EPIDEMIOLOGY

Neighborhood characteristics and breast tumor methylation: using epigenomics to explore cancer outcome disparities

Jazib Gohar[1] · Whitney L. Do[2] · Jasmine Miller-Kleinhenz[1] · Karen Conneely[3] · Uma Krishnamurti[4] · Olivia D'Angelo[5] · Keerthi Gogineni[6] · Mylin Torres[7] · Sheryl Gabram-Mendola[8] · Lauren E. McCullough[1]

Abstract

Background Social exposures may drive epigenetic alterations that affect racial disparities in breast cancer outcomes. This study examined the association between neighborhood-level factors and DNA methylation in non-Hispanic Black and White women diagnosed with breast cancer.

Methods Genome-wide DNA methylation was measured using the EPIC array in the tumor tissue of 96 women. Linear regression models were used to examine the association between nine neighborhood-level factors and methylation, regressing β values for each cytosine-phosphate guanine dinucleotide (CpG) site on neighborhood-level factors while adjusting for covariates. Neighborhood data were obtained from the Opportunity Atlas. We used a false discovery rate (FDR) threshold < 0.05, and for CpGs below this threshold, we examined interactions with race. We employed multivariable Cox proportional-hazards models to estimate whether aberrant methylation was associated with all-cause mortality.

Results 26 of the CpG sites were associated with job density or college education (FDR < 0.05). Further exploration of these 26 CpG sites revealed no interactions by race, but a single probe in TMEM204 was associated with all-cause mortality.

Conclusion We identified novel associations between neighborhood-level factors and the breast tumor DNA methylome. Our data are the first to show that dysregulation in neighborhood associated CpG sites may be associated with all-cause mortality. Neighborhood-level factors may contribute to differential tumor methylation in genes related to tumor progression and metastasis. This contributes to the increasing body of evidence that area-level factors (such as neighborhood characteristics) may play an important role in cancer disparities through modulation of the breast tumor epigenome.

Keywords Breast cancer · Social epidemiology · DNA methylation · Racial disparity · Mortality

Introduction

Breast cancer is the most common cancer in American women and one of the leading causes of cancer death [1]; however, substantial disparities exist in mortality by race, socioeconomic status, and neighborhood-level

Disclosures SGM Board Member TurningPoint Breast Cancer Rehabilitation, a 501c3 organization in Atlanta.

✉ Lauren E. McCullough
lauren.mccullough@emory.edu

1 Department of Epidemiology, Emory University Rollins School of Public Health, Atlanta, GA 30322, USA

2 Department of Global Health, Emory University Rollins School of Public Health, Atlanta, GA 30322, USA

3 Department of Human Genetics, Emory University School of Medicine, Atlanta, GA 30322, USA

4 Department of Pathology and Laboratory Medicine, Emory University School of Medicine, Atlanta, GA 30322, USA

5 Department of Surgery, Jackson Memorial Hospital/University of Miami Miller School of Medicine, Miami, FL 33136, USA

6 Department of Medical Oncology, Emory University School of Medicine, Atlanta, GA 30322, USA

7 Department of Radiation Oncology, Emory University School of Medicine, Atlanta, GA 30322, USA

8 Department of Surgery, Emory University School of Medicine, Atlanta, GA 30322, USA

characteristics [2, 3]. While emerging research supports the role of individual and area-level social determinants in breast cancer outcomes, little information is known regarding potential biological mechanisms underlying these associations. Social epigenomics is a growing field that examines how social and environmental experiences may impact the epigenome through histone modification, telomere shortening, and DNA methylation [4]. There is emerging literature that adverse social environments may disproportionately affect minority health through epigenomic perturbations, which could impact disparities in health outcomes [4]. Previous studies have found an association between individual-level SES and DNA methylation among children and adults [5–11]; and at least one study suggests that racial differences in methylation may be mediated, in part, through disparities in childhood SES [12]. This latter point illustrates that the hypothesized underlying mechanism is not biological differences by race, but racial disparities in the lived environment that may drive epigenetic perturbations. Area-level characteristics are an important consideration in disparities research given its multidimensional constituents and vast biologic milieu. For example, even after accounting for individual-level socioeconomic factors, living in a disadvantaged neighborhood is associated with poor health outcomes [12]. Additionally, chronic stress due to unfavorable neighborhood conditions can result in dysregulation of inflammatory and stress reactivity pathways which may, in part, be driven by epigenetic modifications such as DNA methylation [13].

Epigenetic modifications, which result in variation to gene expression without altering the underlying DNA sequence, have been established as biomarkers that can signify environmental factors as potential drivers of disease. DNA methylation, specifically, has become a primary epigenetic mechanism to study due to its influence on gene expression and its responsiveness to lifestyle exposures [14]. Aberrant DNA methylation can result in increased oncogene expression and tumor suppressor gene silencing; a common occurrence in breast carcinogenesis [14, 15]. Exploring the impact of area-level sociodemographic characteristics on methylation may yield mechanistic insight into the role of social stressors on breast cancer progression and mortality disparities.

To date, no study has used an epigenome-wide approach to assess neighborhood-associated methylation and breast cancer prognosis. Thus, the purpose of this study was to conduct an epigenome-wide association study (EWAS) of breast tumor tissue to identify CpG sites associated with several neighborhood-level factors. Additionally, we explored interaction by race and downstream associations with breast cancer prognosis.

Methods

Study population

Study protocol follows the methodology described in Do et al. [16]. Briefly, fresh tumor specimen and clinical data were collected from patients receiving surgery at three metro-Atlanta area hospitals (Emory University Hospital, Emory University Hospital Midtown, and Grady Memorial Hospital). We included 99 non-Hispanic White (NHW) and non-Hispanic Black (NHB) women diagnosed with breast cancer between 2008 and 2017 in this analysis. Eligibility for inclusion included women who were at least 21 years of age, self-reported NHW or NHB, diagnosed with a first-primary stage I, II, or III breast cancer, and received surgery at one of the aforementioned hospitals. Women who were previously diagnosed with breast cancer or did not have a fresh tissue specimen were excluded from this study.

Data collection

Clinical records of women who underwent surgery provided covariate data, including age at diagnosis, race, self-reported smoking status, educational attainment, family history of breast cancer, and zip code data. Tumor characteristics were also obtained from clinical records including estrogen receptor (ER) status, human epidermal growth factor 2 (HER2) status, and progesterone receptor (PR) status; tumor grade; receipt of chemo, radiation, and endocrine therapy; and comorbidities at diagnosis. Updated vital status (through 2/15/2018) was obtained by linking all women to the Georgia Cancer Registry, and cause-specific death abstracted. As described in Do et al. [16], we considered all-cause mortality as our outcome of interest. Given the short follow-up period (median = 3 years) any mortality event would be, in part, driven by underlying breast cancer [17].

