# The proliferative activity of mammary epithelial cells in normal tissue predicts breast cancer risk in premenopausal women

**Sung Jin Huh**[1,2,3,*], **Hannah Oh**[4,5,6,*], **Michael A. Peterson**[1,7], **Vanessa Almendro**[1,2,3], **Rong Hu**[3,4,5,6], **Michaela Bowden**[1,7], **Rosina L. Lis**[1,7], **Maura B. Cotter**[1,7], **Massimo Loda**[1,7], **William T. Barry**[8], **Kornelia Polyak**[1,2,3,*], and **Rulla M. Tamimi**[4,5,*]

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[2]Department of Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

[3]Department of Medicine, Harvard Medical School, Boston, MA 02115, USA

[4]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA

[5]Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA

[6]Department of Nutrition, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA

[7]Center for Molecular Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[8]Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA 02215, USA

[9]Harvard Stem Cell Institute, Cambridge, MA 02138, USA

## Abstract

The frequency and proliferative activity of tissue-specific stem and progenitor cells are suggested to correlate with cancer risk. In this study, we investigated the association between breast cancer risk and the frequency of mammary epithelial cells expressing p27, estrogen receptor (ER), and Ki67 in normal breast tissue. We performed a nested case-control study of 302 women (69 breast cancer cases, 233 controls) who had been initially diagnosed with benign breast disease according to the Nurses' Health Studies. Immunofluorescence for p27, ER, and Ki67 was performed on tissue microarrays constructed from benign biopsies containing normal mammary epithelium and scored by computational image analysis. We found that the frequency of Ki67$^+$ cells was positively associated with breast cancer risk among premenopausal women (odds ratio [OR]=10.1, 95% confidence interval [CI]=2.12–48.0). Conversely, the frequency of ER$^+$ or p27$^+$ cells was inversely, but not significantly, associated with subsequent breast cancer risk (ER$^+$: OR=0.70, 95%

**Corresponding author:** Kornelia Polyak, Dana-Farber Cancer Institute, 450 Brookline Ave., D740C, Boston, MA 02215, or at ; Email: Kornelia_Polyak@dfci.harvard.edu; Rulla Tamimi, Brigham and Women's Hospital, 181 Longwood Ave., Boston, MA 02115, or at ; Email: rulla.tamimi@channing.harvard.edu.
[*]These authors contributed equally to this work.

CI=0.33–1.50; p27[+]: OR=0.89, 95% CI=0.45–1.75). Notably, high Ki67[+]/low p27[+] and high Ki67[+]/low ER[+] cell frequencies were significantly associated with a 5-fold higher risk of breast cancer compared to low Ki67[+]/low p27[+] and low Ki67[+]/low ER[+] cell frequencies, respectively, among premenopausal women (Ki67[hi]/p27[lo]: OR=5.08, 95% CI=1.43–18.1; Ki67[hi]/ER[lo]: OR=4.68, 95% CI=1.63–13.5). Taken together, our data suggest that the fraction of actively cycling cells in normal breast tissue may represent a marker for breast cancer risk assessment, which may therefore impact the frequency of screening procedures in at-risk women.

## Keywords

breast cancer; mammary epithelial progenitor; risk

## INTRODUCTION

Breast cancer is the most common cancer among women in the US and is a major cause of morbidity and mortality globally. It is estimated that in the US alone 232,340 invasive breast cancers and 39,620 breast cancer deaths occur annually (1). Cancer prevention would have the most profound impact on cancer-associated mortality and morbidity. However, the design of cancer preventive strategies requires in-depth knowledge of physiologic processes underlying tumor initiation, availability of biomarkers to identify high-risk individuals and monitor efficacy of preventive interventions, and availability of compounds that effectively decrease risk with minimal side effects. Unfortunately, our understanding of inter-individual differences in normal breast tissue that influence breast cancer risk is still rudimentary, resulting in our currently limited ability to predict and modify these risks. Besides germline mutations in high penetrance cancer-predisposing genes and family history, the most significant determinants of breast cancer risk are reproductive history (e.g., younger age at menarche, nulliparity, older age at first birth), mammographic density, and other factors (e.g., prior breast biopsy for benign lesions, height, exogenous hormones use, postmenopausal obesity) (2).

Neoplastic cell populations descend from single initiating cells, termed "cell-of-origin", and normal tissue stem and progenitor cells are attractive candidates as the cell-of-origin of human tumors. Therefore, breast cancer risk factors may lead to alterations in mammary epithelial stem and progenitor cells before cancer develops, making the assessment of the frequency and characteristics of these cells in normal breast tissue a potential novel biomarker for risk prediction. Several markers have been proposed for the identification of human mammary epithelial progenitors, but currently none of these are suitable for the accurate assessment of these cells in archived tissue samples (3).

Because nulliparous women have higher risk of breast cancer compared to parous women (4), we have previously investigated parity-associated variation in gene expression and epigenetic profiles of four distinct cell types (fibroblasts, luminal, myoepithelial, and progenitor-like cells) in normal breast tissues from nulliparous and parous women including *BRCA1* and *BRCA2* germline mutation carriers (5). We found the most significant differences in epithelial progenitor-like cells, where many genes and pathways important in

self-renewal and differentiation (e.g., p27$^{Kip1}$, Wnt, TGFβ) were less well expressed in parous than in nulliparous women and parous *BRCA1* and *BRCA2* germline mutation carriers. We also found significantly lower numbers of quiescent p27$^+$ and proliferative Ki67$^+$ cells in parous compared to nulliparous women. An exception was parous *BRCA1* and *BRCA2* carriers where these numbers were high. A significant fraction of p27$^+$ cells were also estrogen receptor-positive (ER$^+$) implying hormone responsiveness. Correlating with this, in premenopausal women, the frequency of p27$^+$ cells was higher in the follicular compared to the luteal phase of the menstrual cycle, whereas the opposite was observed for Ki67$^+$ cells. These results suggested that p27 and Ki67 might mark quiescent and actively cycling hormone-responsive mammary epithelial progenitors, a hypothesis supported by our functional studies in tissue slice culture models (5). Thus, we hypothesized that the relative frequencies of Ki67$^+$, p27$^+$, and ER$^+$ cells in normal breast tissue would predict the subsequent risk of breast cancer. The expression and prognostic significance of Ki67, ER, and p27 have been extensively investigated in breast cancer (6,7) and in benign proliferative lesions (8–10), but studies addressing the relevance of Ki67 (11,12) and ER (13) expression in normal mammary epithelium are limited, and p27 has not been analyzed in this context. Therefore, to test our hypothesis, we performed multicolor immunofluorescence analyses of p27, ER, and Ki67 in normal breast tissues of pre- and postmenopausal women with a previous diagnosis of biopsy-confirmed benign breast disease (BBD) and with adjacent normal breast epithelium available and examined the associations with subsequent breast cancer risk in the Nurses' Health Studies.

