIQ) normalization were performed[47], and then we excluded probes located on the X and Y chromosomes, those with single nucleotide polymorphisms (SNP) within the binding region, those able to cross-hybridize to other regions of the genome[48], or those with low variability[49]. After these exclusions, and dropping four samples that had missing covariate data, 706,278 probes were available from 538 samples. Logit transformation of beta values were calculated to get M-values which were used in all downstream analysis[50,51].

Statistical analysis

All statistical analyses were performed in R version 4.1.0. First, we conducted SVA to identify unknown confounding within the NOVI study using the *sva*[52] package. The model matrix included GA, PMA and potential confounders, while the null model excluded GA and PMA. To identify differentially methylated CpGs associated with GA and PMA, we conducted linear regression models using the *gee*[53] package, regressing DNAm levels on GA and PMA (mutually adjusted for), while also adjusting for potential confounders and SVs. Generalized Estimating Equations (GEE) are a type of regression model that accounts for potential non-independence of observations (i.e., multiple births nested within families) as these siblings may have similar methylation patterns and other characteristics. We thus included family ID as a nesting variable in all GEE models. We accounted for multiple comparisons by using Bonferroni corrections ($\alpha = 0.05/706{,}323$). Among CpGs significantly associated with GA or PMA, we then tested whether these age-associated CpGs were in close genomic proximity and formed differentially methylated regions (DMRs) using the dmr*ff*[54] package. We defined DMRs as sets of CpG sites within 1500 bp apart (*maxgap*) with each site having a Bonferroni adjusted p-value < 0.05 (*p.cutoff*) and the same direction of effect[54]. Finally, we ran pathway enrichment analysis on significant differentially methylated genes using the R package *clusterProfiler* which identified pathways defined by GO, KEGG, and Reactome databases[21–23,55,56]. The R package *methylGSA* was then used for pathway enrichment analysis to account for probe bias[24]. Lastly, these tests were adjusted for multiple testing correction using the Benjamini–Hochberg method as is common practice in gene-set enrichment analyses.

Brain-buccal correlation

Much of our work in this cohort has focused on newborn neurobehavior, medical morbidities, and subsequent neurodevelopment associated with epigenetic features. These relationships have been studied using non-invasive collection of neonatal peripheral tissue via cheek swabs to determine whether any of the age-associated changes in buccal cells may be reflective of age-associated changes in the central nervous system (CNS). Since we are studying a peripheral tissue, but much of our work in this cohort focuses on neurodevelopment and neurobehavior, we aimed to understand if any of the age-associated changes in buccal cells may be reflective of age-associated changes in the CNS. Thus, we conducted Pearson correlation tests among the CpGs significantly associated with GA and PMA using a publicly available dataset consisting of paired brain and buccal tissue samples (GEO Accession: GSE111165)[20]. Pathway enrichment analysis, as mentioned above, was conducted on the CpGs with significant brain-buccal correlations for each age metric. Statistical significance was based on CpGs with positive correlations that met the Bonferroni cutoff of ($\alpha = 0.05/706{,}323$). Significant results were then uploaded and analyzed by eFORGE (experimentally-derived Functional element Overlap analysis of ReGions from EWAS), an online tool that identifies cell type specific signaling in epigenomic data[25–27], using reference data from the Consolidated Roadmap Epigenetics- DNase I hypersensitive sites (DHSs) with a strict significant threshold set at $p < 0.01$ and marginal significant threshold set at $p < 0.05$.

Consent to participate

Written informed consent was obtained from all participants and all study procedures and protocols were carried out with approval from the following center's Institutional Review Board *[Children's Mercy Hospital IRB in Kansas City, MO (IRB00004750), Western Institutional Review Board in Puyallup, WA (WIRB20131387), John F. Wolf Human Subjects Committee in Los Angeles, CA (IRB00000389), Spectrum Health Systems, Inc. in Grand Rapids, MI (IRB00009435), Women & Infants Hospital in Providence, RI (IRB00000746), and Wake Forest University Health Sciences in Winston-Salem, NC (IRB00000212)].* All study activities were performed in accordance with relevant guidelines and regulations.

Data availability

The raw and processed DNAm data for NOVI are publicly accessible through NCBI Gene Expression Omnibus (GEO) via accession series GSE128821.

