

RESEARCH

Open Access



Assessing the effect of childbearing on blood DNA methylation through comparison of parous and nulliparous females

Su Chen^{1*}, Miranda Johs¹, Wilfried Karmaus², John W. Holloway³, Parnian Kheirkhah Rahimabad², Jaclyn M. Goodrich⁴, Karen E. Peterson^{4,5}, Dana C. Dolinoy^{4,5}, S. Hasan Arshad^{6,7} and Susan Ewart⁸

Abstract

Background Pregnancy and childbirth have been connected to modified risk of a wide variety of conditions in later life, including neurodegenerative disorders and cancers. The presence, extent, and direction of the effect that childbearing status has on decreasing or increasing the risk of these conditions differs depending on the disease. The mechanisms by which pregnancy and childbirth modify the risk of diseases are still unknown. DNA methylation (DNAm) alterations that occur during pregnancy and persist after childbirth may help us understand this phenomenon.

Results Blood DNAm was available from 89 women (28 parous; 61 nulliparous) at ages 18 and 26 years in the Isle of Wight birth cohort; no significant differences in the population characteristics were present between the analyzed population and the full cohort. We performed an epigenome-wide association study on 389,355 CpGs and identified 184 CpGs to be significantly differentially methylated between parous and nulliparous women after adjusting for confounders and multiple testing. Of these CpGs, 105 had regression coefficients in the same direction in an independent Mexico City based ELEMENT cohort, of which 13 were significant (replication $P < 0.05$). These 13 CpGs were associated with 16 unique genes. DNAm levels tracked with gene expression in 3 of the replicated genes, one of which (*TM2D3*) was differentially expressed in parous vs nulliparous women. Gene disease association analysis identified a network of parous-associated diseases.

Conclusions Our results suggest that pregnancy and childbirth lead to DNAm changes in parous women and these changes persist at least 6 months and up to 8 years postpartum. Parous-related CpG sites may play a role in how childbearing status modifies risk of later life diseases in women. Further studies are needed to explore the linkage and mechanism.

Keywords Pregnancy, Parous, Nulliparous, Childbearing, DNA methylation, CpG

*Correspondence:

Su Chen

suchen@unmc.edu

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Background

Women's reproductive history may be linked to future health outcomes, as bearing children provides protection from some conditions and increases the risk of others. For example, childbearing can decrease the risk of subsequent hormone receptor positive breast, ovarian, and endometrial cancers in parous women (those who have given birth) as compared to nulliparous women (those who have not given birth) [11, 24, 28, 38, 39, 49]. In contrast, even when pregnancy is healthy with no gestational complications, parous women have been shown to have increased risk of subsequent Alzheimer's disease, stroke, and myocardial infarction as compared to their nulliparous counterparts, with parity (the number of pregnancies) also influencing cardiovascular disease risk [15, 59].

Pregnancy has also been linked to changes in DNA methylation (DNAm) – the addition or removal of methyl groups to cytosine-phosphate-guanine (CpG) sites on DNA, frequently in promoter regions – which regulate gene expression. DNAm changes throughout pregnancy are considered as an important part of gene regulation and normal cellular control mechanisms [14, 16]. Fradin et al. investigated DNAm changes in maternal blood of a longitudinal cohort of pregnant women from early pregnancy (up to 18 weeks) to late pregnancy (35 weeks and later) and found a gain of methylation in genes involved in morphogenesis, such as *ezh1*, and a loss of methylation in genes promoting maternal-infant bonding [22]. Gruziova et al. identified 196 CpGs displaying significant longitudinal DNAm change from pre-pregnancy (0.86 to 3.35 years before conception), through early (weeks 10 to 14) and late (weeks 26 to 28) pregnancy, to post-pregnancy (2 to 4 days after delivery); 17 of these CpGs were significant in an independent replication cohort. Many of these CpGs are involved in metabolism-related pathways, with some specific to physiological changes in pregnancy, such as mammary gland development [25]. Other studies have focused on the association of adverse pregnancy conditions, such as gestational diabetes [18, 21] and preeclampsia [3, 26, 53] on DNAm. Finally, patterns of DNAm from blood samples have been associated with various diseases including breast cancer [7, 9, 27, 41, 57, 58] and degenerative neuropathies [1, 56]. Furthermore, neurodegenerative diseases are increasingly being recognized as related to reproductive history [6, 17, 35, 52].

Thus, the elements of the evidence connecting childbearing status, DNAm, and subsequent health outcomes for women may be summarized as: 1) bearing children alters the risk for certain diseases later in life, 2) childbearing changes DNAm, and 3) patterns of DNAm are associated with certain diseases. These connections are predicated on the idea that pregnancy leaves a “footprint” such that pregnancy-associated DNAm changes

are maintained post pregnancy and subsequently impact health outcomes. Indeed, a growing number of studies have examined if the DNAm changes that occur during pregnancy remain from 2–4 days [25], to an average at 10 months [46], to a median at 16.7 years [8] after childbirth. However, knowledge gaps remain as Campagna et al. did not examine DNAm before or during pregnancy. Furthermore, Lin et al. and Gruziova et al. made contradicting conclusions with Lin et al. reporting that DNAm after delivery was overall higher than that during pregnancy in a pilot study of 10 women, while Gruziova et al. showed a decrease in average methylation levels in 91% of identified CpGs over the studied period (before, during and 2–4 days after pregnancy).

There are two major limitations to the current knowledge of childbearing related DNAm changes. First, studies of parous women have not examined nulliparous controls. Second, only a single pilot study [46] has examined if DNAm changes that occur during pregnancy remain beyond the immediate post-partum period; it is important to provide further evidence as to whether these changes are stable and maintained over time or soon return to their prepartum status. This study aimed to fill these two gaps by identifying differential DNAm CpGs between parous and nulliparous women and by evaluating DNAm from pre-pregnancy samples and samples collected at least 6 months after delivery. Specifically, we used data and samples from the Isle of Wight (IOW) birth cohort to conduct an epigenome-wide association study to find differential methylation at two time points in young adulthood between parous and nulliparous women. Many DNA regions experience changes in methylation levels during periods of physiologic changes, such as puberty and pregnancy; in order to focus only on DNAm changes attributable to pregnancy, the methylation comparisons were made between parous and nulliparous women at ages 18 and 26 years, timepoints that were pre- and post-pregnancy for the parous subjects. We assessed replication in the Early Life Exposure in Mexico to Environmental Toxicants (ELEMENT) cohort. Further validation of our results was pursued by assessing gene expression, GOMeth gene ontology, KEGG pathway gene ontology, and differentially methylated regions in the IOW cohort.

Results

Participant characteristics

Characteristics of the IOW study population are provided in Table 1 (left panel). Of the 750 female study subjects, 144 had DNAm samples at both ages 18 and 26 years. Among these 144 women, 89 had information on their childbearing history (“yes” vs “no”) validated from their medical records. Figure 1 depicts how

Table 1 Descriptive characteristics of study populations and analyzed samples (IOW and ELEMENT)

Discovery Cohort: IOW				Replication Cohort: ELEMENT			
Characteristic	Study Population (N = 750)	Analyzed Samples (n = 89)	Test Stat (p-value)	Characteristic	Study Population (N = 275)	Analyzed Samples (n = 54)	Test Stat (p-value)
Age at T1: Mean (SD)	18 (0)	18 (0)	-	Age at T1: Mean (SD)	14.5 (2.2)	16.2 (1.3)	5.61 (<.0001)
Age at T2: Mean (SD)	26 (0)	26 (0)	-	Age at T2: Mean (SD)	16.4 (2.2)	19.2 (1.3)	-9.00 (<.0001)
BMI change (T1 to T2): Mean (SD)	1.58 (16.34)	2.47 (3.49)	0.78 (.4350)	BMI change (T1 to T2): Mean (SD)	1.40 (1.68)	1.15 (3.81)	1.22 (0.2250)
Smoking Cluster at T1: N (%)			4.62 (.2020)	Smoking Cluster at T1: N (%)			20.75 (.0001)
1 – non active and non-passive smokers at T1	185 (32.92%)	21 (23.60%)		1 – non active and non-passive smokers at T1	195 (71.69%)	41 (85.42%)	
2 – passive and non-active smokers at T1	238 (42.35%)	48 (53.93%)		2 – passive and non-active smokers at T1	59 (21.45%)	0	
3 – active and non-passive smokers at T1	7 (1.12%)	1 (1.12%)		3 – active and non-passive smokers at T1	7 (2.55%)	0	
4 – active and passive smokers at T1	132 (21.35%)	19 (21.35%)		4 – active and passive smokers at T1	11 (4.04%)	7 (14.58%)	
SES Status at T1: N (%)			4.25 (.1194)	SES Status at T1: N (%)			6.06 (.1085)
1 – high income, median housing, low education	161 (29.70%)	30 (34.88%)		1 – AMAI Score A/B	14 (5.09%)	0 (0.00%)	
2 – median income, high housing, high education	215 (39.67%)	39 (45.35%)		2 – AMAI Score C	126 (45.82%)	33 (61.11%)	
3 – low income, low housing, median education	166 (30.63%)	17 (19.77%)		3 – AMAI Score D	102 (37.09%)	15 (27.78%)	
				4 – AMAI Score E/F	33 (12.00%)	6 (11.11%)	

analyzed participants were selected from the parent cohorts. Characteristics, including covariates adjusted in the downstream analyses, of the 89 analyzed subjects are also provided and compared with the population in Table 1 to check whether the analytical sample represents the cohort population. There were no significant differences (p -value > 0.05) in the characteristics of the cohort population and analyzed subset (Table 1; left panel). Of those 89 subjects included in analyses, 28 (31.46%) had given birth at least 6 months before age 26 (and after age 18) and 61 (68.54%) had not had a child prior to 26 years. The same characteristics were also compared between the 28 parous and 61 nulliparous women in Table 2 (left panel). There were significant differences

in socioeconomic status (SES) between parous and nulliparous IOW women (p -value = 0.0125, Table 2; left panel). Women with higher income/housing and lower education tended to be nulliparous up to age 26. Smoking status was marginally significantly different (p -value = 0.0634) between parous and nulliparous IOW women, with a higher proportion of nulliparous women being non-smokers (including no environmental tobacco smoke exposure inside the home) at age 18 compared to parous women.

