Published in final edited form as:
Breast Cancer Res Treat. 2012 January ; 131(1): 159–167. doi:10.1007/s10549-011-1702-0.

Traditional Breast Cancer Risk Factors in Relation to Molecular Subtypes of Breast Cancer

Rulla M. Tamimi[1,2], **Graham A. Colditz**[3], **Aditi Hazra**[1,2], **Heather J. Baer**[1,2], **Susan E. Hankinson**[1,2], **Bernard Rosner**[1], **Jonathan Marotti**[4], **James L. Connolly**[4], **Stuart J. Schnitt**[4], and **Laura C. Collins**[4]

[1]Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, 02115

[2]Department of Epidemiology, Harvard School of Public Health, Boston, MA, 02115

[3]Department of Surgery, Washington University School of Medicine, St. Louis, MO 63110

[4]Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02115

Abstract

Background—At least four major categories of invasive breast cancer have been reproducibly identified by gene expression profiling: luminal A, luminal B, HER2-type and basal-like. These subtypes have been shown to differ in their outcome and response to treatment. Whether this heterogeneity reflects the evolution of these subtypes through distinct etiologic pathways has not been clearly defined.

Methods—We evaluated the association between traditional breast cancer risk factors and risk of previously defined molecular subtypes of breast cancer in the Nurses' Health Study. This analysis included 2,022 invasive breast cancer cases for whom we were able to obtain archived breast cancer tissue specimens. Tissue microarrays (TMAs) were constructed and slides were immunostained for estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), cytokeratin 5/6 (CK5/6), and epidermal growth factor receptor (EGFR). Using immunostain results in combination with histologic grade, cases were grouped into molecularly defined subtypes. We used Cox proportional hazards models to estimate hazard ratios (HRs) and 95% confidence intervals (CIs).

Results—We observed differences in the association between risk factors and subtypes of breast cancer. In general, many reproductive factors were most strongly associated with the luminal A subtype, although these differences were not statistically significant. Weight gain since age 18 showed significant differences in its association with molecular subtypes (p-heterogeneity=0.05) and was most strongly associated with the luminal B subtype (p-trend 0.001). Although there was not significant heterogeneity for lactation across subtypes, an inverse association was strongest for basal-like tumors (HR=0.6, 95%CI 0.4–0.8; p-heterogeneity=0.88).

Conclusions—These results support the hypothesis that different subtypes of breast cancer have different etiologies and should not be considered as a single group. Identifying risk factors for less common subtypes such as luminal B, HER2-type and basal-like tumors has important implications for prevention of these more aggressive subtypes.

Corresponding Author: Rulla M. Tamimi, Channing Laboratory, 181 Longwood Avenue, Boston, MA 02115, Telephone: (617) 525-0862, rulla.tamimi@channing.harvard.edu.

Keywords

breast cancer; risk factors; molecular subtype; ER; PR; HER2

INTRODUCTION

Recent studies using microarray technology and unsupervised cluster analysis have provided new insights into the classification of invasive breast cancers [1–4]. These studies have resulted in the identification of several breast cancer subtypes that vary in their gene expression signatures and clinical outcome. The molecularly distinct breast cancer subgroups identified to date include luminal subtypes A and B (both of which are hormone receptor-positive), the HER2-type, and a group known as basal-like cancers [1–4]. Although prognosis and response to treatment has been shown to vary according to these subtypes, it is unclear if classifying breast cancer in this way may help us to understand better the etiology of breast cancer. Examination of breast cancer risk factors in relation to these subtypes may offer the potential to build on traditional classification of tumor types based on ER and PR status [5, 6] and extend insights into etiology.

Immunohistochemical staining of paraffin sections using antibody panels has been shown to be a reasonable, albeit imperfect, surrogate for molecular classification of invasive breast cancers as categorized by gene expression profiling studies [4, 7–10]. Antibodies against estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), cytokeratin 5/6 (CK5/6) and epidermal growth factor receptor (EGFR) have been particularly useful for this purpose in populations in which material for expression profiling is not available or for which studying very large numbers of cases would not be feasible.

