TP53 Pathway Function, Estrogen Receptor Status, and Breast Cancer Risk Factors in the Carolina Breast Cancer Study

Amber N. Hurson[1,2], Mustapha Abubakar[2], Alina M. Hamilton[3], Kathleen Conway[1], Katherine A. Hoadley[4,5], Michael I. Love[5,6], Andrew F. Olshan[1], Charles M. Perou[5], Montserrat Garcia-Closas[2], and Melissa A. Troester[1]

ABSTRACT

Background: TP53 and estrogen receptor (ER) both play essential roles in breast cancer development and progression, with recent research revealing cross-talk between TP53 and ER signaling pathways. Although many studies have demonstrated heterogeneity of risk factor associations across ER subtypes, associations by TP53 status have been inconsistent.

Methods: This case–case analysis included incident breast cancer cases (47% Black) from the Carolina Breast Cancer Study (1993–2013). Formalin-fixed paraffin-embedded tumor samples were classified for TP53 functional status (mutant-like/wild-type-like) using a validated RNA signature. For IHC-based TP53 status, mutant-like was classified as at least 10% positivity. We used two-stage polytomous logistic regression to evaluate risk factor heterogeneity due to RNA-based TP53 and/or ER, adjusting for each other and for PR, HER2, and grade. We then compared this with the results when using IHC-based TP53 classification.

Results: The RNA-based classifier identified 55% of tumors as TP53 wild-type-like and 45% as mutant-like. Several hormone-related factors (oral contraceptive use, menopausal status, age at menopause, and pre- and postmenopausal body mass index) were associated with TP53 mutant-like status, whereas reproductive factors (age at first birth and parity) and smoking were associated with ER status. Multiparity was associated with both TP53 and ER. When classifying TP53 status using IHC methods, no associations were observed with TP53. Associations observed with RNA-based TP53 remained after accounting for basal-like subtype.

Conclusions: This case–case study found breast cancer risk factors associated with RNA-based TP53 and ER.

Impact: RNA-based TP53 and ER represent an emerging etiologic schema of interest in breast cancer prevention research.

Introduction

Many studies evaluating etiologic heterogeneity in breast cancer have found risk factors to be disparately associated with tumor subtypes. Previous studies of heterogeneity by molecular and clinical subtypes have largely focused on subtypes defined by estrogen receptor (ER) status (1–6) or clinical subtypes defined by ER, PR, and HER2 (1, 6–11). A limited number of studies have evaluated RNA-based intrinsic subtypes (12, 13). These schemas emphasize clinically relevant markers that are widely available; however, there may be other important etiologic subtypes. In the Cancer and Steroid Hormone (CASH) Study (14) and the Carolina Breast Cancer Study (CBCS; ref. 15), cross-classification of ER and TP53 status appears to account for more etiologic heterogeneity (as assessed by a global statistical measure of etiologic variance) than other widely accepted clinical multimarker schemes for clinical subtypes. These findings are biologically relevant given the established roles of TP53 and ER in breast cancer development and progression, and evidence of cross-talk between TP53 and ER signaling pathways (16–20).

These previous studies evaluating the TP53- and ER-based schema used IHC methods for TP53 classification, which misclassifies some mutant-like tumors as wild-type-like, particularly for mutations that do not result in protein overexpression or mutations in other genes of the pathway that indirectly suppress TP53 gene expression (21–23). In contrast, RNA approaches detect patterns of loss or activity in the TP53 signaling pathway. Moreover, although the prior studies estimated a unitary measure (D value) to quantify the degree of heterogeneity across all risk factors, we have sought to identify the relative contribution of different tumor markers to the heterogeneity of effects for each risk factor. We use a two-stage logistic regression model (24, 25), which is an efficient method for estimating exposure–disease associations in the presence of tumor subtype heterogeneity across multiple markers, while accounting for multiple comparisons and missing data on tumor markers. This two-stage modeling approach has not been applied to CBCS previously, and heterogeneity has not been assessed for RNA-based TP53 subtypes.

In the current study, breast tumors were classified using a validated RNA signature that aggregates information on the expression of TP53-dependent genes (26). We assessed risk factor heterogeneity across

breast cancer subtypes defined by RNA-based TP53 and IHC-based ER, while accounting for other highly correlated tumor characteristics (i.e., PR, HER2, and tumor grade). Classifying tumors for TP53 functional status using RNA-based methods may reduce misclassification and thereby strengthen etiologic associations.

Materials and Methods

Study population

CBCS is a population-based study that enrolled participants in three phases. The catchment area for phase I (1993–1996) and phase II (1996–2001) spanned 24 counties in North Carolina. The study protocol for phase III (2008–2013) was similar to the prior phases and expanded enrollment to 44 counties. The present study is restricted to invasive breast cancer cases ($N = 4{,}806$ in phases I–III). Study details have been described previously (27). Briefly, incident invasive breast cancers among women 20 to 74 years of age were identified using a rapid case ascertainment system. Black women and those younger than 50 years of age were oversampled. In CBCS, race was self-reported. However, in North Carolina population, self-reported race and genetic ancestry are highly concordant (28). Nonetheless, we herein interpret race as a social construct, which addresses both biological/genetic differences as well as complex social determinants of health. At study enrollment, trained nurses measured body mass index (BMI) and administered a questionnaire to collect data on reproductive and lifestyle risk factors. Risk factor data were collected within 5.5 months of breast cancer diagnosis, on average (29, 30). Clinical characteristics at diagnosis were assessed by collecting medical records. All participants in the CBCS were recruited with written informed consent under a protocol approved by the Institutional Review Board of the School of Medicine, University of North Carolina at Chapel Hill.