Similar characteristics of the ELEMENT female study population ($N=275$) and the DNAm subsample ($n=54$) are detailed in Table 1 (right panel). Participants from ELEMENT were on average younger than IOW

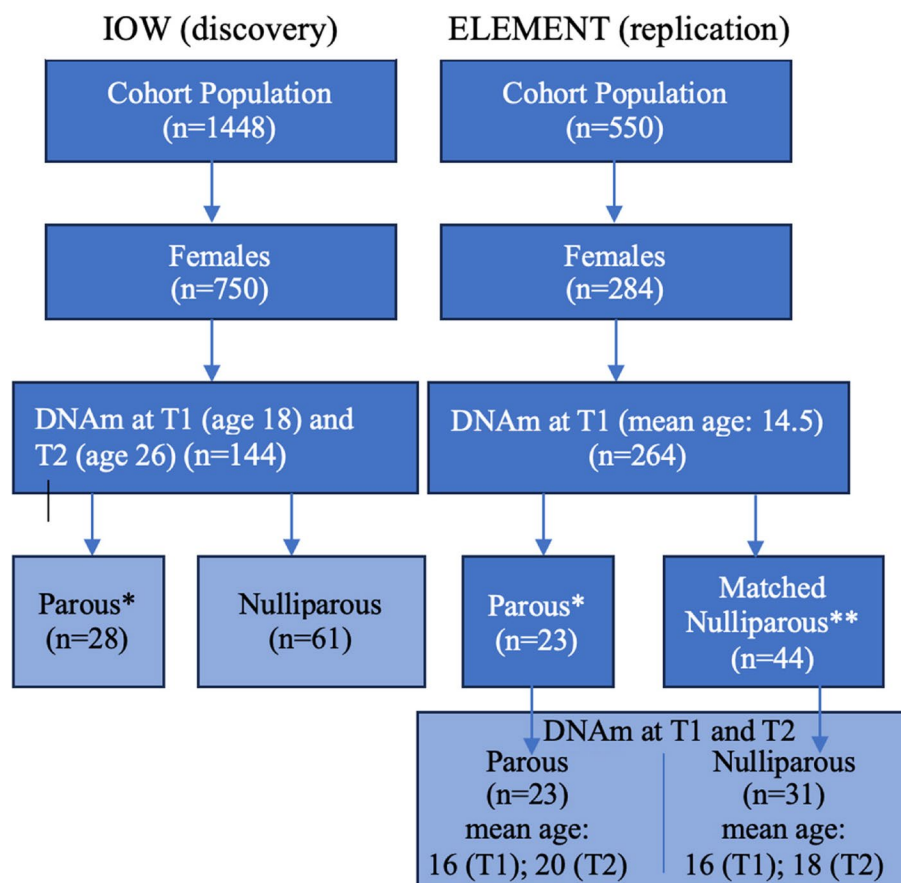


Fig. 1 Flow diagram depicting how analyzed participants were selected from the parent cohorts (*: only included are parous women who were at least 6 months postpartum at the time of DNAm measurement at T2; **: nulliparous women were selected to match the age range at T1 of the parous group cohort of origin as closely as possible. Details can be found in “Childbearing status” of Sect. 5)

participants and their ages of data collection at T1 and T2 had a small range. Due to their younger age, a smaller proportion ($n=23$, ~8% of 275) of ELEMENT participants had given birth at least 6 months before the second DNAm measurement. The nulliparous comparisons ($n=31$) were selected to match parous women with ages as close as possible, which explains that (a) the mean age of analyzed samples at both T1 and T2 was significantly older than the study population at both T1 and T2 (p -values < 0.0001) and (b) the proportion of active and passive smokers was significantly higher in analyzed samples (p -value = 0.0001). As shown in Table 2, the mean ages of parous and nulliparous women in ELEMENT at T1 (pre-pregnancy) were 16.3 and 16.22 yrs old, respectively, and did not differ significantly from each other. The mean age of nulliparous women was significantly younger than parous women at T2 (postpartum) even though we tried to match ages between parous women and nulliparous controls (mean age of nulliparous women: 18.6; mean age of parous women: 20; p -value < 0.0001). Similar to the IOW study, SES was also significantly different

between parous and nulliparous women in the analyzed ELEMENT cohort ($p=0.017$). Women with higher SES at T1 tended to be nulliparous up to T2. Smoking status was not significantly different between parous and nulliparous women in ELEMENT (p -value = 0.5638). In both, the IOW and ELEMENT cohorts, body mass index (BMI) changes were not significantly different between parous and nulliparous women.

ttScreening and linear regression analyses

To identify candidates for parous-related CpGs, a training-testing screening method (ttScreening) was applied to 389,355 CpG sites (overlapped CpGs between HumanMethylation450 and EPIC platforms after quality control) at 26 years of age while adjusting for six cell proportions [51]. In total, 197 CpGs passed ttScreening analysis. Focusing on these CpGs, linear regression analyses were performed thereafter to assess the differential DNAm at age 26 between parous women and their nulliparous counterparts. Cell proportions at age 26, residual DNAm at age 18 (adjusted by cell proportions at age 18), change

Table 2 Descriptive characteristics of parous and nulliparous women (IOW and ELEMENT)

Discovery Cohort: IOW				Replication Cohort: ELEMENT			
Characteristics	Parous (n = 28)	Nulliparous (n = 61)	Test Stat (p-value)	Characteristics	Parous (n = 23)	Nulliparous (n = 31)	Test Stat (p-value)
Age at T1: Mean (SD)	18 (0)	18 (0)	-	Age at T1: Mean (SD)	16.30 (1.44)	16.11 (1.14)	-0.55 (0.59)
Age at T2: Mean (SD)	26 (0)	26 (0)	-	Age at T2: Mean (SD)	20.04 (1.15)	18.63 (1.02)	-4.77 (<.0001)
BMI change (T1 to T2): Mean (SD)	2.98 (3.21)	2.69 (3.61)	0.89 (.3752)	BMI change (T1 to T2): Mean (SD)	1.61 (5.52)	0.82 (1.75)	0.75 (.4568)
Smoking Cluster^b at T1: N (%)				Smoking Cluster^b at T1: N (%)			
1—non active and non-passive smokers at T1	3 (10.71%)	18 (29.51%)	7.28 (.0634)	1—non active and non-passive smokers at T1	19 (86.36%)	22 (91.67%)	0.33 (.5638) ^a
2—passive and non-active smokers at T1	15 (53.57%)	33 (54.10%)		2—passive and non-active smokers at T1	0	0	
3—active and non-passive smokers at T1	1 (3.57%)	0 (0.0%)		3—active and non-passive smokers at T1	0	0	
4—active and passive smokers at T1	9 (32.14%)	10 (16.39%)		4—active and passive smokers at T1	3 (13.64%)	4 (9.34%)	
Household SES at T1: N (%)			10.86 (.0125)	Household SES at T1: N (%)			8.15 (.0170)
1—high income, median housing, low education	5 (17.86%)	25 (40.98%)		1: AMAI Score A/B (highest)	0 (0.00%)	0 (0.00%)	
2—median income, high housing, high education	11 (39.29%)	28 (45.90%)		2: AMAI Score C	9 (39.13%)	24 (77.42%)	
3—low income, low housing, median education	10 (35.71%)	7 (11.48%)		3: AMAI Score D	10 (43.48%)	5 (16.13%)	
				4: AMAI Score E/F (lowest)	4 (17.39%)	2 (6.45%)	

^a The two smoking categories with no subjects were excluded from the chi-square test

^b Smoking clusters were generated from passive and active smoking categories at T1 using K-mean clustering analysis (see “Confounding variables” in Method section for more details)

of BMI from age 18 to 26 yrs, smoking categories (treated as an ordinal variable), and socioeconomic status (treated as a categorical variable) at age 18 were adjusted for in the linear regression model as covariates. 184 CpGs were significantly associated with childbearing (FDR-adjusted p -value < 0.05 from the linear regression). The regression coefficients and p -values of the 184 CpGs are provided in the Appendix (A Table 1).

Replication analyses

The 184 significant CpGs identified in the IOW study were extracted from the ELEMENT DNAm dataset ($n = 54$) and their association with childbearing was estimated using linear regression models, adjusting for equivalent or comparable covariates as in discovery analyses in IOW. Age at T2 was included in the replication analysis since women in ELEMENT subsample demonstrated variability in age. 105 of 184 CpGs had regression coefficients in the same direction as the IOW study, of which 13 were significantly replicated in the ELEMENT study (raw p -value < 0.05). Table 3 provides the regression

coefficients, standard errors (SE) and p -values of these 13 CpGs for the IOW and ELEMENT samples, and their 16 associated unique genes. Boxplots in Fig. 2 demonstrate the mean methylation levels (in beta values) of these 13 replicated CpGs, 7 of which had negative regression coefficients and 6 of which had positive regression coefficients in parous compared with nulliparous women in both cohorts.

