Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

Cancer Epidemiol Biomarkers Prev. Author manuscript; available in PMC 2019 January 01.

Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2018 January ; 27(1): 67–74. doi:10.1158/1055-9965.EPI-17-0404.

Pre-diagnostic smoking is associated with binary and quantitative measures of ER protein and *ESR1* mRNA expression in breast tumors

Eboneé N. Butler, PhD[a], **Jeannette T. Bensen, PhD**[a], **Mengjie Chen, PhD**[b], **Kathleen Conway, PhD**[a], **David B. Richardson, PhD**[a], **Xuezheng Sun, PhD**[a], **Joseph Geradts, MD**[c], **Andrew F. Olshan, PhD**[a], and **Melissa A. Troester, PhD**[a]

[a]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina

[b]Department of Medicine, University of Chicago

[c]Department of Pathology, Dana-Farber Cancer Institute

Abstract

Introduction—Smoking is a possible risk factor for breast cancer and has been linked to increased risk of estrogen-receptor positive (ER+) disease in some epidemiologic studies. It is unknown whether smoking has quantitative effects on ER expression.

Methods—We examined relationships between smoking and ER expression from tumors of 1,888 women diagnosed with invasive breast cancer from a population-based study in North Carolina. ER expression was characterized using binary (+/−) and continuous measures for ER protein, *ESR1* mRNA, and a multigene luminal score (LS) that serves as a measure of estrogen signaling in breast tumors. We used logistic and linear regression models to estimate temporal and dose-dependent associations between smoking and ER measures.

Results—The odds of ER+, *ESR1*+, and LS+ tumors among current smokers (at time of diagnosis), those who smoked 20 or more years, and those who smoked within 5 years of diagnosis were nearly double those of non-smokers. Quantitative levels of *ESR1* were highest among current smokers compared to never smokers overall [mean (log2) = 9.2 vs 8.7, $p < 0.05$] and among ER+ cases; however, we did not observe associations between smoking measures and continuous ER protein expression.

Conclusions—In relationship to breast cancer diagnosis, recent smoking was associated with higher odds of the ER+, *ESR1*+, and LS+ subtype. Current smoking was associated with elevated *ESR1* mRNA levels and an elevated luminal score, but not with altered ER protein.

INTRODUCTION

Epidemiologic studies have demonstrated distinct risk factor profiles for breast cancer subtypes classified according to estrogen-receptor (ER) status (1,2). In a previous analysis from the Carolina Breast Cancer Study, we reported a modest increased risk of ER+ breast cancer in association with pre-diagnostic smoking (3). Several contemporary epidemiologic studies have reported similar associations (3–5). These findings raise the question of whether

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

smoking exposure could be linked to altered estrogen-receptor expression or pathway activity.

If pre-diagnostic smoking modulates ER expression, quantitative levels of ER may differ between breast tumors of smokers and non-smokers. This hypothesis could be tested at the protein level. However clinical IHC assays are tuned to be maximally sensitive for the detection of ER, which leads to saturated signals and suppression of ER's dynamic expression range (6). RNA assays may have a wider dynamic range and therefore may better capture quantitative changes in *ESR1* expression. In addition, multigene scores such as the PAM50 Luminal gene signature (6,7) capture estrogen-signaling across multiple targets, and may offer improved resolution when examining smoking in relation to ER expression. By evaluating quantitative variation in ER expression in relation to smoking, we seek to more directly assess a link between smoking and estrogen-related pathways to breast cancer.

In this large population-based study, we evaluated smoking in association with binary classifications and quantitative measures for ER protein, *ESR1* mRNA, and a multigene luminal score that captures estrogen signaling in tumors; we examined temporal and dose-dependent measures of smoking in relation to each biomarker.

MATERIALS AND METHODS

Study Population

Phase III of the Carolina Breast Cancer Study (CBCS III) is a population-based case-only study that combines epidemiology and molecular biology to examine environmental and genetic risk factors for molecular subtypes of breast cancer (8,9). To be eligible for inclusion, patients must have been female and received a first and primary diagnosis of breast cancer between May 1, 2008 and October 31, 2013. The patient also must have resided in the 44-county North Carolina study region and been between the ages of 20 and 74 at the time of diagnosis. To examine potential differences by age and race, the CBCS employed a randomized recruitment strategy that was designed to oversample young and African American women (10).

Breast cancer cases were identified by a rapid case ascertainment system, implemented through collaboration between Lineberger Comprehensive Cancer Center (LCCC) and the North Carolina Central Cancer Registry (NCCCR). Briefly, CBCS contacted the patient's primary physician to obtain permission to invite the patient into the study, yielding an overall response rate of 70% and a total of 2,998 women. Study participants were asked to consent to a nurse-administered in-person interview that took place in the study participant's home or another pre-arranged location. The average time between study enrollment and interview was 6 months. The nurse administered questionnaire included items on family and personal medical history, reproductive history, smoking, alcohol, diet, medication use and occupational history. Upon consent, the nurse also collected a blood sample and objective anthropometric measurements of height (m), weight (kg), waist (m), and hip (m) circumference. All study activities and protocols were approved by the Institutional Review Board at the University of North Carolina at Chapel Hill School of Medicine. Study

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

participants provided written informed consent and all research activities were conducted in accordance with the U.S. Common Rule.

Study Design

Tumor gene expression analysis—The CBCS includes protein and RNA expression data on genes involved in estrogen-signaling. At the time of interview, investigators asked study participants for written permission to obtain formalin-fixed, paraffin-embedded (FPPE) tumor blocks or tissue slides from the hospital where the diagnostic surgery was to be performed. Tumor blocks were used to construct tissue microarrays (TMAs) for IHC staining, where each patient's tumor was represented by 1 to 4 cores on the microarray. To enrich for tumor cellularity, cores were taken from within a tumor region that was annotated on the tumor block by a pathologist. Hematoxylin and eosin (H&E) slides were constructed for the top, middle, and bottom portion of each block. Cores were excluded if tumor was not included on top and bottom slides. RNA was extracted using the Qiagen RNeasy FFPE kit and protocol applied to separate cores or sections from the same tumor block. The current analysis includes data for 1,888 women analyzed for ER protein level by IHC and 993 women analyzed for RNA expression (Table 1).