Breast tumor markers

Methods for tissue processing and IHC analysis of tumor markers have been described previously (29, 31–33). Briefly, IHC expression of ER, PR, HER2, and TP53 was abstracted from the clinical record for the majority of cases in phases I–II. For the remainder of the cases in phases I–II, and for all cases in phase III, formalin-fixed paraffin-embedded (FFPE) tumor blocks (collected from cases at study enrollment) were requested from the participating pathology laboratories. The tumor blocks were used to generate whole sections for cases in phases I–II and a portion of those in phase III. For the majority of cases in phase III, tumor blocks were used to generate tissue microarrays. IHC staining was completed by the Immunohistochemistry Core Laboratory at UNC and quantified using automated image analysis. Samples with ≥10% positive cells were classified as positive (for ER, PR, and HER2) and mutant-like (for TP53). Concordance between laboratory and clinical record was 93% for ER and HER2, and 88% for PR, as reported in Allott and colleagues (31). At the time of study enrollment for phases I–II, it was not yet the clinical standard of care to classify ER borderline tumors (≥1% and <10% positivity) as ER positive. Thus, many borderline ER-positive tumors in these phases were reported as ER negative. We therefore used a 10% cutoff for ER positivity in all study phases to avoid having differential classification by phase. Additionally, Allott and colleagues have shown that a 10% cutoff for ER positivity has the highest correlation with molecular phenotypes (e.g., intrinsic subtypes; ref. 31). Tumor stage, size, and node status were abstracted from the clinical records. Tumor grade, available only for subjects in CBCS phases I and III, was defined by centralized pathology review.

RNA expression in CBCS has been quantified using NanoString assays on FFPE tumor samples, with tumor tissue slides and cores used for RNA isolation in phase I–II and phase III, respectively (31, 34, 35). A previously validated RNA signature (Supplementary Table S1) that aggregates information on TP53-dependent genes was used to classify TP53 functional status (mutant-like or wild-type-like) based on a similarity-to-centroid approach (26). RNA-based TP53 status was missing for the majority of cases in phase I (70%) and for about half of cases in phases II and III (48% and 53%, respectively). For the earlier phases, missingness is due to biospecimen resource depletion and degradation, but missingness in phase III is due to random sampling of a subset of specimens for molecular analysis. However, because small specimens do not afford adequate tissue for molecular analysis, cases with TP53 status were likely to be larger, later stage, and higher grade (Supplementary Table S2). A research version of the PAM50 predictor was used to classify tumors into intrinsic subtype (34, 36), which was then dichotomized as basal-like or non–basal-like (i.e., luminal A, luminal B, HER2-enriched, or normal-like). For cases in CBCS phase I, two complementary DNA-based methods were used for detecting TP53 gene mutations using FFPE tumor samples. First, single-strand conformational polymorphism (SSCP) analysis was used as a screening procedure to detect mutations in exons 4–8 of the TP53 gene, with subsequent manual radiolabeled sequencing of SSCP positives (37). The Roche p53 Amplichip research test was also used to detect single base-pair substitutions and single base-pair deletions in exons 2–11, as well as splice sites (2 base pairs before and after each exon), in the TP53 gene (38). In a previous paper, Dorsey and colleagues published DNA data on 656 of 861 phase I specimens. Almost all of these ($N = 640$) were also submitted for IHC; however, fewer samples had residual tissue available for RNA ($N = 255$). All assays were carried out by a central laboratory at the University of North Carolina.

Throughout this article, we refer to "TP53" status when making inferences about the mutation status of the TP53 gene, regardless of whether we are using RNA- or protein-based methods. Inferred mutation status (by RNA or IHC methods) is referred to as "mutant-like" or "wildtype-like," with "mutant" and "wild-type" referring to measured DNA mutation status.

Statistical analyses

To compare across technical methods of classifying TP53 status (RNA signature, DNA sequencing, and IHC), we looked at associations between TP53 status (mutant/mutant-like vs. wild-type/wild-type-like) and clinical characteristics. In this analysis, we used generalized linear models (identity link) to estimate relative frequency differences and corresponding 95% confidence intervals (CI), stratified by ER status.

We evaluated heterogeneity of the associations between breast cancer risk factors and tumor markers using a two-stage polytomous logistic regression model to calculate case–case odds ratios (OR) and 95% CIs (24, 25). Associations with the following risk factors were estimated: age at menarche (per 2 years), age at first full-term birth (≥25 vs. <25 years), nulliparity (yes vs. no), multiparity (≥3 vs. <3 births), breastfeeding duration (>4 months vs. never), oral contraceptive use (ever vs. never), menopausal status (postmenopausal vs. premenopausal), age at menopause (<40 vs. >50 years), pre- and postmenopausal BMI (≥30 vs. <25 kg/m^2), estrogen-only hormone therapy use (ever vs. never), estrogen and progesterone hormone therapy use (ever vs. never), smoking status (ever vs. never), alcohol use (ever vs. never), family history of breast cancer in at least one first-degree relative (yes vs. no). There was an intermediate category modeled for breastfeeding duration (<4 months), age at menopause

(≥40 to ≤50 years), and BMI (≥25 to <30). Age at menopause, premenopausal and postmenopausal BMI, and breastfeeding duration are modeled as ordinal (with comparisons shown between the highest and lowest categories), age at menarche was modeled as continuous, and all other variables are modeled as dichotomous. Tumor markers of interest included RNA-based TP53 (mutant-like vs. wild-type-like), ER, PR, and HER2 (positive vs. negative), as well as tumor grade (III vs. I/II). All risk factors of interest were included as predictors, and all tumor markers were included as outcomes, with adjustment for age at diagnosis, race (Black/non-Black), and study phase. We then repeated the analysis using IHC-based TP53 status (mutant-like vs. wild-type-like). As a sensitivity analysis, we assessed risk factor heterogeneity (exposures) by RNA-based TP53 subtype (outcome), adjusting for PAM50-intrinsic subtype (basal-like vs. non–basal-like). The two-stage model handles missing tumor marker data [RNA-based TP53 ($N = 2,456$), PAM50-intrinsic subtype ($N = 2,456$), IHC-based TP53 ($N = 1,603$), tumor grade ($N = 1,169$), HER2 ($N = 374$), PR ($N = 132$), and ER ($N = 114$)] through imputation based on the conditional probability. The model also accounts for multiple comparisons due to the inclusion of multiple outcomes (i.e., tumor markers; ref. 25). All statistical analyses were conducted in R software version 4.0.2 (R Foundation for Statistical Computing).

Results

Demographic and clinical characteristics of cases in the CBCS are found in Supplementary Table S2, demonstrating overrepresentation of younger cases (<50 years of age at diagnosis) and Black cases. The majority of cases were ER positive (68%), PR positive (61%), and HER2 negative (83%). The RNA-based classifier identified 55% of the tumors as TP53 wild-type-like and 45% as mutant-like. After controlling for ER status, stage, race, and age, we did not observe statistically significant differences in the proportion of TP53-mutant-like tumors according to study phase.

