Breast tumor DNA methylation patterns associated with smoking in the Carolina Breast Cancer Study

Kathleen Conway, University of North Carolina
Sharon N. Edmiston, University of North Carolina
Eloise Parrish, University of North Carolina
Christopher Bryant, University of North Carolina
Chiu-Kit Tse, University of North Carolina
Theresa Swift-Scanlan, Virginia Commonwealth University
Lauren McCullough, Emory University
Pei Fen Kuan, SUNY Stony Brook

Journal Title: Breast Cancer Research and Treatment
Volume: Volume 163, Number 2
Publisher: Springer Verlag (Germany) | 2017-06-01, Pages 349-361
Type of Work: Article | Post-print: After Peer Review
Publisher DOI: 10.1007/s10549-017-4178-8
Permanent URL: https://pid.emory.edu/ark:/25593/sqc9m

Final published version: http://dx.doi.org/10.1007/s10549-017-4178-8

Copyright information:

Accessed March 21, 2019 12:06 PM EDT
Breast tumor DNA methylation patterns associated with smoking in the Carolina Breast Cancer Study

Kathleen Conway¹,³, Sharon N. Edmiston³, Eloise Parrish³, Christopher Bryant², Chiu-Kit Tse¹, Theresa Swift-Scanlan⁴,⁵, Lauren E. McCullough⁶, and Pei Fen Kuan⁷

¹Department of Epidemiology, School of Public Health, University of North Carolina, Campus Box 7435, Chapel Hill, NC 27599, USA
²Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill, NC 27599, USA
³Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599, USA
⁴School of Nursing, Virginia Commonwealth University, Richmond, VA 23298, USA
⁵Massey Comprehensive Cancer Center, Virginia Commonwealth University, Richmond, VA 23298, USA
⁶Department of Epidemiology, School of Public Health, Emory University, Atlanta, GA 30322, USA
⁷Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794-3600, USA

Abstract

Purpose—Tobacco smoking is a risk factor in several cancers, yet its roles as a putative etiologic exposure or poor prognostic factor in breast cancer are less clear. Altered DNA methylation contributes to breast cancer development and may provide a mechanistic link between smoking and gene expression changes leading to cancer development or progression.

Methods—Using a cancer-focused array, we examined methylation at 933 CpGs in 517 invasive breast tumors in the Carolina Breast Cancer Study to determine whether methylation patterns differ by exposure to tobacco smoke. Multivariable generalized linear regression models were used to compare tumor methylation profiles between smokers and never smokers, overall, or stratified on hormone receptor (HR) status.

Results—Modest differences in CpG methylation were detected at \( p < 0.05 \) in breast tumors from current or ever smokers compared with never smokers. In stratified analyses, HR− tumors...
from smokers exhibited primarily hypomethylation compared with tumors from never smokers; hypomethylation was similarly detected within the more homogeneous basal-like subtype. Most current smoking-associated CpG loci exhibited methylation levels in former smokers that were intermediate between those in current and never smokers and exhibited progressive changes in methylation with increasing duration of smoking. Among former smokers, restoration of methylation toward baseline (never smoking) levels was observed with increasing time since quitting. Moreover, smoking-related hypermethylation was stronger in HR+ breast tumors from blacks than in whites.

Conclusions—Our results suggest that breast tumor methylation patterns differ with tobacco smoke exposure; however, additional studies are needed to confirm these findings.

Keywords
Breast cancer; Smoking; Breast tumor; Methylation; Epigenetic

Introduction
Cigarette smoking is a well-known risk factor for a number of human cancers, including cancers of the lung [1], head and neck [2], and bladder [3]. The epidemiologic data on active smoking in breast cancer has been mixed [4, 5], but more recent studies support a modest role for smoking in breast cancer development [6–8], most often among women who smoked for a long time [6, 9–12], recently quit [12, 13], or began smoking at an early age or prior to first birth [6, 7, 10, 11, 14]. Several reports have also described stronger associations of smoking with either ER+ [7, 8, 15, 16] or ER− breast cancer [17–20]. In addition to putative etiologic effects, tobacco smoking at the time of breast cancer diagnosis is associated with worse disease-specific and overall survival [21, 22]; however, the biological mechanisms through which smoking might influence the risk of developing or dying from breast cancer are not fully understood.