Gene enrichment analysis and gene-disease association

We conducted gene set enrichment analysis on the 16 genes (associated with 13 replicated CpGs) using GOMeth method, which performs unbiased gene set testing following differential methylation analysis. No KEGG biological pathways survived after FDR adjustment. We then searched these 16 genes in the disease gene database DisGeNET v7.0 (10 of the 16 genes were present in the database) and identified a number of diseases associated with parous-related genes with gene disease association score greater than 0.3 (A Fig. 1). The 10 genes found in the DisGeNET v7.0 include *SEMA3A*, *AKAP13*, *SLC15A2*,

Table 3 CpG sites significantly associated with childbearing history in discovery and replication analyses and with the same coefficient direction

CpGs	Discovery Cohort: IOW				Replication Cohort: ELEMENT			Associated Genes		
	Coef. ^a Parous	SE	Raw p-value	FDR p-value	Coef. ^a Parous	SE	Raw p-value	Chr	Gene Name	UCSC RefGene Group
cg08288130	-0.2552	0.0645	0.0002	0.0009	-0.1625	0.0665	0.0188	8	<i>DOK2</i>	TSS1500
cg13375690	0.2371	0.0609	0.0002	0.0009	0.1422	0.0617	0.0260	8	<i>PDE7A</i>	Body
cg17672798	-0.1740	0.0447	0.0002	0.0009	-0.1445	0.0711	0.0482	10	<i>ADARB2</i>	Body
cg20955022	-0.1778	0.0468	0.0003	0.0010	-0.1732	0.0649	0.0107	7	<i>SEMA3A</i>	TSS1500
cg24737639	0.3731	0.0999	0.0004	0.0010	1.3769	0.5684	0.0197	12	<i>NUP37; PARPBP</i>	TSS1500;TSS200
cg02052068	-0.2273	0.0621	0.0004	0.0011	-0.2269	0.1045	0.0355	5	<i>UGT3A1</i>	TSS1500
cg08285768	0.2687	0.0790	0.0010	0.0017	0.1675	0.0820	0.0474	15	<i>AKAP13</i>	Body
cg16710348	0.3078	0.0950	0.0017	0.0024	0.1801	0.0699	0.0135	3	<i>SLC15A2; SLC15A3</i>	3'UTR
cg18379829	0.1104	0.0412	0.0091	0.0108	0.1542	0.0500	0.0036	1	<i>PLD5; MAP1LC3C</i>	
cg19086905	-0.1789	0.0682	0.0105	0.0123	-0.1631	0.0709	0.0262	15	<i>ADAMTS17</i>	
cg06772580	0.0880	0.0350	0.0139	0.0161	0.1304	0.0405	0.0024	17	<i>TP53I14</i>	Body
cg01832012	-0.1388	0.0559	0.0151	0.0173	-0.0989	0.0473	0.0423	7	<i>TPK1</i>	Body
cg01537571	-0.3006	0.1273	0.0207	0.0231	-0.3071	0.1095	0.0075	15	<i>TM2D3</i>	TSS1500

^a Nulliparous is reference; negative values indicate a decrease in DNAm in parous women

DOK2, *ADAMTS17*, *ADARB2*, *NUP37*, *PLD5*, *TPK1*, and *SLC15A3*, which were most notably linked with neoplasms, mental disorders and nervous system diseases (e.g. Parkinson's disease, schizophrenia, autism), and substance use (e.g. prescription drug abuse, substance use disorder).

RNAseq analysis

Of the 16 genes associated with the 13 identified and replicated parous-related CpGs, 13 genes were present in the gene expression data generated from blood samples at age 26 in the IOW study. We checked the Pearson correlations between the DNAm (M-value) of each of the 13 CpGs and the log-transformed gene expression level of their corresponding genes ($n=165$) and found the methylation levels of cg19086905 (*ADAMTS17*; $r=-0.14$; p -value=0.064) and cg13375690 (*PDE7A*; $r=-0.21$; p -value=0.006) were negatively associated with gene expression levels, while the methylation level of cg01537571 (*TM2D3*; $r=0.16$; p -value=0.037) was positively associated with the gene expression level of *TM2D3*. Considering the non-normality of DNAm and gene expression levels, we also checked Spearman correlations, which provide similar results.

We further tested the differential gene expression at age 26 between parous and nulliparous women ($n=75$; 25 parous vs 50 nulliparous women) for the 16 genes associated with dmCpGs. Only one gene, *TM2D3*, had

significantly lower gene expression levels in parous women compared to the nulliparous counterpart (regression coefficient = -0.17; p -value = 0.0085).

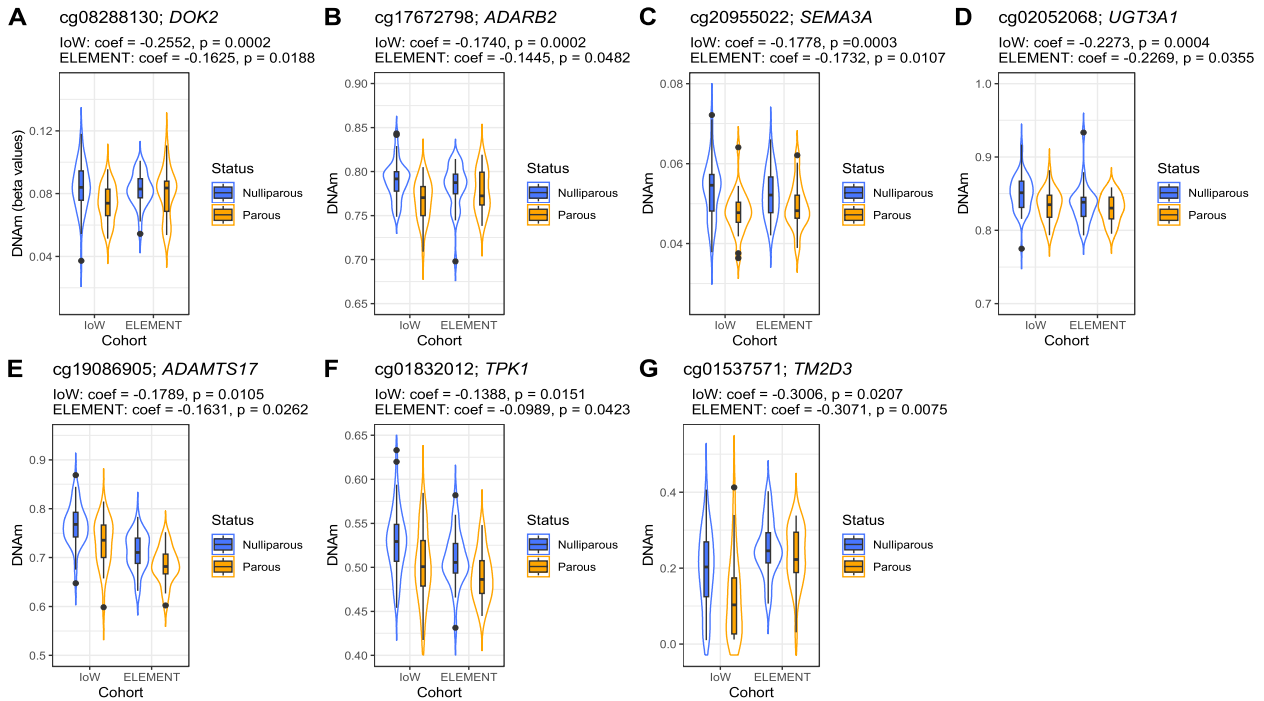
Differentially Methylated Regions (DMRs)

Using the dmrrf R package, we identified 1 DMR that was marginally significantly associated with childbearing (FDR adjusted p -value = 0.1024; Table 4). The DMR contained 1 gene and 12 unique CpG sites as listed in Table 4. None of these 12 CpG sites overlapped with the individual CpG sites identified in the EWAS study using ttScreening and linear regression. However, 5 out of the 12 CpGs were significantly differentially methylated between parous and nulliparous women with raw p -values < 0.05 using the IOW dataset (Table 4; column 8). None of these 5 CpGs were significantly replicated in ELEMENT.

