

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 Melissa A. Troester¹

Affiliations: ¹Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, NC; ²Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD; ³Department of Pathology and Laboratory Medicine, The University of North Carolina, Chapel Hill; ⁴Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC; ⁵Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC; ⁶Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC.

Running title: RNA-based TP53, ER, and breast cancer risk factors

Keywords: breast cancer, TP53, estrogen receptor, RNA signature, risk factors, heterogeneity

Corresponding author:

Melissa Troester

University of North Carolina at Chapel Hill, Department of Epidemiology

253 Rosenau, CB# 7435

Chapel Hill, NC 27599

(919) 966-7408 | troester@unc.edu

Word count: 3,844

Number of figures/tables: 5

C.M.P is an equity stockholder and consultant of BioClassifier LLC; C.M.P is also listed as an inventor on patent applications for the Breast PAM50 Subtyping assay. The other authors declare no potential conflicts of interest.

ABSTRACT

Background: TP53 and estrogen receptor (ER) both play essential roles in breast cancer development and progression, with recent research revealing crosstalk between TP53 and ER signaling pathways. While 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/wildtype-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 to the results when using IHC-based TP53 classification.

Results: The RNA-based classifier identified 55% of tumors as TP53 wildtype-like and 45% as mutant-like. Several hormone-related factors (oral contraceptive use, menopausal status, age at menopause, and pre- and post-menopausal BMI) were associated with TP53 mutant-like status, while 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, there were no associations 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) status¹⁻⁶ 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) Study¹⁴ and the Carolina Breast Cancer Study (CBCS)¹⁵, cross-classification of ER and TP53 status appear to account for more etiologic heterogeneity (as assessed by a global statistical measure of etiologic variance) than other widely accepted clinical multi-marker 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 crosstalk 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 wildtype-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, while 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 modelling 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

The Carolina Breast Cancer Study (CBCS) is a population-based study that enrolled participants in three phases. The catchment area for Phase 1 (1993-1996) and Phase 2 (1996-2001) spanned 24 counties in North Carolina. The study protocol for Phase 3 (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 1-3). Study details have been described previously.²⁷ Briefly, incident invasive breast cancers among women 20-74 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 1-2. For the remainder of the cases in Phases 1-2, and for all cases in Phase 3, 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 1-2 and a portion of those in Phase 3. For the majority of

cases in Phase 3, 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 $\geq 10\%$ 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 et al.³¹ At the time of study enrollment for Phases 1-2, it was not yet the clinical standard of care to classify ER borderline tumors ($\geq 1\%$ 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 et al. has shown that a 10% cutoff for ER positivity has the highest correlation with molecular phenotypes (e.g., intrinsic subtypes).³¹ Tumor stage, size, and node status were abstracted from the clinical records. Tumor grade, available only for subjects in CBCS Phases 1 and 3, 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 1-2 and Phase 3, respectively.^{31,34,35} A previously validated RNA signature (**Table S1**) that aggregates information on TP53-dependent genes was used to classify TP53 functional status (mutant-like or wildtype-like) based on a similarity-to-centroid approach.²⁶ RNA-based TP53 status was missing for the majority of cases in Phase 1 (70%) and for about half of cases in Phases 2 and 3 (48% and 53%, respectively). For the earlier phases, missingness is due to biospecimen resource depletion and degradation, but missingness in Phase 3 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 (**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 1, two complementary DNA-based methods were employed 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 et al. published DNA data on 656 of 861 Phase 1 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 manuscript 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 “wildtype” 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. wildtype/wildtype-like) and clinical characteristics. In this analysis we used generalized linear models (identity link) to estimate relative frequency differences (RFDs) 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 (ORs) and 95% confidence intervals (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/m²), 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. wildtype-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. wildtype-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), 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).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 Carolina Breast Cancer Study are found in **Table S2**, demonstrating over-representation 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 wildtype-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 (**Figure 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 to the other methods. For example, among non-Black women 50 years of age or greater, RNA methods classified 77% of cases as mutant, compared to 45% and 46% by DNA and IHC methods, respectively. Consistent with Figure 1, percent agreement and kappa values for TP53 status classified using the different methods varied by ER status