Epigenetic alterations, such as DNA methylation, are thought to be part of the causal pathway in the development of malignancies, including breast cancer [23, 24]. Methylation does not alter the nucleotide sequence of DNA but can influence gene expression and may provide a mechanistic link between environmental exposures, such as smoking, and altered cellular gene expression leading to the development or progression of cancer [25, 26]. Numerous epigenome-wide studies have reported methylation differences in normal peripheral blood leukocytes (PBL) between current smokers and nonsmokers [27–30], with methylation changes at many loci being reversible after cessation of smoking [28, 29, 31], highlighting the dynamic nature of DNA methylation and the potential for modulation by environmental exposures. In cancers for which smoking is a known risk factor, differences in tumor methylation profiles have been detected between smokers and nonsmokers [32–36], but the effect of tobacco smoke exposure on the epigenetic profile of breast tumors has not been examined.

Studies in the Carolina Breast Cancer Study (CBCS) have reported the associations of smoking and breast cancer mainly among long-term smokers and recent quitters, and observed stronger associations for luminal breast cancer, particularly among African
Americans [9, 16]. Additionally, we observed worse disease-specific survival among breast cancer cases who smoked (unpublished observations). In an effort to better understand the genomic effects of smoking and determine whether breast tumor methylation patterns reflect, in part, this exposure, we examined methylation profiles of invasive breast tumors in the CBCS according to active smoking status using a cancer-focused methylation array.

**Materials and methods**

**CBCS population**

The CBCS is a population-based, case–control study of breast cancer that includes women aged 20–74 years residing in 24 counties of central and eastern North Carolina. Women with a first diagnosis of invasive breast cancer between 1993 and 1996 (phase 1) were identified by the North Carolina Central Cancer Registry through a rapid case ascertainment system. Women diagnosed prior to age 50 and black/African American women were over-sampled to ensure that they comprised roughly half the study sample. Race was self-reported. Additional details of the study design are described elsewhere [37], and case characteristics are provided in Table 1. All aspects of this research were approved by the UNC School of Medicine Institutional Review Board. A total of 861 breast cancer cases were eligible for and consented to participate in the CBCS during phase 1.

**Smoking history**

Risk factor information, including smoking history, was obtained from questionnaires administered to participants by trained nurse interviewers. Active smokers self-reported smoking ≥100 cigarettes, while never smokers smoked <100 cigarettes over their lifetime. On average, interviews were conducted 6 months following case ascertainment. Since a breast cancer diagnosis may influence decisions concerning smoking cessation, current smokers were defined as women who self-reported active smoking at the time of interview and women who reported smoking cessation at the same age (within a year) of case selection [16]. Former smokers quit at any age prior to the age at case selection. These data were used to derive smoking duration in years. All smoking data were recorded as categorical variables.

**Clinical data and tumor tissues**

Clinical data and information on tumor characteristics were obtained from medical records or histopathologic review of tumor tissue. Formalin-fixed paraffin-embedded (FFPE) tumor blocks were obtained from pathology departments at participating hospitals and underwent standardized histopathologic review [38]. The pathologist also encircled the tumor on the hematoxylin and eosin (H&E)-stained slide which was then used to guide manual microdissection of tumor from surrounding non-tumor tissue. Isolated tissue was processed for DNA using a standard proteinase K-based method as previously described [39]. Intrinsic breast tumor subtypes were previously identified using a panel of immunohistochemical markers and included luminal A (ER+ and/or PR+(HR+)/HER2−), luminal B (HR+/HER2+), basal-like (ER−/PR− (HR−)/HER2−/CK5+ or CK6+, or EGFR+), HER2+/HR−, and unclassified (all markers negative) [40].
DNA methylation analysis

DNA methylation profiling was previously accomplished using the Illumina GoldenGate Cancer Panel I methylation bead array [39]. Sodium bisulfite modification of FFPE breast tumor DNA was performed using the EZ DNA Methylation Gold kit according to the manufacturer’s protocol. Bisulfite-converted tumor DNA was analyzed for methylation using the Cancer Panel I array in the Mammalian Genotyping Core facility at UNC. The Cancer Panel I array measures methylation at CpG sites in the promoter (P) or first exon (E) of cancer-related genes, including tumor suppressor genes, oncogenes, DNA repair genes, and others. Array probe names are comprised of the gene symbol followed by the position of the CpG in the P or E upstream from the transcription start site. The methylation level at each CpG site is represented by the beta ($\beta$) value, which ranged from 0 (completely unmethylated) to 1.0 (fully methylated). Probes reported to overlap a single-nucleotide polymorphism (SNP) or repeat [41] were excluded. After quality control filtering as described previously [39], methylation analysis was completed on 517 breast tumors and included 933 CpG loci in 609 genes.