Discussion

Our aim was to identify parous-related DNAm changes that are sustained for at least 6 months after childbirth, as a step toward understanding how childbearing impacts disease risk later in life. To identify CpGs that potentially underlie this relationship, we performed an epigenome-wide association study in the IOW birth cohort to test the hypothesis that parous women have significantly different DNAm as compared to nulliparous women and that these changes persist at least 6 months after giving

CpGs with negative regression coefficients in the IOW and ELEMENT samples (A-G)



CpGs with positive regression coefficients in the IOW and ELEMENT samples (H-M)

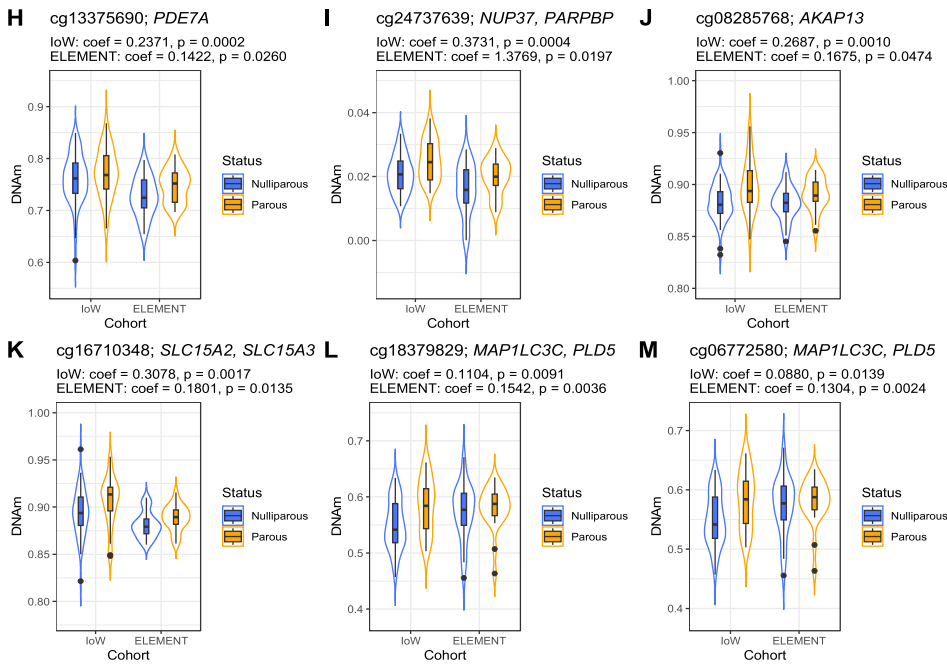


Fig. 2 Violin and boxplots of DNAm (in beta values representing 0 to 1 proportion methylated) measured at age 26 yrs (IOW) and time T2 (ELEMENT) of 13 statistically significant parous-related CpGs with the same coefficient direction in IOW and ELEMENT cohorts

birth. We identified 184 CpGs that were significantly differentially methylated between parous and nulliparous

women. We pursued replication of our findings in a sub-sample of the ELEMENT cohort.

Table 4 DMRs associated with childbearing status

DMR Region	Chr	DMR Start	DMR End	FDR Adjusted p-value	CpGs in the region	UCSC_RefGene_Group	Coef. of DMPs ^a in the region (IOW)	p-value of DMPs in the region (IOW)	Gene Name
1	6	30,687,972	30,688,860	0.1024	cg06993041	TSS200	-0.4332	0.0336*	<i>TUBB</i>
					cg23897356	TSS200	0.0756	0.3762	
					cg02167193	TSS200	-0.064	0.3524	
					cg21585627	TSS200	0.0894	0.4401	
					cg06608112	1stExon;5'UTR	-0.0034	0.9734	
					cg14861518	1stExon	-0.0317	0.3889	
					cg15482861	1stExon	-0.064	0.0247*	
					cg01554625	Body	-0.1374	0.0401*	
					cg21660604	Body	-0.0949	0.0494*	
					cg03130533	Body	-0.1543	0.3274	
					cg01944444	Body	-0.0486	0.6222	
					cg20118306	Body	-0.1767	0.0161*	

* raw p-value < 0.05

^a Differentially methylated positions (DMPs) are the individual CpG sites identified in EWAS study using ttScreening and linear regression

Out of the 184 CpGs identified in the IOW, 105 of these CpGs had regression coefficients in the same direction in the ELEMENT samples, of which 13 were significantly associated in the ELEMENT participants using linear regression while adjusting for the same covariates. Among these 13 parous-related CpGs, 7 had significantly lower levels of methylation in parous versus nulliparous women in both IOW and ELEMENT cohorts. The remaining 6 CpGs had higher levels of methylation in parous women compared to their nulliparous counterparts.

The gene ontology analysis of the 16 genes associated with 13 parous-related CpGs using GOMeth did not identify any significant KEGG biological pathways. We further searched the 16 parous-related genes in the disease gene database DisGeNET v7.0 and identified a gene-disease network of genes that were mainly linked with neoplasms, mental disorders and nervous system diseases (e.g. Parkinson disease, schizophrenia, autism), and substance use (e.g. prescription drug abuse, substance use disorder).

The gene expression data of 13 (out of 16) genes were available from whole blood samples (age 26) in the IOW cohort and were used to further test (i) the association between DNAm of the 13 parous-related CpGs and gene expression of the 13 genes and (ii) the differential gene expression between parous and nulliparous women at age 26. Lower methylation of cg19086905 (*ADAMTS17*) and cg13375690 (*PDE7A*; Body) were linked with higher gene expression levels. Only one gene, TM2 domain containing 3 (*TM2D3*), had significantly lower gene expression level in parous women compared to the

nulliparous counterpart (regression coefficient = -0.1104; p-value = 0.0028). Similarly, the methylation level of cg01537571, which is located within a CpG island and near the transcription start site of *TM2D3*, was significantly lower in parous women than their nulliparous counterparts in both IOW and ELEMENT cohorts. Our association analysis indicated that subjects with higher methylation level of cg01537571 have higher gene expression level at age 26 in IOW study. Genetic variants in *TM2D3* have been reported to be significantly associated with increased risk of developing late-onset Alzheimer's disease through an exome-wide association analysis [34]. Furthermore, increased parity has been shown to be a risk factor for Alzheimer's disease [15], which may be related to the lower methylation of cg01537571 in parous women.

We searched the genes associated with the 13 identified and replicated CpGs through GTEx Portal (<https://gtexpportal.org/home/>) and found that less than half of them (*TM2D3*, *DOK2*, *PDE7A*, *AKAP13*, *SLC15A3*, *TPK1*) are routinely expressed in whole blood samples, with the remainder are primarily expressed in other tissues. With regard to our gene enrichment analysis findings, *TM2D3*, *SLC15A2*, *PLD5*, *ADARB2* and *PARPBP* are expressed in the brain. *MAP1LC3C*, *NUP37* and *ADAMTS17* are expressed in breast tissue. Further studies are needed to investigate how these parous-related genes are differentially expressed across a variety of tissues between parous and nulliparous women.

Among the genes identified in these analyses, *SEMA3A* (semaphorin 3A) is of interest based on its potential role in neoplasias and other diseases that may have a modified

risk in parous women. *SEMA3A* had decreased DNAm in parous women (Fig. 2, Table 3) and was associated with neoplasms as well as mental health and reproductive disorders in the DisGeNET gene disease-class heatmap (A Fig 1). Semaphorins are guidance cue signaling molecules that direct developing neurons and other tissues during embryogenesis. They also persist in adulthood, playing a role in the regulation of neuroplasticity [13]. Changes in *SEMA3A* gene expression levels have been noted in several nervous systems diseases, including schizophrenia (increased in cerebellar Purkinje cells) [23], which was a disorder associated with *SEMA3A* in our gene-disease network. With regard to neoplasia, *SEMA3A* plays paradoxical roles in the tumor micro-environment in different types of tumors [10, 36, 55]. For example, *SEMA3A* promotes tumor progression in hepatocellular carcinoma by means of enhanced proliferation, migration, and invasion and was most commonly upregulated in patients with recurrent hepatocellular carcinoma [32]. In contrast, *SEMA3A* serves as a tumor suppressor in head, neck, and breast cancers [20, 31]. Specifically in breast cancer, *SEMA3A* increases antitumoral M1 macrophage proliferation leading to recruitment and activation of CD8+T cells and NK cells, which repress tumor growth [20]. While more studies are needed to explore the link and mechanism, it may be that parous women may have a lower risk of developing breast cancers through lower DNAm near the transcription start site of *SEMA3A*. Studies investigating whether blood DNAm levels change similarly in blood (assessed here) and breast tissue (target tissue of interest) following childbearing are also needed to understand the implications of this finding.

The key strength of this study is its design that compares DNAm profiles of parous women at two time-points, pre- and post-pregnancy, to those of nulliparous women. Studies that focus on the health effects of pregnancy and parity commonly include data points pre- and post-pregnancy, but less commonly do they have comparable data on nulliparous women. Being able to directly compare methylation at CpGs epigenome-wide between parous and nulliparous women in both the discovery and replication cohorts is a clear asset to advancing our understanding of DNAm as it may relate to long-term health outcomes for women. Furthermore, our study investigated the methylation changes from pre-pregnancy to months or years after childbirth and identified CpGs at which methylation changes persisted. These changes, in turn could be related to later-life diseases if our assertions hold up in future studies. Another strength is our comprehensive approach to validation with DNAm followed by measurement of gene expression, assessment of DMRs, and gene enrichment analyses in the discovery

cohort. The mapping of diseases onto the genes associated with the identified CpGs helps to put our findings into context and provides direction in terms of genes of interest. Additionally, replication of our findings generated from the IOW cohort (98% White in the United Kingdom) in a subsample from an independent cohort from another geographic location (Mexico) and of a different ethnic background (all Hispanic) supports both the internal validity and the generalizability of the original findings. Finally, our study is disease-independent, so as not to limit our outcomes to a predetermined set of conditions. While we were a priori interested in any potential relationships between DNAm and reproductive cancers, we did not anticipate that substance abuse disorders, for example, would be among the outcomes related to the associated genes and thus our results open new avenues to explore.

Our study design has some notable strengths relative to other studies examining the impact of reproductive history on DNAm. Specifically, we elected to use at least a 6-month window following parturition before post-pregnancy samples were collected. In contrast, Gruziova et al. [25] collected samples 2–4 days after delivery, which may be too short an interval to identify DNAm changes that persist. This time in the follow up interval for sample collection may explain the contrasting results in that none of the CpGs identified in this study were among the CpGs identified by Gruziova et al. [25].