Estrogen receptor protein: Immunohistochemical staining of CBCS3 was described in Allott et al. (11). Automated quantification of ER protein was determined by a Genie classifier and the Aperio nuclear v9 algorithm (Aperio Technologies, Vista, CA) (11). We calculated percent positivity for ER as the product of positively stained tumor cells for each core, multiplied by its core-specific weight, summed across all cores per patient (ER WT%). We assigned a cut point of ≥ 10% for 'ER positive' tumors; 1% to < 10% for 'ER borderline' tumors; and < 1% for 'ER negative' tumors. For the ER binary classification, 'ER borderline' tumors were combined with 'ER negative tumors' based on our previous observations that borderline tumors shared other molecular features with ER-negative disease (11).

ESR1 mRNA: *ESR1* was quantified using Nanostring technology (12). Briefly, total *ESR1* mRNA counts were assayed using an *ESR1*-specific molecular probe, which hybridizes to RNA fragments in solution. Hybrids are then fixed to a solid matrix and counted using microscopic imaging, yielding raw mRNA counts. Quality control and data normalization were performed using the NanoStringNorm R package (13). Data were first normalized to the geometric means of 6 internal positive controls and subsequently to the geometric means of 5 reference genes. Normalized *ESR1* counts were log2 transformed, yielding a bimodal Gaussian distribution of the data. We used the mclust R package and an unsupervised analysis to classify the two distributions as *ESR1*– or *ESR1*+, reflecting low and high expression, respectively (14). *ESR1*- tumors had log2 values ranging between 0 to 8.35 and *ESR1*+ tumors had log2 values ranging between 8.38 to 15.64.

PAM50 Subtype and Luminal Score: Breast cancer intrinsic subtype was measured using the RNA-based "PAM50 signature" (7). Differential expression of the 50-gene signature was used to categorize breast cancers into 4 intrinsic subtypes: Luminal A, Luminal B, HER2E, and Basal-like. Each case was classified based upon highest Pearson correlation with a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

centroid defined for each subtype. The PAM50 Luminal gene signature is embedded within the larger signature and includes 8 highly correlated genes associated with Luminal type breast cancers, which are characterized by high ER expression (6,7). The 8 genes include: *BAG1, ESR1, FOXA1, GPR160, NAT1, MAPT, MLPH,* and *PGR*. Each gene was quantified and normalized according to procedures for *ESR1*, as described above. To calculate the Luminal Score (LS), we took the average of the normalized values of the 8 genes. Normalized and transformed values for LS followed a bimodal Gaussian distribution. We used the mclust R package to classify the Luminal Score as LS– or LS+, reflecting low and high scores, respectively. LS– tumors had log2 values ranging between 3.26 to 7.57 and LS+ tumors had log2 values ranging between 7.58 to 11.37. *ESR1* mRNA and the 8 genes embedded in the Luminal Score were assayed along with other genes included in 1 of 3 Nanostring batches or code sets. Samples were randomized to batch and all Nanostring analyses were adjusted for 'code set' in order to minimize potential batch effects.

Smoking Exposure Assessment—Pre-diagnostic history of smoking was obtained during the nurse-administered in-person interview and includes data on smoking duration, frequency, and dose. Women in CBCS were considered ever smokers if they smoked at least 100 cigarettes during their lifetimes. Smoking history was defined as 'ever' or 'never' (history); smoking status defined as 'current', 'former', or 'never' (status); age at smoking initiation measured in years (initiation); smoking duration measured as the total number of years of smoking between initiation and current use or cessation (duration); number of cigarettes smoked per day (dose); and age at smoking cessation, where applicable. Pack-years were defined as a cumulative measure of the number of cigarette packs smoked per day, divided by smoking duration in years. Similarly, pack-decades were defined as cumulative measures of cigarette packs smoked per day, over 10-year intervals.

Covariate Assessment—Potential confounders include: first-degree family history of breast cancer defined as breast cancer diagnosis for mother or a full female sibling (15); alcohol consumption defined as having any history of alcohol use (16–18); ever having breast fed (1); body mass index (BMI kg/m2) (1); parity defined as number of full-term births (1,16); history of oral contraceptive use (19); hormone replacement therapy use (19); menopausal status; meeting physical activity guidelines; age; and race.

Participants were also asked for permission to obtain pathology reports and medical records from the treating facilities. Clinical and pathological data abstracted from medical records and pathology reports included tumor size, stage, and node status; these tumor characteristics were considered as potential confounders of the relationship between smoking and ER expression. For all cases, a single pathologist (JG) determined tumor grade.

Data Analysis

For binary ER, *ESR1*, and LS-defined subtype variables, we used generalized logit models to estimate odds ratios (ORs) and 95% confidence intervals (CIs) for categorical measures of smoking. To evaluate temporal and dose-dependent associations between smoking and subtype, we first estimated the associations via logistic regression for a one unit increase in pack-decades. We compared this cumulative exposure model to an exposure-time-windows

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

model (i.e., piecewise logistic regression model) for three time intervals for time since smoking cessation: 0 – 10 years; 11–20 years; and > 20 years. We used a likelihood ratio test (LRT) to compare the deviances between the two models, the difference of which follows a chi-square distribution with 2 degrees of freedom.