Neighborhood characteristics were collected from the *Opportunity Atlas*, a publicly available atlas of anonymized longitudinal data for nearly every neighborhood and census tract in the USA [18]. Our primary exposures included median rent (2012–2016), job growth rate (2004–2013), median household income (2012–2016), poverty rate (2012–2016), fraction college graduates (2012–2016), fraction non-white (2010), fraction single parents (2012–2016), population density (2010), and job density (2013). Neighborhood values for each primary exposure are estimated averages over the specified time-period. Area-level data were collected from three Census Bureau data sources: the Census 2000 and 2010 short

forms; federal income tax returns in 1989, 1994, 1995, and 1998–2015; and the Census 2000 long form and the 2005–2015 American Community Surveys (ACS). For the purposes of this paper, the use of the *Opportunity Atlas* is akin to obtaining neighborhood characteristics data from the aforementioned sources and are reflective of the years during which data were collected.

Methylation data

Fresh tumor specimens from breast cancer patients were obtained from the Breast Satellite Tissue Bank, Winship Cancer Institute, Emory University, Atlanta, GA, USA. DNA methylation was measured in 99 breast tumor tissue samples using the Illumina Infinium MethylationEPIC Bead-chip (Illumina, San Diego, CA, USA). Methylation assays were performed in accordance with the Infinium HD Methylation Assay protocol. The protocol uses bisulfite treatment of DNA to convert unmethylated cytosines to uracil, allowing identification of methylated versus unmethylated loci. Two site-specific probes then bind to loci-flanking methylated or unmethylated sequences. The fluorescent signal from the methylated probe (M) relative to the total signal of methylated (M) and unmethylated (U) probes combined is the proportion of DNA strands that are methylated for that CpG site [19]. The β-value represents this: $\beta = [M/(M + U)]$. The β-value ranges from 0 to 1, where 1 represents 100% of the cells being methylated at a CpG site. Three samples were removed during pre-processing due to poor performance.

Quality control (QC) was conducted on the data using the *CpGassoc* package in R. Data points with detection p-values > 0.001 or with low signal were set to missing, and CpG sites with missing values (2869) in over 10% of the samples were removed from the dataset. A stricter probe filtering, as suggested by Zhou et al. [19] was employed by filtering out CpG sites including the following: (1) probes with low quality or inconsistent mapping, (2) probes with extension bases inconsistent with the specified color channel or CpG based on mapping, (3) probes with non-unique 30 base pair 3'-subsequences, (4) probes with SNPs in the extension base that causes a color channel switch, and (5) probes with 5 base pair 3'-subsequences that overlap with any SNPs with a global minor allele frequency $> 1\%$. After QC, 758,942 CpG sites remained for evaluation with neighborhood characteristics.

Statistical analysis

Analyses were carried out using R (www.r-project.org/). Neighborhood characteristics considered in this study are reported overall and by race as means and corresponding standard deviations.

Linear regression models were used to assess whether individual mean β-values differed due to neighborhood characteristics, adjusting for model-specific covariates based on a priori knowledge of the literature and causal diagrams. [20–22]. Using the CpGassoc package we regressed β-values for each CpG site on neighborhood-level factors, adjusting for age, race, and smoking status. Additionally, models included a fixed effect for each BeadChip to account for potential chip-to-chip differences in measurement and to adjust for batch effects. Independent regression models were performed for all nine of the neighborhood characteristics of interest and statistical significance defined as a false discovery rate (FDR) q-value < 0.05.

To assess whether the association between neighborhood characteristics and tumor methylation was modified by race, the CpG sites reaching FDR significance in the primary EWAS were tested for interaction. For each of these CpG sites, the β-values were regressed on the identified neighborhood characteristic with an interaction between the neighborhood characteristic and race. All interaction analyses were adjusted for age and chip position, and significance defined as FDR < 0.05.

We used multivariable Cox proportional-hazards models to explore associations between the top neighborhood-associated CpG sites and all-cause mortality, adjusting for (1) age, (2) age and race, and (3) age, race, cancer stage, and ER status.

Results

Demographic characteristics of our study population were reported in Do et al. [16] and are provided in Supplementary Table 1. Briefly, NHB women were, on average, older (mean age = 58 vs 50 years, respectively) and had a higher BMI compared to NHW women (mean BMI = 34.64 vs 29.84 kg/m^2, respectively). No differences between NHB and NHW women with regards to ER negative status (22.9% vs 30.8%, respectively). Neighborhood characteristics, overall and by race, are presented in Table 1. In total, 55 different metro-Atlanta neighborhoods were represented in this analysis, with an average of 1.75 women per neighborhood (values ranging from 1 to 7 women per neighborhood). On average, NHB women resided in neighborhoods with lower rent prices, higher poverty rates, and lower median household incomes compared to neighborhoods of NHW women. Additionally, neighborhoods of NHB women had a greater proportion of non-white and single parent households while having a lower proportion of college-educated households compared to neighborhoods represented by NHW study participants. NHB women also lived in neighborhoods that had a substantially lower job density compared to NHW women (1327 jobs/mi^2 vs 2784 jobs/mi^2).

Table 1 Means (standard deviation) are presented for neighborhood-level factors of the neighborhoods of residence, stratified by race

	Data source	Total	Non-Hispanic Black	Non-Hispanic White
n		96	70	26
Median Rent, 2006–2010 ($)	American Community Survey	958.73 (236.06)	922.73 (219.00)	1050.12 (256.98)
Poverty Rate, 2006–2010 (%)	American Community Survey	26.21 (12.40)	28.96 (16.53)	18.94 (12.40)
Median household income, 2012–2016 ($)	American Community Survey	44,210.53 (20,659.81)	40,956.52 (20,512.14)	52,846.15 (18,805.75)
Fraction Non-White, 2010 (%)	2010 Decennial Census	66.71 (27.50)	72.92 (26.02)	50.93 (25.15)
Fraction single parents, 2006–2010 (%)	American Community Survey	53.47 (23.53)	60.11 (22.14)	36.60 (18.14)
Population density, 2010 (persons/mi^2)	2010 Decennial Census	3,169.84 (2,605.28)	3,270.02 (2,268.96)	2,915.56 (2,605.28)
Job growth rate, 2004–2013 (%)	Longitudinal employer-household dynamics and local area unemployment statistics	1.36 (8.45)	2.21 (9.29)	-0.75 (5.49)
Job density, 2013 (jobs/mi^2)	Longitudinal employer-household dynamics and local area unemployment statistics	1,738.45 (2,383.05)	1,326.68 (1,163.75)	2,783.70 (3,945.40)
Fraction college graduates, 2006–2010 (%)	American Community Survey	41.78 (9.25)	42.73 (9.25)	39.23 (8.93)

In our main effect analysis assessing epigenome-wide association of DNA methylation with the 9 neighborhood-level factors, 26 CpG sites passed the a priori FDR threshold of 0.05. Of these 26 sites, 5 were associated with neighborhood college graduation rates, and 21 with neighborhood job density. These sites are listed in Table 2. Manhattan plots show the distribution of the CpG sites by −log10(p-value) and chromosomal location for college graduation rates and neighborhood job density. (Fig. 1A, B).