There are several limitations to note in our study. First, the nulliparous status of women in the discovery cohort was obtained from the medical record. While unlikely, if a woman received prenatal care and/or gave birth off the Isle of Wight it may not have been recorded in the medical records to which we had access. Second, we excluded women who were within 6 months of childbirth from our study to avoid any transient methylation changes due to childbirth. However, it is unknown how long it takes methylation to stabilize or to diminish postpartum. Third, the sample size of parous women is relatively small (28 parous women in the IOW study and 20 parous women in the ELEMENT study), which would potentially reduce the capability of detecting CpGs that are truly associated with childbearing status. Fourth, the discovery (IOW) and replication (ELEMENT) cohorts are different in terms of population distribution in race/ethnicity, age, and geography, which could limit the power of discovering replicable parous-related epigenetic biomarkers. Fifth, factors/exposures that potentially change during pregnancy may contribute to DNAm changes persisting after childbirth. However, we are unable to assess these effects since these factors/exposures were not measured in nulliparous women at similar times as parous women. Finally, as common practice in epidemiological studies,

blood DNA was used a proxy for changes in other tissues, and it is known that many genes have different methylation and expression profiles that vary by tissue (i.e. breast, brain). Similarly, the gene expression data is limited to whole blood samples in IOW, and as noted above, approximately half of the identified genes are not expressed in this tissue.

Conclusion

Our results suggest that epigenetic changes may be associated with childbearing status in women. If these findings in methylation of blood leukocyte DNA reflect epigenetic changes in other relevant tissues, it may provide a possible mechanism of how childbearing is linked to various future health outcomes. Diseases associated with the identified genes involved cancers, the nervous system, and substance uses. Additional studies are needed in cohorts with follow up in later life to investigate whether and how parous-related DNAm changes are associated with later-life health outcomes in women.

Methods

Discovery cohort – IOW

The Isle of Wight (IOW) is a 3-generation birth cohort established in the United Kingdom to prospectively study the natural history of asthma and allergic conditions [4]. Study participants were enrolled in the birth cohort study between 1989 and 1990 by contacting potential parents (first generation, IOW-F0) while they were carrying the second generation (IOW-F1). The IOW-F1 generation has been followed up for 26 years. IOW-F1 female participants of the birth cohort have been followed through their pregnancy occurring between the years 2011 and 2015. When F1-women became pregnant, information was gathered about their lifestyle and health during pregnancy. The racial identity of most participants (98%) in the F1 generation was white.

Replication cohort – ELEMENT

ELEMENT (Early Life Exposures in Mexico to Environmental Toxicants), is a multi-generation birth cohort of Hispanic women from Mexico City maternity hospitals that has been following participants for nearly three decades with a focus on studying how dietary and environmental toxicant exposures affect women and children. ELEMENT originally consisted of three sequential birth cohorts and multiple early childhood follow-up visits, as previously described [48]. A subset of the original participants (550 mother–child pairs) from the second and third cohorts attended two follow-up visits, approximately two years apart, during adolescence adolescent years. In 2019, ELEMENT launched a third generation study ('E3Gen'), following up ELEMENT participants that had become

parents. Extensive questionnaire (demographic, dietary, and more) and anthropometry data along with blood and urine samples were collected from mothers and their offspring at each study visit.

Childbearing status

For IOW, participants' childbearing status as either parous ("yes" indicating giving at least one live birth) or nulliparous ("no" indicating no reported pregnancy or childbirth) was assessed up to age 26 by their participation in the 3rd generation IOW study (inclusion requires having a child) as well as from the medical records. Parous women who were pregnant or within 6 months postpartum at the time of DNAm measurement at age 26 were excluded from this study. Women with documented early pregnancy termination, infertility treatment, and late miscarriage were excluded from the analyses.

For the ELEMENT replication subsample, participants who had children ages 6 months to 5 years of age in 2019 were invited to attend the E3Gen visit with their child or children. Twenty-three eligible female participants with one or two children each participated; this is referred to as T2 for the parous group. All had epigenetic analyses, e.g., DNAm of blood leukocytes, completed prior to pregnancy at the adolescent T1 visit. For the nulliparous group, women were selected that had a) T1 epigenetic analysis completed, b) attended a second adolescent visit > 2 years later without a known pregnancy in between (their T2 timepoint), and c) did not have missing data for necessary covariates. 44 women met these criteria; from among them 31 were selected to match the age range at T1 of the parous group cohort of origin as closely as possible.

DNA methylation measurements and cell estimation

In both cohorts, peripheral whole blood samples were collected twice. In the IOW cohort study, the first sample (T1) was collected at age 18 years, prior to any pregnancy. The second sample (T2) was collected at age 26 years, which was at least 6 months after childbirth for the parous subjects. DNA was extracted from blood samples using a standard salting procedure [47] or commercial kits (Qiagen, Germantown, MD, USA). DNA concentration was determined using fluorometry (Qubit, Invitrogen™, Thermo Fisher Scientific, Waltham, MA, USA). To measure epigenome-wide methylation levels, age 18 samples were processed on either Illumina Infinium HumanMethylation450 arrays or Infinium MethylationEPIC BeadChips (Illumina, Inc., San Diego, CA, USA). All the age 26 samples were processed on MethylationEPIC BeadChips. Multiple identical control samples were assigned to each bisulfite conversion batch for assessment of assay variability. DNAm data were preprocessed

(including background level correction and quantile normalization) using the CPACOR pipeline for data from both platforms (i.e. HumanMethylation450 and MethylationEPIC) [43, 44]. Specifically, an R package, minfi was used to quantile-normalize the DNAm intensity data [5, 44]. Quantile-normalized intensities were then used to calculate beta (β) values, which represent proportions of intensity of methylated (M) over the sum of methylated and unmethylated (U) sites/probes ($\beta = M / [c + M + U]$, where c is a constant to prevent zero in the denominator if $M + U$ is too small). Beta values close to 0 or 1 tend to suffer from severe heteroscedasticity, and base-2 logit transformed beta values (denoted as M-values) have been demonstrated to perform better in differential analysis of methylation levels [19]. Therefore, methylation levels in the analysis are represented using M-values. As quality control, probes that did not reach a p -value of 10^{-16} in at least 95% of samples were excluded. The same criterion was applied to exclude samples, i.e., samples with p -value $> 10^{-16}$ in at least 95% of the CpGs [44]. CpGs on sex chromosome were excluded. Use of both HumanMethylation450 (containing > 450 K CpGs) and MethylationEPIC (containing > 850 K CpGs) platforms is a consequence of technology evolution (HumanMethylation450 arrays are no longer available). The HumanMethylation450 and MethylationEPIC arrays were initially processed (quality control, quantile normalization of background-corrected intensities) separately for beta estimation and then analysis of 389,355 common CpG sites between 450 K and EPIC was undertaken. Methylation data were quantile normalized using the minfi R package and batch corrected using ComBat in R [37].

In the ELEMENT cohort study, peripheral blood samples were collected via venipuncture and stored in EDTA-preserved after all study visits. For both parous and nulliparous women, the T1 sample for epigenetic analyses came from the early adolescent study visit; participants were ages 11 to 18 years (mean age: 16.3). The T2 sample came from the second adolescent visit for nulliparous subjects and from the E3Gen post-birth visit for parous subjects (age range 16 to 21 years, mean age: 19.23). DNA was isolated from blood leukocytes via the Flexigene kit (Qiagen) and DNAm was quantified via the MethylationEPIC platform. DNAm data were preprocessed using the same pipeline for data as the IOW study.

Since DNAm was analyzed in whole blood, there was a need to assess changes of DNAm while adjusting for changes in cell composition of whole blood. Proportions of the cell types CD4+ T-cells, CD8+ T-cells, natural killer cells, B-cells, neutrophils, monocytes, and eosinophils were estimated using the approach by [29, 33] through estimateCellCounts function from the R package *minfi* [5]. Prior studies suggested that DNA

methylation analyzed in whole blood may be substantially confounded by cell composition effects [30, 33, 42]. Estimated cell compositions were thus included in the analyses as confounders.

Confounding variables

In both discovery (IOW) and replication (ELEMENT) cohorts, confounding variables included BMI, smoking status (including active and second-hand/passive smoking), and socioeconomic status (SES) at both time points when DNAm measurements were collected (denoted as T1 and T2). Active smoking was recorded as “yes” for a current smoker. Second-hand smoking was recorded as “yes” if anyone in the household smokes inside the home. Considering that active and passive smoking and SES may be affected by reproductive status (for example, mothers or family may quit smoking during and after pregnancy; household income may reduce if mothers do not work outside the home beyond a typical maternity leave), smoking status and SES at T2 may serve as mediators between childbearing status and DNAm at T2, and thus were not included as confounding variables. K-mean clustering analysis was applied on the active and second-hand smoking at T1 and grouped female participants into 4 smoking categories (the number of clusters $K=4$ is chosen due to the largest CCC value) interpreted as “non-active and non-passive smokers at T1,” “passive and non-active smokers at T1,” “active and non-passive smokers at T1,” and “active and passive smokers at T1,” which were coded as 1, 2, 3, and 4 respectively and were treated as an ordinal variable in the regression analysis. The K-mean clustering also imputed the missing values in smoking status at T1.

