BMJ Oncology

Associations of DNA methylation in breast tumour subtypes with parity and breastfeeding in a cohort of 1459 Black women: implications for public health

Christine B Ambrosone ,[1] Song Yao,[1] Mark D Long,[2] Chunyu Liu,[3] Jianhong Chen,[1] Warren Davis,[1] Gary Zirpoli,[4] Rochelle Payne-Ondracek,[1] Thaer Khoury,[5] Zhihong Gong,[1] Qiang Hu,[2] Sirinapa Szewczyk,[6] Angela R Omilian,[1] Elisa V Bandera,[7] Song Liu,[2] Lawrence Kushi,[8] Michael J Higgins,[6] Julie R Palmer[9]

To cite: Ambrosone CB, Yao S, Long MD, *et al*. Associations of DNA methylation in breast tumour subtypes with parity and breastfeeding in a cohort of 1459 Black women: implications for public health. *BMJ Oncology* 2025;**4**:e000675. doi:10.1136/bmjonc-2024-000675

MJH and JRP are joint senior authors.

Received 22 November 2024
Accepted 17 February 2025

For numbered affiliations see end of article.

Correspondence to
Dr Christine B Ambrosone; christine.ambrosone@roswellpark.org

ABSTRACT

Objective Having children reduces risk of breast cancer overall, but parity without breastfeeding, more prevalent among black women, increases risk of poor-prognosis oestrogen receptor negative (ER–) breast cancer. We investigated if relationships between parity, breastfeeding and ER subtypes result from epigenetic programming, potentially steering breast progenitor cells to a basal-like phenotype.

Methods and analysis The Illumina MethylationEPIC platform was used to assess genome-wide methylation in formalin-fixed, paraffin-embedded tumours from 1459 Black women with breast cancer. Methylation was evaluated in relation to parity, breastfeeding and breast cancer subtypes in a case-only analysis, with methylation-gene expression pairs tested in a subset of cases. We then performed functional enrichment analysis for probes significantly associated with parity and breastfeeding.

Results Among women who did not breastfeed (n=634), there were 500 significant (p<1e–5) differentially methylated loci (DML) by parity, compared with only five DMLs among women who had breastfed their children (n=568). One of the top DML genes was *FOXA1*, pivotal in governing the luminal lineage of progenitor cells, with a statistically significant interaction (p=0.04) for number of births and breastfeeding. Associations were strongest for ER– disease.

Conclusion In this large study of Black women with breast cancer, we elucidated biological pathways for the observed associations between parity without breastfeeding and breast cancer subtypes, revealing distinct molecular alterations in breast DNA, particularly for ER– tumours. Black women in the USA tend to have more children and are less likely to breastfeed; their breast cancer risk may be reduced by societal systems that promote and support breastfeeding.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Except for suggestive evidence from one small pilot study, it is unknown why women who have children and do not breastfeed are more likely to develop oestrogen receptor (ER) negative (ER–) breast cancer. We formed a consortium and conducted the largest genome-wide study of DNA methylation to date (n=1459 Black women with breast cancer) to determine if methylation is the contributing mechanism and to identify the pathways involved, which could have implications for prevention.

WHAT THIS STUDY ADDS

⇒ in the largest epigenome-wide case-only study of breast tumor DNA, to date, with an inquiry of more than 935 000 CpGs (*5′—C—phosphate—G—3*), regions of DNA where a cytosine nucleotide is followed by a guanine nucleotide along its 5' → 3' direction, this study clearly showed that there are extensive DNA methylation alterations in breast tumours from Black women who do not breastfeed. Results clearly showed that methylation and subsequent suppression of gene expression of loci in a gene pivotal to repressing the basal-like phenotype in breast progenitor cells, FOXA1, is potentially responsible for driving development of ER– breast cancer.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ These findings, showing the molecular basis for the development of ER– breast cancer in Black women, should drive dissemination to practitioners and to pregnant Black women that breastfeeding their children could greatly reduce their risk of more deadly breast cancer subtypes. Policies for employers to provide support for lactating mothers should be encouraged.

OBJECTIVES

Breast cancer death rates are 40% higher for women in the USA who self-identify as Black/African-American and are twofold higher for Black women under age 50.[1] Although social and structural drivers of health likely play an important role in these disparities,Black women are also more likely to have tumours

with poor prognosis phenotypes, particularly those that are oestrogen receptor-negative (ER–) or triple negative breast cancers (TNBC), for which there are few therapeutic targets.[2] Until recently, the underlying causes of the disproportionately high incidence of ER– and TNBC in Black women have been unclear.

Pooling data from the four largest studies of breast cancer in US Black women, theAfrican American Breast Cancer Epidemiology and Risk (AMBER) Consortium showed that, while having children was associated with decreased risk of ER+ breast cancer, parity actually increased risk of ER– disease, with risk greatest among women who did not breastfeed,[3] findings also observed among White women.[4] It is thus likely that higher rates of parity among US Black women and lower prevalence of breastfeeding may contribute to their higher risk of more deadly ER– and TNBC, although limited data indicate that these subtypes are also more prevalent in Black women in West Africa, where breastfeeding is more common.[5] Nonetheless, a clear understanding of the molecular pathways linking these reproductive factors with more aggressive subtypes remains to be elucidated.

One possible mechanism for these associations is through epigenetic reprogramming of breast epithelial cells, driving progenitor cells to a basal-like phenotype. In a pilot study with 383 Black and 350 White women with breast cancer, we previously identified a *FOXA1* promoter proximal CpG as one of the top differentially methylated loci (DML) by race and by ER status.[6] FOXA1 is a pioneer transcription factor regulating the expression of a multitude of downstream genes, including *ESR1,* and is pivotal in governing the luminal lineage of progenitor cells by repressing the basal-like phenotype. Thus, aberrant hypermethylation and reduced expression of *FOXA1* could allow for an increase in undifferentiated luminal cells which, with subsequent genomic alterations, may lead to basal-like breast cancer. Our exploratory analysis suggested that *FOXA1* methylation was positively associated with parity, particularly among women who did not breastfeed.

We conducted the largest epigenome-wide study of breast cancer in Black women, to date, to examine how these reproductive factors affect DNA methylation such that the progenitor cells could progress to a basal-like phenotype.

METHODS AND ANALYSIS

Study population for EPIC array

Patient and public involvement

Breast tumour DNA and detailed risk factor data were obtained from three epidemiological studies: the Women's Circle of Health Study (WCHS)[7] and Follow-up Study (WCHFS),[8] the Black Women's Health Study (BWHS)[9] and the Pathways Study,[10] from which all participants provided informed consent. This was an observational population-based study, not clinical research. Other than participating in the study, patients were not involved in the research, and none of the authors were involved in their clinical care.