Statistical analyses

Statistical analyses were carried out using R (http://www.r-project.org/) or SAS v9.3. To identify CpG loci that showed differential breast tumor methylation between cases who were active smokers versus never smokers, generalized linear regression models (GLM) were fitted to the logit transformed $\beta$-values adjusting for age (continuous), race (African American, white/other), menopausal status, stage (1, 2, 3, 4), body mass index (BMI) (<25, 25 to <30, ≥30 kg/m$^2$), and alcohol consumption (lifetime consumption in grams/week). BMI was considered a potential confounder as BMI differed between smoking groups and obesity (BMI ≥30) was previously associated with tumor methylation differences [42]. Similarly, alcohol consumption differed between smokers and never smokers and was previously correlated with breast tumor methylation patterns [43]. GLM stratified on hormone receptor (HR) status (HR-positive; ER+ and/or PR+, or HR-negative; ER−/PR−) was also performed as tumor methylation profiles are known to differ by HR status [39]. Volcano plots were utilized to display array-wide association patterns of differential methylation between smokers and never smokers, showing the estimated coefficients from the previously mentioned GLMs plotted against the negative logarithm of the raw p-values obtained from the associated Wald significance tests. The Benjamini-Hochberg false discovery rate (FDR) was calculated to adjust for multiple comparisons, and a threshold of <0.05 was considered significant. Spearman correlation coefficients and $p$-values were used to test the relationship between increasing duration of smoking among ever smokers or years since quitting among former smokers and tumor methylation at smoking-associated CpG loci identified from GLM.

Gene ontology analysis

The DAVID Bioinformatics Resources 6.8 Functional Annotation Tool (https://david.ncifcrf.gov/) was used to perform gene-GO term enrichment analysis to identify the most relevant GO terms associated with the genes found to be differentially methylated in breast tumors from smokers versus never smokers. Entrez gene IDs from each gene list were
compared to the background list of 609 genes evaluated from the Illumina Cancer Panel I array after filtering. Genes with more than one CpG site were listed only once in the analysis. We performed functional annotation clustering with default settings. Terms that were significantly enriched (Benjamini-Hochberg corrected \( p < 0.05 \)) are listed.

Correlation between methylation and gene expression

The relationship between methylation and gene expression at CpG loci associated with current (vs. never) smoking was tested using breast tumor data in The Cancer Genome Atlas (TCGA) \[44\] since gene expression data were not available in phase I of the CBCS. Methylation analysis in TCGA was performed using the Illumina Infinium HumanMethylation450 K array (450 K), and gene expression data were generated using RNA sequencing. Due to the small degree of overlap of CpG probes between the 450 K array and the Cancer Panel array used in this study, we examined the relationship between methylation and gene expression only for probes that directly overlapped the two platforms \[39, 42, 45\]. Pearson correlation coefficients were calculated in 581 TCGA breast tumors (from 44 blacks, 431 whites, 29 Asians, 77 of unknown race/ethnicity), or within hormone receptor-positive or hormone receptor-negative subsets, based on RNAseq (Illumina) log2 RSEM gene normalized expression values with methylation beta values for 450 K CpG probes, with significance set at \( p < 0.05 \).

Results

The characteristics of cases from CBCS phase 1 evaluated for tumor methylation according to smoking status are shown in Table 1. Current smokers were younger than never or former smokers \( (p = 0.02) \), smoked for longer \( (p < 0.0001) \), and began smoking at a younger age \( (p = 0.0008) \) compared with former smokers. Smokers were more likely to consume alcohol \( (p < 0.0001) \) and have lower BMI \( (p = 0.05) \) compared with never smokers. Otherwise, case subgroups defined by smoking status did not vary significantly on other demographic or clinical characteristics.

To determine whether tobacco smoke exposure was associated with differences in breast tumor epigenetic profiles, separate multivariable GLMs were used to compare methylation at 933 CpG sites between smokers and never smokers. As illustrated by volcano plots (Fig. 1a), breast tumors of current or ever smokers exhibited both increases and decreases in CpG methylation compared to never smokers at \( p < 0.05 \); similar results were obtained with or without inclusion of stage in the model (Online Resource Fig. S1). We previously observed differences in DNA methylation between HR+ (ER+ and/or PR+) and HR− (ER−/PR−) breast tumors \[39\]; therefore, to minimize the potential contribution of HR subtype heterogeneity to methylation differences, we examined the effect of smoking stratified on HR status. The effect of smoking on methylation differed by HR status, with HR+ breast tumors from smokers exhibiting hypermethylation (Fig. 1b). In contrast, HR− breast tumors from smokers exhibited a somewhat more robust pattern of mostly CpG hypomethylation compared with tumors from never smokers (Fig. 1c) which was stronger among long-term (>20 year) smokers than in those who smoked for shorter duration (Fig. 1d). The HR− breast tumor subset included several intrinsic subtypes based on immunoprofiles (basal-like,
HER2±/HR−, and unclassified), yet smoking was also associated with CpG hypomethylation in the most homogeneous subset of basal-like tumors (Fig. 1e). The divergent patterns of smoking-related differential methylation in HR− versus HR+ breast tumors seen in the volcano plots are summarized numerically in Fig. 1f. Patterns of differential hypo- or hypermethylation in former (vs. never) smokers were similar to those of current smokers (Online Resource Fig. S2).