Received: 8 March 2024; Accepted: 19 July 2024
Published online: 05 August 2024

References

1. Stiles, J. & Jernigan, T. L. The basics of brain development. *Neuropsychol. Rev.* **20**, 327–348. https://doi.org/10.1007/s11065-010-9148-4 (2010).
2. Gilmore, J. H., Knickmeyer, R. C. & Gao, W. Imaging structural and functional brain development in early childhood. *Nat. Rev. Neurosci.* **19**, 123–137. https://doi.org/10.1038/nrn.2018.1 (2018).
3. Smith, L. J., McKay, K. O., van Asperen, P. P., Selvadurai, H. & Fitzgerald, D. A. Normal development of the lung and premature birth. *Paediatr. Respir. Rev.* **11**, 135–142. https://doi.org/10.1016/j.prrv.2009.12.006 (2010).
4. Hernandez, D. G. *et al.* Distinct DNA methylation changes highly correlated with chronological age in the human brain. *Hum. Mol. Genet.* **20**, 1164–1172 (2011).
5. Wikenius, E., Moe, V., Smith, L., Heiervang, E. R. & Berglund, A. DNA methylation changes in infants between 6 and 52 weeks. *Sci. Rep.* **9**, 1–12 (2019).
6. Cao-Lei, L., Laplante, D. P. & King, S. Prenatal maternal stress and epigenetics: Review of the human research. *Curr. Mol. Biol. Rep.* **2**, 16–25. https://doi.org/10.1007/s40610-016-0030-x (2016).
7. Parets, S. E. *et al.* Fetal DNA methylation associates with early spontaneous preterm birth and gestational age. *PLoS ONE* **8**, e67489. https://doi.org/10.1371/journal.pone.0067489 (2013).
8. Provençal, N. *et al.* Glucocorticoid exposure during hippocampal neurogenesis primes future stress response by inducing changes in DNA methylation. *Proc. Natl. Acad. Sci.* **117**, 23280. https://doi.org/10.1073/pnas.1820842116 (2020).