## MATERIALS AND METHODS

### Study design and population

This study is a nested case-control study of breast cancer conducted within the subcohort of women who reported a diagnosis of biopsy-confirmed BBD in the Nurses' Health Study (NHS) and the Nurses' Health Study II (NHSII) cohorts. The NHS is an ongoing cohort study that initiated in 1976, including 121,700 female registered nurses aged 30–55 years. The NHSII is an ongoing cohort study that began in 1989, including 116,430 female registered nurses aged 25–42 years. In both cohorts, participants completed initial mailed, self-administered questionnaires that collected information on their health behaviors, lifestyle factors, reproductive factors, and medical histories. Subsequent biennial follow-up questionnaires were used to assess updated information on established breast cancer risk factors including benign breast disease (BBD), as well as newly diagnosed diseases (e.g., breast cancer). All breast cancers were confirmed via medical record review. Study procedures and follow-up were similar for both cohorts. Details of this nested case-control study and the BBD assessment have been previously described (14,15). Cumulative response rates for both cohorts are >90% and are similar among women regardless of their BBD diagnosis.

In this study, cases were women with a previous biopsy-confirmed BBD who reported a diagnosis of breast cancer (both *in situ* and invasive and cancers) during 1976–1998 for the NHS and 1989–1999 for the NHSII. Controls were women with biopsy-confirmed BBD who remained free of breast cancer at the time the matching case was diagnosed. When

possible, four controls were selected for each case, matched on year of birth and year of benign breast biopsy. For some women, the BBD biopsy may have taken place prior to the start of the study, which was 1976 for NHS and 1989 for NHSII. We attempted to obtain BBD pathology reports and archived biopsy specimens for all cases and controls from their hospital pathology departments; our ability to obtain biopsy blocks did not significantly differ by case and control status. To reduce potential reverse causation due to subclinical tissue change, women were excluded if they had evidence of *in situ* or invasive carcinoma at biopsy or reported a diagnosis of breast cancer within six months of their biopsy (n=24). This investigation was approved by the Institutional Review Board of the Brigham and Women's Hospital. Completion of the self-administered questionnaire was considered to imply informed consent.

### Breast tissue classification and Tissue microarray (TMA) construction

Two study pathologists independently reviewed the hematoxylin and eosin (H&E) slides from the biopsy blocks and completed detailed worksheets on a number of histopathologic features and the subtype of BBD lesion (i.e., non-proliferative, proliferative without atypia, atypical ductal hyperplasia, atypical lobular hyperplasia) (16) in a blinded manner. For tissue microarray construction, a single pathologist reviewed the H&E slides to identify the areas of lesions and normal terminal ductal lobular units (TDLUs) and circled the target areas for TMA coring. The normal TDLUs selected were regions of histopathologically normal tissue that were adjacent to benign lesions (e.g., atypical ductal hyperplasia, usual ductal hyperplasia). Six TMA blocks were constructed in the Dana-Farber Harvard Cancer Center Tissue Microarray Core Facility by obtaining 0.6 mm cores from the targeted area (up to 3 cores for normal TDLU) in each donor block and inserting them into the recipient TMA blocks. We previously evaluated our TMA construction methods and confirmed a high success rate (76%) of capturing normal TDLUs in these TMA blocks (17).

### Multicolor immunofluorescence analyses

5 μm paraffin sections cut from each of the six TMA blocks were deparaffinized in xylene and hydrated in a series of ethanol. After heat-induced antigen retrieval in TRIS-EDTA buffer (pH 9), the samples were blocked with 5% goat serum-phosphate-buffered saline (PBS) and sequentially co-stained with antibodies for p27 (1:00 dilution, BD Biosciences, clone 57/Kip1/p27, mouse monoclonal IgG1) and ER (1:5,000 dilution, clone SP1, rabbit monoclonal, Thermo Scientific RM-9101) or p27 and Ki67 (1:50 dilution, Abcam, #16667, rabbit monoclonal). Procedures for secondary antibody and nuclear staining were performed as previously described (5).

### Spectral Imaging

Stained TMA slides were scanned on a Vectra 2.0.7.1 (PerkinElmer) Imaging workstation with a robotic loader. DAPI was selected for the focus filter. A 4× magnification spectral scan was run with the correct number of TMA rows and columns following the TMA map. The scan of the TMA slide was fully automated. An algorithm learning tool was performed utilizing the InForm 2.0.1 (PerkinElmer) Analysis software package to train for selecting only the appropriate epithelial regions of the tissue and subsequently completed epithelial cell segmentation. The normalized total expression intensity is recorded on a per pixel basis.

The reported mean for a given cell is the average intensity of all the normalized total pixel values in each nucleus. Glands were identified and all target tissue for each particular image was assigned according to the definition from the TMA map. Nuclei finding was enabled by detecting circular objects in the DAPI channel. A two-pixel radius around the nuclei is defined as the cytoplasm. The algorithm was applied to all the images contained within the TMA. To assess the correlation between automated and manual scoring, individual cores from a single immunofluorescence stained-TMA (adjacent slide to the one scored with automated image analysis) were manually scored using the Nikon Ti spinning disk confocal microscope. Manual scoring data were obtained by counting the positive cells in the epithelial regions using the Andor iQ software. The correlation between automated and manual scoring was weak (r=0.13–0.27), which is likely due to a combination of multiple factors including that not the exact same slide and images were scored by the two methods, differences in the type of microscopy (confocal vs. regular fluorescent microscope) and regions scored (confocal images capture smaller areas), and also the criteria for scoring (manual scoring is likely to be more subjective). Given the intensive amount of time required to manually score the individual TMAs, we were only able to score one TMA manually for validation purposes. Additionally, it is unclear that the manual scoring should be considered the gold standard due to difficulties with properly assessing the full tissue area and the more subjective and error-prone nature of scoring. Therefore, our analysis was only based on automated scoring, which is more uniform and was available on all of the TMAs in the cohort.