In Table 3 we provide the results of our interaction analysis, to explore race-specific differences in the 26 CpG sites that were associated with neighborhood characteristics. There were three CpG sites (cg00730549, cg00950813, cg02449575) where the relationship between job density and DNA methylation differed substantially by race (Fig. 2A–C), and one CpG site (cg22544350), where the relationship between college graduation rates and DNA methylation was differential by race (Fig. 2D), although these results were not robust after correcting for 26 comparisons.

Of the 26 CpG sites associated with neighborhood characteristics, we found that eleven were also associated with prognosis (Table 4); most resulting in a weak reduction in overall mortality. These reductions were attenuated in multivariable models accounting for age, race, clinical stage, and ER status. For two CpG sites (cg15375883 and cg15196042) we observed a modest increase in mortality risk, even in multivariable models. After correcting for multiple comparisons using false discovery rate adjustment, cg08214329 (*IFT140/TMEM204*) was associated with prognosis (FDR

q-value = 0.02). The hazard ratio (HR) and corresponding 95% confidence interval (CI) in the fully-adjusted model was HR = 0.93 (95%CI 0.87, 0.98).

Discussion

This study is the first untargeted analysis to examine neighborhood-level factor-associated methylation in breast tumor tissue using the EPIC array. DNA methylation emerged as significantly associated with neighborhood characteristics at 26 CpG sites: 21 with neighborhood job density and 5 with neighborhood college graduation rates. Neighborhood-level factors and differential DNA methylation have been limitedly assessed in the literature. Previous studies have found that DNA methylation may mediate the association between neighborhood-level factors and adverse health outcomes such as inflammation and depression [23–26]. In a subsample of the Multi-ethnic Study of Atherosclerosis (MESA) from 2010 to 2012, neighborhood-level socioeconomic disadvantage and neighborhood social environment were associated with methylation in stress- and inflammation-related genes linked to metabolic disease such as obesity and type-2 diabetes [23]. Neighborhood crime rates were associated with depression, mediated by methylation of the promoter region of the serotonin transporter gene (5-HTT), in a cohort of 99 African American women from the Family and Community Health Study [24]. No study to date has

Table 2 The 26 FDR-significant CpG sites associated with a neighborhood-level factor in breast tumor tissue in EWAS

CpG label	Exposure associated	T.statistic	p-value	FDR	Effect.size	SE	Ref. Genes	Chr
cg00730549	Job density	− 4.99	4.17E−06	0.031	1.74E−07	2.73E−06	TNRC18	7
cg00851060	Job density	− 4.86	6.93E−06	0.039	1.03E−05	8.08E−06		
cg00950813	Job density	− 5.20	1.85E−06	0.019	2.09E−06	5.45E−06	ZNF282	7
cg02449575	Job density	− 5.34	1.09E−06	0.013	4.62E−07	9.74E−07	SPTLC2	14
cg03115690	Job density	− 5.07	3.16E−06	0.027	5.07E−07	4.57E−06		
cg04734977	College	− 5.77	1.70E−07	0.038	− 4.86E−04	8.42E−05	ZNF680	7
cg06520003	Job density	5.00	4.08E−06	0.031	− 3.51E−07	9.03E−07	ACSF2	17
cg07141484	Job density	− 4.90	6.07E−06	0.036	9.02E−07	3.11E−06	TLDC2	20
cg07422416	Job density	4.82	7.98E−06	0.043	− 3.21E−06	4.29E−06	MFHAS1	8
cg07994487	Job density	− 5.05	3.40E−06	0.028	5.23E−06	3.59E−06	IFT140, TMEM204	16
cg08214329	Job density	− 5.05	3.33E−06	0.027	4.47E−06	4.22E−06	IFT140, TMEM204	16
cg09178970	Job density	− 4.88	6.47E−06	0.038	− 1.94E−06	4.07E−06	VPS37B	12
cg09254001	College	− 5.72	2.16E−07	0.038	− 3.64E−04	6.37E−05	IKBIP, APAF1	12
cg09866303	Job density	− 5.07	3.15E−06	0.027	7.25E−07	1.87E−06	RREB1	6
cg14838356	Job density	− 5.42	7.94E−07	0.011	− 8.45E−07	6.20E−06	AGO2	8
cg15196042	College	− 5.68	2.48E−07	0.038	− 4.18E−04	7.35E−05	LOC100129716, ARRDC3	5
cg15375883	Job density	5.57	4.58E−07	0.007	− 3.56E−07	6.58E−06	ST3GAL4	11
cg16088676	Job density	− 4.96	4.84E−06	0.031	− 2.88E−06	4.43E−06		
cg16609534	Job density	− 4.95	4.92E−06	0.031	− 5.58E−07	3.72E−07	AFAP1	4
cg17606115	Job density	− 5.22	1.75E−06	0.019	− 5.71E−06	5.82E−06	GNAI2	3
cg18658674	Job density	− 4.83	7.82E−06	0.043	− 4.69E−07	5.18E−07		
cg20226051	College	− 5.69	2.37E−07	0.038	− 3.69E−04	6.48E−05	MTF1	1
cg20576936	Job density	− 4.98	4.39E−06	0.031	− 1.71E−07	4.07E−07	ZNF627	19
cg21900997	Job density	− 5.29	1.33E−06	0.015	2.18E−08	1.06E−07	SCMH1	1
cg22544350	College	− 5.87	1.14E−07	0.038	− 4.93E−03	8.39E−04	KDM5A, CCDC77	12
cg25994418	Job density	− 4.98	4.33E−06	0.031	1.06E−06	6.00E−06	PLCG1	20

T-statistics, p-value and false discovery rate q-value have been provided. EWAS was adjusted for age, race, smoking status, and batch effects. Reference gene and chromosome number obtained from the Illumina annotation file

linked neighborhood factors to breast tumor methylation and prognosis.