WCHS was first established as a hospital-based study in metropolitan New York City and later expanded to 10 counties in Eastern New Jersey (NJ) and as a follow-up study (WCHFS).[8] African-American/Black women with primary invasive breast cancer, aged 20–75 years, were identified through the NJ State Cancer Registry. In-person interviews were conducted for known and putative breast cancer risk factors. As part of the informed consent process, >95% of participants signed a release for their pathology reports and archived tumour tissues to be provided for use in research. Formalin-fixed, paraffin-embedded (FFPE) tumour blocks (n=655) were obtained from the pathology departments of the treating hospitals.[6]

BWHS is an ongoing cohort study of 59 000 self-identified US Black women, age 21–69 years, who were enrolled in 1995 by completing a detailed health questionnaire, with biennial follow-up. Incident breast cancers are ascertained by self-report, by linkage with state cancer registries of 24 states in which 95% of participants reside and by linkage with the National Death Index. Pathology reports or state cancer registry data were obtained for more than 90% of cases; 54% of cases signed releases for their tumour samples, which were retrieved from 75% of those participants (n=578).[11]

The Pathways Study is a prospective cohort of 4504 women with newly diagnosed incident invasive breast cancer who were members of Kaiser Permanente Northern California[10] at enrolment in 2006–2013. In-person interviews were conducted, and participants signed releases for access to their medical records and clinical tumour specimens, which were retrieved from 227 women who self-reported as African-American/Black.

Tumour tissue specimens

The Roswell Park Biorepository and Laboratory Services Shared Resource received and centrally processed FFPE blocks for all studies and tumour characteristics were abstracted from pathology reports. Sections were cut for H&E staining and scanned into the Aperio ScanScope XT (Aperio Technologies, Vista, California, USA) for viewing by the study pathologist (TK) who circled areas containing high purity of tumour cells for punching of cores for DNA extraction. For hospitals that would not release blocks, unstained full section curls were requested to be used as a source of DNA.

DNA extraction, quality assessment and methylation analysis

DNA was extracted from FFPE using a Qiagen AllPrep DNA/RNA FFPE kit.[12] Yield was quantified using Qubit and fragmentation analysis by Agilent Bioanalyzer and samples were randomised according to study, age at diagnosis, tumour ER status, parity (yes/no) and breastfeeding (yes/no) before shipment to the Center for Inherited Disease Research at Johns Hopkins University where the methylation analysis was performed.

Briefly, DNA samples were treated with sodium bisulfite using an Illumina-recommended bisulfite conversion kit (Zymo EZ DNA Methylation Kit, Zymo Research, Irvine, California, USA), isothermally amplified and then fragmented, precipitated, resuspended and hybridised to Infinium MethylationEPIC BeadChip (V.1.0). The chips were incubated overnight for allele-specific hybridisation and underwent single base extension and staining, with scanning with the Illumina iScan system. The raw image data were processed and analysed using Illumina's GenomeStudio V.2011.1. Preliminary Quality Control was conducted using minfi V.1.34.0 (https://bioconductor.org/packages/release/bioc/html/minfi.html) and ewastools V.1.6 (https://github.com/hhhh5/ewastools) in R V.4.0. The function DetectionP from minfi identified failed positions, using a default detection p value cut-off of 0.01. Ewastools was applied for quality assessment metrics; 95.4% of samples and 100% of methylation control samples passed QC and were released for final analysis. Pre-processing of raw image files was performed using the ChAMP package.[13] After removing 142 370 CpGs that were mapped to multiple genomic locations or single nucleotide polymorphisms (SNPs), or SNPs within 10 bp, 716 606 CpGs remained for analysis. Low-quality probes (detection p value<0.01) were imputed. To adjust for probe design bias (Infinium type-I, type-II), a beta-mixture quantile normalisation method was employed.[14] Reference-based estimation of cell types was performed using EpiDISH[15] with probes previously identified as specific for nine breast tissue resident cell types as reference.[16]

Statistical analysis for epigenome-wide association study

The final sample included genomic data from 1459 tumours which was deposited in dbGaP (Database of Genotypes and Phenotypes) and is available through a controlled access portal.[17] Online supplemental table 1 shows comparisons of the sample to the larger cohorts of WCH, BWHS and Pathways Study. We performed an epigenome-wide association study (EWAS) with linear regression models to relate the log-transformed DNA methylation beta-value of each CpG site with number of births, modelled as an ordinal variable (0, 1, 2, 3 for nulliparous, one, two and three or more births, respectively) with a priori adjustment for study site, age at diagnosis and ER status. Nulliparous cases were included to consider both associations of parity/nulliparity and associations with increasing number of births. EWAS was also conducted to relate methylation status with breastfeeding, modelled as an ordinal variable with four categories: nulliparous, parous with no breastfeeding, <6 months breastfeeding and ≥6 months breastfeeding. Differential associations by breastfeeding were formally tested by including an interaction term (parity×breastfeeding) in the same model along with the main effects of parity and breastfeeding. Epigenome-wide significance level was defined as p≤7e–8 after Bonferroni correction for multiple testing of 716 606 CpGs. The analysis was conducted with all cases combined, as well as stratified by ER status. Global genetic ancestry was estimated using Admixture, based on dense germline genotype data already available for all study participants,[18] for consideration of association with CpGs.

Expression quantitative trait methylation analysis

To reveal relationships between DNA methylation and gene expression, expression quantitative trait methylation (eQTM) analysis was performed using a subset of 408 samples (183 ER−; 225 ER+) for whom RNA sequencing (RNA-seq) profiling was previously performed by the Roswell Park Genomics Shared Resource,[19] with libraries generated using the SureSelect XT RNA Direct Kit (Agilent). RNA-seq counts were normalised using the median ratio method implemented by DESeq2.[20] Variance stabilisation transformed normalised counts (RNA-seq) and M-values (EPIC) were used for correlative analyses. To identify cis-eQTMs (ie, CpGs regulating transcription of neighbouring genes), we considered only methylation probes within 0.5 Mb from the transcribed region of a gene (transcription start site (TSS) to transcription end site (TES)) and tested 10 971 588 methylation-gene expression pairs by Pearson correlation analysis.

Functional enrichment analysis was performed using the enrichR[21] package. Probes significantly associated with parity and breastfeeding were filtered for those found to correlate with proximal gene expression (abs(r)>0.4) and putatively regulated genes compiled for function enrichment analysis. Query pathway sets included Reactome, MSigDB Hallmark and KEGG pathways.

RESULTS

Study population

Characteristics of the study population (n=1459) are shown in table 1, according to ER status. Women with ER− breast cancer (29%) were slightly more likely to be diagnosed before age 50 years than those with ER+ tumours (33% vs 28%) and were more likely to be parous and not to have breastfed.

Global tumour DNA methylation patterns by parity and breastfeeding

Figure 1a shows associations of DNA methylation with number of births (0, 1, 2, 3+) according to breastfeeding status. More CpGs across the epigenome showed differential methylation status in parous women without breastfeeding compared with women who breastfed. The absolute changes in methylation beta value (Δbeta) per birth were significantly larger in the no breastfeeding group than among those who breastfed (median (IQR): 0.012 (0.005–0.025) vs 0.008 (0.003–0.018), p<2.2e–16 figure 1b).