## Statistical analysis

The percentage of positive cells in each tissue was calculated by dividing the total number of positive cells by the total number of cells across all cores. For p27, the average values for two different co-stained slides were used as a summary measure. The percentage of positive cells for each marker was then categorized into tertiles. To ensure a reliable sampling of mammary epithelial cells, we excluded tissues if they had low total cell count (<100 cells/case) or did not have at least one core of normal TDLUs with evaluable staining. The number of women included in the analysis varied by marker due to variability in assay and TMA section cut used in staining. A total of 302 women were included in the analyses [69 cases (women with subsequent breast cancer diagnosis) and 231 controls for p27, 65 cases and 219 controls for ER, and 52 cases and 179 controls for Ki67 analysis].

To avoid losing data due to incomplete matched sets (not all benign breast biopsies for eligible cases and controls were obtained), unconditional logistic regression was performed, adjusting for matching factors (age at biopsy, year at biopsy, time since biopsy), to estimate odds ratios (OR) and 95% confidence intervals (CI) for developing subsequent breast cancer by tertiles of marker expression. Multivariate models additionally adjusted for BBD subtype (nonproliferative, proliferative without atypia, proliferative with atypia), a potential confounder. Further adjustment for other breast cancer risk factors did not change the results; thus, results were not shown here. We performed a test for trend by modeling exposures (p27, ER, and Ki67 expression) as continuous variables (percentage of positive cells). We also estimated OR and 95% CI for cross-classified groups defined by high or low expression (above or below the median) of two different tissue markers (Ki67 and p27, Ki67

and ER, p27 and ER). In subgroup analyses, we repeated the analyses after restricting to premenopausal women. Because the sample size was too small for postmenopausal women, the analyses were not repeated among postmenopausal women. We also tested for interactions by the age of the specimens using likelihood ratio tests to assess whether the associations vary by this variable. Lastly, we repeated the analyses with different thresholds for defining positive cells. All statistical tests used a two-sided type I error of 5 %. Analyses were conducted with SAS version 9 (SAS Institute).

## RESULTS

### Expression pattern of p27, ER, and Ki67 in non-malignant breast tissues

To explore whether the relative frequency of mammary epithelial cells positive for Ki67, p27, or ER in normal breast tissues may be associated with subsequent risk of breast cancer, we performed multi-color immunofluorescence analyses of Ki67, p27, and ER on tissue microarrays (TMAs) composed of normal mammary epithelium (terminal ductal lobular units, TDLUs) and benign breast disease (BBDs) (see Materials and methods and Table 1) collected from women in the Nurses' Health Study (NHS and NHSII cohorts). BBD samples included non-proliferative lesions, proliferative lesions without atypia, atypical ductal hyperplasia, and atypical lobular hyperplasia. Normal mammary epithelium was selected from biopsies with confirmed BBDs; each compartment was independently evaluated and scored. Two slides were used from each TMA, one stained for the combination of p27 and Ki67 and another for p27 and ER since we were not able to co-stain all three markers on the same slide. We used automated image analysis to identify epithelial cells based on nuclear staining and to score the three markers (Figure 1A). We determined cellular positivity for Ki67, p27, and ER by subtracting the cytoplasmic background signal from the nuclear signal of each cell, since all three markers are nuclear proteins. We set an intensity signal threshold for each marker on each TMA slide to define positive cells. The frequency of p27$^+$ cells in the p27/ER and the p27/Ki67 co-stained slides were moderately correlated (Spearman r=0.68, 95% CI=0.60–0.75) among 211 women who had p27 staining on both slides (Figure 1B). Representative immunofluorescence images are depicted in Figure 1C and Figure S1. The frequencies of p27$^+$, ER$^+$, and Ki67$^+$ cells in different types of BBD tissues and normal TDLUs are summarized in Figure S2 and Tables S1–S3.

We found extensive variability in the frequency of positive cells both within and between different histologic groups. Among BBD tissues, proliferative lesions with atypical hyperplasia had higher median frequencies of cells positive for each of the three markers (18% p27$^+$, 10% ER$^+$, 0.51% Ki67$^+$) than non-proliferative lesions (10% p27$^+$, 0.95% ER$^+$, 0.17% Ki67$^+$). With the exception of atypical hyperplasia, the percentages of Ki67$^+$ cells in the BBD lesions we analyzed were lower than previously reported by some groups (8–10,18). This discrepancy might be due to the differences in the sensitivity of the assays used for the identification of the positive cells (immunofluorescence and automated image analysis/scoring by us and immunohistochemistry and manual scoring by others) and the characteristics of the cohorts (e.g., phase of menstrual cycle, parity).

In normal breast tissues (normal TDLUs), the median frequencies of p27$^+$, ER$^+$, and Ki67$^+$ cells were 15% (range 0%–72%), 6% (range 0%–71%), and 0.85% (range 0%–30%),

respectively. The frequencies of p27$^+$ and Ki67$^+$ cells in normal TDLUs were not correlated (Spearman r=0.05, 95% CI= −0.08 to 0.18) whereas the percentage of p27$^+$ and ER$^+$ cells was slightly positively correlated (Spearman r=0.21, 95% CI= 0.10 to 0.32); the correlations were slightly stronger among postmenopausal women (r=0.17, 95% CI= −0.14 to 0.46 between p27 and Ki67, r=0.33, 95% CI= 0.07 to 0.55 between p27 and ER) (Table S4).

### Associations with subsequent breast cancer risk

Next, we determined potential associations between the percentage of cells positive for each of the markers in the normal mammary epithelium and subsequent breast cancer risk in the NHS and NHS II cohorts. The mean age at benign biopsy was 44.3 years among 302 women included in our analyses. The median time between the benign biopsy and breast cancer diagnosis was 8.6 years (range 1.1–22.5 years). Compared with controls, cases were more likely to have ever used oral contraceptives, have ever smoked, and have a family history of breast cancer (Table 1).

Overall, the frequencies of p27$^+$ and ER$^+$ cells in normal breast tissues were not significantly associated with the risk of subsequent breast cancer when comparing women at the highest vs. the lowest tertiles of cell fractions ( 25.5% vs. < 8.0% p27: OR=0.89, 95% CI=0.45–1.75, p-trend=0.97; 10.0% vs. < 3.0% ER: OR=0.70, 95% CI=0.33–1.50, p-trend=0.84) (Table 2). The frequency of Ki67$^+$ cells was positively but not significantly associated with subsequent breast cancer risk ( 2.0% vs. <0.28% Ki67: OR=1.92, 95% CI=0.83–4.43, p-trend=0.14). When we restricted our analysis to premenopausal women, the positive association with percentage of Ki67$^+$ cells was statistically significant ( 2.2% vs. <0.5% Ki67: OR=10.1, 95% CI=2.12–48.0, p-trend=0.02) (Table 3). The associations with p27$^+$ and ER$^+$ cellular frequencies remained non-significant among premenopausal women. Results were similar when testing for associations with the frequencies of double positive (i.e., p27$^+$Ki67$^+$ and p27$^+$ER$^+$) cells.