We found that area-level job density and college graduation rates associate with differential methylation in breast tumors. Job density may reflect several relevant exposures with varied impacts on health, including surface transportation, which has been associated with outdoor air pollution and differential DNA methylation [27]. It could also be a proxy for residential development, access to green space and health care which are associated with positive health outcomes [28]. Though none of these exposures have been examined in the context of DNA methylation. College graduation rates could also proxy for several exposures including educational attainment, specific occupations, social interactions, or engagement in healthy behaviors [31]. While college graduation may represent economic prosperity, our study found no association with median household income. Low educational attainment has been linked to differential methylation [29] and increases in mortality [30, 31], though the mechanisms underlying this association are not

well understood. Shiftwork [34] and manufacturing occupations [35, 36] have also been shown to adversely impact the DNA methylome, and some studies report methylation as a product of occupational exposures may increase the risk of various cancers, including esophageal [37], lung [38], and bladder carcinomas [39]. While job density and college graduate rates may have several mechanisms of action leading to differential methylation, including those related to stress [32], this study provides initial evidence to support the important role they may play in influencing the breast tumor methylome and potentially adverse outcomes in response to methylation.

While many of the CpGs modulated in response to neighborhood characteristics are within the transcription start site or body of genes that have been implicated in carcinogenesis [40–44], here we highlight two. A single probe on Chr 16 (cg08214329) with overlapping promoters for *IFT140* and *TMEM204* was associated with mortality among women with breast cancer. *TMEM204* is expressed in all cancers with low specificity (Median range 3.1–20.5 FPKM,
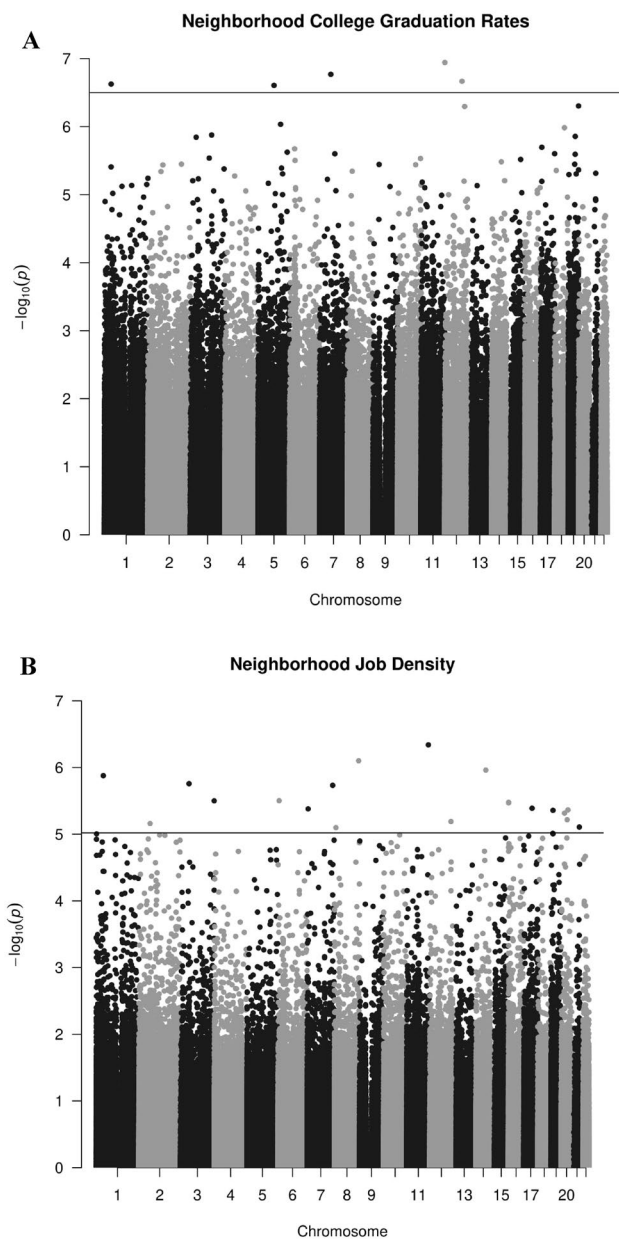
Fig. 1 Manhattan plots of CpG sites for **A** neighborhood college graduation rates and **B** neighborhood job density. FDR < 0.05 CpG sites are those above the solid black line

RNAseq TCGA data), and low expression has been associated with unfavorable liver cancer outcomes, but favorable survival in melanoma (https://www.proteinatlas.org/ENSG0000131634-TMEM204/pathology), consistent with our findings. We observed improved breast cancer survival (~7% reduction in mortality) with every 1-unit increase in methylation.

Of the 26 probes identified as being modulated by neighborhood factors, only one (cg00950813, Chr 7) interacted with race and was associated with all-cause

Table 3 Interaction assessment between race and neighborhood-level factor of the 26 FDR-significant neighborhood-level factor-associated CpG sites

CpG label	Associated neighborhood characteristic	*p*-value	FDR
cg00730549	Job density	0.013*	0.065
cg00851060	Job density	0.21	1.00
cg00950813	Job density	0.032*	0.28
cg02449575	Job density	0.024*	0.21
cg03115690	Job density	0.95	1.00
cg04734977	College	0.79	1.00
cg06520003	Job density	0.068	0.59
cg07141484	Job density	0.14	0.67
cg07422416	Job density	0.98	1.00
cg07994487	Job density	0.32	1.00
cg08214329	Job density	0.32	1.00
cg09178970	Job density	0.40	0.94
cg09254001	College	0.84	1.00
cg09866303	Job density	0.70	0.34
cg14838356	Job density	0.14	0.62
cg15196042	College	0.59	0.95
cg15375883	Job density	0.68	1.00
cg16088676	Job density	0.25	1.00
cg16609534	Job density	0.37	1.00
cg17606115	Job density	0.43	1.00
cg18658674	Job density	0.63	1.00
cg20226051	College	0.74	1.00
cg20576936	Job density	0.89	1.00
cg21900997	Job density	0.095	0.61
cg22544350	College	0.027*	0.23
cg25994418	Job density	0.23	0.53