The heterogeneity of breast tumors may be a reflection of distinct etiologic pathways. For example, risk factors that influence estrogen levels (i.e, circulating hormones, postmenopausal hormone use and postmenopausal adiposity), are associated with ER-positive, but not ER- negative breast cancer [5, 6, 11]. The additional markers used to classify the molecular phenotypes beyond ER and PR may allow for greater refinement of tumor subtypes and allow for identification of distinct etiologic pathways. Because luminal A and B tumor subtypes are both hormone receptor positive and together represent the majority of breast cancer cases it is not surprising that the majority of identified breast cancer risk factors are those associated with hormonal exposures. By examining risk factors in relation to more homogenous subtypes we may identify novel risk factors for less common subtypes of breast cancer. In the current study, we examine the association between traditional breast cancer risk factors and the risk of the previously defined molecular subtypes of breast cancer.

MATERIALS AND METHODS

Study population

Study Design and Population—The Nurses' Health Study was initiated in 1976 when 121,700 U.S. registered nurses ages 30–55 returned an initial questionnaire. The cohort has been followed by mailed questionnaires biennially to update exposure information and ascertain non-fatal incident diseases. Information on body mass index (BMI), reproductive history, age at menopause, and postmenopausal hormone (PMH) use as well as diagnosis of cancer and other diseases are updated every two years through questionnaires. The follow-up rate among this cohort was over 90% through 1996[12].

Breast cancer case confirmation

All women reporting incident diagnoses of cancer were asked for permission to review their medical records to confirm the diagnosis and to classify cancers as in situ or invasive, by histologic type, size, and presence or absence of metastases. To identify cases of cancer in nonrespondents who died, death certificates for all deceased participants and medical records for the incident cancers were obtained. Following medical record review, 99% of self-reported breast cancers were confirmed.

Breast cancer tissue block collection and tissue microarray (TMA) construction

Detailed description of the tissue block collection and TMA construction have been described previously [13, 14]. Briefly, we constructed tissue microarrays from 3,093 cancers and positive lymph nodes from 2,897 participants. TMAs were constructed in the Dana Farber Harvard Cancer Center Tissue Microarray Core Facility, Boston, MA. Three 0.6-mm cores were obtained from each breast cancer and were inserted into the recipient TMA blocks. We excluded from the current analysis participants with positive lymph nodes only (n=25), lobular carcinoma in situ (n=31), in situ carcinomas with both ductal and lobular features (n=13), ductal carcinoma in situ (n=272), and additional rare tumor types including malignant phyllodes tumors, neuroendocrine carcinoma and angiosarcoma (n=10).

Immunohistochemical analysis

We performed immunohistochemical staining for ER, PR, HER2, CK5/6, and EGFR on 5-μm paraffin sections cut from the TMA blocks. Immunostains for each marker were performed in a single staining run on a Dako Autostainer (Dako Corporation, Carpinteria, CA). These particular biomarkers were selected for analysis because they have been commonly used as surrogates to classify invasive breast cancers according to their molecular phenotypes [4, 7–10]. Sources and dilutions of the primary antibodies used in this study are listed in Appendix 1. Immunostains for ER, PR, HER2, CK 5/6 and EGFR were performed using methods described in detail previously [13, 15]. Appropriate positive and negative controls were included in all staining runs.

Immunostained TMA slides were evaluated for ER and PR expression, HER2 protein over-expression, and expression of CK5/6 and EGFR in each core. Tumor cells that showed any nuclear staining for ER or PR were considered ER-positive or PR-positive respectively, whereas all ER-negative or PR-negative cases showed complete absence of tumor cell staining in all tissue cores. Of note, low positive ER or PR (1–10% of tumor cell nuclei staining) and positive ER or PR (>10% of tumor cell nuclei staining) were collapsed into a single “positive” category for the purposes of this analysis. Tumor cells were considered positive for HER2 protein over-expression when more than 10% of the cells showed moderate or strong membrane staining (2+ and 3+). The results of analyses in which HER2 positivity was defined as 3+ were very similar to those presented with a definition of 2+ and 3+. Cases were considered basal CK-positive or EGFR-positive if any cytoplasmic and/or membranous staining was detected in the tumor cells, even if focal. These latter criteria are similar to those previously used for scoring these markers in invasive basal-like cancers [4, 7, 8].