9. Breton, C. V. *et al.* Small-magnitude effect sizes in epigenetic end points are important in children's environmental health studies: The children's environmental health and disease prevention research center's epigenetics working group. *Environ. Health Perspect.* **125**, 511–526. https://doi.org/10.1289/EHP595 (2017).
10. McEwen Lisa, M. *et al.* The PedBE clock accurately estimates DNA methylation age in pediatric buccal cells. *Proc. Natl. Acad. Sci.* **117**, 23329–23335. https://doi.org/10.1073/pnas.1820843116 (2020).
11. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, 1–20 (2013).
12. Graw, S. *et al.* NEOage clocks-epigenetic clocks to estimate post-menstrual and postnatal age in preterm infants. *Aging* **13**, 23527 (2021).
13. Liu, L. *et al.* Global, regional, and national causes of under-5 mortality in 2000–15: An updated systematic analysis with implications for the sustainable development goals. *Lancet* **388**, 3027–3035. https://doi.org/10.1016/S0140-6736(16)31593-8 (2016).
14. UNICEF, W., World Bank, UN-DESA population division. *Levels and trends in child mortality: report 2017*, <http://www.who.int/maternal_child_adolescent/documents/levels_trends_child_mortality_2017/en/> (2018).
15. Kramer, B. W., Niklas, V. & Abman, S. Bronchopulmonary dysplasia and impaired neurodevelopment—What may be the missing link?. *Am. J. Perinatol.* **39**, S14–S17. https://doi.org/10.1055/s-0042-1756677 (2022).
16. Majnemer, A. *et al.* Severe bronchopulmonary dysplasia increases risk for later neurological and motor sequelae in preterm survivors. *Dev. Med. Child Neurol.* **42**, 53–60. https://doi.org/10.1017/s001216220000013x (2000).
17. Piyasena, C. *et al.* Dynamic changes in DNA methylation occur during the first year of life in preterm infants. *Front. Endocrinol.* **7**, 158 (2016).
18. Schuster, J. *et al.* Effect of prematurity on genome wide methylation in the placenta. *BMC Med. Genet.* **20**, 116. https://doi.org/10.1186/s12881-019-0835-6 (2019).
19. Wang, X. M. *et al.* Comparison of DNA methylation profiles associated with spontaneous preterm birth in placenta and cord blood. *BMC Med. Genom.* **12**, 1. https://doi.org/10.1186/s12920-018-0466-3 (2019).
20. Braun, P. R. *et al.* Genome-wide DNA methylation comparison between live human brain and peripheral tissues within individuals. *Transl. Psychiatr.* **9**, 47. https://doi.org/10.1038/s41398-019-0376-y (2019).
21. Kanehisa, M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci.* **28**, 1947–1951. https://doi.org/10.1002/pro.3715 (2019).
22. Kanehisa, M., Furumichi, M., Sato, Y., Kawashima, M. & Ishiguro-Watanabe, M. KEGG for taxonomy-based analysis of pathways and genomes. *Nucleic Acids Res.* **51**, D587–D592. https://doi.org/10.1093/nar/gkac963 (2023).
23. Kanehisa, M. & Goto, S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30. https://doi.org/10.1093/nar/28.1.27 (2000).
24. Ren, X. & Kuan, P. F. methylGSA: A bioconductor package and Shiny app for DNA methylation data length bias adjustment in gene set testing. *Bioinformatics* **35**, 1958–1959. https://doi.org/10.1093/bioinformatics/bty892 (2019).
25. Breeze, C. E. Cell type-specific signal analysis in EWAS. *bioRxiv* https://doi.org/10.1101/2021.05.21.445209 (2021).
26. Breeze, C. E. *et al.* eFORGE: A tool for identifying cell type-specific signal in epigenomic data. *Cell Rep.* **17**, 2137–2150. https://doi.org/10.1016/j.celrep.2016.10.059 (2016).
27. Breeze, C. E. *et al.* eFORGE v2.0: Updated analysis of cell type-specific signal in epigenomic data. *Bioinformatics* **35**, 4767–4769. https://doi.org/10.1093/bioinformatics/btz456 (2019).
28. Wheater, E. N. W. *et al.* DNA methylation and brain dysmaturation in preterm infants. *medRxiv* https://doi.org/10.1101/2021.04.08.21255064 (2021).
29. Greer, C., Troughton, R. W., Adamson, P. D. & Harris, S. L. Preterm birth and cardiac function in adulthood. *Heart* **108**, 172–177. https://doi.org/10.1136/heartjnl-2020-318241 (2022).
30. Sparrow, S. *et al.* Epigenomic profiling of preterm infants reveals DNA methylation differences at sites associated with neural function. *Transl. Psychiatr.* **6**, e716. https://doi.org/10.1038/tp.2015.210 (2016).
31. Scarborough, J. *et al.* Symptomatic and preventive effects of the novel phosphodiesterase-9 inhibitor BI 409306 in an immune-mediated model of neurodevelopmental disorders. *Neuropsychopharmacology* **46**, 1526–1534. https://doi.org/10.1038/s41386-021-01016-3 (2021).
32. Harms, J. F., Menniti, F. S. & Schmidt, C. J. Phosphodiesterase 9A in brain regulates cGMP signaling independent of nitric-oxide. *Front. Neurosci.* **13**, 837. https://doi.org/10.3389/fnins.2019.00837 (2019).
33. Benton, M. C. *et al.* Methylome-wide association study of whole blood DNA in the Norfolk Island isolate identifies robust loci associated with age. *Aging* **9**, 753–768. https://doi.org/10.18632/aging.101187 (2017).
34. Shi, L. *et al.* DNA methylation markers in combination with skeletal and dental ages to improve age estimation in children. *Forensic Sci. Int. Genet.* **33**, 1–9. https://doi.org/10.1016/j.fsigen.2017.11.005 (2018).
35. Safran, M. *et al. Practical Guide to Life Science Databases* (Springer, 2021).
36. Cameron, V. A. *et al.* DNA methylation patterns at birth predict health outcomes in young adults born very low birthweight. *Clin. Epigenet.* **15**, 47. https://doi.org/10.1186/s13148-023-01463-3 (2023).
37. Parets, S. E., Bedient, C. E., Menon, R. & Smith, A. K. Preterm birth and its long-term effects: Methylation to mechanisms. *Biology* **3**, 498–513. https://doi.org/10.3390/biology3030498 (2014).
38. O'Shea, T. M. *et al.* The ELGAN study of the brain and related disorders in extremely low gestational age newborns. *Early Hum. Dev.* **85**, 719–725. https://doi.org/10.1016/j.earlhumdev.2009.08.060 (2009).
39. McElrath, T. F. *et al.* Pregnancy disorders that lead to delivery before the 28th week of gestation: An epidemiologic approach to classification. *Am. J. Epidemiol.* **168**, 980–989. https://doi.org/10.1093/aje/kwn202 (2008).
40. Everson, T. M. *et al.* Serious neonatal morbidities are associated with differences in DNA methylation among very preterm infants. *Clin. Epigenet.* **12**, 151. https://doi.org/10.1186/s13148-020-00942-1 (2020).
41. 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. https://doi.org/10.2217/epi-2018-0037 (2018).
42. Natarajan, G. *et al.* Effect of inborn vs. outborn delivery on neurodevelopmental outcomes in infants with hypoxic-ischemic encephalopathy: Secondary analyses of the NICHD whole-body cooling trial. *Pediatr. Res.* **72**, 414–419. https://doi.org/10.1038/pr.2012.103 (2012).
43. Ambalavanan, N. *et al.* Predictors of death or bronchopulmonary dysplasia in preterm infants with respiratory failure. *J. Perinatol.* **28**, 420–426. https://doi.org/10.1038/jp.2008.18 (2008).
44. Redpath, S. *et al.* Do transport factors increase the risk of severe brain injury in outborn infants <33 weeks gestational age?. *J. Perinatol.* **40**, 385–393. https://doi.org/10.1038/s41372-019-0447-1 (2020).
45. Singmann, P. *et al.* Characterization of whole-genome autosomal differences of DNA methylation between men and women. *Epigenet. Chromatin* **8**, 43. https://doi.org/10.1186/s13072-015-0035-3 (2015).
46. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, 1724–1735. https://doi.org/10.1371/journal.pgen.0030161 (2007).
47. Teschendorff, A. E. *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* **29**, 189–196. https://doi.org/10.1093/bioinformatics/bts680 (2013).
48. Pidsley, R. *et al.* Critical evaluation of the illumina methylationEPIC beadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* **17**, 208. https://doi.org/10.1186/s13059-016-1066-1 (2016).