Among premenopausal women, those with high Ki67$^+$ and low p27$^+$ cell fraction had a 5.1-fold higher breast cancer risk compared with those with low Ki67$^+$ and low p27$^+$ cell frequencies (95% CI=1.43–18.1) (Table 4). Similarly, women with high Ki67$^+$ and low ER$^+$ cell frequencies had a 4.7-fold higher breast cancer risk compared with women with low Ki67$^+$ and low ER$^+$ cell fraction (95% CI=1.63–13.5). Results were similar when we used different thresholds for defining stain-positive cells (data not shown). Additionally, results did not vary by the age of the specimens (data not shown), suggesting that sample storage-related artifacts are not likely to impact our results.

## DISCUSSION

The total number of cell division of tissue-specific stem or progenitor cells, and thus, the relative frequencies of proliferative cells in tissues, is thought to correlate with the lifetime risk of cancer but experimental data supporting this model are scarce (19). In line with this, the frequencies of mammary luminal epithelial progenitors were found to be higher in normal breast tissues of *BRCA1* germline mutation carriers with high risk of breast cancer (20). Similarly, we have previously shown that the frequencies of Ki67$^+$ and p27$^+$ cells are higher in the normal breast tissues of nulliparous women who have higher risk of breast

cancer compared to parous ones and also higher in *BRCA1* and *BRCA2* mutation carriers compared to non-carriers (5). We also found that a significant fraction of p27+ cells is also ER+ implying their regulation by hormonal signals.

Ki67 is a proliferative marker expressed only in actively cycling cells (7,21), whereas ER mediates estrogen signaling and is required for the normal growth and differentiation of the mammary epithelium (22,23). Although Ki67 expression in breast tumors is often used to predict cancer prognosis and response to therapy (7,24,25), its role in normal breast tissue and the relation of this to breast cancer risk is unknown. A number of prior studies have analyzed Ki67 and ER expression in premalignant breast lesions with the goal of dissecting abnormalities relevant to functional alterations in ER signaling and breast cancer risk (8–10,18,26,27). The main limitations of these studies are the small sample sizes, and not all evaluated subsequent risk of breast cancer. Overall, although not entirely consistent, it appears that in premalignant breast lesions such as atypical hyperplasia, higher fraction of Ki67+ cells is associated with increased risk of breast cancer. Similar observations were made for the expression of ER where the fraction of ER+ cells is increased significantly in proliferative lesions with atypia (26,27). These studies suggest that for women with atypia, evaluation of these markers may better predict those who will develop breast cancer. However, atypia is found in ~10% of breast biopsies (14). In the normal mammary epithelium, ER and Ki67 was found to display mutually exclusive expression patterns (26,28), which is lost in hyperplastic benign lesions that are associated with increased breast cancer risk (29,30). However, none of these prior studies analyzing Ki67 and ER in normal breast tissues have evaluated associations between ER and Ki67 in normal mammary epithelium and subsequent risk of breast cancer (median of 8.6 years later) as was done in this study.

p27 is a cyclin-dependent kinase inhibitor and a marker of growth arrested/quiescent cells (31). Genetic deletion of p27 in animal models has been associated with increased growth and premature depletion of tissue-specific progenitor cells (32–34), implying a role for p27 in the regulation of stem/progenitor cell proliferation and pool size. p27 is also a prognostic marker in breast cancer as breast tumors with low p27 levels have worse prognosis (6,35). However, the expression of p27 has not been analyzed in normal mammary epithelium with respect to breast cancer risk. Here, we investigated potential associations between the fraction of cells positive for Ki67, p27, and ER and subsequent breast cancer risk.

In this nested case-control study of breast cancer among women with a previous BBD biopsy, the frequencies of p27+ and ER+ cells in normal breast TDLUs were not significantly associated with subsequent breast cancer risk. However, the percentage of Ki67+ cells was significantly positively associated with breast cancer risk among premenopausal women. We acknowledge that our study has several limitations. First, given the small sample size, we were not able to estimate the effect sizes with great power as indicated by our wide confidence intervals. As we continue to recruit and collect biopsy samples from more recent BBD patients in the cohort, future analyses may allow more precise estimation of the association sizes. Second, when constructing TMAs, we collected three cores of normal tissue from each biopsy sample; thus, our tissue marker expression data may not represent the whole biopsy slide due to sampling variability in scoring. Third,

although we were able to reduce inter-observation variations in staining interpretation using computer-assisted automated image analysis, we observed low correlations between manual-read and machine-read data. However, in sensitivity analyses where we adjusted thresholds for defining positive cells, similar results were found with different thresholds confirming the robustness of our results. Additionally, our biopsy samples were not timed within menstrual cycle among premenopausal women, and prior data showed variations in the relative fraction of cells positive for ER (13,27,36–38), p27 (5), and Ki67 (5) depending on the phase of menstrual cycle. Given the limited data we had available, we were not able to test for differences in associations between various phases of the menstrual cycle. Lastly, we have limited ability to generalize our results to healthy population without BBD because normal tissues of BBD patients may possibly act differently from those of healthy women.

Nevertheless, our results support the hypothesis that the proliferative cell fraction within the normal breast tissue predicts subsequent risk of breast cancer. Currently, there are no molecular assays used to predict breast cancer risk in the general population. Breast cancer risk prediction models presently available do a reasonable job of predicting the number of breast cancers that may occur in the population; however they are unable to predict well which individual women are likely to develop breast cancer (39). The results of this study suggest a new paradigm of predicting breast cancer risk by incorporating molecular markers such as Ki67. These findings could have a major impact on the care of women who undergo a breast biopsy that reveals histologically normal breast tissue. Incorporation of such a marker could help stratify women into high- and low-risk groups and could be the basis for more personalized screening strategies.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## List of abbreviations

**NHS**     Nurses' Health Study

**ADH**     atypical ductal hyperplasia

## References

1. Society AC. Breast Cancer facts & Figures 2013–2014. Atlanta: American Cancer Society, Inc.; 2013.

2. Chen WY, Colditz GA. Risk factors and hormone-receptor status: epidemiology, risk-prediction models and treatment implications for breast cancer. Nat Clin Pract Oncol. 2007; 4(7):415–23. [PubMed: 17597706]