The proportion of TP53 mutant/mutant-like tumors varies according to RNA-, DNA-, and IHC-based approaches (**Fig. 1**). Given the fact that the distribution of tumors differs by age and race, proportions are stratified by these factors when comparing across technical methods of TP53 classification. Among ER-positive cases, the proportion of cases identified as TP53 mutant was similar across classification methods in each demographic group. Among ER-negative cases, however, the RNA signature classified a higher proportion of cases as mutant compared with the other methods. For example, among non-Black

women 50 years of age or greater, RNA methods classified 77% of cases as mutant, compared with 45% and 46% by DNA and IHC methods, respectively. Consistent with **Fig. 1**, percent agreement and kappa values for TP53 status classified using the different methods varied by ER status (Supplementary Tables S3 and S4). Agreement of RNA and DNA-based TP53 status is generally high, ranging across demographic groups from 73% to 75% among ER positive cases and 73% to 82% among ER-negative cases (Supplementary Table S3). Agreement of RNA and IHC-based TP53 status varies by ER status, with high agreement among ER-positive cases (ranging from 70% to 76% across demographic groups) and relatively low agreement among ER-negative cases (ranging from 54% to 57%; Supplementary Table S4).

The distributions of TP53 expression score are shown by type of DNA mutations (structural and functional), stratified by race (**Fig. 2**). Tumors with a score greater than zero are classified as TP53 mutant-like. Compared with Black cases, non-Black cases had a higher proportion of DNA-based wild-type tumors and a lower proportion of nonsense and indel mutations. Also, the majority of TP53 wild-type tumors among non-Black cases (73%) showed no loss of pathway function by RNA-based classifier, whereas among Black cases, only about 60% of TP53 wild-type tumors showed normal TP53 pathway function. Non-Black cases had more tumors with subtle TP53 missense changes that do not result in loss of TP53 pathway function, whereas nonsense and indel mutations are nearly all associated with RNA-based TP53 mutant-like status, regardless of race. When further stratifying missense mutations and single base-pair substitutions by hotspot/non-hotspot mutations, 100% of hotspot mutations among Black cases were classified as TP53 mutant-like, compared with 88% among non-Black cases.

We assessed the associations of DNA-, IHC-, and RNA-defined TP53 status with clinical factors, stratified by ER status (**Fig. 3**; Supplementary Table S5). Generally, associations with clinical factors were of higher magnitude and significance when classifying TP53 status using the RNA signature. For example, all three classification methods revealed associations with PR status and tumor grade, but RNA-based classification showed the largest differences in TP53 prevalence. Further, RNA was the only method to capture a statistically significant association between TP53 status and Black race (regardless of ER status) and tumor stage (among ER positive cases). For some clinical factors, the magnitude of effect when using DNA-based TP53 classification was similar to that for RNA-based, notably age at diagnosis (among ER-positive cases) and tumor size (among ER-negative cases).
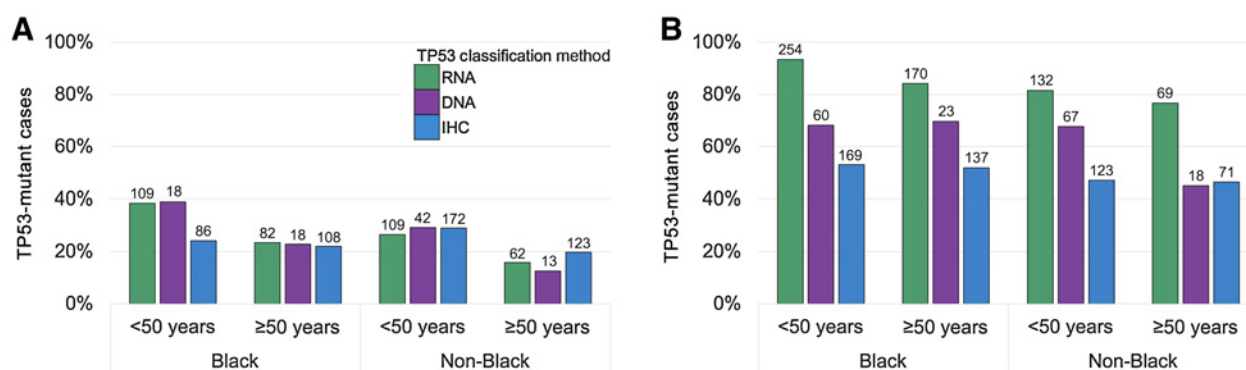

Figure 1.
Proportion of breast cancer cases classified as TP53 mutant/mutant-like across three classification methods, by age and race categories among (**A**) ER-positive cases and (**B**) ER-negative cases. Data on RNA- and IHC-based TP53 classification include subjects from CBCS phases I–III. Data on DNA-based TP53 classification include subjects from CBCS phase I.
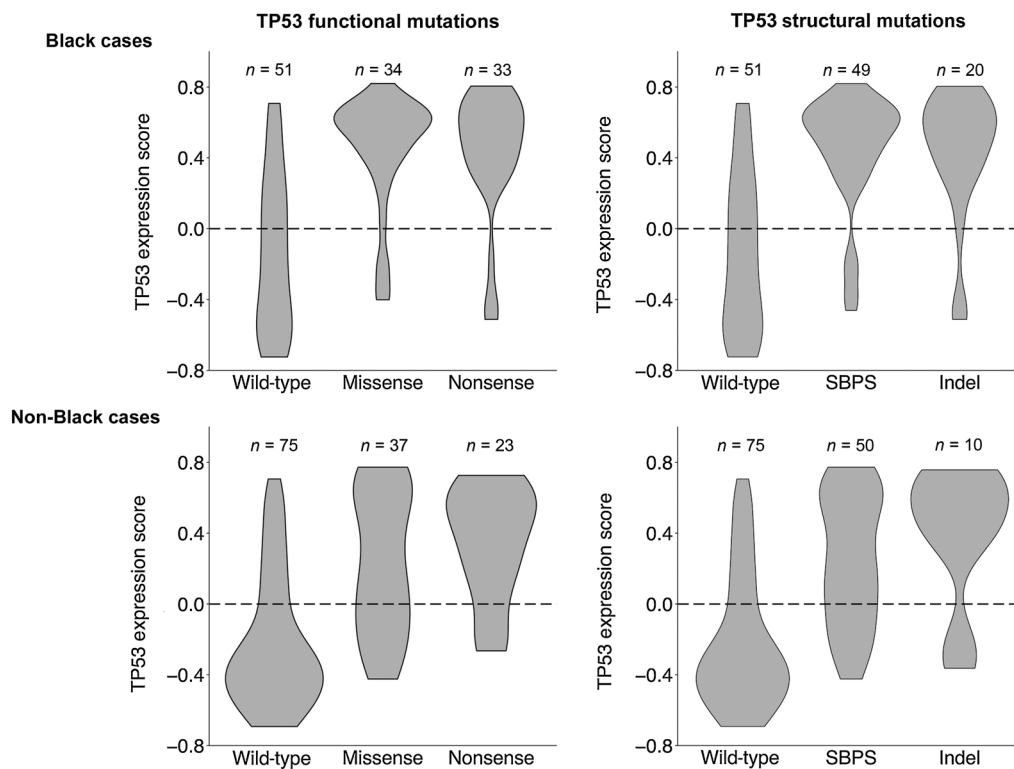
Figure 2.
Distribution of TP53 expression score across functional and structural DNA mutation types, stratified by race, in CBCS phase I. The TP53 expression score is the correlation to the TP53 mutant-like centroid. Tumors with a correlation greater than or less than zero are classified as TP53 mutant-like or wild-type-like, respectively. Indel, insertion or deletion; SBPS, single base-pair substitution.