To evaluate the hypothesis that odds of ER, *ESR1*, and LS-positive subtypes vary with time since smoking exposure, we used a generalized logit model with a lognormal latency function to calculate time weighted exposure estimates for the 40-year period preceding breast cancer diagnosis. The latency period between smoking exposure and breast cancer occurrence is thought to be as much as 40 years and the lognormal distribution has been used to describe variation in disease risk with time since exposure (20). Specifically, the lognormal latency function can be used to describe the rise, peak, and decline in risk or log-odds with respect to time since exposure. The highest weights are assigned during the time interval where smoking is associated with the greatest odds of ER+, *ESR1*+, or LS+ breast cancer and may signify the most etiologically-relevant time interval for smoking exposure. The macro used to model the lognormal latency function is described in Richardson (2009) (21).

We used linear regression to model the relationship between continuous measures of ER *ESR1*, LS and categorical measures of smoking, adjusted for age, race, and Nanostring batch (where applicable). We calculated the estimated value of continuous biomarker expression for each individual, based on coding of the smoking exposure and covariates (age, race, and Nanostring batch) pattern. Expression levels for each biomarker were described according to interquartile range and visualized using box plots within categories of smoking.

All analyses were conducted using SAS 9.4 (SAS Institute Inc, Cary, NC) and R version 3.3.3.

RESULTS

Figure 1 illustrates the relationships between categorical and continuous measures for luminal score, *ESR1* mRNA, and ER protein. Figure 1A shows distinct clusters for *ESR1* and LS, reflecting low and high expression for each. We compared binary classifications for ER protein as measured by IHC to those for *ESR1* and LS and observed moderate to good values for sensitivity (se) and specificity (sp) (ER vs. *ESR1*: se=92%, sp=86% and ER vs. LS: se=89%, sp=85%). Weighted percent ER protein (ER WT%) was positively correlated with log2 values for *ESR1* mRNA (r=0.70, $p < 0.01$< Figure 1B), however, *ESR1* mRNA appeared to have a greater dynamic range compared to ER WT%. Among ER+ tumors, quantitative protein values tended to saturate the upper end of the percentage range.

Categorical measures of pre-diagnostic smoking history, dose, and duration were associated with increased odds of ER+, *ESR1*+, and LS+ subtypes (Supplemental Figures 1, 2, and 3). ER+ breast tumors were most common among ever smokers compared to never smokers (OR = 1.51 95% CI: 1.15, 1.97) as shown in Supplemental Figure 1. When stratified by smoking status at time of diagnosis, current smokers were twice as likely to be ER+ compared to never smokers (OR=1.89 95% CI: 1.33, 2.70); former smokers had slightly

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

elevated odds of ER+ breast cancer (OR = 1.25 95% CI: 0.91, 1.73). Smoking duration of 20 years or more had elevated odds of ER+ breast cancer (OR = 1.79 95% CI: 1.26, 2.56). We observed more modestly elevated odds of the ER+ subtype for shorter duration of smoking. Women who smoked <1/2 or 1/2 to 1 packs of cigarettes per day had increased odds of ER+ breast cancer [(OR = 1.48 95% CI: 1.04, 2.10) and (OR = 1.57 95% CI: 1.09, 2.26), respectively]. However, for the highest category for smoking dose (> 1 pack/day), we observed slightly weaker odds for ER+ tumors (OR = 1.44 95% CI: 087, 2.37). With respect to 'time since smoking cessation', smoking within 5 years of breast cancer diagnosis was associated with a 60% increased odds of having ER+ breast cancer (OR 1.59 95% CI 1.15, 2.20). In general, we observed similar patterns of association between smoking measures and the *ESR1*+ and LS+ subtypes (Supplemental Figures 2 and 3). Notably, the magnitudes of the ORs were slightly higher for the RNA-based measures, particularly for smoking duration and time since smoking exposure.

We also evaluated the distribution of Luminal A, Luminal B, and Basal-like intrinsic subtypes with respect to smoking status at time of diagnosis. We observed a higher frequency of Luminal A vs. Basal-like tumors among current and former smokers [Current: 64.3% vs. Never: 54.2%, Frequency Difference (95% CI) = 10.2% (3.0%, 20.1%) and Former: 66.3% vs. Never: 54.2%, Frequency Difference (95% CI): 12.1% (3.2%, 21.1%)]. When adjusted for age and race, relative frequency of Luminal A tumors remained elevated, but the difference estimates were slightly attenuated. We observed no substantial difference in the proportion of Luminal B vs. Basal-like breast cancer according to smoking status.

Recency of smoking appeared to alter several of the estrogen-related biomarkers we examined (Table 2). Our cumulative exposure models suggest that a 1-unit increase in pack-decades was associated with a 10% to 18% increase in the odds of having a 'positive' subtype: ER+ (OR = 1.09 95% CI: 0.99, 1.20), *ESR1*+ (OR = 1.18 95% CI: 1.04, 1.34), and LS+ (OR = 1.18 95% CI: 1.04, 1.35). Moreover, for the exposure time-windows models, total pack-decades smoked within 10 years of a breast cancer diagnosis was associated with the greatest odds of having ER+, *ESR1*+, or LS+ breast cancer when compared to exposure accumulated between 10 and 20 or greater than 20 years prior to diagnosis. However, results from our likelihood ratio test suggest that the exposure time-windows model provides improved fit over the cumulative exposure model for LS-defined subtypes (p=0.04), but did not substantially improve data fit for the ER (p=0.63) and *ESR1* subtypes (p=0.27).