Because methylation differences in current smokers are expected to be more stable than in former smokers based on studies in peripheral blood leukocytes (PBLs), we have focused on CpG loci differing in HR+ or HR− breast tumors from current smokers. A list of loci significant at p <0.05 is given in Online Resource Table S1. The mean percent change in methylation in HR− tumors at CpG loci differing between current and never smokers was +/-11% (range 0.8–39%). Two candidate genes (CDKN1A and IL12B) showed significant current smoking-related differential methylation after FDR correction (FDR adjusted p <0.05) in HR− breast tumors. Box plots in Fig. 2 illustrate methylation levels at top CpG loci related to current smoking in HR− cases; these include several genes previously found to differ in PBLs between current smokers and nonsmokers (RARA, ACVR1, SEPT9, CDKN1A, MCC, ATP10A, and PIK3R1). Most current smoking-related CpG loci in HR− tumors exhibited methylation levels in former smokers that were intermediate between those in current and never smokers (Fig. 2a), and exhibited progressive changes in methylation with increasing duration of smoking among ever smokers (Fig. 2b; Online Resource Table S2). Among former smokers, restoration of methylation toward the baseline (never smoking) level was observed with increasing time since quitting (Fig. 2c). For a number of genes differentially methylated in current smokers (ACVR1, CYP2E1, FGF1, EPHA2, KRT5, MCF2, PTHR1, RARRES1, SEPT9, SNCG, TRIM29, VBP1 in HR− cases, and COL1A1 and TAL1 in HR+ cases), two probes independently detected similar patterns of smoking-associated differential methylation (Online Resource Table S1). Boxplots for additional current smoking-related CpGs are provided in Online Resource Fig. S3.

Analysis of TCGA breast tumor data indicated that 39 current smoking-associated CpGs were common to the Illumina methylation platforms used in TCGA and CBCS. Approximately half of these CpGs showed significant inverse correlations between methylation and gene expression, overall or within HR-based subtypes (Online Resource Table S3), suggesting that smoking-related methylation differences may translate to alterations in tumor gene expression.

Recent studies in the CBCS indicated that smoking was more strongly associated with the development of luminal (HR+) breast cancers among blacks [16], but worse long-term survival among black smokers and smokers with ER− breast tumors (unpublished observations).

Therefore, we also examined the effect of current smoking on methylation profiles of HR+ or HR− breast tumors stratified by race. HR− breast tumors in blacks and whites similarly exhibited hypomethylation among current smokers (Fig. 3a, b), whereas among HR+ cases, black smokers exhibited a stronger pattern of hypermethylation (Fig. 3c, d); these patterns of smoking-related differential methylation are summarized in Fig. 3e. At the level of
individual probes, 19% of current smoking-related CpGs in HR− tumors from black cases overlapped with those of whites (Fig. 3f; Online Resources Table S4 and Table S5), including loci in the *CDM, FGF1, FRK, PIK3R1, SNURF,* and *TRIM29* genes, as shown in boxplots in Fig. 3g. In black smokers, three candidate CpGs in *RIPK4, SMO,* and *MAF* exhibited hypermethylation in HR+ tumors and were significant after FDR correction (Online Resource Table S5).

DAVID gene ontology analysis of genes differentially methylated in current versus never smokers showed that in HR− cases, smoking was associated with significant terms such as ‘response to organic substance,’ ‘response to stress,’ and ‘response to chemical stimulus’ (Online Resource Table S6). In HR+ tumors, significant GO terms were related mainly to regulation of RNA processes or transcription.