Next, we used two-stage models to assess the contribution of ER and TP53 to etiologic heterogeneity, considering each TP53 classification method separately. Given RNA-based mutant-like status, several statistically significant associations with risk factors were identified with ER and TP53, whereas fewer associations were identified with PR, HER2, and grade (**Fig. 4A**; Supplementary Table S6). Adjusting for the effects of the other markers, TP53 was significantly associated with several hormone-related factors, such as oral contraceptive use, menopausal status, age at menopause, and pre- and postmenopausal BMI (**Fig. 4A**). Relative to TP53 wild-type-like, the risk of a TP53 mutant-like tumor was higher among women who ever used oral contraceptives [OR (95% CI) = 1.38 (1.02–1.87)] as well as among premenopausal women with a BMI of 30 kg/m^2 or greater [1.28 (1.11–1.47)]. Conversely, the risk of a TP53 mutant-like tumor was lower among postmenopausal women generally [0.71 (0.54–0.93)], and specifically among those with an age at menopause less than 40 years [0.87 (0.77–0.97)] or a BMI of 30 kg/m^2 or greater [0.86 (0.75–0.98); **Fig. 5A**; Supplementary Table S6]. ER status was associated with certain reproductive factors (age at first birth and parity) and smoking status (**Fig. 4A**). Nulliparous women, those 25 years of age or greater at first birth, and those who ever smoked were at higher risk of an ER-positive tumor compared with negative (**Fig. 5A**; Supplementary Table S6). Multiparity was independently associated with both ER and TP53 (**Fig. 4A**). Having three or more births was associated with higher risk of TP53 mutant-like compared with wild-type-like, as well as of ER positive compared with negative (**Fig. 5A**; Supplementary Table S6).

Unlike RNA-based TP53, no statistically significant associations were observed between risk factors and IHC-based TP53 status

(**Figs. 4B** and **5B**; Supplementary Table S7). There were a greater number of risk factor associations with ER and tumor grade in this model than with TP53. Grade was associated with several factors that were previously observed to be associated with RNA-based TP53 (including menopausal status, age at menopause, and premenopausal BMI). The associations between ER status and age at first birth, nulliparity, and smoking status that had been observed when adjusting for RNA-based TP53 persisted when adjusting for IHC-based TP53, with additional associations observed with premenopausal BMI and alcohol use. Finally, we performed a sensitivity analysis to assess the influence of adjustment for PAM50-intrinsic subtype (basal-like vs. non–basal-like; **Fig. 4C** and Supplementary Table S8). The associations between risk factors and RNA-based TP53 were unaffected.

Discussion

This study showed that RNA-based TP53 and ER status are both related to breast cancer risk factors and can thus define etiologically relevant subtypes of breast cancer. TP53 was most strongly associated with hormonal factors and BMI, whereas ER was mostly associated with nulliparity and smoking. Prior analyses examining heterogeneity of the effects of BMI across subtypes defined by ER status (without accounting for TP53 status) in Black and White women have been somewhat inconsistent (6, 39, 40). However, in line with the current study, parity has consistently shown differential associations by ER status (6, 41–44). The consistency of the RNA-based TP53 effects after adjustment for basal-like status suggested this may be an alternative etiologic schema with value in parallel to the intrinsic subtypes that are
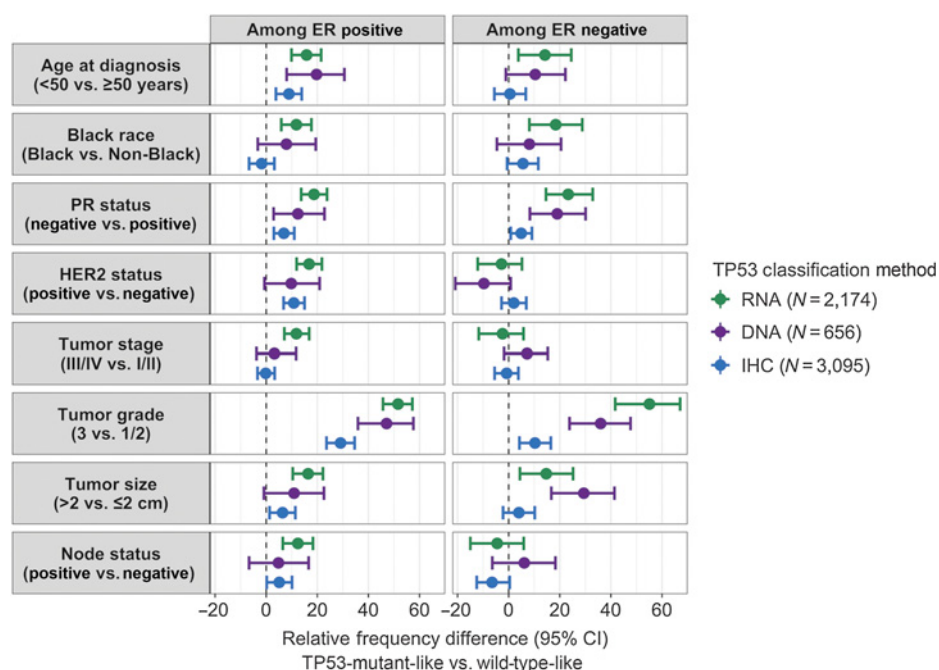
Figure 3.
Associations of clinicopathology variables with RNA-, DNA-, and IHC-defined TP53 status, stratified by ER status. Data on RNA- and IHC-based TP53 classification includes subjects from CBCS phases I–III. Data on DNA-based TP53 classification includes subjects from CBCS phase I. See Supplementary Table S4 for details. CI, confidence interval; ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; IHC, immunohistochemistry; PR, progesterone receptor.