*significant at $p < 0.05$

mortality. cg00950813 is located in the body of *ZNF282* and encodes the zinc finger protein 282 (ZNF282). The zinc finger proteins are involved in a variety of cellular mechanisms and have been implicated in tumorigenesis and cancer progression [45]. *ZNF282* has been found to have an oncogenic role and promote tumorigenesis in esophageal and breast tissue [46, 47]. *ZFP282* is a co-activator of estrogen receptor α and thus is required for estrogen-dependent breast cancer cell growth [50], suggesting differential expression by breast cancer subtype [48]. In addition, *ZNF282* has a small ubiquitin-like modifying (SUMO) role in estrogen signaling and breast tumorigenesis in mouse models and the SUMO pathway is hyperactivated in breast cancer [47]. Given the important role of *ZNF282* in influencing ER positivity, we examined whether methylation at this site was associated with ER status, adjusting for age, race, and chip. In our post hoc case-case comparison we found that every 0.01 unit increase in methylation was associated with an 11%
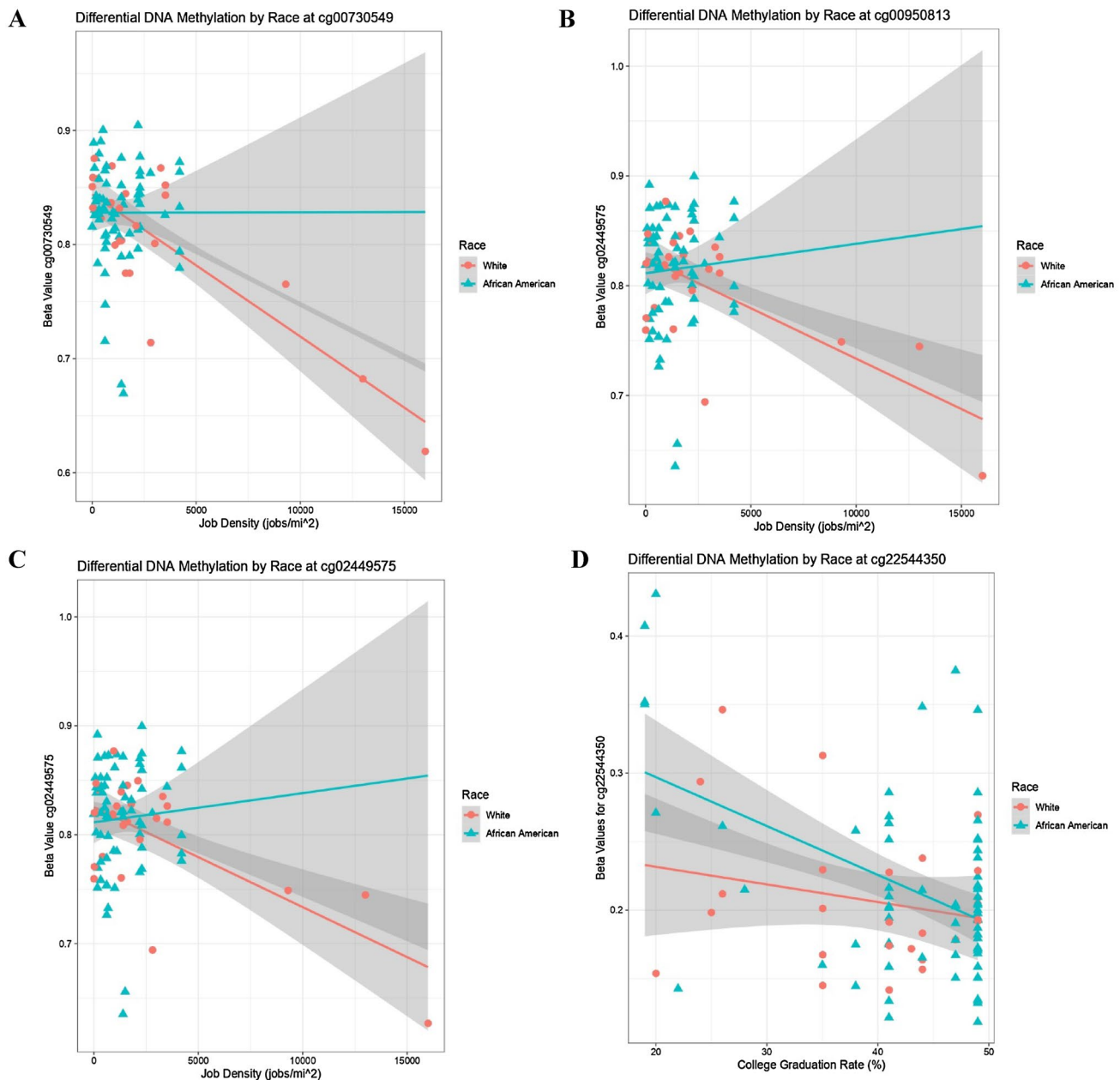
Fig. 2 Scatterplots and regression lines depicting β-values by neighborhood-level factors excluding methylation outliers, examining CpG sites that exhibited interaction by race at $p < 0.05$

increased odds of ER positive status compared to ER negative status (Odds Ratio 1.11; 95%CI 1.04, 1.24).

In this site, there was an inverse association between methylation and job density, which appears to be primarily driven by NHW women. The observed interaction by race is unlikely a result of differential ER positivity (77% ER positivity among NHB vs. 69% among NHW), but due to lower methylation among three NHW women living in communities with substantial job density. It is unknown whether this interaction would remain had our study included NHB women residing in areas with high job density, and upon

removing the three NHW women with job density > 5000 jobs/mi², the interaction did not persist. Our findings underscore the potential importance of sociodemographic characteristics in observed molecular differences by race.

We also found an inverse association between methylation at cg00950813 and mortality; each 1% increase in the β-value was associated with a 5% decrease in all-cause mortality; consistent with the putative role of *ZNF282* in breast carcinogenesis. It is well established that ER status plays a critical role in breast cancer prognosis, where women with ER negative tumors have poorer outcomes compared to

Table 4 Hazard ratios models of the 26 FDR-significant neighborhood-level factor-associated CpG sites adjusted for (1) age, (2) age and BMI, and (3) age, clinical stage, and ER negative status