**Discussion**

The results of this study suggest that exposure to tobacco smoke is associated with modest differences in breast tumor methylation patterns. Current smokers with HR− breast tumors showed primarily CpG hypomethylation compared with never smokers, whereas smokers with HR+ breast tumors showed hypermethylation, and these changes were more pronounced with longer duration of smoking but were diminished with time since smoking cessation. Smoking-related differential methylation was not obviously explained by intrinsic subtype composition as these were very similar among the smoker and never smoker groups, and smoking-associated hypomethylation was also detected when analyses were restricted to the most homogeneous basal-like subtype. Importantly, we also adjusted for alcohol use and BMI as these differed between smokers and nonsmokers in the CBCS and previously were found to be associated with epigenetic differences in breast tumors [42, 43]. Interestingly, gene ontology analysis revealed different enrichment terms associated with smoking in HR− and HR+ tumors, with smoking-related terms in HR− tumors related to response to organic chemicals and stress, while those in HR+ tumors were associated with transcriptional regulation, suggesting that these tumor subsets may both be associated with smoking but via different pathways. Smoking-related epigenetic differences also varied somewhat by race in that the hypermethylation in HR+ breast tumors was more pronounced in black smokers, which may be consistent with epidemiologic findings from CBCS indicating that smoking was associated with the development of luminal breast cancer mainly among black women [16].

Several array-based studies have reported smoking-related methylation differences in cancers for which smoking is a major risk factor, including cancers of the lung [32, 33], bladder [34–36], and head and neck [46, 47]. In epigenome-wide studies of smoking in PBLs [27–30], hypomethylation was the predominant effect, with CpGs in the *AHRR* and/or *F2RL3* genes showing reduced methylation in current smokers. Although these genes were not included on the Cancer Panel I array, several others previously reported to show changes in methylation in current smokers also differed in breast tumors [28–30, 48]. One recent study investigated smoking-associated differences in breast tumor promoter methylation at 13 candidate genes and found that current smoking (vs. non-current smoking) was associated with hypomethylation of *DAPK* [49]; however, ours is the first to report on the
relationship between tobacco smoke exposure and breast tumor methylation profiles using an array-based approach.

Tobacco smoke is comprised more than 7000 chemicals, including 69 established carcinogens [50, 51], and is considered to be a complete carcinogen with the ability to act during tumor initiation, promotion, and progression [51, 52]. Nicotine, the primary addictive constituent of cigarette smoke, may have in vitro genotoxic [53], promotional [54], and anti-apoptotic effects [55] in breast tumor cells. Thus, the pro-carcinogenic effects of smoking in breast epithelium could be mediated via alterations in DNA methylation throughout the continuum of the disease process, both early in breast cancer initiation and/or later during tumor growth and progression [51, 52]. A number of mechanisms have been posited for how smoking alters DNA methylation, including induction of DNMT1 expression [56], recruitment of DNMT1 and subsequent methylation of CpGs adjacent to sites of DNA damage [57, 58], modulation of gene expression by nicotine, including of CDKN1A [59, 60], alteration of expression and DNA binding stability of transcription factors, e.g., Sp1, which can protect CpG sites from methylation [61, 62], and hypoxia, which can lead to increases in the methyl donor, S-adenosylmethionine [63]. Smoking is also associated with reduced serum folate levels [64, 65], which could result in impaired DNA methylation, especially in rapidly growing HR− tumors. Indeed, triple-negative (including basal-like) breast tumors frequently overexpress the folate receptor, FOLR1, which is thought to provide a growth advantage to tumor cells, particularly in a low-folate environment [66].

Genes found to be differentially methylated in the tumors of smokers have varied roles in breast or other cancers. CDKN1A (cyclin-dependent kinase inhibitor, p21/WAF1) is a key negative regulator of the cell cycle and cell proliferation [67]. CDKN1A hypermethylation in breast tumors was associated with reduced mRNA and protein expression [68, 69], and in endometrial tumors, p21 protein expression was inversely associated with current smoking [70]. The pro-inflammatory cytokine, IL12, of which IL12B (p40) is a subunit, functions in cell-mediated immunity and host immune responses to tumors [71]. Genetic variants in IL12B are associated with breast cancer risk [72] and survival in chemotherapy-treated ER− breast cancer patients [73]. Other smoking-associated genes have roles in DNA repair (OGG1, BRCA1, FANCE, and TDG), signal transduction (FRK, TYRO3, EPHA2, EPHB1, MST1R, and ACVR1), cell cycle regulation (CDKN1A, MAP3K8, and MAPK9), inflammation and immune function (IL12B, IL13, IL6, and HLA-F), xenobiotic or drug metabolism (CYP2E1 and UGT1A1), or function as growth factors (FGF1, HGF, and PDGFRA) or transcription factors (TRIM29, RARA).