now widely studied. Although TP53 is not widely used as an etiologic marker, one of the advantages of an ER/P53 defined schema is that TP53 has a well-known role in DNA repair, whereas ER has well-known receptor-mediated effects. Thus, incorporating both markers might reflect two important mechanisms in breast cancer.

Despite the important biological role of TP53 in breast tumors (30% have a mutation), few studies have defined TP53 as a key etiologic marker. This may be because of inconsistencies between studies, particularly those that use IHC classification methods. Other than one study observing heterogeneity with regard to smoking status (45), no prior studies have observed heterogeneity of the effects of

any environmental or reproductive risk factors across breast cancer subtypes defined by IHC-based TP53 (as a single marker; refs. 11, 33, 46, 47). It is possible that the null associations are due to misclassification; they may also be due to distinct TP53 biology that is captured by the different measures. The IHC-based TP53 classification method captures missense mutations but is a poor surrogate for deletions and insertions, as well as nonsense and frameshift mutations; in contrast, RNA-based methods capture downstream transcriptional activity—making RNA methods more sensitive to pathway changes caused by these mutation types (26, 32, 48). Misclassification alone, however, may not be a sufficient explanation for the differences. For
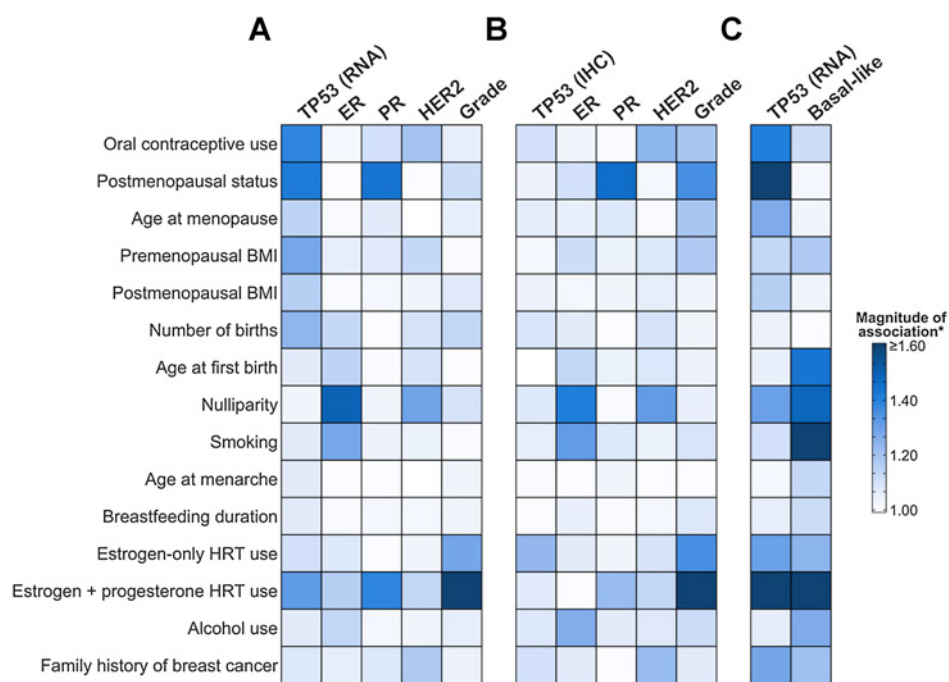