3. Inman JL, Robertson C, Mott JD, Bissell MJ. Mammary gland development: cell fate specification, stem cells and the microenvironment. Development. 2015; 142(6):1028–42. [PubMed: 25758218]

4. Colditz GA, Rosner BA, Chen WY, Holmes MD, Hankinson SE. Risk factors for breast cancer according to estrogen and progesterone receptor status. J Natl Cancer Inst. 2004; 96(3):218–28. [PubMed: 14759989]

5. Choudhury S, Almendro V, Merino VF, Wu Z, Maruyama R, Su Y, et al. Molecular Profiling of Human Mammary Gland Links Breast Cancer Risk to a p27(+) Cell Population with Progenitor Characteristics. Cell stem cell. 2013; 13(1):117–30. [PubMed: 23770079]

6. Guan X, Wang Y, Xie R, Chen L, Bai J, Lu J, et al. p27(Kip1) as a prognostic factor in breast cancer: a systematic review and meta-analysis. J Cell Mol Med. 2010; 14(4):944–53. [PubMed: 19298520]

7. Yerushalmi R, Woods R, Ravdin PM, Hayes MM, Gelmon KA. Ki67 in breast cancer: prognostic and predictive potential. Lancet Oncol. 2010; 11(2):174–83. [PubMed: 20152769]

8. Santisteban M, Reynolds C, Barr Fritcher EG, Frost MH, Vierkant RA, Anderson SS, et al. Ki67: a time-varying biomarker of risk of breast cancer in atypical hyperplasia. Breast Cancer Res Treat. 2010; 121(2):431–7. [PubMed: 19774459]

9. Shaaban AM, Sloane JP, West CR, Foster CS. Breast cancer risk in usual ductal hyperplasia is defined by estrogen receptor-alpha and Ki-67 expression. Am J Pathol. 2002; 160(2):597–604. [PubMed: 11839580]

10. Hameed O, Ghali VS, Tartter PI, Mizrachi H. Immunohistochemical staining for cyclin D1 and Ki-67 aids in the stratification of atypical ductal hyperplasia diagnosed on breast core biopsy. Am J Clin Pathol. 2005; 124(6):862–72. [PubMed: 16416735]

11. Suzuki R, Atherton AJ, O'Hare MJ, Entwistle A, Lakhani SR, Clarke C. Proliferation and differentiation in the human breast during pregnancy. Differentiation. 2000; 66(2–3):106–15. [PubMed: 11100901]

12. Feuerhake F, Sigg W, Hofter EA, Unterberger P, Welsch U. Cell proliferation, apoptosis, and expression of Bcl-2 and Bax in non-lactating human breast epithelium in relation to the menstrual cycle and reproductive history. Breast Cancer Res Treat. 2003; 77(1):37–48. [PubMed: 12602903]

13. Khan SA, Rogers MA, Khurana KK, Meguid MM, Numann PJ. Estrogen receptor expression in benign breast epithelium and breast cancer risk. J Natl Cancer Inst. 1998; 90(1):37–42. [PubMed: 9428781]

14. Collins LC, Baer HJ, Tamimi RM, Connolly JL, Colditz GA, Schnitt SJ. The influence of family history on breast cancer risk in women with biopsy-confirmed benign breast disease: results from the Nurses' Health Study. Cancer. 2006; 107(6):1240–7. [PubMed: 16902983]

15. Tamimi RM, Colditz GA, Wang Y, Collins LC, Hu R, Rosner B, et al. Expression of IGF1R in normal breast tissue and subsequent risk of breast cancer. Breast Cancer Res Treat. 2011; 128(1):243–50. [PubMed: 21197570]

16. Page DL, Dupont WD, Rogers LW, Rados MS. Atypical hyperplastic lesions of the female breast. A long-term follow-up study. Cancer. 1985; 55(11):2698–708. [PubMed: 2986821]

17. Collins LC, Wang Y, Connolly JL, Baer HJ, Hu R, Schnitt SJ, et al. Potential role of tissue microarrays for the study of biomarker expression in benign breast disease and normal breast tissue. Appl Immunohistochem Mol Morphol. 2009; 17(5):438–41. [PubMed: 19363445]

18. Lee S, Mohsin SK, Mao S, Hilsenbeck SG, Medina D, Allred DC. Hormones, receptors, and growth in hyperplastic enlarged lobular units: early potential precursors of breast cancer. Breast Cancer Res. 2006; 8(1):R6. [PubMed: 16417654]

19. Tomasetti C, Vogelstein B. Cancer etiology. Variation in cancer risk among tissues can be explained by the number of stem cell divisions. Science. 2015; 347(6217):78–81. [PubMed: 25554788]

20. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med. 2009; 15(8):907–13. [PubMed: 19648928]

21. van Dierendonck JH, Keijzer R, van de Velde CJ, Cornelisse CJ. Nuclear distribution of the Ki-67 antigen during the cell cycle: comparison with growth fraction in human breast cancer cells. Cancer Res. 1989; 49(11):2999–3006. [PubMed: 2720660]

22. Anderson E. The role of oestrogen and progesterone receptors in human mammary development and tumorigenesis. Breast Cancer Res. 2002; 4(5):197–201. [PubMed: 12223124]

23. Dickson RB, Stancel GM. Estrogen receptor-mediated processes in normal and cancer cells. J Natl Cancer Inst Monogr. 2000; (27):135–45. [PubMed: 10963625]

24. Dowsett M, Smith IE, Ebbs SR, Dixon JM, Skene A, Griffith C, et al. Short-term changes in Ki-67 during neoadjuvant treatment of primary breast cancer with anastrozole or tamoxifen alone or combined correlate with recurrence-free survival. Clin Cancer Res. 2005; 11(2 Pt 2):951s–8s. [PubMed: 15701892]

25. Urruticoechea A, Smith IE, Dowsett M. Proliferation marker Ki-67 in early breast cancer. J Clin Oncol. 2005; 23(28):7212–20. [PubMed: 16192605]

26. Krishnamurthy S, Sneige N. Molecular and biologic markers of premalignant lesions of human breast. Adv Anat Pathol. 2002; 9(3):185–97. [PubMed: 11981114]

27. Shoker BS, Jarvis C, Clarke RB, Anderson E, Hewlett J, Davies MP, et al. Estrogen receptor-positive proliferating cells in the normal and precancerous breast. Am J Pathol. 1999; 155(6): 1811–5. [PubMed: 10595909]

28. Ricketts D, Turnbull L, Ryall G, Bakhshi R, Rawson NS, Gazet JC, et al. Estrogen and progesterone receptors in the normal female breast. Cancer Res. 1991; 51(7):1817–22. [PubMed: 2004366]