(**Tables S3-S4**). Agreement of RNA and DNA-based TP53 status is generally high, ranging across demographic groups from 73-75% among ER positive cases and 73-82% among ER negative cases (**Table S3**). Agreement of RNA and IHC-based TP53 status varies by ER status, with high agreement among ER positive cases (ranging from 70-76% across demographic groups) and relatively low agreement among ER negative cases (ranging from 54-57%) (**Table S4**).

The distributions of TP53 expression score are shown by type of DNA mutations (structural and functional), stratified by race (**Figure 2**). Tumors with a score greater than zero are classified as TP53 mutant-like. Compared to Black cases, non-Black cases had a higher proportion of DNA-based wildtype tumors and a lower proportion of nonsense and indel mutations. Also, the majority of TP53 wildtype 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 wildtype 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, while 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 to 88% among non-Black cases.

We assessed the associations of DNA-, IHC-, and RNA-defined TP53 status with clinical factors, stratified by ER status (**Figure 3, 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).

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,

while fewer associations were identified with PR, HER2, and grade (**Figure 4A, 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 post-menopausal BMI (**Figure 4A**). Relative to TP53 wildtype-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 30kg/m² 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 30kg/m² or greater [0.86 (0.75, 0.98)] (**Figure 5A, Table S6**). ER status was associated with certain reproductive factors (age at first birth and parity) and smoking status (**Figure 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 to negative (**Figure 5A, Table S6**). Multiparity was independently associated with both ER and TP53 (**Figure 4A**). Having three or more births was associated with higher risk of TP53 mutant-like compared to wildtype-like, as well as of ER positive compared to negative (**Figure 5A, Table S6**).

Unlike RNA-based TP53, no statistically significant associations were observed between risk factors and IHC-based TP53 status (**Figures 4B and 5B, 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, **Figure 4C, 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, while 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 now widely studied. While TP53 is not widely used as an etiologic marker, one of the advantages of an ER/TP53 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 regards 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).^{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 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 study¹⁵ observed heterogeneity of the effects of nulliparity and a case-case study⁴⁹ 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 Study¹⁴ and CBCS¹⁵, 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 to 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. While 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 HRT use by tumor grade, have been previously reported in different populations.⁵⁰ Compared to 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 employ 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 odds ratios reported here cannot be directly interpreted as indicative of either a deleterious (for odds

ratios above one) or protective (for odds ratios 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 2-3, 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 non-differential misclassification. It is also important to note that while 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 biologic 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.

ACKNOWLEDGEMENTS

We would like to thank the CBCS participants for their generous participation, as well as the study staff. We would like to thank Haoyu Zhang for providing support related to the implementation and interpretation of the two-stage model. We also acknowledge Lin Wu and Roche Molecular Systems for providing p53 AmpliChip analysis of p53 mutations.

Research reported in this publication was supported by the National Institutes of Health under award number P30ES010126 (M.A. Troester), U01CA179715 (M.A. Troester), U54CA156733 (M.A. Troester), R01CA253450 (M.A. Troester and K.A. Hoadley), and F31CA257388 (A.M. Hamilton). M. Garcia-Closas, M. Abubakar, and A.N. Hurson are supported by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics (Z01CP010119). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. C.M. Perou and M.A. Troester are supported by a grant from UNC Lineberger Comprehensive Cancer Center, which is funded by the University Cancer Research Fund of North Carolina, the Susan G Komen Foundation (OGUNC1202) and the National Cancer Institute Specialized Program of Research Excellence (SPORE) in Breast Cancer (NIH/NCI P50-CA58223). This research recruited participants &/or obtained data with the assistance of Rapid Case Ascertainment, a collaboration between the North Carolina Central Cancer Registry and UNC Lineberger. We are grateful to CBCS participants and study staff.