CpG Label	Adjusted for age HR (95% CI)	Adjusted for age and race HR (95% CI)	Adjusted for age, race, clinical stage, and ER status (95% CI)
cg00730549	0.94 (0.88, 1.01)	0.94 (0.87, 1.01)	0.94 (0.86, 1.03)
cg00851060	0.96 (0.92, 1.00)	0.96 (0.92, 1.00)	0.98 (0.93, 1.03)
cg00950813	0.96 (0.93, 0.99)	0.95 (0.92, 0.99)	0.95 (0.91, 1.00)
cg02449575	0.96 (0.90, 1.04)	0.97 (0.90, 1.04)	0.95 (0.87, 1.04)
cg03115690	0.93 (0.88, 0.99)	0.93 (0.87, 0.99)	0.97 (0.90, 1.05)
cg04734977	0.95 (0.66, 1.39)	0.95 (0.65, 1.40)	0.84 (0.52, 1.36)
cg06520003	1.02 (0.97, 1.11)	1.02 (0.94, 1.11)	0.98 (0.90, 1.07)
cg07141484	0.91 (0.85, 0.98)	0.91 (0.85, 0.98)	0.91 (0.84, 0.99)
cg07422416	1.00 (0.97. 1.04)	1.00 (0.97, 1.04)	1.00 (0.95, 1.05)
cg07994487	0.94 (0.91, 0.98)	0.94 (0.91, 0.98)	0.95 (0.91, 1.00)
cg08214329	0.92 (0.88, 0.97)	0.92 (0.88, 0.97)	0.93 (0.87, 0.98)
cg09178970	0.96 (0.75, 1.24)	0.97 (0.75, 1.25)	1.10 (0.81, 1.49)
cg09254001	0.97 (0.55, 1.71)	0.96 (0.55, 1.70)	0.79 (0.38, 1.66)
cg09866303	0.97 (0.93, 1.02)	0.97 (0.93, 1.02)	0.97 (0.92, 1.02)
cg14838356	0.98 (0.94, 1.01)	0.97 (0.94, 1.01)	0.97 (0.92, 1.02)
cg15196042	1.58 (0.96, 2.58)	1.58 (0.97, 2.60)	1.97 (1.00, 3.87)
cg15375883	1.20 (0.99, 1.44)	1.19 (0.99, 1.44)	1.27 (1.04, 1.56)
cg16088676	0.93 (0.89, 0.98)	0.93 (0.89, 0.97)	0.94 (0.89, 1.00)
cg16609534	0.89 (0.80, 0.99)	0.89 (0.80, 0.99)	0.88 (0.79, 0.99)
cg17606115	0.88 (0.80, 0.98)	0.88 (0.80, 0.98)	0.89 (0.79, 1.00)
cg18658674	0.97 (0.93, 1.00)	0.97 (0.93, 1.00)	0.98 (0.94, 1.03)
cg20226051	1.08 (0.71, 1.65)	1.07 (0.70, 1.64)	1.12 (0.63, 2.00)
cg20576936	0.98 (0.93, 1.04)	0.98 (0.93, 1.04)	0.96 (0.90, 1.03)
cg21900997	0.91 (0.82, 1.00)	0.91 (0.81, 1.02)	0.90 (0.78, 1.02)
cg22544350	0.97 (0.90, 1.03)	0.96 (0.90, 1.03)	0.92 (0.84, 1.01)
cg25994418	1.00 (0.96, 1.05)	1.00 (0.95, 1.05)	0.98 (0.93, 1.03)

women with ER positive tumors [49], yet we found no interaction between cg00950813 and ER with mortality, and little modulation of its effect in multivariable models adjusting for ER status. Given that this site is located in the gene body, concluding the impact on gene expression is not straightforward. Gene body methylation is frequently cited as an indicator of an active gene [50]. However, a recent large scale EWAS has shown that regardless of the location in the gene, methylation is primarily associated with decreased expression [51]. Given the oncogenic role of *ZNF282*, our findings would align with this—such that increased methylation would lead to decreased gene expressed thereby reducing all-cause mortality.

The limitations of our study should be acknowledged. While we included only 96 women in this study, we were able to identify area-level differential methylation at several CpG sites even after correcting for multiple comparisons. While the effect sizes may not be robust, the goal of this research was to understand whether the DNA methylome may be one mechanism by which the social environment influences carcinogenesis. The ubiquity of exposure may suggest that population-level strategies to

improve adverse neighborhood characteristics may have a large impact on cancer outcomes, particularly for vulnerable groups. More robust studies are needed to validate our findings. Our study population may not reflect the known subtype distribution among non-Hispanic White women who are infrequently diagnosed with triple-negative disease. While this limits external validity, internal validity is enhanced by having a relatively balanced distribution of subtypes by race. We performed 9 different EWAS in total, each independently FDR corrected. While this approach is less conservative, given the novelty of our study hypothesis we aimed to reduce the likelihood of false negatives and inform future work in this area. Further, many of our exposures were correlated, and all 26 significant CPGs were further evaluated for associations with all-cause mortality in multivariable analyses again FDR adjusted. Additionally, based on our a priori criteria, we examined interaction by race for only those CpG sites that reached FDR significance in the primary EWAS. It is possible that significant interaction by race may exist in other CpG sites that did not reach FDR significance and is an important consideration for future studies. Confounding

by unmeasured factors may also potentially influence our results, although we anticipate most factors would be downstream of neighborhood characteristics and thus on the causal pathway. Finally, we focus exclusively on the DNA methylome, upstream from our outcome of interest. In doing so, we negate other layers of the epigenome and downstream impacts on gene expression.

This study took an untargeted approach to examine potential epigenetic mechanisms underlying the association between neighborhood-level factors and breast cancer prognosis in a diverse population of women that underwent breast cancer surgery in metro-Atlanta. Our preliminary results suggest that the behavioral and biological responses consequent to neighborhood stressors have the potential to modulate the breast tumor epigenome and subsequent outcomes. Our results require replication in other race and ethnic groups in diverse regions across the USA. Importantly, this research highlights that if the characteristics perpetuating poor outcomes in minority communities are alleviated, outcomes can also be improved. Future efforts should focus on replicating the methodology in a larger, equally diverse patient population to validate these preliminary findings and further explore the epigenetic mechanisms identified in this study.

Declarations

Conflict of interest No conflict of interests to disclose.