49. Logue, M. W. *et al.* The correlation of methylation levels measured using Illumina 450K and EPIC BeadChips in blood samples. *Epigenomics* **9**, 1363–1371. https://doi.org/10.2217/epi-2017-0078 (2017).
50. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinform.* **11**, 587. https://doi.org/10.1186/1471-2105-11-587 (2010).
51. Xie, C. *et al.* Differential methylation values in differential methylation analysis. *Bioinformatics* **35**, 1094–1097. https://doi.org/10.1093/bioinformatics/bty778 (2019).
52. 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. https://doi.org/10.1093/bioinformatics/bts034 (2012).
53. Carey V.J., Lumley T, & Ripley B.D. gee: Generalized Estimation Equation Solver, http://CRAN.R-project.org/package=gee, R package version 4.13–18 (2012),
54. Suderman M, Staley J.R., French R, Arathimos R, Simpkin A, Tilling K. Dmrff: Identifying differentially methylated regions efficiently with power and control. BioRxiv. Dec 31:508556. (2018)
55. Wu, T. *et al.* ClusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141. https://doi.org/10.1016/j.xinn.2021.100141 (2021).
56. Yun, G., Wang, L. G., Han, Y. & Hen, Q. Y. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS J. Integr. Biol.* **16**, 284–287. https://doi.org/10.1089/omi.2011.0118 (2012).

Disclaimer
The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions
TME, CJM, and BML initiated, acquired the funding for, and, along with KMH, designed this investigation. KMH performed the statistical analyses and KMH, TME, AH, KNC and MC interpreted the results. AAB, BSC, JC, JH, JAH, ECM, CRN, SLP, LMS, SAD, LMD and TMO coordinated data collection. KMH and TME drafted the manuscript. All authors contributed to interpretation of the results and revisions to the manuscript. All authors read and approved the final manuscript.

Funding
The Neonatal Neurobehavior and Outcomes in Very Preterm Infants (NOVI) cohort was also supported by R01HD072267, R01HD084515, and P30 ES019776. Dr. Camerota was supported by a career development award (K01MH129510).

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-024-68071-w.

Correspondence and requests for materials should be addressed to T.M.E.