Current smoking was associated with the greatest odds of the LS+ breast cancer subtype. Figure 2 illustrates variation with time of exposure for the association between pack-decades and LS+ breast cancer for the 40-year period preceding breast cancer diagnosis. Our latency model with lognormal weighted exposures demonstrated increased odds of the LS+ subtype for pre-diagnostic smoking proximal to time of diagnosis. A likelihood ratio test comparing the lognormal latency model to the cumulative exposure model for the same 40-year period did not suggest that our latency model provided improved fit for the data (LRT = 4.2, 2 df). However, the dose-response parameter estimate in our latency model was statistically significant, thereby suggesting the peak in odds proximal to diagnosis may be the most etiologically relevant time point for smoking and ER+ breast cancer occurrence.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

RNA levels were also more quantitatively sensitive to differences in smoking history. Supplemental Tables 1 and 2 present estimated biomarker expression values for ER protein, *ESR1* mRNA, and the luminal score, adjusted for age, race, and Nanostring batch (where applicable). In general, ER protein levels did not vary across smoking exposures for breast cancer cases overall or when restricted to ER+ cases. Compared to never smokers, we observed the highest levels of *ESR1* mRNA and the highest luminal scores among current smokers [(mean (log2) = 9.2 vs. 8.7, $p < 0.05$) and (mean (log2) = 8.3 vs. 7.9, $p < 0.05$), respectively]. When restricted to ER+ breast cancer cases, we still observed higher levels of *ESR1* among current smokers, however the luminal score association was attenuated. Figures 3 and 4 visualize estimated expression values for *ESR1* and LS among 'Never', 'Former', and 'Current' smokers. We explored whether the luminal score and *ESR1* levels varied in association with smoking after stratification by age. We found that while *ESR1* and Luminal Scores were slightly higher among older women – consistent with higher rates of ER positive disease in older women – the general patterns of association with smoking status were similar by age. Likewise, we did not see evidence of effect modification or confounding by race.

DISCUSSION

Findings from our study lend quantitative support to the hypothesis that smoking could be linked to estrogen-mediated pathways in breast tumors. In our case-only study of nearly 2,000 patients, we observed increased odds of the ER+ subtype for temporal and dose-dependent measures of smoking. We also demonstrate quantitative changes in ER-related tumor subtypes characterized by *ESR1* mRNA and a multigene luminal score (LS). Increased odds of ER+, *ESR1*+, and LS+ subtypes was most apparent among women who were self-reported current smokers at time of diagnosis. Logistic regression models with latency parameters allowed us to simultaneously model dose, duration, and time of exposure to demonstrate that the most etiologically relevant period for smoking and ER-defined breast cancer may be during pre-diagnostic smoking closest in time to diagnosis. In addition, we observed that current smoking was associated with increased quantitative levels for *ESR1*, but not ER protein, which may suggest that RNA more sensitively captures biological differences when compared to ER protein expression.

Contemporary epidemiologic studies have demonstrated positive associations between smoking and ER+ breast cancer with estimates ranging between 10%–50% increased risk among current or former smokers (3,4,22,23). Our case-only analysis in CBCS Phase III demonstrated that relative to non-smokers the odds of having ER+ vs. ER- breast cancer was approximately double among current smokers. These findings are consistent with our previous case-control analysis in the Carolina Breast Cancer Study Phase I and II, which also showed increased risk of ER+ disease among smokers and heterogeneity of ORs for the Luminal (ER+) and Basal-like (ER-) subtypes (3,8). In that study, we observed a positive association between smoking and ER+ risk but no association between smoking and the ER- subtype– a pattern that has been observed in other studies performed in US populations (4,22). However, in contrast, other studies of smoking and breast cancer risk in Swedish, Swiss, and Australian populations have demonstrated positive associations between smoking and the ER- breast cancer subtype (24–26). These conflicting observations may reflect

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

temporal differences in exposure, behavioral patterns, or may also be an artifact of differing methods used to assay ER expression (e.g., ligand-binding, immunoreactivity). Varying methods used to assay ER protein expression may also result in different thresholds for ER-positivity. Thus, a careful investigation of the relationship between smoking and ER-defined breast cancer subtypes should consider era, methodological approaches, and characteristics of population of interest.

In both clinical and research settings, immunohistochemistry has been used as the standard for estrogen-receptor quantification in breast tumors (27). IHC is highly sensitive for the detection of ER protein, serving as an excellent marker to guide clinical decision making. However, protein saturation may preclude studying subtler, quantitative differences in association with etiologic factors. Our study addresses this potential limitation by using *ESR1* mRNA counts to characterize breast tumors as *ESR1*+ and *ESR1*-. Unlike ER protein expression values for percent positivity, the log2 transformed *ESR1* mRNA counts in our study follow a bimodal Gaussian distribution, identifying two distinct classes of breast tumors reflecting low and high expression of *ESR1*. Based on *ESR1*+ subtypes, where current smoking, long smoking duration of more than 20 years, and smoking within 5 years of a breast cancer diagnosis was associated with 2 to 3 times the odds of having a positive (+) subtype.

In addition, our study benefits by the incorporation of a multigene luminal score embedded in the PAM50 signature, used to classify breast tumors according to intrinsic subtype (6,7). The 8 genes included in the luminal score are highly correlated with Luminal subtypes, which are characterized by high estrogen-receptor expression. Multigene scores may offer improved resolution over single gene markers as they are often predictive, prognostic, and may have etiologic relevance by capturing additional dimensions of estrogen response not captured by a single gene. We observed similar patterns of association between smoking and the ER+, *ESR1*+, and LS+ subtypes. Notably, however, the magnitudes of association were slightly higher for the *ESR1* and LS mRNA classifications.

Although the prevalence of cigarette smoking has steadily decreased since the 1950s, approximately 50% of women in the United States report a history of ever smoking and 14% are self-reported current smokers (28). For protracted exposures in studies of etiology, it is important to evaluate measures of dose, duration, and temporality to fully evaluate associations with the outcome. Women with the longest smoking histories in our study were older compared to never smokers and were also most likely to be self-reported current smokers at time of diagnosis. As such, traditional metrics for smoking in studies of cancer etiology are confounded by age, dose, and duration of exposure thereby creating a challenge in understanding how combination of dose and timing influence biomarker expression in breast tumors.