References

1. How Common Is Breast Cancer?: Breast Cancer Statistics (2020) https://www.cancer.org/cancer/breast-cancer/about/how-common-is-breast-cancer.html
2. Gerend MA, Pai M (2008) Social determinants of Black-White disparities in breast cancer mortality: a review. Cancer Epidemiol Biomark Prev 17(11):2913–2923. https://doi.org/10.1158/1055-9965.epi-07-0633
3. Huo D, Hu H, Rhie SK, Gamazon ER, Cherniack AD, Liu J et al (2017) Comparison of breast cancer molecular features and survival by African and European ancestry in the cancer genome AtlasBreast cancer molecular features and survival by African and European AncestryBreast cancer molecular features and survival by African and European ancestry. JAMA Oncol 3(12):1654–1662. https://doi.org/10.1001/jamaoncol.2017.0595
4. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS (2012) Factors underlying variable DNA methylation in a human community cohort. Proc Natl Acad Sci USA 109(Supplement_2):17253–17260. https://doi.org/10.1073/pnas.1121249109
5. Lam LL, Emberly E, Fraser HB, Neumann SM, Chen E, Miller GE, Kobor MS (2012) Factors underlying variable DNA methylation in a human community cohort. Proc Natl Acad Sci USA 109(Supplement_2):17253–17260. https://doi.org/10.1073/pnas.1121249109
6. Mcdade TW, Ryan CP, Jones MJ, Hoke MK, Borja J, Miller GE et al (2019) Genome-wide analysis of DNA methylation in relation to socioeconomic status during development and early adulthood. Am J Phys Anthropol 169(1):3–11. https://doi.org/10.1002/ajpa.23800
7. Needham BL, Smith JA, Zhao W, Wang X, Mukherjee B, Kardia SLR et al (2015) Life course socioeconomic status and DNA methylation in genes related to stress reactivity and inflammation: the multi-ethnic study of atherosclerosis. Epigenetics 10(10):958–969. https://doi.org/10.1080/15592294.2015.1085139
8. Bush NR, Edgar RD, Park M, Macisaac JL, Mcewen LM, Adler NE et al (2018) The biological embedding of early-life socioeconomic status and family adversity in childrens genome-wide DNA methylation. Epigenomics 10(11):1445–1461. https://doi.org/10.2217/epi-2018-0042
9. Coker ES, Gunier R, Huen K, Holland N, Eskenazi B (2018) DNA methylation and socioeconomic status in a Mexican-American birth cohort. Clin Epigenet. https://doi.org/10.1186/s13148-018-0494-z
10. Fiorito G, Polidoro S, Dugué P-A, Kivimaki M, Ponzi E, Matullo G et al (2017) Social adversity and epigenetic aging: a multi-cohort study on socioeconomic differences in peripheral blood DNA methylation. Sci Rep. https://doi.org/10.1038/s41598-017-16391-5
11. Stringhini S, Polidoro S, Sacerdote C, Kelly RS, Veldhoven KV, Agnoli C et al (2015) Life-course socioeconomic status and DNA methylation of genes regulating inflammation. Int J Epidemiol 44(4):1320–1330. https://doi.org/10.1093/ije/dyv060
12. Pan Y, Wang X, Treiber F, Harshfield G, Su S (2019) Differential DNA methylation between African Americans and European Americans: the role of childhood socioeconomic status and relations to health disparity. FASEB J 33(S1):593–515
13. Smith JA, Zhao W, Wang X, Ratliff SM, Mukherjee B, Kardia SLR et al (2017) Neighborhood characteristics influence DNA methylation of genes involved in stress response and inflammation: the Multi-Ethnic Study of Atherosclerosis. Epigenetics 12(8):662–673. https://doi.org/10.1080/15592294.2017.1341026

14. Kulis M, Esteller M (2010) DNA methylation and cancer. Adv Genet 70:27–56. https://doi.org/10.1016/B978-0-12-380866-0.60002-2

15. Herman JG, Baylin SB (2003) Gene silencing in cancer in association with promoter hypermethylation. N Engl J Med 349(21):2042–54. https://doi.org/10.1056/NEJMra023075

16. Do WL, Conneely K, Gabram-Mendola S et al (2020) Obesity-associated methylation in breast tumors: a possible link to disparate outcomes? Breast Cancer Res Treat 181:135–144. https://doi.org/10.1007/s10549-020-05605-6s

17. He W, Sofie Lindström L, Hall P, Czene K (2017) Cause-specific mortality in women with breast cancer in situ. Int J Cancer 140(11):2414–2421. https://doi.org/10.1002/ijc.30413

18. Chetty R, Friedman J, Hendren N, Jones M, Porter S (2018) The opportunity atlas: mapping the childhood roots of social mobility. Harvard University Press, Cambridge. https://doi.org/10.3386/W25147

19. Bock C (2012) Analysing and interpreting DNA methylation data. Nat Rev Genet 13(10):705–19. https://doi.org/10.1038/nrg3273

20. Zhou W, Laird PW, Shen H (2017) Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. Nucleic Acids Res 45(4):e22. https://doi.org/10.1093/nar/gkw967

21. Draheim KM, Chen H-B, Tao Q, Moore N, Roche M, Lyle S (2010) ARRDC3 suppresses breast cancer progression by negatively regulating integrin β4. Oncogene 29(36):5032–5047. https://doi.org/10.1038/onc.2010.250

22. Howards PP, Schisterman EF, Poole C, Kaufman JS, Weinberg CR (2012) "Toward a clearer definition of confounding" revisited with directed acyclic graphs. Am J Epidemiol 176(6):506–511. https://doi.org/10.1093/aje/kws127

23. Giurgescu C, Nowak AL, Gillespie S, Nolan TS, Anderson CM, Ford JL et al (2019) Neighborhood environment and DNA methylation: implications for cardiovascular disease risk. J Urban Health 96(S1):23–34. https://doi.org/10.1007/s11524-018-00341-1

24. Lei M-K, Beach SR, Simons RL, Philibert RA (2015) Neighborhood crime and depressive symptoms among African American women: genetic moderation and epigenetic mediation of effects. Soc Sci Med 146:120–128. https://doi.org/10.1016/j.socscimed.2015.10.035

25. Lei M-K, Simons RL, Beach SRH, Philibert RA (2017) Neighborhood disadvantage and biological aging: using marginal structural models to assess the link between neighborhood census variables and epigenetic aging. J Gerontol. https://doi.org/10.1093/geronb/gbx015

26. Swartz JR, Hariri AR, Williamson DE (2016) An epigenetic mechanism links socioeconomic status to changes in depression-related brain function in high-risk adolescents. Mol Psychiatry 22(2):209–214. https://doi.org/10.1038/mp.2016.82

27. Baccarelli A, Wright RO, Bollati V, Tarantini L, Litonjua AA, Suh HH, Zanobetti A, Sparrow D, Vokonas PS, Schwartz J (2009) Rapid DNA methylation changes after exposure to traffic particles. Am J Respir Crit Care Med 179(7):572–578. https://doi.org/10.1164/rccm.200807-1097OC

28. Jennings V, Bamkole O (2019) The relationship between social cohesion and urban green space: an avenue for health promotion. Int J Environ Res Public Health 16(3):452. https://doi.org/10.3390/ijerph16030452

29. van Dongen J, Bonder MJ, Dekkers KF et al (2018) DNA methylation signatures of educational attainment. npj Sci Learn 3:7. https://doi.org/10.1038/s41539-018-0020-2

30. Buckles K, Hagemann A, Malamud O, Morrill M, Wozniak A (2016) The effect of college education on mortality. J Health Econ 50:99–114. https://doi.org/10.1016/j.jhealeco.2016.08.002

31. Krueger PM, Tran MK, Hummer RA, Chang VW (2015) Mortality attributable to low levels of education in the United States.