Strengths of this study include the large size and population-based design of the CBCS and inclusion of cases and breast tumors with relatively complete risk factor, histopathologic and clinical data. Limitations include the use of intrinsic subtyping based on a panel of IHC protein expression markers rather than gene expression subtyping (due to lack of mRNA availability from archival FFPE specimens), which may have resulted in some misclassification, mainly among luminal breast cancers [74]. However, smoking-associated differential hypomethylation observed among HR− tumors was also seen within the more accurately classified basal-like subtype [74], reducing the likelihood that the alterations in methylation in HR− tumors were due to subtype heterogeneity rather than smoking.
Moreover, while cell-type heterogeneity is often an inherent problem in tissue analysis, the breast tumors in CBCS were manually dissected to enrich for tumor cells. Although the data were collected on a first generation methylation array that over-sampled genes in cancer-related pathways, many genes on the platform had strong coverage for the best-studied methylation sites in breast cancer research. Additionally, as this study focused on CpG methylation in gene promoters, the results may uniquely reflect effects of smoking in these genomic regions.

Our results indicate that tobacco smoke exposure is associated with modest differences in breast tumor epigenetic patterns and that these vary with HR status and race. Additional studies are needed to confirm our findings and to clarify the mechanisms underlying the effects of smoking on tumor methylation.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

We thank the participants and staff of the CBCS for their ongoing support of the study.

**Funding:** This research was supported by Grants to KC from Susan G. Komen Foundation (Grant #KG081397) and from the University Cancer Research Fund of the Lineberger Comprehensive Cancer Center at the University of North Carolina at Chapel Hill. The Carolina Breast Cancer Study was also funded by the University Cancer Research Fund of North Carolina and the National Cancer Institute Specialized Program of Research Excellence (SPORE) in Breast Cancer (NIH/NCI P50-CA58223).

**References**


59. Satta R, Maloku E, Zhubi A, Pipiri F, Hajos M, Costa E, Guidotti A. Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in


Fig. 1.
Patterns of breast tumor differential methylation in smokers compared with never smokers. Generalized linear regression models (GLM) (logit link) adjusted for age, race, menopausal status, stage, BMI, and alcohol consumption were used to identify CpG loci differentially methylated in breast tumors in smokers versus never smokers. Volcano plots display array-wide patterns of breast tumor differential methylation in current or ever smokers among: a all cases, b HR+ cases, c HR− cases, d HR− cases according to smoking duration (long-term >20 years or shorter-term ≤20 years) among ever smokers, or e basal-like cases. Each volcano plot displays the negative log of unadjusted p-values for differences in β (proportion DNA methylated) at each probe on the y axis versus the correlation coefficient for methylation at each CpG locus on the x axis. Probes that fall above the broken line are significant at p < 0.05. Probes hypomethylated in smokers have negative coefficients, while probes hypermethylated in smokers have positive coefficients. f Bar graph summarizing numbers of differentially hypomethylated or hypermethylated CpG probes at p < 0.05 in
ever smokers, current smokers, long-term (>20 year), or shorter-term (<20 year) smokers versus never smokers among all cases or by HR status
Fig. 2.
Boxplots illustrating methylation levels at top CpGs differing in methylation between current and never smokers with HR− breast tumors. 

**A Smoking Status**
- Methylation differences according to smoking status in current (C) (n = 37) or former (F) (n = 34) smokers compared with never smokers (N) (n = 92).

**B Smoking Duration in Years (ever smokers)**
- Methylation with increasing duration of smoking in years among ever smokers: N (n = 92), ≤10 (n = 25), 11–20 (n = 16), or >20 years (n = 29).

**C Years Since Quitting (former smokers)**
- Methylation with years since quitting among former smokers: C (n = 37), ≤10 (n = 18), 11+ (n = 16), N (n = 92). Never or current smokers are included for reference in some plots as appropriate.

*Boxplot boundaries indicate the interquartile range, and mean and median are indicated by the black bar and diamond, respectively.*
Fig. 3.
Patterns of breast tumor differential methylation in black or white current smokers versus never smokers with HR+ or HR− breast cancer. Generalized linear regression (GLM) models adjusted for age, race, menopausal status, stage, BMI, and alcohol consumption were used to compare breast tumor methylation beta values between current and never smokers. a Volcano plots display array-wide patterns of breast tumor differential methylation in current versus never smokers among a black HR− cases, b black HR+ cases, c white HR− cases, and d white HR+ cases. Probes that fall above the broken line are significant at \( p < 0.05 \). Probes hypomethylated in smokers have negative coefficients and probes hypermethylated in smokers have positive coefficients. e Bar graph summarizing numbers of differentially hypomethylated or hypermethylated CpG probes in black and white current smokers with HR− or HR+ breast tumors. f Venn diagrams summarizing the overlap of CpG probes differentially methylated in current smokers among black and white cases with HR− or HR+ breast tumors.