29. Iqbal M, Davies MP, Shoker BS, Jarvis C, Sibson DR, Sloane JP. Subgroups of non-atypical hyperplasia of breast defined by proliferation of oestrogen receptor-positive cells. J Pathol. 2001; 193(3):333–8. [PubMed: 11241412]

30. Clarke RB, Howell A, Potten CS, Anderson E. Dissociation between steroid receptor expression and cell proliferation in the human breast. Cancer Res. 1997; 57(22):4987–91. [PubMed: 9371488]

31. Polyak K, Lee MH, Erdjument-Bromage H, Koff A, Roberts JM, Tempst P, et al. Cloning of p27Kip1, a cyclin-dependent kinase inhibitor and a potential mediator of extracellular antimitogenic signals. Cell. 1994; 78(1):59–66. [PubMed: 8033212]

32. Fero ML, Rivkin M, Tasch M, Porter P, Carow CE, Firpo E, et al. A syndrome of multiorgan hyperplasia with features of gigantism, tumorigenesis, and female sterility in p27(Kip1)-deficient mice. Cell. 1996; 85(5):733–44. [PubMed: 8646781]

33. Kiyokawa H, Kineman RD, Manova-Todorova KO, Soares VC, Hoffman ES, Ono M, et al. Enhanced growth of mice lacking the cyclin-dependent kinase inhibitor function of p27(Kip1). Cell. 1996; 85(5):721–32. [PubMed: 8646780]

34. Cheng T, Rodrigues N, Dombkowski D, Stier S, Scadden DT. Stem cell repopulation efficiency but not pool size is governed by p27(kip1). Nat Med. 2000; 6(11):1235–40. [PubMed: 11062534]

35. Tan P, Cady B, Wanner M, Worland P, Cukor B, Magi-Galluzzi C, et al. The cell cycle inhibitor p27 is an independent prognostic marker in small (T1a,b) invasive breast carcinomas. Cancer Res. 1997; 57(7):1259–63. [PubMed: 9102210]

36. Chung K, Hovanessian-Larsen LJ, Hawes D, Taylor D, Downey S, Spicer DV, et al. Breast epithelial cell proliferation is markedly increased with short-term high levels of endogenous estrogen secondary to controlled ovarian hyperstimulation. Breast Cancer Res Treat. 2012; 132(2): 653–60. [PubMed: 22083231]

37. Taylor D, Pearce CL, Hovanessian-Larsen L, Downey S, Spicer DV, Bartow S, et al. Progesterone and estrogen receptors in pregnant and premenopausal non-pregnant normal human breast. Breast Cancer Res Treat. 2009; 118(1):161–8. [PubMed: 19205874]

38. Khan SA, Sachdeva A, Naim S, Meguid MM, Marx W, Simon H, et al. The normal breast epithelium of women with breast cancer displays an aberrant response to estradiol. Cancer Epidemiol Biomarkers Prev. 1999; 8(10):867–72. [PubMed: 10548314]

39. Anothaisintawee T, Teerawattananon Y, Wiratkapun C, Kasamesup V, Thakkinstian A. Risk prediction models of breast cancer: a systematic review of model performances. Breast Cancer Res Treat. 2012; 133(1):1–10. [PubMed: 22076477]
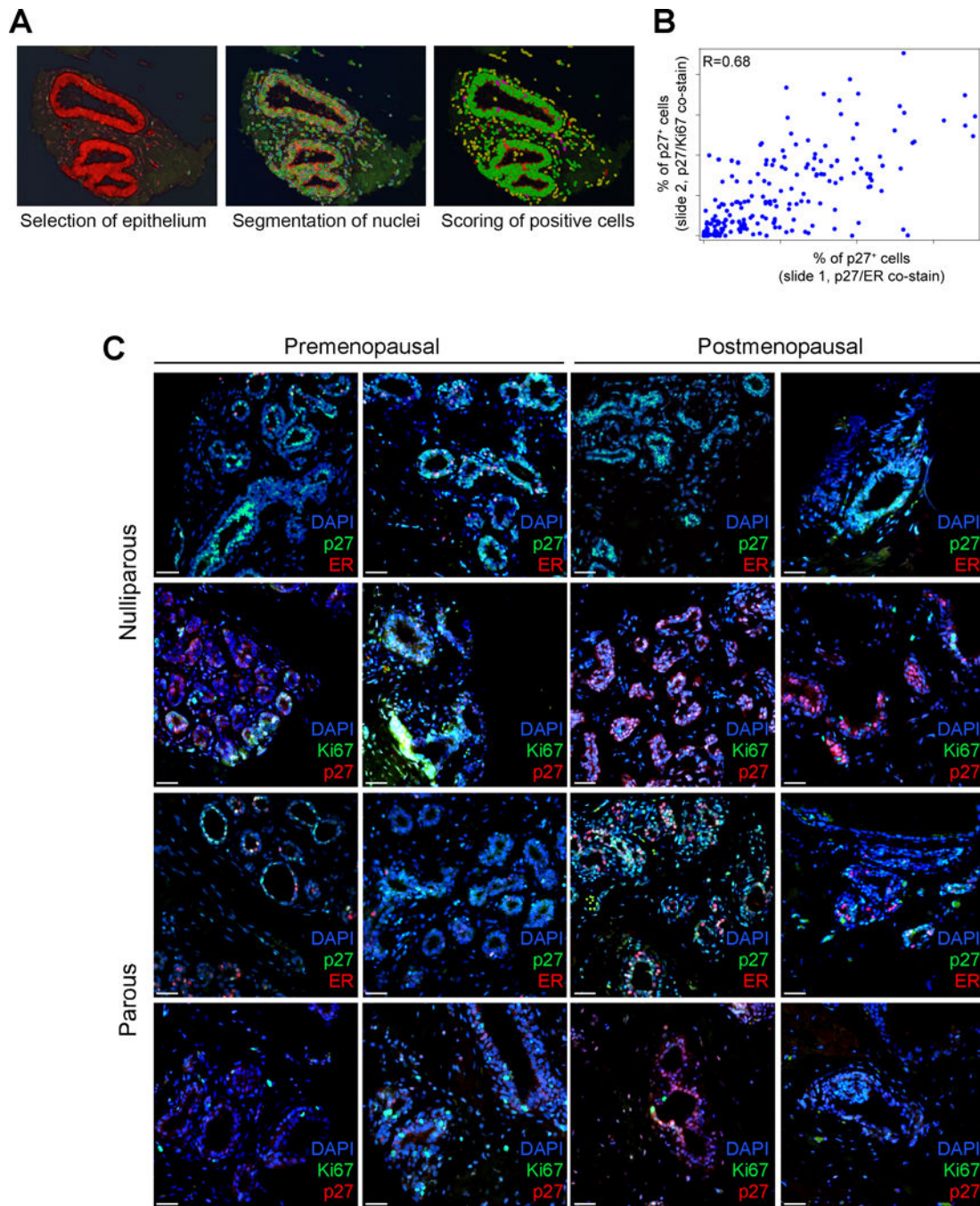
**Figure 1. Imaging analysis of the normal breast tissue in the TMAs**
**A**, Tissue and subsequently completed cell segmentation in normal breast tissue using the inForm software. Segmented normal epithelium (red), cell (green) and stained positive cells (pink). Only cells segmented in the normal epithelium were analyzed. **B**, Spearman correlation between the percentage of p27[+] cells on p27/ER and p27/Ki67 co-stained slides (r=0.68; N=211). **C**, Representative double-immunofluorescence images of staining for