32. Bakusic J, Schaufeli W, Claes S, Godderis L (2017) Stress, burnout and depression: a systematic review on DNA methylation mechanisms. J Psychosom Res 92:34–44. https://doi.org/10.1016/j.jpsychores.2016.11.005

33. Hout M (2012) Social and economic returns to college education in the United States. Ann Rev Sociol 38(1):379–400. https://doi.org/10.1146/annurev.soc.012809.102503

34. Bollati V, Baccarelli A, Sartori S, Tarantini L, Motta V, Rota F, Costa G (2010) Epigenetic effects of shiftwork on blood DNA methylation. Chronobiol Int 27(5):1093–1104. https://doi.org/10.3109/07420528.2010.490065

35. Wright RO, Schwartz J, Wright RJ, Bollati V, Tarantini L, Park SK, Hu H, Sparrow D, Vokonas P, Baccarelli A (2010) Biomarkers of lead exposure and DNA methylation within retrotransposons. Environ Health Perspect 118(6):790–795. https://doi.org/10.1289/ehp.0901429

36. Kile ML, Fang S, Baccarelli AA, Tarantini L, Cavallari J, Christiani DC (2013) A panel study of occupational exposure to fine particulate matter and changes in DNA methylation over a single workday and years worked in boilermaker welders. Environ Health 12(1):47. https://doi.org/10.1186/1476-069X-12-47

37. Ganji S, Miotto E, Callegari E, Sayehmiri K, Fereidooni F, Yazdanbod M, Rastgar-Jazii F, Negrini M (2010) Associations of risk factors obesity and occupational airborne exposures with CDKN2A/p16 aberrant DNA methylation in esophageal cancer patients. Dis Esophagus 23(7):597–602. https://doi.org/10.1111/j.1442-2050.2010.01059.x

38. Ali AHK, Kondo K, Namura T, Senba Y, Takizawa H, Nakagawa Y, Toba H, Kenzaki K, Sakiyama S, Tangoku A (2011) Aberrant DNA methylation of some tumor suppressor genes in lung cancers from workers with chromate exposure. Mol Carcinog 50(2):89–99. https://doi.org/10.1002/mc.20697

39. Wilhelm-Benartzi CS, Christensen BC, Koestler DC, Andres Houseman E, Schned AR, Karagas MR, Kelsey KT, Marsit CJ (2011) Association of secondhand smoke exposures with DNA methylation in bladder carcinomas. Cancer Causes Control 22(8):1205–1213. https://doi.org/10.1007/s10552-011-9788-6

40. Arakaki AKS, Pan WA, Lin H, Trejo J (2018) The α-arrestin ARRDC3 suppresses breast carcinoma invasion by regulating G protein-coupled receptor lysosomal sorting and signaling. J Biol Chem 293(9):3350–3362. https://doi.org/10.1074/jbc.RA117.001516

41. Hou J, Wu J, Dombkowski A et al (2012) Genomic amplification and a role in drug-resistance for the KDM5A histone demethylase in breast cancer. Am J Transl Res 4(3):247–256

42. Roa-de La Cruz L, Martínez-Morales P, Morán-Cruz I et al (2018) Expression analysis of ST3GAL4 transcripts in cervical cancer cells. Mol Med Rep 18(1):617–621. https://doi.org/10.3892/mmr.2018.8938

43. Ji D, Zhong X, Jiang X et al (2018) The role of long non-coding RNA AFAP1-AS1 in human malignant tumors. Pathol Res Pract 214(10):1524–1531. https://doi.org/10.1016/j.prp.2018.08.014

44. Raymond JR Jr, Appleton KM, Pierce JY, Peterson YK (2014) Suppression of GNAI2 message in ovarian cancer. J Ovarian Res 7:6. https://doi.org/10.1186/1757-2215-7-6

45. Cassandri M, Smirnov A, Novelli F et al (2017) Zinc-finger proteins in health and disease. Cell Death Discov 3:17071. https://doi.org/10.1038/cddiscovery.2017.71

46. Yeo S-Y, Ha SY, Yu EJ, Lee K-W, Kim JH, Kim S-H (2014) ZNF282 (Zinc finger protein 282), a novel E2F1 co-activator, promotes esophageal squamous cell carcinoma. Oncotarget. https://doi.org/10.18632/oncotarget.2630

47. Yu EJ, Kim S-H, Kim MJ, Seo W-Y, Song K-A, Kang M-S et al (2012) SUMOylation of ZFP282 potentiates its positive

PLoS ONE 10(7):e0131809. https://doi.org/10.1371/journal.pone.0131809

effect on estrogen signaling in breast tumorigenesis. Oncogene 32(35):4160–4168. https://doi.org/10.1038/onc.2012.420

48. Chen L, Zeng T, Pan X, Zhang YH, Huang T, Cai YD (2019) Identifying methylation pattern and genes associated with breast cancer subtypes. Int J Mol Sci 20(17):4269. https://doi.org/10.3390/ijms20174269

49. Baretta Z, Olopade OI, Huo D (2015) Heterogeneity in hormone-receptor status and survival outcomes among women with synchronous and metachronous bilateral breast cancers. Breast 24(2):131–136. https://doi.org/10.1016/j.breast.2014.12.001

50. Smith ZD, Meissner A (2013) DNA methylation: roles in mammalian development. Nat Rev Genet 14(3):204–220. https://doi.org/10.1038/nrg3354

51. Kennedy EM, Goehring GN, Nichols MH, Robins C, Mehta D, Klengel T, Eskin E, Smith AK, Conneely KN (2018) An integrated -omics analysis of the epigenetic landscape of gene expression in human blood cells. BMC Genomics 19(1):476. https://doi.org/10.1186/s12864-018-4842-3

52. Chu QD, Burton G, Glass J, Smith MH, Li BD (2010) Impact of race and ethnicity on outcomes for estrogen receptor-negative breast cancers: experience of an academic center with a charity hospital. J Am Coll Surg 210(5):585–594. https://doi.org/10.1016/j.jamcollsurg.2010.01.025