Table 1 Characteristics of the study population with methylation results and epidemiological data by oestrogen receptor (ER) status (n=1459 Black women with breast cancer)

	ER– breast cancer n=417 (29%)				ER+ breast cancer n=1042 (71%)			
	WCHS (n=169)	BWHS (n=173)	Pathways (n=75)	All ER–	WCHS (n=486)	BWHS (n=405)	Pathways (n=151)	All ER+
Age at diagnosis								
<50	63 (37)	50 (29)	24 (32)	**137 (33)**	156 (32)	104 (26)	34 (23)	**294 (28)**
≥50	106 (63)	123 (71)	51 (68)	**280 (67)**	330 (68)	301 (74)	117 (77)	**748 (72)**
Parity								
Nulliparous	21 (12)	30 (17)	8 (11)	**59 (14)**	84 (17)	92 (23)	19 (13)	**195 (19)**
Parous	148 (88)	143 (83)	67 (89)	**358 (86)**	402 (83)	313 (77)	132 (87)	**847 (81)**
Number of births (n=1205)								
1	39 (26)	43 (30)	16 (24)	**98 (27)**	111 (23)	116 (29)	36 (23)	**263 (31)**
2	57 (39)	61 (43)	17 (25)	**135 (38)**	129 (27)	110 (27)	40 (30)	**279 (33)**
≥3	52 (35)	39 (27)	34 (51)	**125 (35)**	162 (33)	87 (21)	56 (42)	**305 (36)**
Breastfeeding among parous women								
Ever	58 (39)	64 (45)	40 (61)	**162 (46)**	168 (42)	160 (51)	78 (60)	**406 (48)**
Never	90 (61)	78 (55)	26 (39)	**194 (54)**	234 (58)	153 (49)	53 (40)	**440 (52)**
Unknown	0	1	1	**2**	0	0	1	**1**
Breastfeeding duration								
<6 months	20 (36)	33 (54)	15 (38)	**68 (44)**	56 (35)	69 (45)	32 (41)	**157 (40)**
≥6 months	35 (64)	28 (46)	25 (63)	**88 (56)**	103 (65)	85 (55)	46 (59)	**234 (60)**
Unknown	3	3	0	**6**	9	6	0	**15**

BWHS, Black Women's Health Study; WCHS, Women's Circle of Health Study.

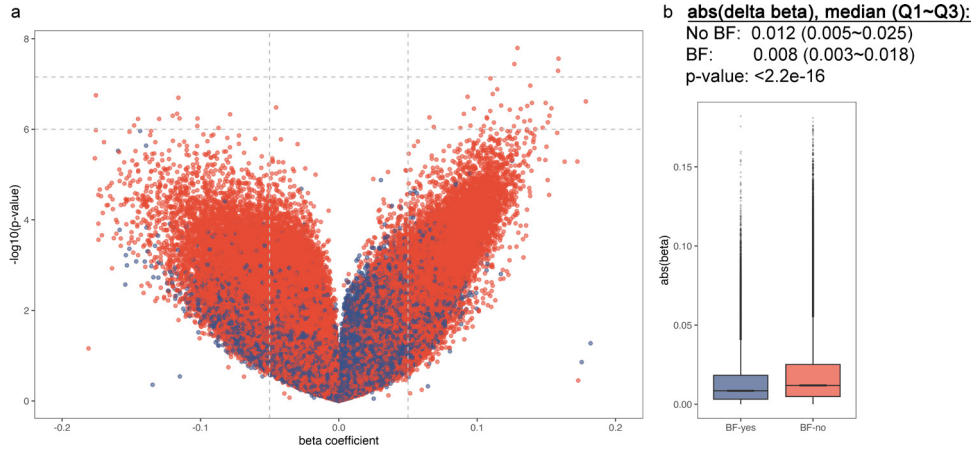

Figure 1 Global breast tumour DNA methylation by parity and breastfeeding. Individual CpG in association with the number of births (0=nulliparous, 1=1 birth, 2=2 births, 3=3 or more births) was analysed as an ordinal variable using a linear model with adjustment for study, age at diagnosis and tumour oestrogen receptor status. The analyses were performed in women with and without breastfeeding separately, and in each scenario, nulliparous women (number of births=0) were included. (a) Volcano plot of –log10-transformed p values (y-axis) against regression coefficients (x-axis, change in methylation beta value, or Δbeta, per birth) for those derived from the models of women with breastfeeding (blue) and without breastfeeding (red). The two horizontal dashed lines denote the p value threshold of 0.05 and 7e–8 after Bonferroni correction. The two vertical dashed lines denote regression coefficients of –0.1 and 0.1. (b) Boxplots of the absolute values of regression coefficients (y-axis) by women with and without breastfeeding. The median with first (Q1) and third (Q3) quartiles of the absolute values of the regression coefficients are shown, with p values between the two groups derived from the t-test.

Abbreviations: BF breastfeeding; CpG, 5'—C—phosphate—G—3

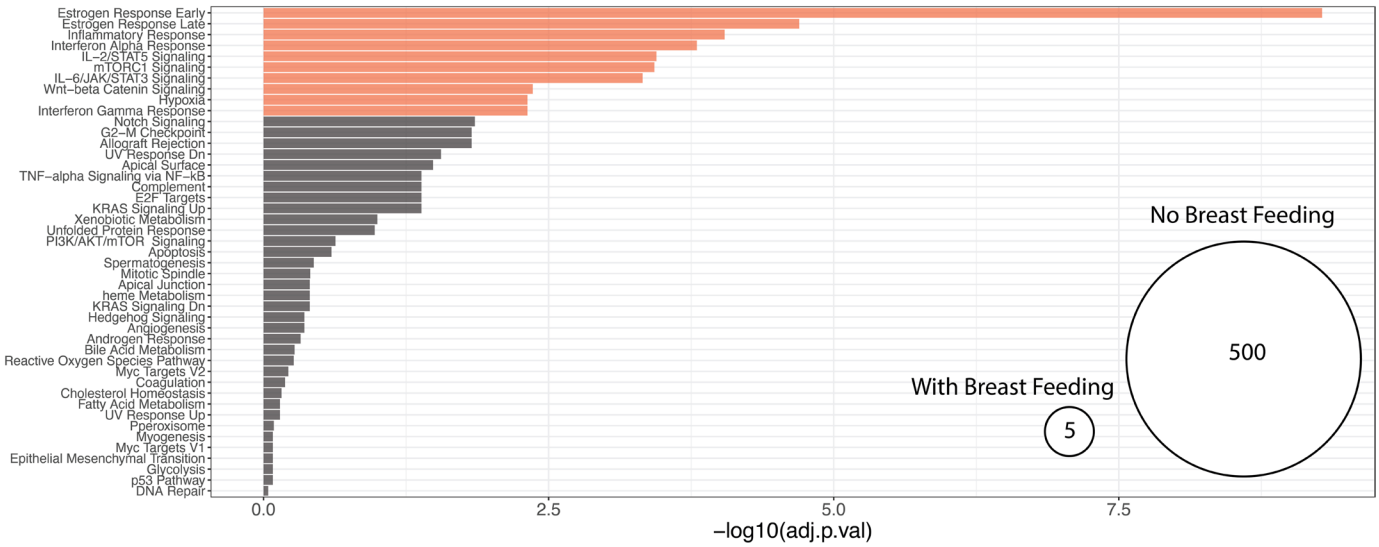

Figure 2 Functional pathway enrichment of DNA methylation changes associated with parity in women who did not breastfeed. Based on a p value threshold of 1e–5, the Venn diagram shows that there were 500 CpGs in significant association with the number of births in women without breastfeeding (blue), and only five significant CpGs in women with breastfeeding (red), whereas no overlapping CpGs were found between the two analyses. In expression quantitative trait methylation analysis based on matched DNA methylation data and transcriptomic data from 408 of the tumours, a total of 293 top genes were identified with a p value threshold of 1e–5 and r>0.4 or r<–0.4. These 293 genes were then subject to functional pathway enrichment analysis using the Broad HALLMARK pathway set. Significant pathways (adjusted p<0.05) are highlighted in pink, and overlapping genes identified from each pathway noted.