Breast Cancer Res Treat. Author manuscript; available in PMC 2018 June 01.
breast cancers. **g** Boxplots of selected smoking-related CpG loci illustrating the range of beta values in HR– breast tumors from black or white current or former smokers versus never smokers
Table 1

Selected breast cancer case characteristics by smoking status

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Current smokers (N = 125) No. (%)</th>
<th>Former smokers (N = 124) No. (%)</th>
<th>Never smokers (N = 268) No. (%)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Case characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age, mean (SD)</td>
<td>46.9 (10.6)</td>
<td>51.0 (11.3)</td>
<td>50.4 (12.5)</td>
<td>0.009</td>
</tr>
<tr>
<td>Age 50+ years</td>
<td>35 (28.0)</td>
<td>53 (42.7)</td>
<td>111 (41.4)</td>
<td>0.02</td>
</tr>
<tr>
<td>&lt;50 years</td>
<td>90 (72.0)</td>
<td>71 (57.3)</td>
<td>157 (58.6)</td>
<td></td>
</tr>
<tr>
<td><strong>Race</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White/Other*</td>
<td>76 (60.8)</td>
<td>77 (62.1)</td>
<td>148 (55.2)</td>
<td>0.35</td>
</tr>
<tr>
<td>Black</td>
<td>49 (39.2)</td>
<td>47 (37.9)</td>
<td>120 (44.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Menopausal status</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Postmenopausal</td>
<td>55 (44.0)</td>
<td>59 (47.6)</td>
<td>128 (47.8)</td>
<td>0.77</td>
</tr>
<tr>
<td>Premenopausal</td>
<td>70 (56.0)</td>
<td>65 (52.4)</td>
<td>140 (52.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Alcohol consumption</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>16 (12.8)</td>
<td>21 (16.9)</td>
<td>113 (42.2)</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>Ever</td>
<td>109 (87.2)</td>
<td>103 (83.1)</td>
<td>155 (57.8)</td>
<td></td>
</tr>
<tr>
<td><strong>Body mass index (BMI)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;25 (normal)</td>
<td>60 (48.0)</td>
<td>39 (32.2)</td>
<td>86 (33.0)</td>
<td>0.05</td>
</tr>
<tr>
<td>25 to &lt;30 (overweight)</td>
<td>29 (23.2)</td>
<td>36 (29.8)</td>
<td>80 (30.7)</td>
<td></td>
</tr>
<tr>
<td>&gt;30 (obese)</td>
<td>36 (28.8)</td>
<td>46 (38.0)</td>
<td>95 (36.4)</td>
<td></td>
</tr>
<tr>
<td><strong>Smoking characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cigarettes/day</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;1/2 pack</td>
<td>32 (25.6)</td>
<td>51 (41.5)</td>
<td>–</td>
<td>0.02</td>
</tr>
<tr>
<td>1/2–1 pack</td>
<td>49 (39.2)</td>
<td>44 (35.8)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>&gt;1 pack</td>
<td>44 (35.2)</td>
<td>28 (22.8)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Duration of smoking (yrs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;10</td>
<td>19 (15.2)</td>
<td>59 (48.0)</td>
<td>–</td>
<td>&lt; 0.0001</td>
</tr>
<tr>
<td>11–20</td>
<td>22 (17.6)</td>
<td>29 (23.6)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>&gt;20</td>
<td>84 (67.2)</td>
<td>35 (28.5)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Time since quitting (yrs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤10</td>
<td>–</td>
<td>54 (43.6)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>11+</td>
<td>–</td>
<td>70 (56.4)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>Age at initiation (yrs)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;15</td>
<td>31 (24.8)</td>
<td>13 (10.5)</td>
<td>–</td>
<td>0.0008</td>
</tr>
<tr>
<td>16–20</td>