Classification of Molecular Subtypes

Immunostained TMA sections were reviewed under a microscope and visually scored for each individual tissue core as described above. We classified a case as positive if there was staining in any of the three cores from that case and negative if there was no immunostaining present. Based on RNA expression data [1–3] and previous large scale epidemiologic studies [16, 17] we used a panel of immunohistochemical markers to classify

the tumors into molecular subtypes. Cases that were ER-positive and/or PR-positive and HER2-negative and histologic grade 1 and 2 were classified as luminal A cancers; cases that were either a.) ER-positive and/or PR-positive and HER2-positive or b.) ER-positive and/or PR-positive, HER2-negative and histologic grade 3 were classified as luminal B; cases that were ER-negative, PR-negative, and HER2-positive were classified as HER2 type; and cases that were negative for ER, PR, and HER2 and positive for CK 5/6 and/or EGFR were categorized as basal-like. Cases that lacked expression of all 5 markers were considered "unclassified". Of the invasive tumors on tissue microarrays, 2249 could be classified into one of these 5 molecular subtypes.

Statistical Analysis

Information on breast cancer risk factors was obtained from biennial questionnaires. Women who reported a diagnosis of cancer other than nonmelanoma skin cancer were excluded at baseline and from subsequent follow-up analysis. Person-time for each participant was calculated from the date of return of the 1976 questionnaire to the date of breast cancer diagnosis, date of any other cancer diagnosis (not including nonmelanoma skin cancer), death from any cause or June 1, 1998, which ever came first.

The primary analysis used incidence rates with person-months in the denominator.. Person-time for each participant was calculated from the date of return of the 1976 questionnaire to the date of breast cancer diagnosis, date of any other cancer diagnosis (not including nonmelanoma skin cancer), death from any cause or June 1, 1998, whichever came first. The primary analysis used incidence rates with person-months in the denominator. For each woman, person-months were allocated to each exposure category, beginning in 1976 and updated every two years. We used Cox proportional hazards models to estimate hazard ratios (RRs) and 95% confidence intervals (CIs). Multivariate analysis included age at menopause, family history of breast cancer in a first-degree relative, personal history of benign breast disease (BBD), body mass index (BMI) at age 18, weight change since age 18, age at menarche, parity/age at first birth, alcohol consumption, menopausal status/PMH use, lactation, and smoking. These variables were considered because they are either well established risk factors for breast cancer [18] or have been reported to be associated with a particular molecular phenotype of breast cancer[16, 17].

To determine if the association between exposures is differentially associated with tumor subtypes, we used competing risks models[19, 20]. Specifically, this approach uses the data augmentation method described by Lunn and McNeil [21] to create a separate observation for each subject for each type of outcome and then stratifies on event type, allowing for estimation of separate associations of each risk factor with the relative hazard of each type of outcome [20]. Likelihood ratio tests were conducted to compare models assuming different associations of exposures with each subtype of tumor to models assuming the same association with all types; a significant p value for this test of heterogeneity would indicate that the associations are different for the different tumor subtypes. All analyses were conducted with SAS version 9.2 (SAS, Cary, North Carolina). All statistical tests were two-sided and p-values <0.05 were considered statistically significant. This study was approved by the Committee on the Use of Human Subjects in Research at Brigham and Women's Hospital.

RESULTS

During the course of follow-up, 2,022 invasive breast cancer cases were identified that could be classified into one of five molecular phenotypes. Based on immunostaining data from the five markers used in conjunction with histologic grade (which was used as a surrogate for Ki67 proliferation index), 1,267 tumors were classified as luminal A; 321 were luminal B;

113 were HER2; 226 were basal-like and 95 tumors were unclassifiable (ER-/PR-/HER2-/EGFR-/CK5/6-). In addition, 3,549 breast cancers occurred among women for whom we were unable to classify into these phenotypes due to lack of either tissue availability or staining results for one or more of the five markers. Compared with women with tissue specimens, the women for whom we were unable to obtain specimens were very similar with respect to breast cancer risk factors and tumor characteristics [13].

The mean age at diagnosis ranged from 55.3 years for women with basal like tumors to 58.0 years for women with luminal A tumors (Table 1). In general, the luminal A tumors were smaller, less likely to have nodal involvement, were of lower stage and were more often of lower histologic grade than the other four molecular subtypes. The luminal B and basal like tumors were more often high grade, while the HER2 type was most likely to be stage III/IV (33.0%) and the unclassified type was most likely to be metastatic at diagnosis (8.1%) relative to the other subtypes.