Abbreviations: CpG, 5'—C—phosphate—G—3

Functional enrichment analysis of molecular pathways impacted by DNA methylation and parity

Based on a p value threshold of 1e–5, there were 500 CpGs associated with the number of births in the no breastfeeding group and only five CpGs in the breastfeeding group, with no overlap between the two sets (figure 2). Moderate correlation (r>0.4 or r<–0.4) between methylation at a given CpG and the expression of a gene within 0.5 Mb was observed for 293 of the 500 CpGs that were differentially methylated between nulliparous and parous women who had not breastfed. Analysis of these genes revealed significant (adjusted p<0.05) enrichment of several molecular pathways including those for both early and late oestrogen response, inflammatory response and several immune-related pathways (interferon alpha and gamma response), as well as several signalling axes such as IL-2/STAT, mTORC1, IL-6/JAK/STAT3, and Wnt-beta pathways.

Specific DML in relation to parity and breastfeeding

Manhattan plots (figure 3) show relationships of epigenome-wide methylation with number of births and breastfeeding duration, and table 2 details the top five significant CpGs. In all cases combined, methylation of cg27648238 in *ENPP2* on chromosome 18 was associated with the number of births at the epigenome-wide significance level (Δbeta per birth=0.059, p=3.3e–08, figure 3a, table 2), with no differences according to breastfeeding status. There were no significant CpGs associated with breastfeeding duration.

When stratified by breastfeeding, there were no significant associations of parity with methylation among women who breastfed. However, among women who did not breastfeed, four CpGs were significantly associated with the number of births (figure 3b, table 2). These included the intergenic cg24322780 (Δbeta per birth=0.129, p=1.6e–08), cg08919180 in *TNFSF4* (Δbeta per birth=0.127, p=3.6e–08), cg05387815 in *ZNF578* (Δbeta per birth=0.158, p=3.1e–08) and cg00955911 in *FOXA1* (Δbeta per birth=0.159, p=2.8e–08) (table 2). Among women who did not breastfeed, there was a linear increase in methylation level with increasing number of births for all four CpGs (online supplemental figure 1). In the breastfeeding group, associations were weaker and significant only at a nominal level (p<0.05). For potential interactions, we tested the four CpGs that were significantly associated with the number of births among women who did not breastfeed. Of those four CpGs, the only significant interaction was for CpG cg00955911 in *FOXA1*, with Δbeta=0.159 in the group who did not breastfeed versus Δbeta=0.078 in those who did breastfeed (p interaction=0.04). The addition of cell composition to the regression models did not alter results.

As shown in online supplemental table 2, although all five epigenome-wide significant CpGs associated with number of births showed similar associations when stratified by tumour ER status, for three of the four CpGs, the Δbetas were 50–90% higher in ER– than ER+ cases.