Figure 4.
Risk factor associations with breast tumor markers, when (**A**) classifying TP53 functional status using the RNA signature, (**B**) classifying TP53 status using immunohistochemistry staining, and (**C**) accounting for basal-like intrinsic subtype. Associations with each tumor marker have been adjusted for the associations of all other tumor markers, as well as age at diagnosis, race, and study phase. See Supplementary Tables S6–S8 for sample sizes, ORs (95% CI), and P values for heterogeneity. Age at menopause, premenopausal and postmenopausal BMI, and breastfeeding duration are modeled as ordinal variables (with comparisons shown between the highest and lowest categories), age at menarche is modeled as continuous, and all other risk factors are modeled as dichotomous variables. *The magnitude of association was calculated as the odds ratio with the lowest risk category as the referent. HER2, human epidermal growth factor receptor 2; HRT, hormone replacement therapy; PR, progesterone receptor.
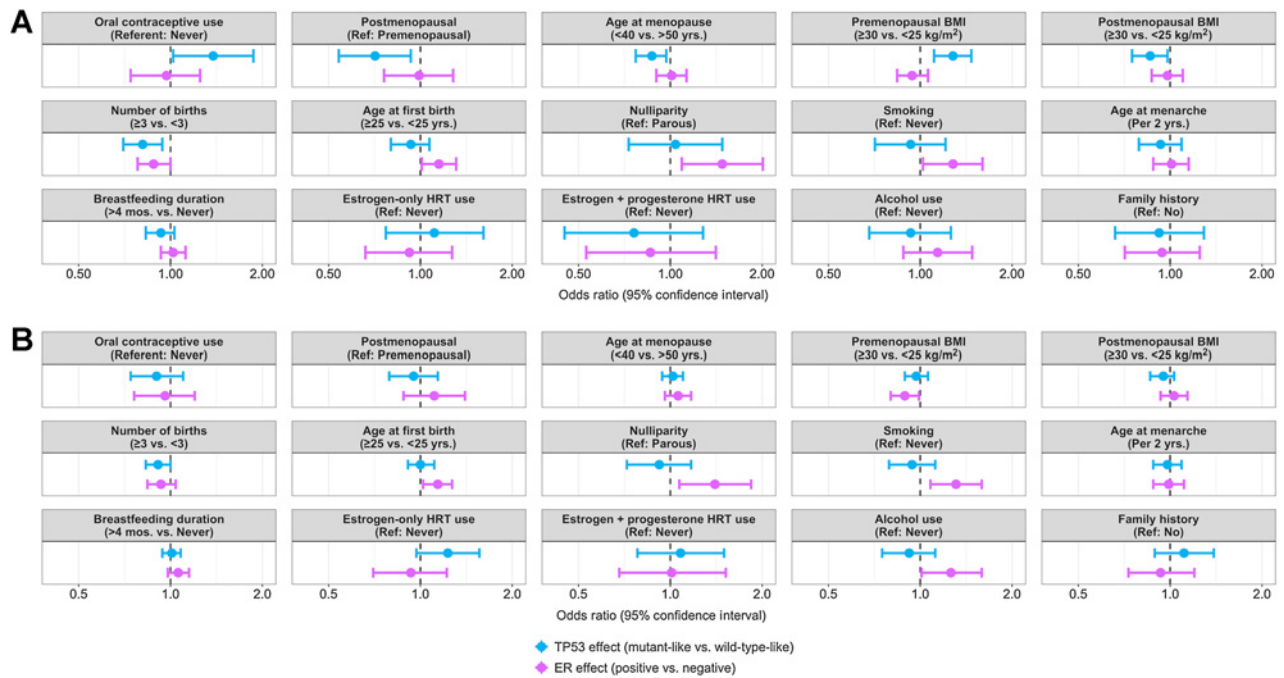
Figure 5.
Case–case ORs and 95% CIs for breast cancer risk factor associations with (**A**) RNA-based TP53 status and ER status, as well as with (**B**) IHC-based TP53 status and ER status. ORs for TP53 and ER are mutually adjusted for each other, as well as for PR, HER2, tumor grade, age at diagnosis, race, and study phase. Age at menopause, premenopausal and postmenopausal BMI, and breastfeeding duration are modeled as ordinal variables (with comparisons shown between the highest and lowest categories), age at menarche is modeled as continuous, and all other risk factors are modeled as dichotomous variables. HRT, hormone replacement therapy.

example, it is important to consider the complex relationship between ER and TP53. Only three prior studies of risk factor heterogeneity by IHC-based TP53 subtypes have stratified by ER status (11, 15, 49). A case–control study (15) observed heterogeneity of the effects of nulliparity and a case–case study (49) found heterogeneity of the effects of parity and breastfeeding across TP53 subtypes within luminal A-like cases as defined by IHC. Otherwise, the risk profiles were similar among the cross-classified tumor subtypes. These findings are similar to our results showing no heterogeneity of epidemiologic risk factors by IHC-based TP53 status when accounting for ER and other tumor markers; however, novel associations are reported in our study when using the RNA-based subtype.

Two prior studies, using data from the CASH Study (14) and CBCS (15), assessed which individual markers (ER, PR, HER2, and IHC-based TP53) or combinations of these markers showed the greatest evidence for etiologic heterogeneity. The magnitude of heterogeneity was quantified using a single measure that captures the extent to which the subtypes differ with respect to a profile of given risk factors. In both populations, ER status provided a stronger heterogeneity signal compared with PR, HER2, or IHC-based TP53. Both studies also found that subtypes formed by ER and IHC-based TP53 explained a higher degree of etiologic heterogeneity than the widely accepted IHC-defined intrinsic subtypes.

The present study builds on this work by using a two-stage model to address the question of whether effects for individual risk factors differ across levels of each individual tumor marker, while adjusting for multiple correlated tumor features. Although the prior studies estimated a unitary measure to quantify the degree of heterogeneity across all risk factors, we have sought to identify the relative contribution

of different tumor markers to the heterogeneity of effects for each risk factor. To this end, we found that RNA-based TP53 and IHC-based ER accounted for more heterogeneity of risk factor associations, with specific risk factor profiles for the two markers. We also observed associations between select risk factors and PR, HER2, and grade that were independent of ER and TP53. Some of these, such as heterogeneity of the effects of hormone replacement therapy use by tumor grade, have been previously reported in different populations (50). Compared with the prior analyses, the present study has more than doubled the sample size of breast cancer cases and directly measured heterogeneity by estimating case–case comparisons. The strength of a case–case approach is that it is statistically efficient. It is a substantial advantage to use case–case methods in a context such as this, where the case–control associations have been previously reported (15). Unlike most case–control analyses, here the associations with each tumor marker have been adjusted for the associations of all the other tumor markers (TP53, ER, PR, HER2, and grade). This is important because a key assumption for interpreting a case–case odds ratio as evidence of etiologic heterogeneity is that it is not affected by markers of progression. The case–case ORs reported here cannot be directly interpreted as indicative of either a deleterious (for ORs above one) or protective (for ORs below one) association, rather case–case odds ratio represents a measure of heterogeneity.

Limitations of our analyses included the lack of DNA mutation data for participants in CBCS phases II–III, which prevented us from evaluating risk factor heterogeneity across subtypes defined by DNA-based TP53 status. Additionally, RNA-based TP53 status was missing for about half of participants with complete risk factor data

(*N* = 2,456). As with most studies, specimens available for analysis tended to be larger tumors with more aggressive features. Nonetheless, relative to resources like TCGA, the CBCS has much higher prevalence of smaller, low-grade tumors. Another limitation was that as risk factors were measured at the time of breast cancer diagnosis, reporting could be related to the disease but is unlikely to be differential with respect to the tumor characteristics. This could result in nondifferential misclassification. It is also important to note that whereas our models accounted for multiple outcomes (i.e., tumor markers), we did not account for multiple covariates (i.e., risk factors). Lastly, although about half of the cases were Black women, numbers were small when stratifying cases by both race and tumor subtype. It will be important for future studies to compare TP53 effects across different ancestries, races, and ethnicities.