We observed differences in the association between breast cancer risk factors and molecular subtypes of breast cancer (Table 2). In general, reproductive risk factors for breast cancer including age at menarche, parity, and age at first birth tended to be most strongly associated with the luminal A subtype, although there was no evidence of statistical heterogeneity across the subtypes. For example, compared to nulliparous women, having three or more children was inversely associated with luminal A breast cancer (HR=0.7, 95%CI 0.5–1.0; p-trend=0.01). However, this inverse association was not observed among the other subtypes.

We found that hormonal exposures in later adult life exhibited the greatest heterogeneity in association with these molecular subtypes. Weight gain since age 18 was positively associated with both luminal A (p-trend=0.05) and B tumor (p-trend=0.001) subtypes, but not with the ER- subtypes (p-heterogeneity=0.05). The association between weight gain since age 18 and luminal B tumors was significantly stronger than the association with luminal A tumors (p-heterogeneity=0.0007). There was suggestive evidence that the association between postmenopausal hormone use and breast cancer may also vary by subtype (p-heterogeneity=0.08). We observed a significant association between estrogen only hormone therapy (HR=1.4, 95%CI 1.1–1.7) and estrogen plus progestin therapy (HR=1.5, 95%CI 1.2–2.0) and risk of luminal A tumors, but not luminal B tumors (P-heterogeneity=0.004). Estrogen plus progestin therapy was also associated with basal-like (HR=1.8, 95% CI 1.0–3.4) and unclassified (HR=2.9, 95% CI 1.1–7.6) tumors, but not luminal B or HER2-type.

Family history of breast cancer was differentially associated with breast cancer subtypes (p=0.01). Having one first degree relative with breast cancer was significantly associated with luminal A and B subtypes only, while having two first degree relatives with breast cancer was associated with an increased risk of luminal A (HR=2.3, 95% CI 1.3–4.2), HER2-type (HR=2.5, 95% CI 0.3–18.1), and basal-like (HR=2.9, 95% CI 0.7–11.7) tumors, and not the others.

There was little evidence that the association between the other risk factors we considered differed by molecular subtypes. Having a prior benign breast disease was associated with a 20–70% increased risk of all breast cancer subtypes (p-heterogeneity=0.69). BMI at age 18 was inversely associated with luminal A, basal-like and unclassified subtypes of breast cancer (p-heterogeneity=0.49). Although there were no significant differences in the association between lactation across molecular subtypes, an inverse association was strongest for the basal-like tumors (HR=0.6, 95%CI 0.4–0.9; p-heterogeneity=0.88).

DISCUSSION

In this study of over 2,000 breast cancer cases, we found significant differences in the association between breast cancer risk factors and molecular subtypes of tumors. As expected many reproductive risk factors including age at menarche, parity, age at first birth, age at menopause, and postmenopausal hormone use were associated with luminal A tumors, the most common type of breast of cancer. In general, hormonally related risk factors in later adult life demonstrated the most heterogeneity. The association between postmenopausal hormone use and luminal A tumors was significantly stronger than the relation with luminal B tumors (the other ER+ subtype). Interestingly, weight gain since age 18 was more strongly associated with luminal B tumors than luminal A tumors.

Unexpectedly, we found that some hormonal factors were associated with hormone receptor negative subtypes. For example, age at menopause was significantly associated with the HER2-type and unclassified subtype, and current estrogen plus progestin hormone use was strongly associated with both basal-like and unclassified tumor types.

A number of other studies have evaluated the association between breast cancer risk factors and tumor subtypes, although only a handful have evaluated markers beyond ER, PR and HER2. The Polish Breast Cancer Study (n=804 breast cancer cases) also found that most established breast cancer risk factors were associated with luminal A tumors [16]. However, they reported that age at menarche was more strongly inversely associated with basal-like tumors than luminal A tumors (p-heterogeneity =0.0009), which we did not observe in the current study. Similar to our study, the Polish Breast Cancer Study found that having a family history of breast cancer was a risk factor for almost all subtypes although the magnitude of the effect was greatest for basal-like and HER2-type breast cancers.