REFERENCES

1. Yang XR, Chang-Claude J, Goode EL, et al. Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. *Journal of the National Cancer Institute*. 2011;103(3):250-263.
2. Ambrosone CB, Zirpoli G, Ruszczyk M, 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 : CCC*. 2014;25(2):259-265.
3. Aktipis CA, Ellis BJ, Nishimura KK, Hiatt RA. Modern reproductive patterns associated with estrogen receptor positive but not negative breast cancer susceptibility. *Evolution, medicine, and public health*. 2014;2015(1):52-74.
4. Anderson WF, Pfeiffer RM, Wohlfahrt J, Ejertsen B, Jensen MB, Kroman N. Associations of parity-related reproductive histories with ER± and HER2± receptor-specific breast cancer aetiology. *International journal of epidemiology*. 2017;46(1):86-95.
5. Figueroa JD, Davis Lynn BC, Edusei L, et al. Reproductive factors and risk of breast cancer by tumor subtypes among Ghanaian women: A population-based case-control study. *International journal of cancer*. 2020;147(6):1535-1547.
6. Benefield HC, Zirpoli GR, Allott EH, et al. Epidemiology of Basal-like and Luminal Breast Cancers among Black Women 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*. 2021;30(1):71-79.
7. Brouckaert O, Rudolph A, Laenen A, et al. Reproductive profiles and risk of breast cancer subtypes: a multi-center case-only study. *Breast cancer research : BCR*. 2017;19(1):119.
8. Gaudet MM, Gierach GL, Carter BD, et al. Pooled Analysis of Nine Cohorts Reveals Breast Cancer Risk Factors by Tumor Molecular Subtype. *Cancer research*. 2018;78(20):6011-6021.
9. Lambertini M, Santoro L, Del Mastro L, et al. Reproductive behaviors and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis of epidemiological studies. *Cancer treatment reviews*. 2016;49:65-76.
10. Holm J, Eriksson L, Ploner A, et al. Assessment of Breast Cancer Risk Factors Reveals Subtype Heterogeneity. *Cancer research*. 2017;77(13):3708-3717.
11. Ma H, Wang Y, Sullivan-Halley J, et al. Use of four biomarkers to evaluate the risk of breast cancer subtypes in the women's contraceptive and reproductive experiences study. *Cancer research*. 2010;70(2):575-587.
12. Kwan ML, Bernard PS, Kroenke CH, et al. Breastfeeding, PAM50 tumor subtype, and breast cancer prognosis and survival. *Journal of the National Cancer Institute*. 2015;107(7).
13. Kwan ML, Kroenke CH, Sweeney C, 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. *Statistics in medicine*. 2013;32(29):5039-5052.
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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2019.
16. Coates AS, Millar EK, O'Toole SA, 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 research : BCR*. 2012;14(6):R143.
17. Berger C, Qian Y, Chen X. The p53-estrogen receptor loop in cancer. *Current molecular medicine*. 2013;13(8):1229-1240.
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(8):2147-2156.
19. Silwal-Pandit L, Vollen HK, Chin SF, et al. TP53 mutation spectrum in breast cancer is subtype specific and has distinct prognostic relevance. *Clinical cancer research : an official journal of the American Association for Cancer Research*. 2014;20(13):3569-3580.