In the present study, we use cumulative and time-varying (latency) models to simultaneously evaluate dose, duration, and timing of exposure; we observed that pre-diagnostic smoking proximal to time of diagnosis may be associated with increased odds of ER+, *ESR1*+, and LS+ subtypes. We also observed higher quantitative levels of *ESR1* among current smokers and women who smoked within 5 years of breast cancer diagnosis. Thus, the timing of a

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

woman's smoking exposure relative to date of diagnosis may be key to understanding the relationship between smoking and ER-defined breast tumors. Future studies of smoking and breast cancer risk may benefit from statistical methods that can be used to elucidate associations for exposures confounded by time.

There also may be clinical implications for changes in quantitative estrogen pathway genes. Prospective studies of breast cancer survivors have suggested that smoking exposure prior to diagnosis may influence survival outcomes, presumably through reduced efficacy of ER-targeted therapies (29–31). Researchers have also suggested that fluctuations in endogenous estrogens may influence intrinsic subtyping in premenopausal women (32); so it is plausible that an exogenous exposure like smoking, which could modulate estrogen-receptor expression, may also have implications for intrinsic subtyping and subsequent treatment decisions. At present, IHC biomarkers have the greatest clinical application in breast cancer, though high sensitivity IHC assays may have somewhat saturated signals, limiting our ability to assess quantitative changes in protein. Notably, we did not observe associations between smoking and quantitative ER protein expression. RNA measures for ESR1 may prove useful in evaluating quantitative changes, however the feasibility of implementing such measures in clinical settings remains unknown. The current findings suggest that in research settings designed to understand breast cancer heterogeneity in relation to exposure history, quantitative levels of RNA may have value. As genomic tests become more widely used, sensitivity of these tests to smoking behavior other exposures will be important to understand.

With the unique compilation of population-based observational and biomarker data, CBCS is an ideal resource to examine the association between smoking and ER-defined breast cancer subtypes. Findings from our study add a unique contribution to the body of literature by considering multiple methods to characterize ER-defined breast tumors and by incorporating measures of time, duration and dose to identify etiologically relevant exposure periods. We also suggest that RNA measures may provide improved resolution of gene expression for studies seeking to evaluate the etiology of ER+ breast tumors. Future work should seek to examine smoking in relation to other proposed biomarkers of breast carcinogenesis and should evaluate other patient exposures as possible modulators of clinical and/or genomic tumor characteristics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

Research reported in this publication was supported by the P50-CA058223, U01-CA179715, and the University Cancer Research Fund, University of North Carolina at Chapel Hill. The work was also supported by F31CA200336 from the National Cancer Institute of the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

References

1. Millikan RC, Newman B, Tse CK, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat. 2008; 109(1):123–139. [PubMed: 17578664]
2. Trivers KF, Lund MJ, Porter PL, et al. The epidemiology of triple-negative breast cancer, including race. Cancer causes & control : CCC. 2009; 20(7):1071–1082. [PubMed: 19343511]
3. Butler EN, Tse CK, Bell ME, Conway K, Olshan AF, Troester MA. Active smoking and risk of Luminal and Basal-like breast cancer subtypes in the Carolina Breast Cancer Study. Cancer causes & control : CCC. 2016; 27(6):775–786. [PubMed: 27153846]
4. Kawai M, Malone KE, Tang MT, Li CI. Active smoking and the risk of estrogen receptor-positive and triple-negative breast cancer among women ages 20 to 44 years. Cancer. 2014; 120(7):1026–1034. [PubMed: 24515648]
5. Nyante S, Gierach G, Dallal C, et al. Cigarette smoking and postmenopausal breast cancer risk in a prospective cohort. British journal of cancer. 2014; 110(9):2339–2347. [PubMed: 24642621]
6. Nielsen TO, Parker JS, Leung S, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010; 16(21):5222–5232. [PubMed: 20837693]
7. Parker JS, Mullins M, Cheang MC, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009; 27(8):1160–1167. [PubMed: 19204204]
8. Newman B, Moorman PG, Millikan R, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat. 1995; 35(1):51–60. [PubMed: 7612904]
9. McGee SA, Durham DD, Tse CK, Millikan RC. Determinants of breast cancer treatment delay differ for African American and White women. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2013; 22(7):1227–1238.
10. Weinberg CR, Sandler DP. Randomized recruitment in case-control studies. American journal of epidemiology. 1991; 134(4):421–432. [PubMed: 1877602]
11. Allott EH, Cohen SM, Geradts J, et al. Performance of Three-Biomarker Immunohistochemistry for Intrinsic Breast Cancer Subtyping in the AMBER Consortium. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2016; 25(3):470–478.
12. Geiss GK, Bumgarner RE, Birditt B, et al. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature biotechnology. 2008; 26(3):317–325.
13. Waggott, DM. R package version 1.1.21. 2015. NanoStringNorm: Normalize NanoString miRNA and mRNA Data.
14. Fraley, C., Raftery, AE., Murphy, TB., Scrucca, L. Technical Report No 597. Department of Statistics, University of Washington; 2012. Normal Mixture Modeling for Model-Based Clustering, Classification, and Density Estimation. (mclust Version 4 for R)
15. Yang XR, Sherman ME, Rimm DL, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2007; 16(3):439–443.
16. Tamimi RM, Colditz GA, Hazra A, et al. Traditional breast cancer risk factors in relation to molecular subtypes of breast cancer. Breast Cancer Res Treat. 2012; 131(1):159–167. [PubMed: 21830014]
17. Williams LA, Olshan AF, Tse CK, Bell ME, Troester MA. Alcohol intake and invasive breast cancer risk by molecular subtype and race in the Carolina Breast Cancer Study. Cancer causes & control : CCC. 2016; 27(2):259–269. [PubMed: 26705260]
18. Park SY, Kolonel LN, Lim U, White KK, Henderson BE, Wilkens LR. Alcohol consumption and breast cancer risk among women from five ethnic groups with light to moderate intakes: the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Multiethnic Cohort Study. International journal of cancer. 2014; 134(6):1504–1510. [PubMed: 24037751]