p27/ER and p27/Ki67 of normal breast tissues from the indicated groups of women. Nuclei are visualized by DAPI (blue) stain. Scale bar: 25 μm.

**Table 1**

**Age-standardized characteristics of the study population at benign breast biopsy by breast cancer case-control status in the Nurses' Health Study and the Nurses' Health Study II**

Values are means (SD) or percentages and are standardized to the age distribution of the study population. Values of polytomous variables may not sum to 100% due to rounding

| | Controls (N=233) | Cases[a] (N=69) |
|---|---|---|
| **Stained for p27 expression in normal breast TDLUs** | | |
| Percentage of p27[+], median (IQR) [*] | 15.4 (25.0) | 16.1 (26.0) |
| **Stained for ER expression in normal breast TDLUs** | | |
| Percentage of ER[+], median (IQR) [*] | 6.4 (12.0) | 5.0 (10.8) |
| **Stained for Ki67 expression in normal breast TDLUs** | | |
| Percentage of Ki67[+], median (IQR) [*] | 0.77 (2.5) | 1.15 (2.7) |
| Age, years [*] | 44.6 (8.8) | 43.5 (9.0) |
| Type of benign lesion | | |
| -Non-proliferative, % | 28.2 | 23.1 |
| -Proliferative without atypia, % | 55.7 | 49.9 |
| -Proliferative with atypical hyperplasia, % | 16.1 | 27.0 |
| Height, inches | 64.6 (2.7) | 64.3 (2.1) |
| Average body size at ages 5–10 years | | |
| -Level 1 (most lean), % | 32.3 | 32.0 |
| -Level 1.5–2, % | 30.1 | 35.9 |
| -Level 2.5–3, % | 16.3 | 9.2 |
| -Level 3.5–4, % | 13.3 | 19.2 |
| -Level 4.5+ (most overweight), % | 8.0 | 3.7 |
| BMI at age 18 years, kg/m$^2$ | 21.1(3.1) | 20.9(2.2) |
| BMI at biopsy, kg/m$^2$ | 23.1(5.8) | 21.9(5.8) |
| Age at menarche, yrs | 12.6(1.4) | 12.3(1.3) |
| Parous at biopsy, % | 94.0 | 92.9 |
| Parity at biopsy (among parous women) | 3.1(1.5) | 3.2(1.6) |
| Age at first birth, years (among parous women) | 24.6(3.4) | 25.2(3.0) |
| Premenopausal, % | 74.9 | 74.2 |
| Age at menopause (among postmenopausal women at biopsy) | 45.5(7.2) | 47.2(4.4) |
| Ever oral contraceptive use | 51.0 | 61.2 |
| Ever Smoking | 50.4 | 66.3 |
| Family history of breast cancer, % | 17.0 | 29.8 |
| Adolescent alcohol drinker, % | 52.3 | 56.4 |
| Cumulative average lifetime alcohol consumption[b], g/d | 4.7(7.1) | 3.8(3.6) |
| Cumulative average adult physical activity[c], MET-hr/wk | 15.9(19.1) | 11.2(9.9) |

[a]Among 69 breast cancer cases, 54 cases were invasive cancers and 18 cases were *in situ* cancers. Among these 69 cases, 18 cases were ipsilateral, 25 cases were contralateral, and 8 cases were bilateral with respect to the index benign biopsies (the remaining 18 cases were from the unknown side of the breast).

[*]Value is not age adjusted.

Abbreviations- BMI: body mass index, IQR: interquartile range, TDLU: terminal duct lobular unit

**Table 2**

Odds ratios (95% confidence interval) of developing subsequent breast cancer according to tertiles of mean percentage of p27[+], ER[+], and Ki67[+] in normal breast tissue in the Nurses' Health Study and the Nurses' Health Study II.

| | Case | Control | Model 1[a]<br>OR (95% CI) | Model 2[b]<br>OR (95% CI) |
|---|---|---|---|---|
| **Percentage of p27[+] cells** | | | | |
| < 8.0 % | 25 | 75 | 1.0 (Ref) | 1.0 (Ref) |
| 8.0 – 25.4 % | 21 | 78 | 0.73 (0.37, 1.46) | 0.86 (0.42, 1.76) |
| 25.5 % | 23 | 78 | 0.84 (0.43, 1.63) | 0.89 (0.45, 1.75) |
| Per 1% increase | | | 1.00 (0.98, 1.02) | 1.00 (0.99, 1.02) |
| p-trend | | | 0.99 | 0.97 |
| **Percentage of ER[+] cells** | | | | |
| < 3.0 % | 24 | 72 | 1.0 (Ref) | 1.0 (Ref) |
| 3.0 – 9.9 % | 22 | 71 | 0.98 (0.49, 1.94) | 0.90 (0.45, 1.80) |
| 10.0 % | 19 | 76 | 0.86 (0.42, 1.77) | 0.70 (0.33, 1.50) |
| Per 1% increase | | | 1.00 (0.98, 1.03) | 1.00 (0.97, 1.02) |
| p-trend | | | 0.86 | 0.84 |
| **Percentage of Ki67[+] cells** | | | | |
| < 0.28 % | 13 | 64 | 1.0 (Ref) | 1.0 (Ref) |
| 0.28 – 2.0 % | 18 | 60 | 1.66 (0.72, 3.85) | 1.63 (0.70, 3.80) |
| 2.0 % | 21 | 55 | 1.94 (0.85, 4.44) | 1.92 (0.83, 4.43) |
| Per 1% increase | | | 1.05 (0.98, 1.13) | 1.05 (0.98, 1.13) |
| p-trend | | | 0.15 | 0.14 |
| **Percentage of p27[+]ER[+] cells** | | | | |
| < 1.5 % | 22 | 72 | 1.0 (Ref) | 1.0 (Ref) |
| 1.5 – 5.4 % | 26 | 68 | 1.33 (0.67, 2.61) | 1.23 (0.62, 2.45) |
| 5.5 % | 17 | 79 | 0.77 (0.37, 1.60) | 0.69 (0.33, 1.46) |
| Per 1% increase | | | 1.00 (0.96, 1.04) | 1.00 (0.96, 1.04) |
| p-trend | | | 0.91 | 0.97 |
| **Percentage of p27[+]Ki67[+] cells** | | | | |
| < 0.16 % | 22 | 82 | 1.0 (Ref) | 1.0 (Ref) |
| 0.16 – 1.4 % | 11 | 40 | 1.11 (0.47, 2.61) | 1.24 (0.52, 2.96) |
| 1.5 % | 19 | 57 | 1.29 (0.61, 2.71) | 1.36 (0.64, 2.90) |
| Per 1% increase | | | 1.03 (0.86, 1.22) | 1.05 (0.88, 1.25) |
| p-trend | | | 0.76 | 0.62 |