Analyses of the joint effects of ER and RNA-based TP53 status with breast cancer risk factors suggest that cross-classification of these two markers may be an important etiologic schema in breast cancer prevention research. These results are compelling given the established role of estrogen-dependent risk factors as well as DNA-repair–mediated effects of TP53 in breast cancer etiology. Considering the biological role of these two separate pathways and their established interaction, it is biologically intuitive that they could be strong markers for etiologic heterogeneity, as both pathways would appear to have independent effects and may have joint effects on risk. Given these etiologic differences and the strength of the RNA-based method for increasing the magnitude of the effects observed, future work should evaluate the prognostic implications of different classification methods.

Authors' Disclosures

C.M. Perou reports personal fees from Bioclassifier LLC outside the submitted work; in addition, Dr Perou has a patent for U.S. Patent No. 12,995,459 issued to Bioclassifier. No disclosures were reported by the other authors.

Authors' Contributions

A.N. Hurson: Conceptualization, formal analysis, visualization, methodology, writing–original draft, writing–review and editing. **M. Abubakar:** Writing–review and editing. **A.M. Hamilton:** Writing–review and editing. **K. Conway:** Writing–review and editing. **K.A. Hoadley:** Writing–review and editing. **M.I. Love:** Writing–review and editing. **A.F. Olshan:** Writing–review and editing. **C.M. Perou:** Writing–review and editing. **M. Garcia-Closas:** Supervision, writing–review and editing. **M.A. Troester:** Conceptualization, supervision, methodology, writing–review and editing.

References

1. Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. J Natl Cancer Inst 2011; 103:250–63.
2. Ambrosone CB, Zirpoli G, Ruszczyk M, Shankar J, Hong CC, McIlwain D, et al. Parity and breastfeeding among African-American women: differential effects on breast cancer risk by estrogen receptor status in the Women's Circle of Health Study. Cancer Causes Control 2014;25:259–65.
3. Aktipis CA, Ellis BJ, Nishimura KK, Hiatt RA. Modern reproductive patterns associated with estrogen receptor positive but not negative breast cancer susceptibility. Evol Med Public Health 2014;2015:52–74.
4. Anderson WF, Pfeiffer RM, Wohlfahrt J, Ejlertsen B, Jensen MB, Kroman N. Associations of parity-related reproductive histories with ER± and HER2± receptor-specific breast cancer aetiology. Int J Epidemiol 2017;46:86–95.
5. Figueroa JD, Davis Lynn BC, Edusei L, Titiloye N, Adjei E, Clegg-Lamptey JN, et al. Reproductive factors and risk of breast cancer by tumor subtypes among Ghanaian women: A population-based case–control study. Int J Cancer 2020; 147:1535–47.
6. Benefield HC, Zirpoli GR, Allott EH, Shan Y, Hurson AN, Omilian AR, et al. Epidemiology of basal-like and luminal breast cancers among black women in the AMBER consortium. Cancer Epidemiol Biomarkers Prev 2021;30:71–79.
7. Brouckaert O, Rudolph A, Laenen A, Keeman R, Bolla MK, Wang Q, et al. Reproductive profiles and risk of breast cancer subtypes: a multi-center case-only study. Breast Cancer Res 2017;19:119.
8. Gaudet MM, Gierach GL, Carter BD, Luo J, Milne RL, Weiderpass E, et al. Pooled analysis of nine cohorts reveals breast cancer risk factors by tumor molecular subtype. Cancer Res 2018;78:6011–21.
9. Lambertini M, Santoro L, Del Mastro L, Nguyen B, Livraghi L, Ugolini D, et al. Reproductive behaviors and risk of developing breast cancer according to tumor subtype: a systematic review and meta-analysis of epidemiological studies. Cancer Treat Rev 2016;49:65–76.
10. Holm J, Eriksson L, Ploner A, Eriksson M, Rantalainen M, Li J, et al. Assessment of breast cancer risk factors reveals subtype heterogeneity. Cancer Res 2017;77: 3708–17.
11. Ma H, Wang Y, Sullivan-Halley J, Weiss L, Marchbanks PA, Spirtas R, et al. Use of four biomarkers to evaluate the risk of breast cancer subtypes in the women's contraceptive and reproductive experiences study. Cancer Res 2010;70:575–87.
12. Kwan ML, Bernard PS, Kroenke CH, Factor RE, Habel LA, Weltzien EK, et al. Breastfeeding, PAM50 tumor subtype, and breast cancer prognosis and survival. J Natl Cancer Inst 2015;107:djv087.
13. Kwan ML, Kroenke CH, Sweeney C, Bernard PS, Weltzien EK, Castillo A, et al. Association of high obesity with PAM50 breast cancer intrinsic subtypes and gene expression. BMC Cancer 2015;15:278.
14. Begg CB, Zabor EC, Bernstein JL, Bernstein L, Press MF, Seshan VE. A conceptual and methodological framework for investigating etiologic heterogeneity. Stat Med 2013;32:5039–52.
15. Benefield HC, Zabor EC, Shan Y, Allott EH, Begg CB, Troester MA. Evidence for etiologic subtypes of breast cancer in the Carolina Breast Cancer Study. Cancer Epidemiol Biomarkers Prev 2019;28:1784–91.
16. Coates AS, Millar EK, O'Toole SA, Molloy TJ, Viale G, Goldhirsch A, et al. Prognostic interaction between expression of p53 and estrogen receptor in patients with node-negative breast cancer: results from IBCSG Trials VIII and IX. Breast Cancer Res 2012;14:R143.
17. Berger C, Qian Y, Chen X. The p53-estrogen receptor loop in cancer. Curr Mol Med 2013;13:1229–40.
18. Caleffi M, Teague MW, Jensen RA, Vnencak-Jones CL, Dupont WD, Parl FF. p53 gene mutations and steroid receptor status in breast cancer. Clinicopathologic correlations and prognostic assessment. Cancer 1994;73:2147–56.

19. Silwal-Pandit L, Vollan HKM, Chin S-F, Rueda OM, McKinney S, Osako T, et al. TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. Clinical Cancer Res 2014;20:3569–80.

20. Bailey ST, Shin H, Westerling T, Liu XS, Brown M. Estrogen receptor prevents p53-dependent apoptosis in breast cancer. PNAS 2012;109:18060–5.