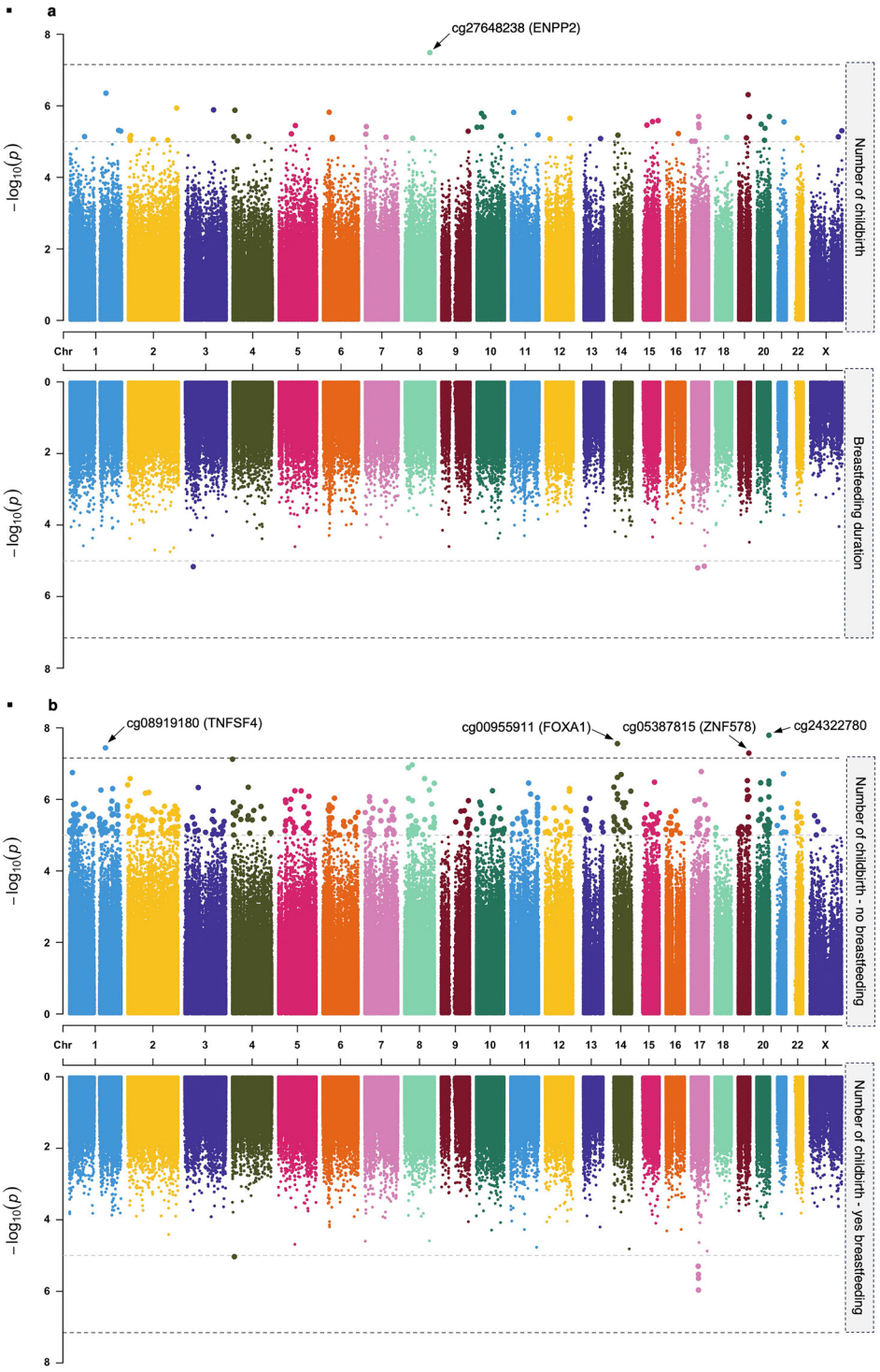

Figure 3 Epigenome-wide significant CpGs associated with number of births according to breastfeeding status. (a) Mirrored Manhattan plots of epigenome-wide associations of breast tumour DNA methylation with number of births (upper panel) and with duration of breastfeeding (lower panel) are shown. Number of births (0=nulliparous, 1=1 birth, 2=2 births, 3=3 or more births) and duration of breastfeeding ((0=nulliparous, 1=parous without breastfeeding, 2=parous and <6 months breastfeeding, 3=parous and with 6 or more months of breastfeeding) were each tested as an ordinal variable in linear regression models with adjustment for study, age at diagnosis and tumour oestrogen receptor (ER) status. Epigenome-wide significance was determined by p≤7e−8 with Bonferroni correction for 715 k CpGs tested. (b) Mirrored Manhattan plots of epigenome-wide associations of breast tumour DNA methylation with number of births (0=nulliparous, 1=1 birth, 2=2 births, 3=3 or more births) analysed as an ordinal variable in linear regression models with adjustment for study, age at diagnosis and tumour ER status in women without breastfeeding (upper panel) and women with breastfeeding (lower panel). For all plots, the two dashed horizontal lines indicate p value thresholds of 1e−5 (the cut-off used to select CpGs and genes for pathways enrichment analysis) and 7e−8, with gene symbols linked to the CpGs significant at 7e−8 annotated.
Abbreviations: CpG, 5'—C—phosphate—G—3

Table 2 Epigenome-wide significant CpGs associated with number of births according to breastfeeding status

				All cases		No breastfeeding group		Breastfeeding group		P interaction (number of births × breastfeeding)
CpG ID	chr	Position	Gene	Δbeta	P value	Δbeta	P value	Δbeta	P value	
cg27648238	8	119673360	ENPP2	0.059	3.26E−08	0.066	5.46E−07	0.056	2.57E−05	0.76
cg24322780	20	59412933		0.089	1.99E−06	0.129	1.61E−08	0.067	3.09E−03	0.28
cg08919180	1	173205510	TNFSF4	0.096	4.42E−07	0.127	3.62E−08	0.077	5.95E−04	0.14
cg05387815	19	52492997	ZNF578	0.113	2.01E−06	0.158	5.11E−08	0.075	9.80E−03	0.49
cg00955911	14	37592782	FOXA1	0.104	6.62E−06	0.159	2.75E−08	0.078	5.71E−03	0.04

Number of births (0, 1, 2, 3+) was tested as an ordinal variable in linear regression models with adjustment for study, age at diagnosis and tumour oestrogen receptor status. Epigenome-wide significance was determined by p≤7e−8 with Bonferroni correction for 715k CpGs tested. Abbreviations: CpG, 5'—C—phosphate—G—3

Identification of top CpG associations across studies

For validation purposes, we repeated analyses of the five significant CpGs across the three studies, combining data from BWHS and the Pathways Study (the two smaller studies) and comparing to results from WCHS. As shown in online supplemental table 3, results were consistent across these two data sets.

African ancestry in relation to top DMLs

The median estimated proportion of African ancestry in the study sample was 0.82 (range: 0.30–1.00, data not shown). Per cent African ancestry was not correlated with any of the top five CpGs (strongest r=0.06, p=0.03 with cg08919180), indicating a negligible impact of ancestral admixture on the detected associations between parity and DNA methylation.

Genomic context of top CpGs, gene expression and breast cancer subtypes

Focusing on the top DMLs by parity as shown in table 2, we performed eQTM analysis for relationships between DNA methylation and gene expression. The CpG cg27648238 resides in an evolutionarily conserved region, yet the area is devoid of known epigenetic regulatory elements (online supplemental figure 2). Although the nearest gene is *ENPP2,* there was no correlation with messenger RNA (mRNA) expression of *ENPP2*. Instead, the methylation levels of cg27648238 were most significantly and inversely correlated with expression of *COL14A1* (r=−0.26, adjusted p=2.3e−6). This gene is further downstream from cg27648238 and encodes for a minor isoform of collagen of the extracellular matrix, involved in the regulation of the tumour microenvironment (TME). mRNA expression of *COL14A1* was lower in ER− than in ER+ tumours (online supplemental figure 3).

cg24322780 is in an intergenic region on chromosome 20; several nearby CpGs in this region were also differentially methylated according to parity, with slightly higher p values (online supplemental figure 2). cg24322780 is 112kb downstream of the nearest known protein gene, EDN3. No prominent epigenetic regulatory elements are known to be in this region. eQTM analysis identified several genes whose expression was positively correlated with the CpG methylation levels, including *ZNF831* (r=0.36, adjusted p=4.3e−12) which is another 50kb upstream. mRNA expression of *ZNF831* was higher in ER− than in ER+ tumours (online supplemental figure 3). cg08919180 resides in the first exon of *TNFSF4* on chromosome 1 and is the only significant CpG at this locus (online supplemental figure 2). In silico functional annotation identified this region as a DNase hypersensitive site with potential interactions with regulatory elements. However, eQTM analysis showed no correlation of cg08919180 with expression of *TNFSF4* (r=−0.08, adjusted p=0.23); instead, the most significant correlation was with *ANKRD45,* which is 464kb downstream (r=0.26, adjusted p=1.5e−6) and plays a role in cell division and proliferation. mRNA expression levels for *ANKRD45* were higher in ER− than in ER+ tumours (online supplemental figure 3).

cg05387815 is in the third intron of *ZNF578* in a CpG island on chromosome 19, and the locus was identified as an enhancer in functional annotation (online supplemental figure 2). eQTM analysis showed no correlation of cg05387815 methylation with mRNA expression of *ZNF578* (r=0.05, adjusted p=0.56) but the most significant eQTM association was with another zinc finger transcription family member *ZNF83* (r=−0.26, adjusted p=1.6e−6), which is approximately 120kb 3' downstream. ER− breast cancers had lower levels of *ZNF83* than ER+ tumours (online supplemental figure 3).

Parity, breastfeeding and *FOXA1* methylation and expression

In EWAS, five CpGs in *FOXA1* were differentially methylated according to parity among women who did not breastfeed. One of the top five identified, cg00955911 is in the first intron of *FOXA1* on chromosome 14. This is in the same CpG island and approximately 1.1kb upstream of the top *FOXA1* CpG identified in our previous study, cg04932551,[6] which is also the second most significant CpG at this locus in the current study (Δbeta per birth=0.179, p=2.4e−07). In addition, a third cg03772350 in the same CpG island and two other CpGs (cg16539957, cg19578835) in a nearby CpG island in the 5' direction all emerged at a suggestive significance level (p<1e−5)

(online supplemental figure 2). In eQTM analysis, the methylation level of cg00955911 was in strong inverse correlation with *FOXA1* mRNA expression (r=−0.75, adjusted p<2.2e−16); levels were markedly lower in ER− tumours (online supplemental figure 3).