19. Kwan ML, Kushi LH, Weltzien E, et al. Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. Breast cancer research : BCR. 2009; 11(3):R31. [PubMed: 19463150]
20. Terry PD, Miller AB, Rohan TE. Cigarette smoking and breast cancer risk: a long latency period? International journal of cancer. 2002; 100(6):723–728. [PubMed: 12209614]
21. Richardson DB. Latency models for analyses of protracted exposures. Epidemiology. 2009; 20(3): 395–399. [PubMed: 19262389]
22. Gaudet MM, Gapstur SM, Sun J, Diver WR, Hannan LM, Thun MJ. Active smoking and breast cancer risk: original cohort data and meta-analysis. J Natl Cancer Inst. 2013; 105(8):515–525. [PubMed: 23449445]
23. Kabat GC, Kim M, Phipps AI, et al. Smoking and alcohol consumption in relation to risk of triple-negative breast cancer in a cohort of postmenopausal women. Cancer causes & control : CCC. 2011; 22(5):775–783. [PubMed: 21360045]
24. Manjer J, Malina J, Berglund G, Bondeson L, Garne JP, Janzon L. Smoking associated with hormone receptor negative breast cancer. International journal of cancer. 2001; 91(4):580–584. [PubMed: 11251985]
25. Cooper JA, Rohan TE, Cant EL, Horsfall DJ, Tilley WD. Risk factors for breast cancer by oestrogen receptor status: a population-based case-control study. Br J Cancer. 1989; 59(1):119–125. [PubMed: 2757918]
26. Morabia A, Bernstein M, Ruiz J, Heritier S, Diebold Berger S, Borisch B. Relation of smoking to breast cancer by estrogen receptor status. International journal of cancer. 1998; 75(3):339–342. [PubMed: 9455790]
27. Yaziji H, Taylor CR, Goldstein NS, et al. Consensus recommendations on estrogen receptor testing in breast cancer by immunohistochemistry. Applied immunohistochemistry & molecular morphology : AIMM. 2008; 16(6):513–520. [PubMed: 18931614]
28. Agaku IT, King BA, Dube SR. Centers for Disease C, Prevention. Current cigarette smoking among adults - United States, 2005–2012. MMWR Morbidity and mortality weekly report. 2014; 63(2):29–34. [PubMed: 24430098]
29. Daniell HW. Estrogen receptors, breast cancer, and smoking. The New England journal of medicine. 1980; 302(26):1478.
30. Kakugawa Y, Kawai M, Nishino Y, et al. Smoking and survival after breast cancer diagnosis in Japanese women: A prospective cohort study. Cancer science. 2015; 106(8):1066–1074. [PubMed: 26052951]
31. Persson M, Simonsson M, Markkula A, Rose C, Ingvar C, Jernstrom H. Impacts of smoking on endocrine treatment response in a prospective breast cancer cohort. Br J Cancer. 2016; 115(3): 382–390. [PubMed: 27280635]
32. Bernhardt SM, Dasari P, Walsh D, Townsend AR, Price TJ, Ingman WV. Hormonal Modulation of Breast Cancer Gene Expression: Implications for Intrinsic Subtyping in Premenopausal Women. Frontiers in oncology. 2016; 6:241. [PubMed: 27896218]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Impact

A multigene luminal score and single-gene *ESR1* mRNA may capture tumor changes associated with smoking.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

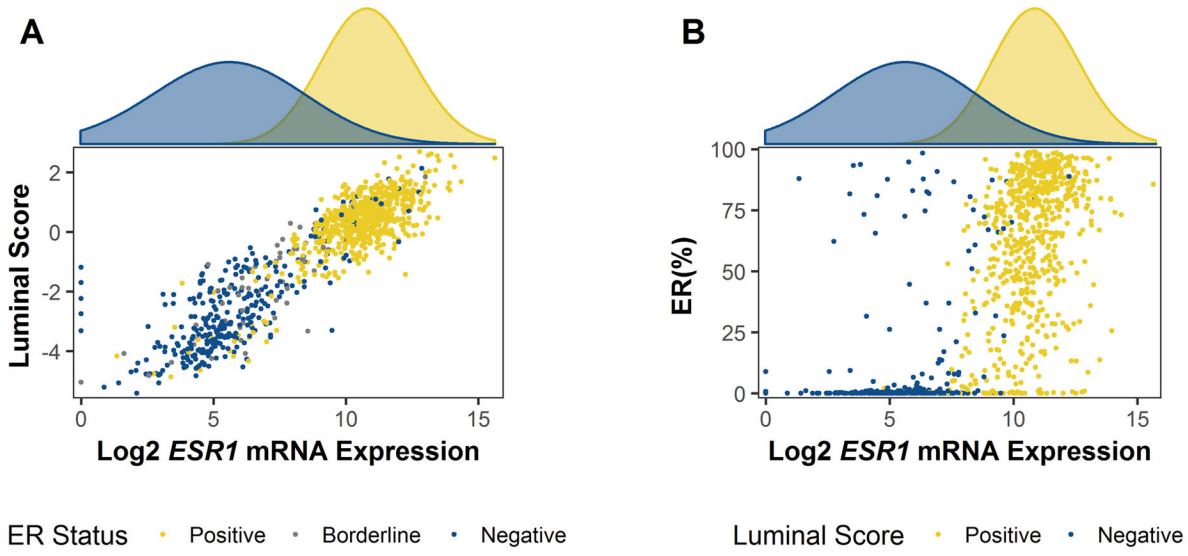

Figure 1. Correlations between ER protein, *ESR1* mRNA, and the multigene Luminal Score