In the IOW study, SES categories were generated by K-mean clustering using education, housing information and income, which were extracted from the questionnaire. Education was recorded as “School,” “6th Form College,” “Further Education,” and “Other.” Housing information was recorded in ordinal data as “Rented Council/Housing Assoc.,” “Rented Private,” “Lives with Parents,” “Owned Private,” and “Other.” Income was also recorded in the ordinal groups as “Less than £12,000,” “£12,000 to £17,999,” “£18,000 to £29,999,” “£30,000 to £41,999,” “Greater than £42,000,” and “Prefer not to say.” In the cluster analysis, “Other” in education and housing information, and “Prefer not to say” in the income were treated as missing values. In the replication study with ELEMENT cohort data, socio-economic status is measured through a scale created by the Mexican Association of Marketing Research (AMAI) that classifies a household’s socioeconomic status based on 6 variables: level of education of head of household, number of bedrooms, number of bathrooms, employed people over the age of

14, number of vehicles, and access to internet [12]. The standardized criteria for socio-economic levels in Mexico "AMAI 13×6" was utilized for analyses. "AMAI 13×6" rule classified the households in Mexico into 6 levels: A (highest), B, C, D, E, and F (lowest).

The change of BMI from T1 to T2, smoking categories at T1, and socioeconomic status at T1 were included in the model as covariates. Age of the participants at T2 was also included as a covariate in the replication analysis with the ELEMENT data.

Gene expression measurement

RNAseq was used to assess gene expression in blood samples collected at age 26 years. Total RNA was extracted using PAXgene Blood RNA kits (Qiagen, Germantown, PA, USA) and brought up in PAXgene Blood miRNA PreAnalytic elution buffer. RNA was quantified by Qubit dsDNA High Sensitivity assays (Invitrogen™, Thermo Fisher Scientific, Waltham, MA, USA) and RIN scores were determined on a 2100 Bioanalyzer instrument (Agilent Technologies, Santa Clara, CA, USA). Sequencing libraries were prepared using the Illumina TruSeq Stranded mRNA Library Preparation Kits with IDT for Illumina Unique Dual Index barcode primers following the manufacturer's recommendations (Illumina, Inc., San Diego, CA, USA). Completed libraries were assessed for quantity and quality using a combination of Qubit dsDNA High Sensitivity assays and either an Advanced Analytical Fragment Analyzer High Sensitivity DNA NGS or Agilent 4200 TapeStation High Sensitivity DNA 1000 assays. Libraries were pooled in equimolar amounts for multiplexed sequencing. Pools were quantified using the KAPA Library Quantification qPCR kits (Roche Diagnostics Corp., Indianapolis, IN, USA) or Invitrogen Collibri Quantification qPCR kits. Each pool was loaded onto one lane of either an Illumina HiSeq 4000 flow cell, an Illumina NovaSeq 6000 S4 flow cell, or an Illumina NovaSeq 6000 SP flow cell and sequencing was performed in a 2×75 bp or 2×150 bp paired end format. Base calling was done by Illumina Real Time Analysis (RTA) (v2.7.7 or v3.4.4) and output of RTA was demultiplexed and converted to FastQ format with Illumina Bcl2fastq (v2.19.1 or v2.20.0).

HISAT2 (v2.1.0) aligner was used to map reads against the Human Genome GRCh37 version 75 [40] and alignment files were produced in the Sequence Alignment Map (SAM) format. Files were converted into the Binary Alignment Map (BAM) format using SAMtools (v1.3.1) [45]. The number of reads mapped to each gene in the same reference genome used for alignment were counted using HTSe (v0.00.1) [2]. The countToFPKM package (<https://github.com/AAlhendi1707/countToFPKM>) was used to calculate normalized read count

FPKM (Fragments Per Kilobase of transcript per Million mapped reads) and their log transformed values were used for data analysis.

Statistical analysis

The study aimed to examine whether DNAm changes are associated with childbearing status between ages 18 and 26 years old. A set of analyses was carried out to achieve this goal. First, ttScreening approach was applied to screen for significant CpG sites with differential DNAm (M-values) at age 26 that are potentially associated with childbearing status up to age 26 while adjusting for composition of cell types CD8+ T cells, CD4+ T cells, natural killer cells, B cells, monocytes, and granulocytes at age 26. ttScreening is a screening approach that repeatedly splits samples into training and test data and utilizes the training and testing data to filter out uninformative DNAm sites [51]. CpGs showing statistical significance in at least 50% of randomly-split "training and testing" datasets were selected as potentially outcome-associated CpGs and were then considered for further downstream analysis. Second, focusing on the potential significant CpGs identified by ttScreening, linear regressions were implemented with DNAm (M-values) at age 26 as the dependent variable, childbearing status as the predictor, and cell-adjusted DNAm at age 18 or T1 (i.e. residuals from linear regressions of DNAm at age 18 on the six cell compositions at age 18), six cell compositions at age 26, BMI changes from age 18 to 26 yrs, smoking categories at age 18, and socioeconomic status at age 18 as covariates. Third, in addition to identification of individual parous-associated CpGs, we performed regional DNAm analyses using DMRff [54] to identify differentially methylated regions (DMRs) associated with childbirth in women. We first conducted EWAS using linear regression with DNAm in M values at age 26 as a dependent variable, childbearing status as an independent variable, and with the following covariates: 1) cell-adjusted DNAm at age 18 (i.e. residuals from linear regressions of DNAm at age 18 on the six cell compositions at age 18), 2) six cell compositions at age 26, 3) BMI changes from age 18 to 26, 4) smoking categories at age 18, and 5) SES at age 18. We then performed regional DNAm analysis using DMRff with the regression coefficients, standard errors, and *p*-values of childbearing status generated from the EWAS analysis. Fourth, we conducted gene ontology analyses of differentially methylated individual CpGs using the GOMeth function from the R package MissMethyl [50]. Fifth, we searched the disease gene database, DisGeNET v7.0, which is a platform containing large collections of genes and variants associated with health status, to characterize genes and related molecular pathways of the CpGs of interest. Sixth, differential gene expression

analysis was performed on the genes (mapping from the identified CpGs in the Second Step) to further test (i) the association between DNAm of identified CpGs and gene expression of the corresponding genes at age 26 using Pearson's and Spearman's correlation and (ii) the differential gene expression at age 26 between parous and nulliparous women using linear regressions.

Replication analysis

To replicate the findings identified in the discovery cohort (IOW), linear regressions were implemented with DNAm at T2 as the dependent variable, childbearing status as the predictor, and cell-adjusted DNAm at T1 (i.e. residuals from linear regressions of DNAm at T1 on age and cell composition at T1), six cell compositions and age at T2, BMI changes between T1 and T2, smoking categories at T1 and T2, as well as socioeconomic status (SES) at T2 as covariates.

Abbreviations

CpG	Cytosine-phosphate-guanine
DMPs	Differentially Methylated Positions
DMRs	Differentially Methylated Regions
DNAm	DNA methylation
ELEMENT	Early Life Exposure in Mexico to ENvironmental Toxicants
IOW	Isle of Wight
SE	Standard Errors
SES	Socioeconomic status

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s43682-024-00025-9>.

Supplementary Material 1.

Supplementary Material 2.

Acknowledgements

Not applicable.

Authors' contributions

SC and SE lead the study concept and design. SHA, JH, WK contributed to the study design and SHA, JH, WK and SE generated data from the IOW cohort. JG, KP, and DD contributed data from the ELEMENT cohort. SC and MJ analyzed the data. SC, MJ, PKR and SE drafted the manuscript. All authors contributed to the data interpretation and approved the final version of the manuscript.

Funding

The Isle of Wight study received support from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) (R21HD099614), and National Institute of Allergy and Infectious Diseases (NIAID) (R01AI091905; R01AI121226). The ELEMENT study received funding from the U.S. Environmental Protection Agency (US EPA) (grant: RD83543601) and National Institute for Environmental Health Sciences (NIEHS) (award numbers: P01 ES02284401, 1U2C ES026553, R24 ES028502, R35 ES031686, and K01HL151673).

Availability of data and materials

Information on data access is available from this link: <https://allergyresearch.org.uk/studies/birth-cohort/#cohort-data-use>

Declarations

Ethics approval and consent to participate

All study protocols of the IOW cohort were approved by the local research ethics committee (NRES Committee South Central—Hampshire B, U.K.) and the University of Nebraska Medical Center Institutional Review Board. Written consents were obtained from all participants or their parents at recruitment and all follow-ups.

The institutional review boards at the Mexico National Institute of Public Health and the University of Michigan approved all research protocols of the ELEMENT cohort. Informed consent was obtained from adult participants; parental consent and participant assent was obtained during childhood study visits.

Consent for publication

All authors read and approved the final manuscript for publication.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Biostatistics, College of Public Health, University of Nebraska Medical Center, Omaha, NE 68198, USA. ²Division of Epidemiology, Biostatistics, and Environmental Health, School of Public Health, University of Memphis, Memphis, TN 38152, USA. ³Human Development and Health, Faculty of Medicine, University of Southampton, Southampton, UK. ⁴Department of Environmental Health Sciences, University of Michigan School of Public Health, Ann Arbor, MI, USA. ⁵Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor, MI, USA. ⁶Clinical and Experimental Sciences, Faculty of Medicine, University of Southampton, Southampton SO17 1BJ, UK. ⁷David Hide Asthma and Allergy Research Centre, Newport, Isle of Wight PO30 5TG, UK. ⁸College of Veterinary Medicine, Michigan State University, East Lansing, MI, USA.