Because of the key role of FOXA1 in regulating oestrogen response elements, we analysed the quantitative breast tumour mRNA expression of *ESR1* in a subset of samples in relation to *FOXA1* methylation and ER status. There was a strong positive correlation between mRNA levels of *FOXA1* and *ESR1*, consistent between ER+ and ER− tumours (r=0.79, p<2.2e−16; online supplemental figure 4a). Methylation of cg00955911 in *FOXA1* was inversely correlated with ESR1 mRNA levels (r=−0.66, p<2.2e−16; online supplemental figure 4b), and 95% of all ER+ tumours were hypomethylated based on a beta-value <0.25. Moreover, there was a trend of decreasing mRNA expression of FOXA1 and ESR1 with increasing number of births in women who had not breastfed, but not in women who breastfed (online supplemental figure 4c,d), consistent with the methylation findings.

As noted in the methods, we estimated cell types using EpiDISH, with DNA methylation probes previously identified as specific for nine breast tissue resident cell types used as reference and compared distributions of cell type estimates across groups. There were no significant differences in cell types according to parity and breastfeeding (online supplemental figure 5).

CONCLUSIONS

In this study of breast cancer in 1459 Black women, results clearly demonstrate that having children causes extensive methylation changes in breast tissue, but that breastfeeding reduces these changes, likely restoring the breast DNA methylome to its pre-pregnant state. Our earlier research with a smaller sample set indicated that these effects were observed primarily in tumours that were ER−, but in this large study, we observed associations with parity and breastfeeding in both ER− and ER+ tumours, although associations were stronger for ER− disease.

Results from functional pathway enrichment of the 500 CpGs that were differentially methylated according to parity in women who had not breastfed showed that oestrogen and inflammatory and immune (interferon α, interleukin-2) responses were the top pathways involved in these pregnancy-related changes. These results confirm the central role of oestrogen signalling in breast tumourigenesis, even in tumours that are ER−. Results also support the involvement of inflammation and wound healing processes in breast tumours from parous women who do not breastfeed.

Pregnancy and subsequent birth and lactation bring profound physiological changes to the mammary gland. In pregnancy, the breast undergoes massive proliferation and differentiation from the prepregnant ductal state to the milk-producing lobuloalveolar state.[22–24] In women who breastfeed, the mammary gland goes through orderly and gradual involution during and after weaning, characterised by non-inflammatory tissue remodelling and return to the prepregnant state. In women who do not breastfeed, however, sudden cessation of lactation results in rapid tissue remodelling accompanied by inflammation and wound healing-like processes, shown to lead to a temporary increased risk of often aggressive postpartum breast cancer.[24] Although most breast cancers are not considered post partum, there may exist a long-term memory effect in the mammary tissue shaped by potent hormonal exposure and tissue remodelling that may influence a woman's later risk of breast cancer.

One likely mechanism for this memory effect is epigenetic reprogramming, which has been shown in animal studies to be fundamental to most differentiation processes in the mammary gland.[25–28] We had previously hypothesised that parity changes the methylation state of luminal progenitor cells, believed to be the cell-of-origin for both ER+ and ER− breast tumours and shifts the differentiation trajectory of these cells from luminal to 'basal-like' lineage, which eventually leads to the development of ER− breast cancer.[29] Breastfeeding following birth may reduce that increased risk of ER− breast cancer by counterbalancing the methylation effect of parity on the luminal progenitor cells; this protection may be lost with lack of or very short duration of breastfeeding.

In epigenome-wide analyses of DNA methylation in relation to parity and breastfeeding, there were numerous loci that were differentially methylated, although few were statistically significant at the genome-wide level. For the top five CpGs that were significant in relation to parity, with linear relationships with number of births, it was clear that the absolute changes in methylation beta values were greatest among women who did not breastfeed.

Of those DMLs, one of the top five was a CpG in *FOXA1*, with the association observed primarily in women who did not breastfeed. FOXA1 is a pioneer transcription factor pivotal in governing the luminal lineage of progenitor cells. It regulates the expression of a multitude of downstream genes, including *ESR1*, that favour the luminal over the basal phenotype of progenitor cells. Indeed, in a mechanistic mouse model study, we found that heterozygous deletion of *Foxa1* resulted in a 125% increase in the proportion of luminal progenitor cells in the mammary gland, with a concomitant decrease in the fraction of mature luminal cells (ER+).[30] These results are consistent with the hypothesis that abnormal or persistent methylation-mediated repression of *FOXA1* and potentially of other genes that regulate mammary epithelial cell differentiation by reproductive factors can lead to aberrant, maturation-arrested luminal progenitor cells which, following transformation by additional alterations, could lead to ER− breast cancer.

These findings replicate and confirm earlier analysis with a much smaller sample set using the 450K methylation array,[6] not included in this analysis. Additionally, we previously found inverse associations between *FOXA1* methylation and FOXA1 protein expression with

increasing number of births, more apparent in women who did not breastfeed,[31] with results similar between ER+ and ER– tumour subtypes. Similarly, in our analysis of methylation of the top loci overall and in women who did not breastfeed by ER status, relationships were observed in both groups. This suggests that the effects of breastfeeding on the expression of FOXA1 and subsequently ESR1 in breast tissues may manifest on a continuous scale rather than qualitatively different by the binary tumour ER status.

Notably, FOXA1 proximal cg00955911 was hypomethylated in almost all ER+ tumours (95%) based on a beta-value <0.25. Since ER+ tumours have significantly higher FOXA1 expression than found in paired adjacent normal tissues, as shown in our previous study,[6] our new findings suggest that hypomethylation and subsequent upregulation of FOXA1 and its transcription target ESR1 in premalignant breast tissue may be necessary for the development of ER+ breast cancer. Thus, hypermethylation of FOXA1 with parity may contribute to its association with lower risk of ER+ breast cancer.

The role of the additional CpGs that were in the top five loci differentially methylated according to parity is less straightforward. Genomic context and gene expression analyses showed that one locus, cg27648238, was significantly associated with the number of births regardless of breastfeeding and inversely correlated with COL14A1 expression, which is involved in the regulation of the TME. Consistent with our findings of lower levels of gene expression in ER– tumours, a transcriptomic study with 60 breast cancer tissues also found that COL14A1 was downregulated in aggressive, basal and HER2+ subtypes.[32]

The other CpGs that were significant only among women who did not breastfeed included cg24322780, which is in an intergenic region. However, several genes were correlated with methylation of this locus, including *ZNF831,* which belongs to the large zinc finger transcription factor family, is associated with immune activity and stem cell regulation in breast cancer,[33] and was recently identified as a novel transcriptional suppressor and putative breast tumour suppressor gene by inhibiting the STAT3/BCL2 signalling pathway.[34] Its downregulation was linked with stem cell phenotype and higher recurrence risk in breast cancer, and it appears to control a network which corresponds to immune response in the basal subtype. However, we found higher expression of ZNF831 in the more aggressive ER– tumours in our data. Another top locus, cg05387815, also maps to a zinc finger gene, *ZNF83,* that showed lower expression in ER– tumours, although its role in breast cancer has not been characterised in the literature.