20. Bailey ST, Shin H, Westerling T, Liu XS, Brown M. Estrogen receptor prevents p53-dependent apoptosis in breast cancer. *Proceedings of the National Academy of Sciences of the United States of America*. 2012;109(44):18060-18065.
21. Greenblatt MS, Bennett WP, Hollstein M, Harris CC. Mutations in the p53 tumor suppressor gene: clues to cancer etiology and molecular pathogenesis. *Cancer research*. 1994;54(18):4855-4878.
22. MacGeoch C, Barnes DM, Newton JA, et al. p53 protein detected by immunohistochemical staining is not always mutant. *Dis Markers*. 1993;11(5-6):239-250.
23. Tsuda H, Hirohashi S. Association among p53 gene mutation, nuclear accumulation of the p53 protein and aggressive phenotypes in breast cancer. *International journal of cancer*. 1994;57(4):498-503.
24. Chatterjee N. A Two-Stage Regression Model for Epidemiological Studies with Multivariate Disease Classification Data. *Journal of the American Statistical Association*. 2004;99(465):127-138.
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.
26. Troester MA, Herschkowitz JI, Oh DS, et al. Gene expression patterns associated with p53 status in breast cancer. *BMC cancer*. 2006;6:276.
27. Newman B, Moorman PG, Millikan R, et al. The Carolina Breast Cancer Study: integrating population-based epidemiology and molecular biology. *Breast cancer research and treatment*. 1995;35(1):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 biology*. 2020;21(1):42.
29. Carey LA, Perou CM, Livasy CA, et al. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *Jama*. 2006;295(21):2492-2502.
30. Millikan RC, Newman B, Tse CK, et al. Epidemiology of basal-like breast cancer. *Breast cancer research and treatment*. 2008;109(1):123-139.
31. 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.
32. Williams LA, Butler EN, Sun X, 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, et al. Environmental factors in relation to breast cancer characterized by p53 protein expression. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2002;11(9):829-835.
34. Troester MA, Sun X, Allott EH, et al. Racial Differences in PAM50 Subtypes in the Carolina Breast Cancer Study. *Journal of the National Cancer Institute*. 2018;110(2).
35. Bhattacharya A, Hamilton AM, Furberg H, et al. An approach for normalization and quality control for NanoString RNA expression data. *Brief Bioinform*. 2020.
36. 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.
37. Conway K, Edmiston SN, Cui L, et al. Prevalence and spectrum of p53 mutations associated with smoking in breast cancer. *Cancer research*. 2002;62(7):1987-1995.
38. Baker L, Quinlan PR, Patten N, et al. p53 mutation, deprivation and poor prognosis in primary breast cancer. *British journal of cancer*. 2010;102(4):719-726.
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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2007;16(9):1795-1802.
40. Schoemaker MJ, Nichols HB, Wright LB, et al. Association of Body Mass Index and Age With Subsequent Breast Cancer Risk in Premenopausal Women. *JAMA oncology*. 2018;4(11):e181771.

41. Palmer JR, Viscidi E, Troester MA, et al. Parity, lactation, and breast cancer subtypes in African American women: results from the AMBER Consortium. *Journal of the National Cancer Institute*. 2014;106(10).
42. Warner ET, Tamimi RM, Boggs DA, et al. Estrogen receptor positive tumors: do reproductive factors explain differences in incidence between black and white women? *Cancer causes & control : CCC*. 2013;24(4):731-739.
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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2017;26(2):270-277.
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 epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 2011;20(9):1883-1891.
45. Gammon MD, Hibshoosh H, Terry MB, et al. Cigarette smoking and other risk factors in relation to p53 expression in breast cancer among young women. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology*. 1999;8(3):255-263.
46. van der Kooy K, Rookus MA, Peterse HL, van Leeuwen FE. p53 protein overexpression in relation to risk factors for breast cancer. *American journal of epidemiology*. 1996;144(10):924-933.
47. Furberg H, Millikan RC, Geradts J, et al. Reproductive factors in relation to breast cancer characterized by p53 protein expression (United States). *Cancer causes & control : CCC*. 2003;14(7):609-618.
48. Comprehensive molecular portraits of human breast tumours. *Nature*. 2012;490(7418):61-70.
49. Abubakar M, Guo C, Koka H, 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, et al. Etiology of hormone receptor positive breast cancer differs by levels of histologic grade and proliferation. *International journal of cancer*. 2018;143(4):746-757.