21. Greenblatt MS, Bennett WP, Hollstein M, Harris CC. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. Cancer Res 1994;54:4855–78.

22. Macgeoch C, Barnes DM, Newton JA, Mohammed S, Hodgson SV, Ng M, et al. p53 protein detected by immunohistochemical staining is not always mutant. Dis Markers 1993;11:239–50.

23. Tsuda H, Hirohashi S. Association among p53 gene mutation, nuclear accumulation of the p53 protein and aggressive phenotypes in breast cancer. Int J Cancer 1994;57:498–503.

24. Chatterjee N. A two-stage regression model for epidemiological studies with multivariate disease classification data. J Am Statist Assoc 2004;99:127–38.

25. Zhang H, Zhao N, Ahearn TU, Wheeler W, García-Closas M, Chatterjee N. A mixed-model approach for powerful testing of genetic associations with cancer risk incorporating tumor characteristics. Biostatistics 2020;22:772–88.

26. Troester MA, Herschkowitz JI, Oh DS, He X, Hoadley KA, Barbier CS, et al. Gene expression patterns associated with p53 status in breast cancer. BMC Cancer 2006;6:276.

27. Newman B, Moorman PG, Millikan R, Qaqish BF, Geradts J, Aldrich TE, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. Breast Cancer Res Treat 1995;35:51–60.

28. Bhattacharya A, García-Closas M, Olshan AF, Perou CM, Troester MA, Love MI. A framework for transcriptome-wide association studies in breast cancer in diverse study populations. Genome Biol 2020;21:42.

29. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006;295:2492–502.

30. Millikan RC, Newman B, Tse C-K, Moorman PG, Conway K, Smith LV, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 2008;109: 123–39.

31. Allott EH, Cohen SM, Geradts J, Sun X, Khoury T, Bshara W, et al. Performance of three-biomarker immunohistochemistry for intrinsic breast cancer subtyping in the AMBER consortium. Cancer Epidemiol Biomarkers Prev 2016;25:470–8.

32. Williams LA, Butler EN, Sun X, Allott EH, Cohen SM, Fuller AM, et al. TP53 protein levels, RNA-based pathway assessment, and race among invasive breast cancer cases. NPJ Breast Cancer 2018;4:13.

33. Furberg H, Millikan RC, Geradts J, Gammon MD, Dressler LG, Ambrosone CB, et al. Environmental factors in relation to breast cancer characterized by p53 protein expression. Cancer Epidemiol Biomarkers Prev 2002;11:829–35.

34. Troester MA, Sun X, Allott EH, Geradts J, Cohen SM, Tse C-K, et al. Racial differences in PAM50 subtypes in the Carolina Breast Cancer Study. J Natl Cancer Inst 2018;110:176–82.

35. Bhattacharya A, Hamilton AM, Furberg H, et al. An approach for normalization and quality control for NanoString RNA expression data. Brief Bioinform 2021; 22:bbaa163.

36. Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, et al. Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol 2009;27:1160–7.

37. Conway K, Edmiston SN, Cui L, Drouin SS, Pang J, He M, et al. Prevalence and spectrum of p53 mutations associated with smoking in breast cancer. Cancer Res 2002;62:1987–95.

38. Baker L, Quinlan PR, Patten N, Ashfield A, Birse-Stewart-Bell LJ, McCowan C, et al. p53 mutation, deprivation and poor prognosis in primary breast cancer. Br J Cancer 2010;102:719–26.

39. Palmer JR, Adams-Campbell LL, Boggs DA, Wise LA, Rosenberg L. A prospective study of body size and breast cancer in black women. Cancer Epidemiol Biomarkers Prev 2007;16:1795–802.

40. Premenopausal Breast Cancer Collaborative Group, Schoemaker MJ, Nichols HB, Wright LB, Brook MN, Jones ME, et al. Association of body mass index and age with subsequent breast cancer risk in premenopausal women. JAMA Oncol 2018;4:e181771.

41. Palmer JR, Viscidi E, Troester MA, Hong C-C, Schedin P, Bethea TN, et al. Parity, lactation, and breast cancer subtypes in African American women: results from the AMBER Consortium. J Natl Cancer Inst 2014;106:dju237.

42. Warner ET, Tamimi RM, Boggs DA, Rosner B, Rosenberg L, Colditz GA, et al. Estrogen receptor positive tumors: do reproductive factors explain differences in incidence between black and white women? Cancer Causes Control 2013;24: 731–9.

43. Bertrand KA, Bethea TN, Adams-Campbell LL, Rosenberg L, Palmer JR. Differential patterns of risk factors for early-onset breast cancer by ER status in African American women. Cancer Epidemiol Biomarkers Prev 2017;26: 270–7.

44. Palmer JR, Boggs DA, Wise LA, Ambrosone CB, Adams-Campbell LL, Rosenberg L. Parity and lactation in relation to estrogen receptor negative breast cancer in African American women. Cancer Epidemiol Biomarkers Prev 2011;20:1883–91.

45. Gammon MD, Hibshoosh H, Terry MB, Bose S, Schoenberg JB, Brinton LA, et al. Cigarette smoking and other risk factors in relation to p53 expression in breast cancer among young women. Cancer Epidemiol Biomarkers Prev 1999;8:255–63.

46. van der Kooy K, Rookus MA, Peterse HL, van Leeuwen FE. p53 protein overexpression in relation to risk factors for breast cancer. Am J Epidemiol 1996;144:924–33.

47. Furberg H, Millikan RC, Geradts J, Gammon MD, Dressler LG, Ambrosone CB, et al. Reproductive factors in relation to breast cancer characterized by p53 protein expression (United States). Cancer Causes Control 2003;14:609–18.

48. Comprehensive molecular portraits of human breast tumours. Nature 2012;490: 61–70.

49. Abubakar M, Guo C, Koka H, Sung H, Shao N, Guida J, et al. Clinicopathological and epidemiological significance of breast cancer subtype reclassification based on p53 immunohistochemical expression. NPJ Breast Cancer 2019;5:20.

50. Abubakar M, Chang-Claude J, Ali HR, Chatterjee N, Coulson P, Daley F, et al. Etiology of hormone receptor positive breast cancer differs by levels of histologic grade and proliferation. Int J Cancer 2018;143:746–57.