The fourth top CpG methylated by parity, cg08919180, is in a coding region of *TNFSF4,* but methylation status was not associated with gene expression; rather, expression was linked to *ANKRD45,* which plays a role in cell division and proliferation, but no known involvement in breast cancer,[33] although we did observe higher expression levels in ER– tumours.

This is the largest study, to date, to examine molecular mechanisms underlying the higher prevalence of ER– breast cancer in Black women, with a focus on methylation of breast tumour DNA in relation to parity without breastfeeding. Data and samples were derived from large collaborative studies with harmonised epidemiological data. To account for potential systematic differences, models were adjusted for study site and validation analyses were conducted separately with each study. Extensive work was done to optimise extraction of high-quality DNA from FFPE tumour tissues, all of which was done in one laboratory. Tumour tissues were not available for all the cases, leaving a possibility of selection bias; however, characteristics of those included are similar to those in the overall studies. There is also a possibility that the DNA methylation patterns do not represent the entire tumour, since DNA was extracted from selected cores taken from areas rich with neoplastic epithelial cells as well as curls cut from tissue blocks. Finally, there is always a possibility for uncontrolled confounding. For genome-wide analysis, we sought the most parsimonious model and therefore did not adjust for epidemiological variables that might impact methylation but are likely complex and may vary by CpGs.

We acknowledge that this study cannot denote causality but rather elucidates mechanisms underlying relationships between parity and breastfeeding and breast cancer subtypes. Because carcinogenesis might impact epigenetic processes, use of normal breast tissue would be optimal. However, breast tissue is largely adipose tissue and not necessarily representative of the epithelial cells that transform into tumour cells, and 'normal adjacent' tissue may contain epithelial cells that have already undergone neoplastic transformation.

Despite study limitations, our EWAS findings are biologically plausible, especially because the results replicate those identified in an earlier smaller study. Elucidation of biological mechanisms is a critical step toward establishing causality, which itself supports the importance of interventions to improve access to the conditions that make breastfeeding possible for all women.

Author affiliations
[1]Cancer Prevention and Control, Roswell Park Cancer Institute, Buffalo, New York, USA
[2]Biostatistics and Bioinformatics, Roswell Park Cancer Institute, Buffalo, New York, USA
[3]Biostatistics, Boston University, Boston, Massachusetts, USA
[4]Slone Epidemiology, Boston University, Boston, Massachusetts, USA
[5]Pathology, Roswell Park Cancer Institute, Buffalo, New York, USA
[6]Molecular and Cellular Biology, Roswell Park Cancer Institute, Buffalo, New York, USA
[7]Cancer Epidemiology and Health Outcomes, Rutgers The State University of New Jersey, New Brunswick, New Jersey, USA
[8]Division of Research, Kaiser Permanente, Oakland, California, USA
[9]Slone Epidemiology Center, Boston University, Boston, Massachusetts, USA

X Christine B Ambrosone @cambros1

Acknowledgements We are grateful to Lucas Mendicino who worked with the data and was involved in aspects of the research. We gratefully acknowledge all

the participants and patients for their donation of their time to participate in the research and for their provision of their tissue samples.

Contributors CBA is corresponding author. She was responsible for the conception and design of the study and acquiring funding. She was integral to the acquisition analysis and interpretation of the data. She largely drafted the initial work. She has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. CBA is responsible for the overall content as guarantor. She accepts full responsibility for the finished work and/or the conduct of the study, had access to the data, and controlled the decision to publish. SY made substantial contributions to design of the study and to analysis and interpretation of the data. He has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. MDL made substantial contributions to the design of the analytical plan, the analysis of the data and interpretation of results. He has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. CL made substantial contributions to the design of the analytical plan, the analysis of the data, and interpretation of results. She has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. JC played a large role in extraction of DNA from FFPE samples, analysis of the data, reporting of results and interpretation of the findings. He has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature. WD oversaw the acquisition of samples from the participating studies and supervised the DNA extractions, plating of samples and working with CIDR for the methylation assays. He has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. GZ was integral to the design of the study and coordination of the work. He made substantial contributions to the design of the analytical plan, the analysis of the data, and interpretation of results. He has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. RP-O was responsible for coordination and tracking of samples and DNA from the participating studies and worked with CIDR for the methylation assays. She was responsible for developing approaches to extraction of DNA from FFPE tumours. She has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature.TK was the study pathologist. He was integral to identifying area in tumours for extraction of DNA and lent his expertise in breast cancer pathology throughout the study. He has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. ZG was responsible for the conception and design of the study and acquiring funding. She was integral to the acquisition, analysis and interpretation of the data. She has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. QH was integral to the design of the study and to the analysis and interpretation of the data. He has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. SS made substantial contributions to the development of study hypotheses and the design of the study, as well as the acquisition of the data. She has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. ARO made substantial contributions to the design of the study, and to the analysis and interpretation of the data. She has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature. EVB was key to the conduct of the parent study, overseeing case identification, contact, and interviewing, as well as obtaining funding for that study. She has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature. SL oversaw the pipeline for the genomic data to be analyzed, working with bioinformatics and biostatistics staff for the preparation, analysis and interpretation of the data. He has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature. LK is the PI of one of the contributing studies and was responsible for provision of data and samples for these analyses. He has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature. MJH was responsible for the conception and design of the study, acquiring the funding and coordinating the research. He was integral to the analysis and interpretation of the data and the drafting of the manuscript. He has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved, and the resolution documented in the literature. JRP is PI of one of the contributing studies and was responsible for the conception and design of the study. She was integral to analysis and interpretation of the data. She has approved the submitted version and has agreed both to be personally accountable for the author's own contributions and to ensure that questions related to the accuracy or integrity of any part of the work, even ones in which the author was not personally involved, are appropriately investigated, resolved and the resolution documented in the literature.

Funding R01CA225947 (Ambrosone, Higgins, Palmer, MPIs); X01HG009373 (Ambrosone, Palmer, Higgins, MPIs); P01 CA151135 (Ambrosone, Palmer, Olshan, MPIs); R01 CA10059 (Ambrosone); R01 CA185623 (Bandera, Hong, Demissie, MPIs); R01CA105274 (Kushi), U01CA195566 (Kushi, Ambrosone MPIs); U01 CA164974 (Palmer, Rosenberg MPIs); Breast Cancer Research Foundation BCRF-24-004 (Ambrosone), Susan G. Komen Foundation SAC180086 (Palmer); Karin Grunebaum Foundation for Cancer Research (Palmer); Roswell Park Biorepository and Laboratory Services (BLS), a CCSG shared resource, supported by P30 CA16056 (Johnson).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by Boston University Medical Campus IRB protocol H-38432, Kaiser Permanente IRB Net #: 1263171, Roswell Park Comprehensive Cancer Center IRB number BDR#095817, Rutgers Cancer Institute of New Jersey IRB number: Pro2017000069. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. Data are available upon reasonable request. Genomic data have been deposited in

dbGaP - Study Accession: phs002688.v1.p1. Questionnaire data may be available upon reasonable request.