[a]Model 1 (Matching factors only): age at diagnosis (continuous, years), calendar year of biopsy (<1970, 1970–1980, 1980+), time since biopsy (continuous, years).

[b]Model 2: Model 1 + type of benign lesion (nonproliferative, proliferative without atypia, proliferative with atypia).

**Table 3**

**Odds ratios (95% confidence interval) of developing subsequent breast cancer according to tertiles of the percentage of p27[+], ER[+], and Ki67[+] in normal breast tissue among premenopausal women in the Nurses' Health Study and the Nurses' Health Study II**

Adjusted for matching factors only: age at diagnosis (continuous, years), calendar year of biopsy (<1970, 1970–1980, 1980+), time since biopsy (continuous, years).

| | Case | Control | OR (95% CI) |
|---|---|---|---|
| **Percentage of p27[+] cells** | | | |
| < 9.0 % | 18 | 58 | 1.0 (Ref) |
| 9.0 – 24.9 % | 17 | 52 | 0.97 (0.44, 2.12) |
| 25.0 % | 14 | 62 | 0.72 (0.32, 1.61) |
| Per 1% increase | | | 0.99 (0.97, 1.02) |
| p-trend | | | 0.60 |
| **Percentage of ER[+] cells** | | | |
| < 2.9 % | 17 | 53 | 1.0 (Ref) |
| 2.9 – 9.1 % | 17 | 52 | 1.06 (0.47, 2.35) |
| 9.2 % | 12 | 59 | 0.74 (0.32, 1.74) |
| Per 1% increase | | | 0.99 (0.96, 1.03) |
| p-trend | | | 0.68 |
| **Percentage of Ki67[+] cells** | | | |
| < 0.5 % | 2 | 54 | 1.0 (Ref) |
| 0.5 – 2.1 % | 17 | 41 | 12.1 (2.56, 57.4) |
| 2.2 % | 16 | 42 | 10.1 (2.12, 48.0) |
| Per 1% increase | | | 1.10 (1.01, 1.19) |
| p-trend | | | *0.02* |
| **Percentage of p27[+]ER[+] cells** | | | |
| < 1.6 % | 17 | 53 | 1.0 (Ref) |
| 1.6 – 5.2 % | 17 | 52 | 1.52 (0.69, 3.32) |
| 5.2 % | 12 | 59 | 0.61 (0.25, 1.50) |
| Per 1% increase | | | 0.98 (0.92, 1.04) |
| p-trend | | | 0.52 |
| **Percentage of p27[+]Ki67[+] cells** | | | |
| < 0.1 % | 16 | 54 | 1.0 (Ref) |
| 0.1 – 0.5 % | 20 | 48 | 9.71 (2.65, 35.6) |
| 0.5 % | 10 | 62 | 3.47 (1.43, 8.42) |
| Per 1% increase | | | 1.31 (1.01, 1.70) |
| p-trend | | | *0.04* |

**Table 4**

**Odds ratios (95% confidence interval) of developing subsequent breast cancer according to cross-classified groups of Ki67 and p27, Ki67 and ER, p27 and ER in normal breast tissue among premenopausal women in the Nurses' Health Study and the Nurses' Health Study II**

The fraction of p27[+], ER[+], and Ki67[+] in normal breast tissue were dichotomized around the median (above vs. below median percentage of stain-positive cells). Odds ratio and 95% confidence intervals were estimated using matching factors only model: age at diagnosis (continuous, years), calendar year of biopsy (<1970, 1970–1980, 1980+), time since biopsy (continuous, years).

| Tissue expression | | Case | Control | OR (95% CI) |
|---|---|---|---|---|
| **Ki67 and p27 (cross-classified)** | | | | |
| Low Ki67 (<1.15 %) | Low p27 (<16.6 %) | 4 | 39 | 1.0 (Ref) |
| | High p27 (16.6+ %) | 6 | 35 | 1.99 (0.49, 8.0) |
| High Ki67 (1.15+ %) | Low p27 (<16.6 %) | 14 | 30 | *5.08 (1.43, 18.1)* |
| | High p27 (16.6+ %) | 11 | 31 | 3.5 (0.98, 12.5) |
| **Ki67 and ER (cross-classified)** | | | | |
| Low Ki67 (<1.15 %) | Low ER (<5.3 %) | 6 | 43 | 1.0 (Ref) |
| | High ER (5.3+ %) | 3 | 26 | 1.08 (0.23, 5.01) |
| High Ki67 (1.15+ %) | Low ER (<5.3 %) | 20 | 33 | *4.68 (1.63, 13.5)* |
| | High ER (5.3+ %) | 3 | 25 | 0.90 (0.20, 4.08) |
| **p27 and ER (cross-classified)** | | | | |
| Low ER (<5.3 %) | Low p27 (<16.6 %) | 18 | 61 | 1.0 (Ref) |
| | High p27 (16.6+ %) | 18 | 43 | 1.51 (0.69, 3.31) |
| High ER (5.3+ %) | Low p27 (<16.6 %) | 6 | 18 | 1.27 (0.43, 3.79) |
| | High p27 (16.6+ %) | 4 | 40 | 0.39 (0.12, 1.25) |