Figure 1A includes a scatterplot showing the relationship between estrogen-receptor immunohistochemistry status, *ESR1* mRNA expression (log2), and luminal score (median-centered) (n=993). ER positive breast tumors are colored yellow (≥10% weighted positivity); ER borderline tumors are colored gray (1% to < 10% weighted positivity); and ER negative tumors are colored dark blue (< 1% weighted positivity). Figure 1B includes a scatterplot showing the relationship between ER weighted percent positivity (%), *ESR1* mRNA expression (log2), and luminal score binary classifications (i.e., LS+ and LS–). ER IHC and *ESR1* mRNA values were positively correlated (r=0.70, p <0.01). An expectation-maximization (EM) algorithm identified two distinct clusters for *ESR1* expression (*ERS1–,* dark blue; *ESR1+,* yellow).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

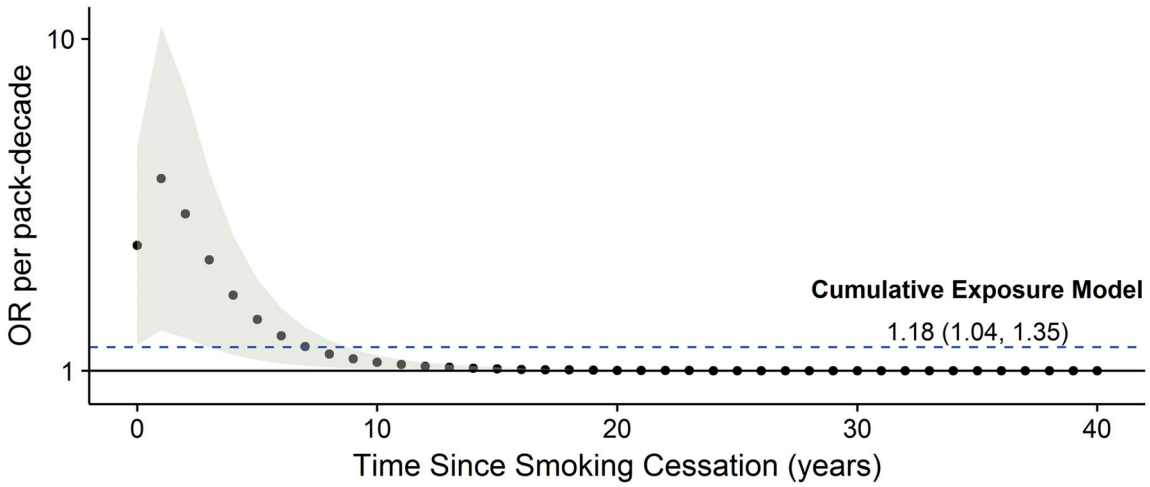

Figure 2. Odds of LS+ breast cancer with time since smoking cessation

Figure 2 displays variation with time since smoking cessation in the association between pack-decades of cigarettes smoked and luminal score positive (LS+) breast cancer. Logistic regression models were adjusted for age and race (n=993). The dashed blue line indicates the estimated odds ratio for cumulative smoking exposure (pack-decades) for the model described in Table 2 (OR and 95% CI = 1.18 (1.04, 1.35)). The solid dark gray dots indicate point estimates for the association between pack-decades and LS+ breast cancer for each year preceding breast cancer diagnosis over a period of 40 years, with exposure time points weighted using a lognormal distribution. The light gray bands represent 95% confidence intervals surrounding point estimates.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

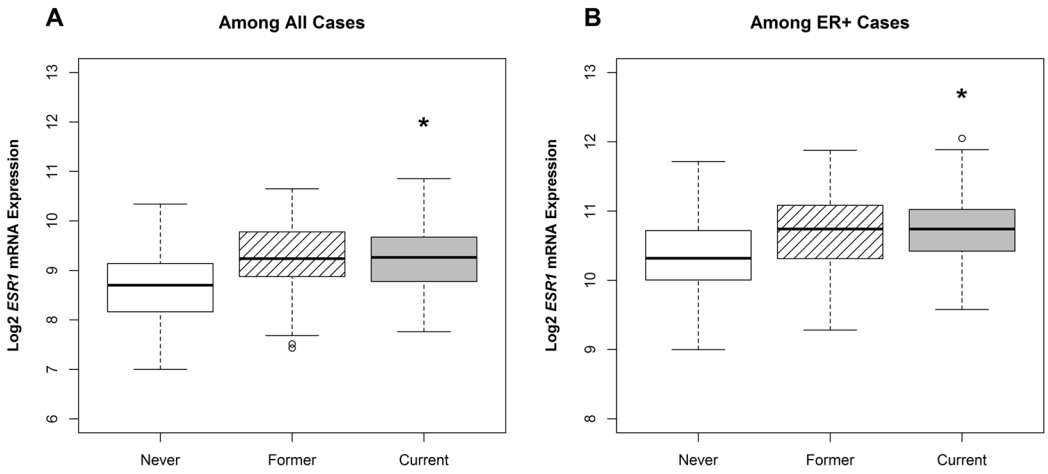

Figure 3. Pre-diagnostic smoking status and the distribution of *ESR1* mRNA

Figure 3 includes boxplots displaying estimated expression values for *ESR1* among (A) all cases (n=986) and (B) among ER+ cases (n=644), by pre-diagnostic smoking status. *ESR1* values were estimated from a linear regression model adjusted for age, race, and Nanostring batch. ***** = $p < 0.05$, where 'Never' smokers serve as the referent group. Estimated expression values are based on cases with complete covariate data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