Received: 11 February 2024 Accepted: 28 March 2024

Published online: 10 April 2024

References

- Acha B, Corroza J, Sanchez-Ruiz de Gordo J, Cabello C, Robles M, Mendez-Lopez I, Macias M, Zueco S, Roldan M, Urdanoz-Casado A, Jerico I, Erro ME, Alcolea D, Lleo A, Blanco-Luquin I, Mendioroz M, i, B. S. G. Association of blood-based DNA methylation markers with late-onset Alzheimer disease: a potential diagnostic approach. *Neurology*. 2023;101(23):e2434–47. <https://doi.org/10.1212/WNL.0000000000207865>.
- Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics*. 2015;31(2):166–9. <https://doi.org/10.1093/bioinformatics/btu638>.
- Anderson CM, Ralph JL, Wright ML, Linggi B, Ohm JE. DNA methylation as a biomarker for preeclampsia. *Biol Res Nurs*. 2014;16(4):409–20. <https://doi.org/10.1177/1099800413508645>.
- Arshad SH, Patil V, Mitchell F, Potter S, Zhang H, Ewart S, Mansfield L, Venter C, Holloway JW, Karmaus WJ. Cohort profile update: the Isle of Wight Whole population Birth Cohort (IOWBC). *Int J Epidemiol*. 2020;49(4):1083–4. <https://doi.org/10.1093/ije/dyaa068>.
- Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363–9. <https://doi.org/10.1093/bioinformatics/btu049>.
- Bae JB, Lipnicki DM, Han JW, Sachdev PS, Kim TH, Kwak KP, Kim BJ, Kim SG, Kim JL, Moon SW, Park JH, Ryu SH, Youn JC, Lee DY, Lee DW, Lee SB, Lee JJ, Jhoo JH, Skoog I. ... for Cohort Studies of Memory in an International, C. Parity and the risk of incident dementia: a COSMIC study. *Epidemiol Psychiatr Sci* 2020;29:e176. <https://doi.org/10.1017/S2045796020000876>
- Bodelon C, Ambatipudi S, Dugue PA, Johansson A, Sampson JN, Hicks B, Karlins E, Hutchinson A, Cuenin C, Chajes V, Southey MC, Romieu I, Giles GG, English D, Polidoro S, Assumma M, Baglietto L, Vineis P, Severi G, Garcia-Closas M. Blood DNA methylation and breast cancer risk: a

- meta-analysis of four prospective cohort studies. *Breast Cancer Res* 2019;21(1):62. <https://doi.org/10.1186/s13058-019-1145-9>
8. Campagna MP, Xavier A, Stankovich J, Maltby VE, Slee M, Yeh WZ, Kilpatrick T, Scott RJ, Butzkueven H, Lechner-Scott J, Lea RA, Jokubaitis VG. Parity is associated with long-term differences in DNA methylation at genes related to neural plasticity in multiple sclerosis. *Clin Epigenetics*. 2023;15(1):20. <https://doi.org/10.1186/s13148-023-01438-4>.
 9. Capparuocia L, Tamagnone L. Semaphorin signaling in cancer cells and in cells of the tumor microenvironment—two sides of a coin. *J Cell Sci*. 2009;122(Pt 11):1723–36. <https://doi.org/10.1242/jcs.030197>.
 10. Carrasco A. The amai system of classifying households by socio-economic level <http://www.Esomar.Org>. 2002.
 11. Carulli D, de Winter F, Verhaagen J. Semaphorins in Adult Nervous System Plasticity and Disease. *Front Synaptic Neurosci*. 2021;13:672891. <https://doi.org/10.3389/fnsyn.2021.672891>.
 12. Chen S, Mukherjee N, Janjanam VD, Arshad SH, Kurukulaaratchy RJ, Holmway JW, Zhang H, Karmaus W. Consistency and variability of DNA methylation in women during puberty, young adulthood, and pregnancy. *Genet Epigenet*. 2017;9:1179237X17721540. <https://doi.org/10.1177/1179237X17721540>.
 13. Colucci M, Cammarata S, Assini A, Croce R, Clerici F, Novello C, Mazzella L, Dagnino N, Mariani C, Tanganelli P. The number of pregnancies is a risk factor for Alzheimer's disease. *Eur J Neurol*. 2006;13(12):1374–7. <https://doi.org/10.1111/j.1468-1331.2006.01520.x>.
 14. Das J, Maitra A. Maternal DNA methylation during pregnancy: a review. *Reprod Sci*. 2021;28(10):2758–69. <https://doi.org/10.1007/s43032-020-00456-4>.
 15. de Lange AG, Kaufmann T, van der Meer D, Maglanoc LA, Alnaes D, Moberget T, Douaud G, Andreassen OA, Westlye LT. Population-based neuroimaging reveals traces of childbirth in the maternal brain. *Proc Natl Acad Sci U S A*. 2019;116(44):22341–6. <https://doi.org/10.1073/pnas.1910666116>.
 16. Dias S, Willmer T, Adam S, Pfeiffer C. The role of maternal DNA methylation in pregnancies complicated by gestational diabetes. *Front Clin Diabetes Healthc*. 2022;3:982665. <https://doi.org/10.3389/fcdhc.2022.982665>.
 17. Du P, Zhang X, Huang CC, Jafari N, Kibbe WA, Hou L, Lin SM. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587. <https://doi.org/10.1186/1471-2105-11-587>.
 18. Enquobahrie DA, Moore A, Muhie S, Tadesse MG, Lin S, Williams MA. Early pregnancy maternal blood DNA methylation in repeat pregnancies and change in gestational diabetes mellitus status—a pilot study. *Reprod Sci*. 2015;22(7):904–10. <https://doi.org/10.1177/1933719115570903>.
 19. Fradin D, Tost J, Busato F, Mille C, Lachaux F, Deleuze JF, Aptér G, Benachi A. DNA methylation dynamics during pregnancy. *Front Cell Dev Biol*. 2023;11:1185311. <https://doi.org/10.3389/fcell.2023.1185311>.
 20. Gilabert-Juan J, Saez AR, Lopez-Campos G, Sebastia-Ortega N, Gonzalez-Martinez R, Costa J, Haro JM, Callado LF, Meana JJ, Nacher J, Sanjuan J, Molto MD. Semaphorin and plexin gene expression is altered in the prefrontal cortex of schizophrenia patients with and without auditory hallucinations. *Psychiatry Res*. 2015;229(3):850–7. <https://doi.org/10.1016/j.psychres.2015.07.074>.
 21. Gleicher N. Why are reproductive cancers more common in nulliparous women? *Reprod Biomed Online*. 2013;26(5):416–9. <https://doi.org/10.1016/j.rbmo.2013.01.007>.
 22. Gruziova O, Merid SK, Chen S, Mukherjee N, Hedman AM, Almqvist C, Andolf E, Jiang Y, Kere J, Scheynius A, Soderhall C, Ullemar V, Karmaus W, Melen E, Arshad SH, Pershagen G. DNA methylation trajectories during pregnancy. *Epigenet Insights*. 2019;12:2516865719867090. <https://doi.org/10.1177/2516865719867090>.
 23. Harville EW, Mishra PP, Kahonen M, Raitoharju E, Marttila S, Raitakari O, Lehtimäki T. Reproductive history and blood cell DNA methylation later in life: the young Finns study. *Clin Epigenetics*. 2021;13(1):227. <https://doi.org/10.1186/s13148-021-01215-1>.
 24. Hinkula M, Pukkala E, Kyronen P, Kauppila A. Grand multiparity and the risk of breast cancer: population-based study in Finland. *Cancer Causes Control*. 2001;12(6):491–500. <https://doi.org/10.1023/a:1011253527605>.
 25. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, Wiencke JK, Kelsey KT. DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics*. 2012;13:86. <https://doi.org/10.1186/1471-2105-13-86>.
 26. Houseman EA, Kim S, Kelsey KT, Wiencke JK. DNA methylation in whole blood: uses and challenges. *Curr Environ Health Rep*. 2015;2(2):145–54. <https://doi.org/10.1007/s40572-015-0050-3>.
 27. Hu ZQ, Zhou SL, Zhou ZJ, Luo CB, Chen EB, Zhan H, Wang PC, Dai Z, Zhou J, Fan J, Huang XW. Overexpression of semaphorin 3A promotes tumor progression and predicts poor prognosis in hepatocellular carcinoma after curative resection. *Oncotarget*. 2016;7(32):51733–46. <https://doi.org/10.18632/oncotarget.10104>.
 28. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 2014;15(2):R31. <https://doi.org/10.1186/gb-2014-15-2-r31>.
 29. Jakobsdottir J, van der Lee SJ, Bis JC, Chouraki V, Li-Kroeger D, Yamamoto S, Grove ML, Naj A, Vronskaya M, Salazar JL, DeStefano AL, Brody JA, Smith AV, Amin N, Sims R, Ibrahim-Verbaas CA, Choi SH, Satizabal CL, Lopez OL, van Duijn CM. Rare functional variant in TM2D3 is associated with late-onset Alzheimer's disease. *PLoS Genet* 2016;12(10):e1006327. <https://doi.org/10.1371/journal.pgen.1006327>
 30. Jett S, Malviya N, Schelbaum E, Jang G, Jahan E, Clancy K, Hristov H, Pahlajani S, Niotis K, Loeb-Zeitlin S, Havryliuk Y, Isaacson R, Brinton RD, Mosconi L. Endogenous and exogenous estrogen exposures: how women's reproductive health can drive brain aging and inform Alzheimer's prevention. *Front Aging Neurosci*. 2022;14:831807. <https://doi.org/10.3389/fnagi.2022.831807>.
 31. Jiang J, Zhang F, Wan Y, Fang K, Yan ZD, Ren XL, Zhang R. Semaphorins as Potential Immune Therapeutic Targets for Cancer. *Front Oncol*. 2022;12:793805. <https://doi.org/10.3389/fonc.2022.793805>.
 32. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118–27. <https://doi.org/10.1093/biostatistics/kxj037>.
 33. Jung AY, Ahearn TU, Behrens S, Middha P, Bolla MK, Wang Q, Arndt V, Aronson KJ, Augustinsson A, Beane Freeman LE, Becher H, Brenner H, Canzian F, Carey LA, Consortium, C. T. S., Czene K, Eliassen AH, Eriksson M, Evans DG, ... Chang-Claude J. Distinct reproductive risk profiles for intrinsic-like breast cancer subtypes: pooled analysis of population-based studies. *J Natl Cancer Inst* 2022;114(12):1706–1719. <https://doi.org/10.1093/jnci/djac117>
 34. Katz TA. Potential Mechanisms underlying the protective effect of pregnancy against breast cancer: a focus on the IGF pathway. *Front Oncol*. 2016;6:228. <https://doi.org/10.3389/fonc.2016.00228>.
 35. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. *Nat Methods*. 2015;12(4):357–60. <https://doi.org/10.1038/nmeth.3317>.
 36. Kresovich JK, Xu Z, O'Brien KM, Shi M, Weinberg CR, Sandler DP, Taylor JA. Blood DNA methylation profiles improve breast cancer prediction. *Mol Oncol*. 2022;16(1):42–53. <https://doi.org/10.1002/1878-0261.13087>.
 37. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS. Reply to Suderman et al.: Importance of accounting for blood cell composition in epigenetic studies. *Proc Natl Acad Sci USA*. 2013;110(14):E1247. <https://doi.org/10.1073/pnas.1222104110>.
 38. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, Afzal U, Schulz R, Scott J, Jarvelin MR, Elliott P, McCarthy MI, Kooner JS, Chambers JC. Erratum to: A coherent approach for analysis of the illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol*. 2016;17:73. <https://doi.org/10.1186/s13059-016-0934-z>.
 39. Lehne B, Drong AW, Loh M, Zhang W, Scott WR, Tan ST, Afzal U, Scott J, Jarvelin MR, Elliott P, McCarthy MI, Kooner JS, Chambers JC. A coherent approach for analysis of the Illumina HumanMethylation450 BeadChip improves data quality and performance in epigenome-wide association studies. *Genome Biol*. 2015;16(1):37. <https://doi.org/10.1186/s13059-015-0600-x>.
 40. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing, S. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9. <https://doi.org/10.1093/bioinformatics/btp352>.
 41. Lin MW, Tsai MH, Shih CY, Tai YY, Lee CN, Lin SY. Comparison of DNA methylation changes between the gestation period and the after-delivery state: a pilot study of 10 women. *Front Nutr*. 2022;9:829915. <https://doi.org/10.3389/fnut.2022.829915>.

42. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 1988;16(3):1215. <https://doi.org/10.1093/nar/16.3.1215>.
43. Perng W, Tamayo-Ortiz M, Tang L, Sanchez BN, Cantoral A, Meeker JD, Dolinoy DC, Roberts EF, Martinez-Mier EA, Lamadrid-Figueroa H, Song PXX, Ettinger AS, Wright R, Arora M, Schnaas L, Watkins DJ, Goodrich JM, Garcia RC, Solano-Gonzalez M, ... Peterson KE. Early Life Exposure in Mexico to Environmental Toxicants (ELEMENT) project. *BMJ Open* 2019;9(8):e030427. <https://doi.org/10.1136/bmjopen-2019-030427>
44. Phipps AI, Buist DS, Malone KE, Barlow WE, Porter PL, Kerlikowske K, Li CI. Reproductive history and risk of three breast cancer subtypes defined by three biomarkers. *Cancer Causes Control.* 2011;22(3):399–405. <https://doi.org/10.1007/s10552-010-9709-0>.
45. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics.* 2016;32(2):286–8. <https://doi.org/10.1093/bioinformatics/btv560>.
46. Ray MA, Tong X, Lockett GA, Zhang H, Karmaus WJ. An efficient approach to screening epigenome-wide data. *Biomed Res Int.* 2016;2016:2615348. <https://doi.org/10.1155/2016/2615348>.
47. Schelbaum E, Loughlin L, Jett S, Zhang C, Jang G, Malviya N, Hristov H, Pahlajani S, Isaacson R, Dyke JP, Kamel H, Brinton RD, Mosconi L. Association of reproductive history with brain MRI biomarkers of dementia risk in midlife. *Neurology.* 2021;97(23):e2328–39. <https://doi.org/10.1212/WNL.00000000000012941>.
48. Suderman M, Staley JR, French R, Arathimos R, Simpkin A, Tilling K. dmrff: identifying differentially methylated regions efficiently with power and control. *bioRxiv* 2018;2018. <https://doi.org/10.1101/508556>.
49. Sun Y, Zhu J, Yang Y, Zhang Z, Zhong H, Zeng G, Zhou D, Nowakowski RS, Long J, Wu C, Wu L. Identification of candidate DNA methylation biomarkers related to Alzheimer's disease risk by integrating genome and blood methylome data. *Transl Psychiatry.* 2023;13(1):387. <https://doi.org/10.1038/s41398-023-02695-w>.
50. Tang Q, Cheng J, Cao X, Surowy H, Burwinkel B. Blood-based DNA methylation as biomarker for breast cancer: a systematic review. *Clin Epigenetics.* 2016;8:115. <https://doi.org/10.1186/s13148-016-0282-6>.
51. Tang Q, Holland-Letz T, Slynko A, Cuk K, Marme F, Schott S, Heil J, Qu B, Golatta M, Bewerunge-Hudler M, Sutter C, Surowy H, Wappenschmidt B, Schmutzler R, Hoth M, Bugert P, Bartram CR, Sohn C, Schneeweiss A, Burwinkel B. DNA methylation array analysis identifies breast cancer associated RPTOR, MGRN1 and RAPSN hypomethylation in peripheral blood DNA. *Oncotarget* 2016;7(39):64191–64202. <https://doi.org/10.18632/oncotarget.11640>
52. Troisi R, Bjorge T, Gissler M, Grotmol T, Kitahara CM, Myrtevit Saether SM, Ording AG, Skold C, Sorensen HT, Trabert B, Glimelius I. The role of pregnancy, perinatal factors and hormones in maternal cancer risk: a review of the evidence. *J Intern Med.* 2018;283(5):430–45. <https://doi.org/10.1111/joim.12747>.
53. Wallerius M, Wallmann T, Bartish M, Ostling J, Mezheyeuski A, Tobin NP, Nygren E, Pangigadde P, Pellegrini P, Squadrito ML, Ponten F, Hartman J, Bergh J, De Milito A, De Palma M, Ostman A, Andersson J, Rolny C. Guidance Molecule SEMA3A Restricts Tumor Growth by Differentially Regulating the Proliferation of Tumor-Associated Macrophages. *Cancer Res.* 2016;76(11):3166–78. <https://doi.org/10.1158/0008-5472.CAN-15-2596>.
54. Wang T, Li P, Qi Q, Zhang S, Xie Y, Wang J, Liu S, Ma S, Li S, Gong T, Xu H, Xiong M, Li G, You C, Luo Z, Li J, Du L, Wang C. A multiplex blood-based assay targeting DNA methylation in PBMCs enables early detection of breast cancer. *Nat Commun.* 2023;14(1):4724. <https://doi.org/10.1038/s41467-023-40389-5>.
55. Wang Z, Chen J, Zhang W, Zheng Y, Wang Z, Liu L, Wu H, Ye J, Zhang W, Qi B, Wu Y, Song X. Axon guidance molecule semaphorin3A is a novel tumor suppressor in head and neck squamous cell carcinoma. *Oncotarget.* 2016;7(5):6048–62. <https://doi.org/10.18632/oncotarget.6831>.
56. Xing Z, Alman AC, Kirby RS. Parity and risk of cardiovascular disease in women over 45 years in the United States: national health and nutrition examination survey 2007–2018. *J Womens Health (Larchmt).* 2022;31(10):1459–66. <https://doi.org/10.1089/jwh.2021.0650>.
57. Xu Z, Sandler DP, Taylor JA. Blood DNA methylation and breast cancer: a prospective case-cohort analysis in the sister study. *J Natl Cancer Inst.* 2020;112(1):87–94. <https://doi.org/10.1093/jnci/djz065>.
58. Yeung KR, Chiu CL, Pidsley R, Makris A, Hennessy A, Lind JM. DNA methylation profiles in preeclampsia and healthy control placentas. *Am J Physiol Heart Circ Physiol.* 2016;310(10):H1295–1303. <https://doi.org/10.1152/ajpheart.00958.2015>.
59. Zhang X, Klamer B, Li J, Fernandez S, Li L. A pan-cancer study of class-3 semaphorins as therapeutic targets in cancer. *BMC Med Genomics.* 2020;13(Suppl 5):45. <https://doi.org/10.1186/s12920-020-0682-5>.