FIGURE LEGENDS

Figure 1. Proportion of breast cancer cases classified as TP53 mutant/mutant-like across three classification methods, by age and race categories among (A) estrogen receptor (ER) positive cases and (B) ER negative cases. Data on RNA- and IHC-based TP53 classification includes subjects from CBCS Phases 1-3. Data on DNA-based TP53 classification includes subjects from CBCS Phase 1. IHC = immunohistochemistry

Figure 2. Distribution of TP53 expression score across functional and structural DNA mutation types, stratified by race, in CBCS Phase 1. 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 wildtype-like, respectively. CBCS = Carolina Breast Cancer Study, Indel = insertion or deletion, SBPS = single base pair substitution

Figure 3. Associations of clinico-pathology variables with RNA-, DNA-, and IHC-defined TP53 status, stratified by estrogen receptor (ER) status. Data on RNA- and IHC-based TP53 classification includes subjects from CBCS Phases 1-3. Data on DNA-based TP53 classification includes subjects from CBCS Phase 1. See Table S4 for details. CI = confidence interval, ER = estrogen receptor, HER2 = human epidermal growth factor receptor 2, IHC = immunohistochemistry, PR = progesterone receptor

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 Tables S6-S8 for sample sizes, odds ratios (95% confidence intervals) 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 is the exponentiation of the absolute value of the log(odds ratio). BMI = body mass index, HER2 =

human epidermal growth factor receptor 2, HRT = hormone replacement therapy, IHC = immunohistochemistry, PR = progesterone receptor

Figure 5. Case-case odds ratios and 95% confidence intervals for breast cancer risk factor associations with (A) RNA-based TP53 status and estrogen receptor (ER) status, as well as with (B) IHC-based TP53 status and ER status. Odds ratios 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. BMI = body mass index, HRT = hormone replacement therapy, IHC = immunohistochemistry

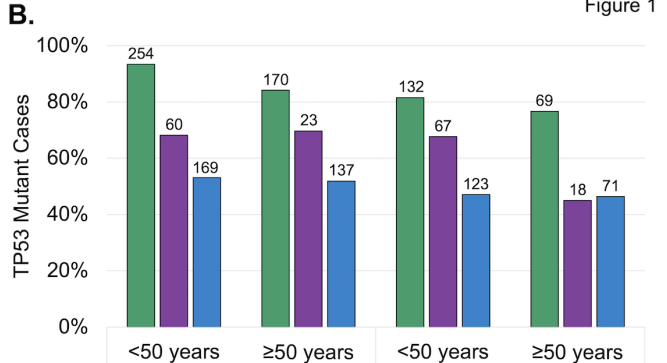
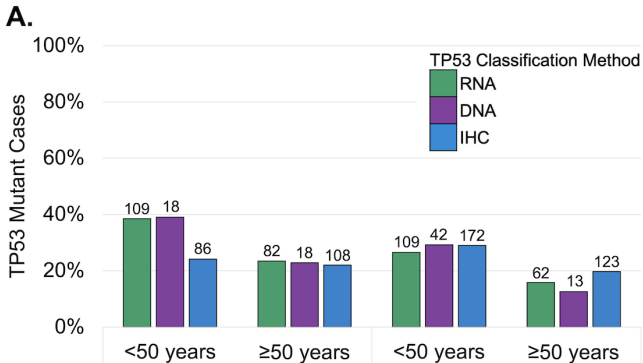
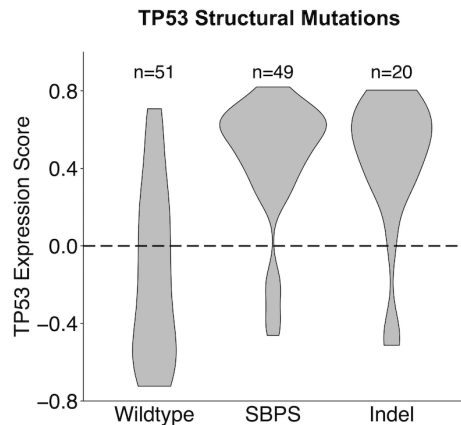
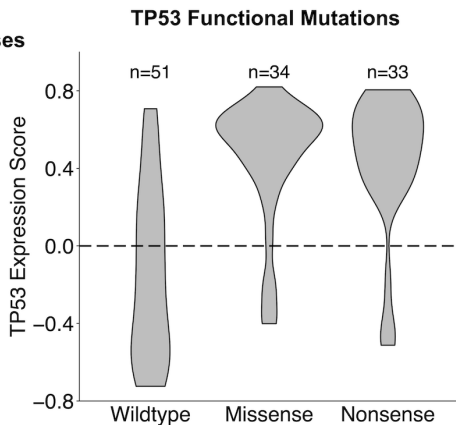
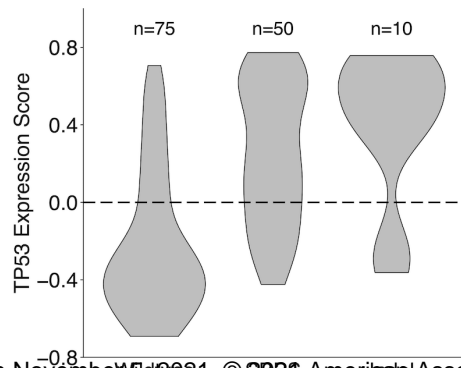
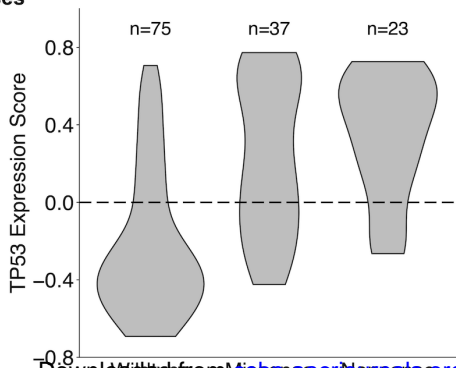