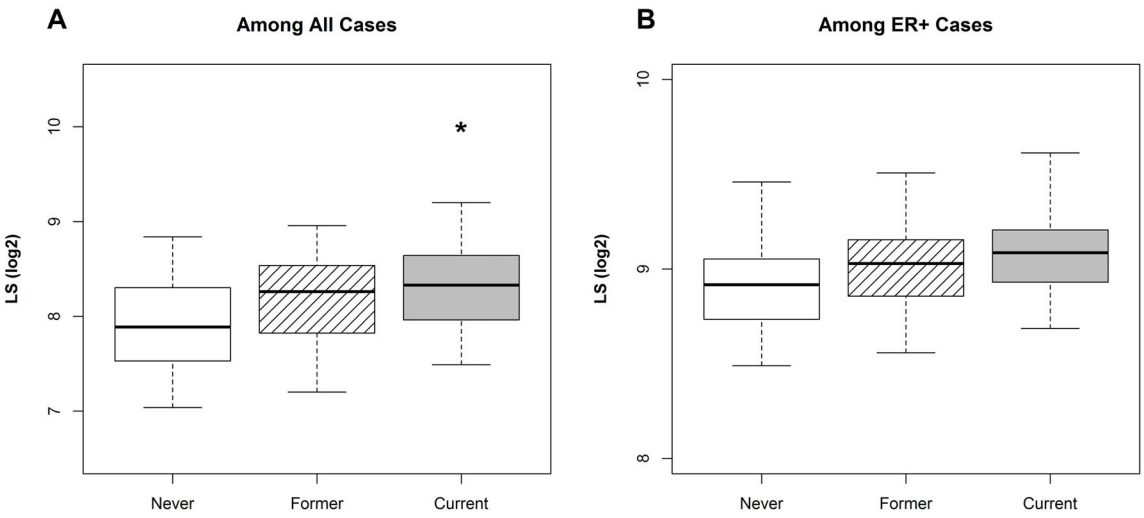

Figure 4. Pre-diagnostic smoking status and the distribution of Luminal Score values

Figure 4 includes boxplots displaying the distribution of the luminal score (LS) among (A) all cases (n=986) and (B) among those who were ER+ (n=644), by smoking status. LS values were estimated from a linear regression model adjusted for age, race, and Nanostring batch. ***** = $p < 0.05$, where 'Never' smokers serve as the referent group. Estimated expression values are based on cases with complete covariate data.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 1

Age, race, and smoking characteristics of CBCS III study participants included in immunohistochemistry analysis (Overall: N=1,888) and the subset of study participants sampled for Nanostring analysis (Nanostring Sampled: n=993).

	Overall		Nanostring Sampled	
Characteristics	**n**	**%**	**n**	**%**
Race				
AA	957	50.7	488	49.1
Non-AA	931	49.3	505	50.9
Age				
<50	994	52.7	506	51
≥50	894	47.3	487	49
Smoking History				
Never	1036	54.9	523	52.7
Ever	851	45.1	469	47.3
Missing	1		1	
Smoking Status				
Never	1036	54.9	523	52.7
Former	502	26.6	260	26.2
Current	349	18.5	209	21.1
Missing	1		1	
Duration of smoking				
Never	1036	54.9	523	52.7
≤10 years	243	12.9	143	14.4
11–20 years	202	10.7	98	9.88
> 20 years	405	21.5	228	23
Missing	2		1	
Amount smoked				
Never	1036	54.9	523	52.7
< 1/2 pack	332	17.6	184	18.5
1/2–1 pack	340	18.0	187	18.9
> 1 pack	179	9.5	98	9.88
Missing	1		1	
Time Since Smoking Cessation[a]				
Never	1036	54.9	523	52.7
< 5 years	440	23.3	260	26.2
5–10 years	63	3.3	31	3.13
11–20 years	136	7.2	67	6.75
> 20 years	212	11.2	111	11.2
Missing	1		1	

Abbreviations: AA-African American

[a] – Time since smoking cessation with respect to date of diagnosis.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Table 2

Estimated odds ratios and 95% confidence intervals for cumulative smoking exposure and ER-defined breast cancer subtypes. Associations are described as the trend per pack-decade[a] overall and within intervals for time since smoking cessation.

	ER	*ESR1*	LS
	n=1,163/537 (+/−) OR per pack-decade (95% CI)	**n=608/304 (+/−) OR per pack-decade (95% CI)**	**n=619/293 (+/−) OR per pack-decade (95% CI)**
Cumulative Exposure	1.09 (0.99, 1.20)	1.18 (1.04, 1.34)	1.18 (1.04, 1.35)
Time since smoking cessation[b]			
0–10 years prior	1.54 (0.74, 3.19)	2.15 (0.82, 5.64)	2.99 (1.11, 8.08)
11–20 years prior	0.83 (0.37, 1.89)	0.87 (0.30, 2.55)	0.79 (0.26, 2.35)
> 20 years prior	1.08 (0.88, 1.32)	1.07 (0.83, 1.39)	1.01 (0.78, 1.31)
Test of heterogeneity			
LRT, 2 df[c]	0.94	2.59	6.39
p-value	0.63	0.27	0.04

Abbreviations: + Positive subtype. – Negative subtype. ER-Estrogen-receptor. IHC-Immunohistochemistry. LRT- Likelihood Ratio Test. LS-Luminal Score.

[a]- Smoking exposure was modeled as the number of packs smoked per decade. Odds ratios and 95% confidence intervals were derived from unconditional logistic regression models, adjusted for: Nanostring batch, age, race, menopausal status, parity, breastfeeding, family history of breast cancer, alcohol use, body mass index (kg/m^2), physical activity, oral contraceptive use, hormone replacement therapy use, node status, stage, tumor size, and tumor grade.

[b]- Time since smoking cessation with respect to date of diagnosis.

[c]- LRT comparing cumulative and exposure-time-windows model, with 2 degrees of freedom.

Note. Odds ratios were estimated using cases with complete covariate data.