ORCID iD

Christine B Ambrosone http://orcid.org/0000-0003-1717-9943

REFERENCES

1 Giaquinto AN, Sung H, Miller KD, *et al*. Breast Cancer Statistics, 2022. *CA Cancer J Clin* 2022;72:524–41.
2 DeSantis CE, Ma J, Gaudet MM, *et al*. Breast cancer statistics, 2019. *CA Cancer J Clin* 2019;69:438–51.
3 Palmer JR, Viscidi E, Troester MA, *et al*. Parity, Lactation, and Breast Cancer Subtypes in African American Women: Results from the AMBER Consortium. *J Natl Cancer Inst* 2014;106.
4 Fortner RT, Sisti J, Chai B, *et al*. Parity, breastfeeding, and breast cancer risk by hormone receptor status and molecular phenotype: results from the Nurses' Health Studies. *Breast Cancer Res* 2019;21:40.
5 Anyigba CA, Awandare GA, Paemka L. Breast cancer in sub-Saharan Africa: The current state and uncertain future. *Exp Biol Med (Maywood)* 2021;246:1377–87.
6 Espinal AC, Buas MF, Wang D, *et al*. FOXA1 hypermethylation: link between parity and ER-negative breast cancer in African American women? *Breast Cancer Res Treat* 2017;166:559–68.
7 Ambrosone CB, Ciupak GL, Bandera EV, *et al*. Conducting Molecular Epidemiological Research in the Age of HIPAA: A Multi-Institutional Case-Control Study of Breast Cancer in African-American and European-American Women. *J Oncol* 2009;2009:871250.
8 Bandera EV, Demissie K, Qin B, *et al*. The Women's Circle of Health Follow-Up Study: a population-based longitudinal study of Black breast cancer survivors in New Jersey. *J Cancer Surviv* 2020;14:331–46.
9 Palmer JR, Wise LA, Horton NJ, *et al*. Dual Effect of Parity on Breast Cancer Risk in African-American Women. *JNCI Journal of the National Cancer Institute* 2003;95:478–83.
10 Kwan ML, Ambrosone CB, Lee MM, *et al*. The Pathways Study: a prospective study of breast cancer survivorship within Kaiser Permanente Northern California. *Cancer Causes Control* 2008;19:1065–76.
11 Davis Lynn BC, Lord BD, Cora R, *et al*. Associations between quantitative measures of TDLU involution and breast tumor molecular subtypes among breast cancer cases in the Black Women's Health Study: a case-case analysis. *Breast Cancer Res* 2022;24:86.
12 Ondracek RP, Chen J, Marosy B, *et al*. Results and lessons from dual extraction of DNA and RNA from formalin-fixed paraffin-embedded breast tumor tissues for a large Cancer epidemiologic study. *BMC Genomics* 2022;23:614.
13 Tian Y, Morris TJ, Webster AP, *et al*. ChAMP: updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* 2017;33:3982–4.
14 Teschendorff AE, Marabita F, Lechner M, *et al*. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics* 2013;29:189–96.
15 Teschendorff AE, Breeze CE, Zheng SC, *et al*. A comparison of reference-based algorithms for correcting cell-type heterogeneity in Epigenome-Wide Association Studies. *BMC Bioinformatics* 2017;18:105.
16 Gao Y, Widschwendter M, Teschendorff AE. DNA Methylation Patterns in Normal Tissue Correlate more Strongly with Breast Cancer Status than Copy-Number Variants. *EBioMedicine* 2018;31:243–52.
17 Ambrosone CB, Long MD, Liu C. Data from: associations of dna methylation in breast tumor subtypes with parity and breastfeeding in a cohort of 1,459 black women: implications for public health [dbGaP Repository]. n.d. Available: https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002688.v1.p1
18 Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;19:1655–64.
19 Barnard ME, Wang X, Petrick JL, *et al*. Psychosocial stressors and breast cancer gene expression in the Black Women's Health Study. *Breast Cancer Res Treat* 2024;204:327–40.
20 Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;15:550.
21 Kuleshov MV, Jones MR, Rouillard AD, *et al*. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;44:W90–7.
22 Watson CJ. Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. *Breast Cancer Res* 2006;8:203.
23 Stein T, Morris JS, Davies CR, *et al*. Involution of the mouse mammary gland is associated with an immune cascade and an acute-phase response, involving LBP, CD14 and STAT3. *Breast Cancer Res* 2004;6:R75–91.
24 Borges VF, Lyons TR, Germain D, *et al*. Postpartum Involution and Cancer: An Opportunity for Targeted Breast Cancer Prevention and Treatments? *Cancer Res* 2020;80:1790–8.
25 Katz TA, Liao SG, Palmieri VJ, *et al*. Targeted DNA Methylation Screen in the Mouse Mammary Genome Reveals a Parity-Induced Hypermethylation of Igf1r That Persists Long after Parturition. *Cancer Prev Res (Phila)* 2015;8:1000–9.
26 Huh SJ, Clement K, Jee D, *et al*. Age- and pregnancy-associated DNA methylation changes in mammary epithelial cells. *Stem Cell Reports* 2015;4:297–311.
27 Dos Santos CO, Dolzhenko E, Hodges E, *et al*. An epigenetic memory of pregnancy in the mouse mammary gland. *Cell Rep* 2015;11:1102–9.
28 Smith ZD, Meissner A. DNA methylation: roles in mammalian development. *Nat Rev Genet* 2013;14:204–20.
29 Ambrosone CB, Higgins MJ. Relationships between Breast Feeding and Breast Cancer Subtypes: Lessons Learned from Studies in Humans and in Mice. *Cancer Res* 2020;80:4871–7.
30 Sribenja S, Maguire O, Attwood K, *et al*. Deletion of *Foxa1* in the mouse mammary gland results in abnormal accumulation of luminal progenitor cells: a link between reproductive factors and ER-/TNBC breast cancer? *Am J Cancer Res* 2021;11:3263–70.
31 Cheng T-YD, Yao S, Omilian AR, *et al*. FOXA1 Protein Expression in ER+ and ER– Breast Cancer in Relation to Parity and Breastfeeding in Black and White Women. *Cancer Epidemiol Biomarkers Prev* 2020;29:379–85.
32 Malvia S, Chintamani C, Sarin R, *et al*. ABERRANT EXPRESSION OF COL14A1, CELRS3, and CTHRC1 IN BREAST CANCER CELLS. *Exp Oncol* 2023;45:28–43.
33 da Silveira WA, Palma PVB, Sicchieri RD, *et al*. Transcription Factor Networks derived from Breast Cancer Stem Cells control the immune response in the Basal subtype. *Sci Rep* 2017;7:2851.
34 Fan J, Zhang Z, Chen H, *et al*. *Zinc finger protein 831* promotes apoptosis and enhances chemosensitivity in breast cancer by acting as a novel transcriptional repressor targeting the *STAT3/Bcl2* signaling pathway. *Genes Dis* 2024;11:430–48.