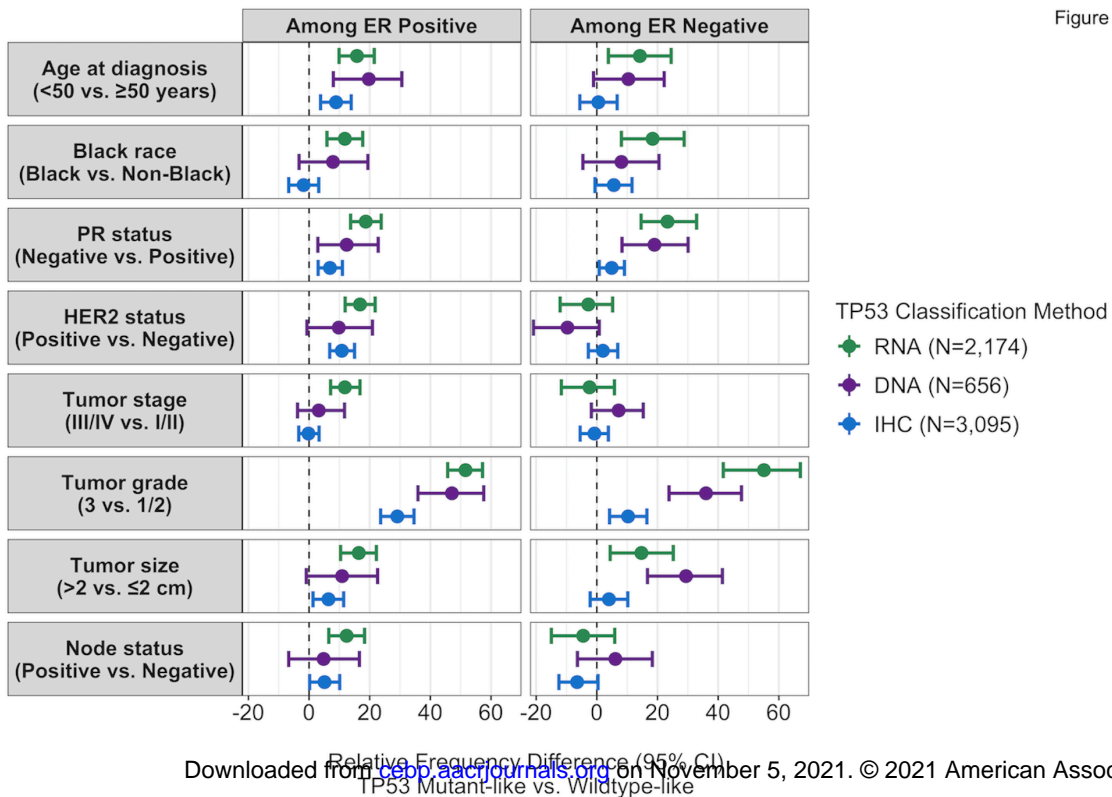
Figure 2

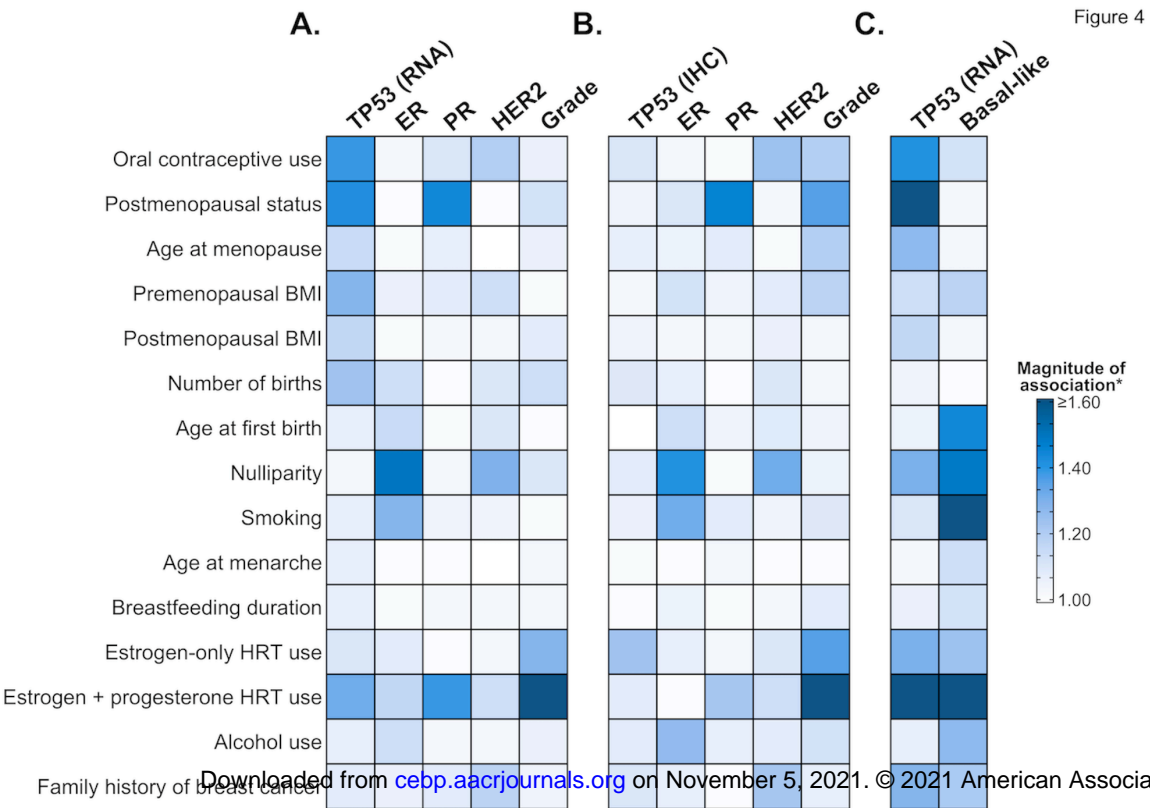
Black Cases



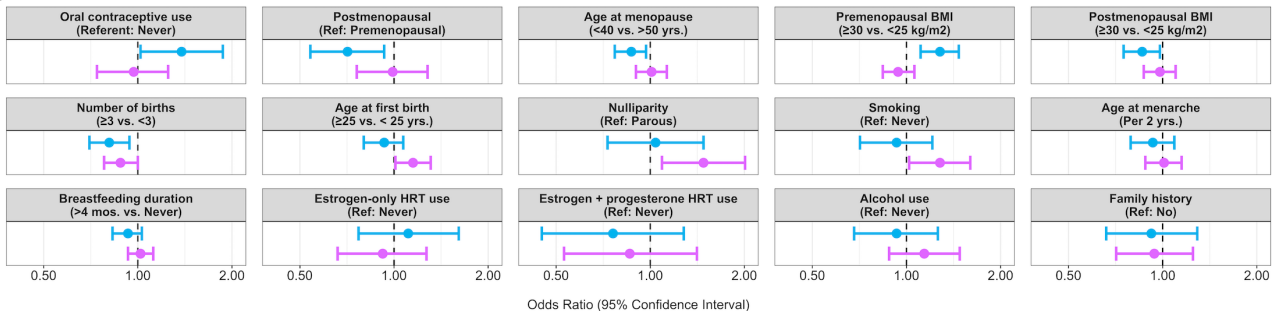
Non-Black Cases



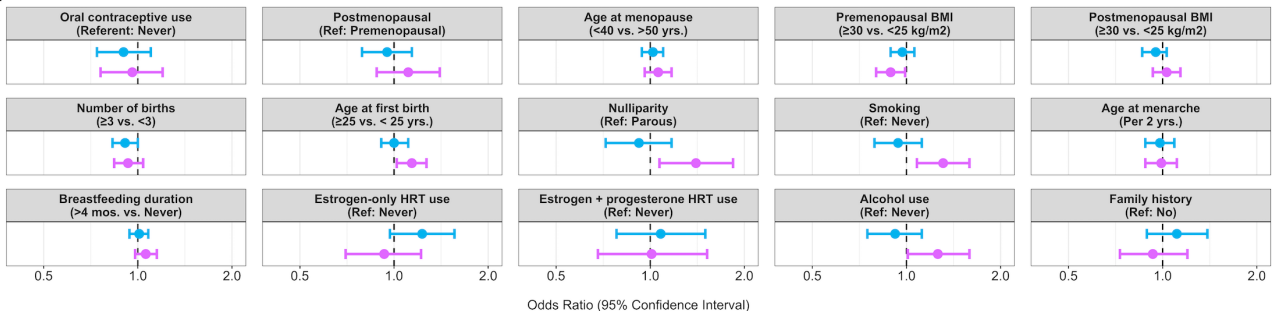




A.



B.



Cancer Epidemiology, Biomarkers & Prevention

AACR American Association
for Cancer Research

TP53 pathway function, estrogen receptor status, and breast cancer risk factors in the Carolina Breast Cancer Study

Amber N Hurson, Mustapha Abubakar, Alina M Hamilton, et al.

Cancer Epidemiol Biomarkers Prev Published OnlineFirst November 4, 2021.

Updated version	Access the most recent version of this article at: doi: 10.1158/1055-9965.EPI-21-0661
Supplementary Material	Access the most recent supplemental material at: http://cebp.aacrjournals.org/content/suppl/2021/11/02/1055-9965.EPI-21-0661.DC1
Author Manuscript	Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://cebp.aacrjournals.org/content/early/2021/11/04/1055-9965.EPI-21-0661 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.