<td>54 (43.2)</td>
<td>81 (65.3)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td>&gt;20</td>
<td>40 (32.0)</td>
<td>30 (24.2)</td>
<td>–</td>
<td></td>
</tr>
<tr>
<td><strong>Clinical characteristics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage (\text{a})</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>45 (39.5)</td>
<td>45 (39.8)</td>
<td>88 (34.7)</td>
<td>0.52</td>
</tr>
<tr>
<td>II</td>
<td>51 (44.7)</td>
<td>55 (48.7)</td>
<td>139 (54.7)</td>
<td></td>
</tr>
<tr>
<td>Characteristic</td>
<td>Current smokers (N = 125) No. (%)</td>
<td>Former smokers (N = 124) No. (%)</td>
<td>Never smokers (N = 268) No. (%)</td>
<td>p-value</td>
</tr>
<tr>
<td>---------------------</td>
<td>-----------------------------------</td>
<td>----------------------------------</td>
<td>---------------------------------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>15 (13.2)</td>
<td>9 (8.0)</td>
<td>21 (8.3)</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>3 (2.6)</td>
<td>4 (3.5)</td>
<td>6 (2.4)</td>
<td></td>
</tr>
<tr>
<td>ER/PR status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER+/PR+</td>
<td>59 (49.6)</td>
<td>69 (58.1)</td>
<td>122 (46.6)</td>
<td>0.38</td>
</tr>
<tr>
<td>ER+/PR−</td>
<td>15 (12.6)</td>
<td>9 (7.6)</td>
<td>24 (9.2)</td>
<td></td>
</tr>
<tr>
<td>ER−/PR+</td>
<td>8 (6.7)</td>
<td>7 (5.9)</td>
<td>24 (9.2)</td>
<td></td>
</tr>
<tr>
<td>ER−/PR−</td>
<td>37 (31.1)</td>
<td>34 (28.6)</td>
<td>92 (35.1)</td>
<td></td>
</tr>
<tr>
<td>Intrinsic subtype</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All cases</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luminal A</td>
<td>54 (54.0)</td>
<td>52 (53.6)</td>
<td>106 (49.1)</td>
<td>0.99</td>
</tr>
<tr>
<td>Luminal B</td>
<td>15 (15.0)</td>
<td>14 (14.4)</td>
<td>36 (16.7)</td>
<td></td>
</tr>
<tr>
<td>Basal-like</td>
<td>19 (19.0)</td>
<td>20 (20.6)</td>
<td>47 (21.8)</td>
<td></td>
</tr>
<tr>
<td>HER2+/ER−</td>
<td>6 (6.0)</td>
<td>5 (5.2)</td>
<td>15 (6.9)</td>
<td></td>
</tr>
<tr>
<td>Unclassified</td>
<td>6 (6.0)</td>
<td>6 (6.2)</td>
<td>12 (5.5)</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luminal A</td>
<td>19 (52.8)</td>
<td>18 (47.4)</td>
<td>44 (46.3)</td>
<td>0.79</td>
</tr>
<tr>
<td>Luminal B</td>
<td>4 (11.1)</td>
<td>4 (10.5)</td>
<td>14 (14.7)</td>
<td></td>
</tr>
<tr>
<td>Basal-like</td>
<td>9 (25.0)</td>
<td>12 (31.6)</td>
<td>25 (26.3)</td>
<td></td>
</tr>
<tr>
<td>HER2+/ER−</td>
<td>4 (11.1)</td>
<td>3 (7.9)</td>
<td>6 (6.3)</td>
<td></td>
</tr>
<tr>
<td>Unclassified</td>
<td>0 (0)</td>
<td>1 (2.6)</td>
<td>6 (6.3)</td>
<td></td>
</tr>
<tr>
<td>White/other</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Luminal A</td>
<td>35 (54.7)</td>
<td>34 (57.6)</td>
<td>62 (51.2)</td>
<td>0.83</td>
</tr>
<tr>
<td>Luminal B</td>
<td>11 (17.2)</td>
<td>10 (17.0)</td>
<td>22 (18.2)</td>
<td></td>
</tr>
<tr>
<td>Basal-like</td>
<td>10 (15.6)</td>
<td>8 (13.6)</td>
<td>22 (18.2)</td>
<td></td>
</tr>
<tr>
<td>HER2+/ER−</td>
<td>2 (3.1)</td>
<td>2 (3.4)</td>
<td>9 (7.4)</td>
<td></td>
</tr>
<tr>
<td>Unclassified</td>
<td>6 (9.4)</td>
<td>5 (8.5)</td>
<td>6 (5.0)</td>
<td></td>
</tr>
</tbody>
</table>

*The white/other cases included 291 Caucasians, 3 American Indians, 6 Asians, and 1 other.

According to the AJCC breast tumor staging guidelines. The differences between column totals and numbers within some categories are due to missing data in some cases for some variables.