The Carolina Breast Cancer Study (CBCS) (n=1424 breast cancer cases) also examined the association between risk factors for molecular subtypes of breast cancer in a case-control study of both Caucasian and African American women[17]. Millikan *et al.* found that increasing parity was associated with reduced risk of luminal A tumors and an increased risk of basal-like tumors. Our results with respect to parity are consistent with this finding. In addition, The CBCS reported an inverse association between lactation and basal-like tumors. Although there was no significant heterogeneity between lactation and subtype in our study, we did find a strong inverse association between lactation and basal-like tumors. For women with total breast feeding of 4+ months, we found a 40% reduced risk of basal-like breast cancer, which is in line with the 30% reduced risk observed in the CBCS.

In addition, studies examining risk factors in relation to tumors classified with information on ER, PR and HER2 only have also been conducted. A combined study of the LACE and Pathways studies within Kaiser Permanente Northern California examined breast cancer risk factors in relation to subtypes defined by ER, PR, and HER2. In this study of 2544 invasive breast cancer cases, Kwan et al [22]found that relative to luminal A cases (ER+ and/or PR+/HER2-), luminal B cases (ER+ and/or PR+/HER2+) were less likely to consume alcohol and use HRT. Breast feeding for at least four months was associated with a lower risk of triple negative cases (ER-/PR-/HER2-) compared with luminal A. Similarly, two other studies Phipps et al [23](n=1130 total cases) and Gaudet et al [24](n=890 total cases), also reported an inverse association between breastfeeding for 6 or more months and triple negative breast tumors.

Of interest, a number of risk factors in our study did not demonstrate heterogeneity across tumor subtypes including age at menarche, BMI at age 18, previous BBD, and alcohol consumption. It is possible that these factors are having a similar effect on risk across the different subtypes and this may be indicating how these factors are affecting breast cancer

etiology. For example, having a prior BBD may indicate having early proliferative lesions which could have developed through a number of different pathways. BBD is believed to be a general marker of breast cancer risk, and thus may reflect the culmination of many risk factors and not be specific to any one pathway. It is also possible that we may not have had enough power to detect the difference across subtypes for some exposures.

Our classification of tumor subtypes was similar although not identical to those used in previous epidemiologic studies [16, 17]. Both of the prior studies utilized immunohistochemical markers to define molecular subtypes, while we also incorporated histologic grade. Others have shown that the distinction between luminal A and B tumors can be refined by adding the proliferation marker Ki67 to ER, PR, and HER2[25]. Given that Ki67 data were not available for our cases, we used histologic grade as a surrogate for proliferation rate given the close correlation between proliferation rate and histologic grade. Thus, our definitions for luminal A and B are different than the two previous studies, but more in keeping with the most recently proposed classification scheme[25]. This may limit our ability to compare across studies and explain some of the differences observed.

The results of the current study taken together with the previous studies suggest that many of the traditional breast cancer risk factors are associated with luminal A tumors, and that there may be some differences with other subtypes. The associations with luminal A tumors is not surprising since these are the most common tumor type and may also reflect the selection of exposures we have focused in the current study. We have chosen to examine traditional breast cancer risk factors; these were initially identified because they are the risk factors shown to be most commonly associated with breast cancer. The majority of these risk factors have also been shown to be associated with the most common type of breast cancer, namely ER+ breast cancers. While it is reassuring that these risk factors are associated with the luminal A subtype, it does not help us to further our understanding of risk factors associated with ER- subtypes. Adding the additional markers and dividing the cases into smaller subsets limits our power to detect associations for the rare subtypes. One of the most consistent findings from this study and other studies is the inverse association between breast feeding and triple negative or basal like tumors. This finding supports the hypothesis that molecular classification of tumors may help us to better understand etiology and/or provide insights into the mechanisms by which these less common molecular subtypes develop.

This study has a number of strengths including the large study population with over 2,000 invasive breast cancer cases, the prospectively collected nature of the exposure variables, and uniform staining and scoring of molecular markers. Despite the large number of cases, we were still limited by the number of less common subtypes in particular the HER2-type and the unclassified tumors. Thus, our power to detect significant differences by these subtypes was limited. In addition, we lacked adequate power to examine these associations among premenopausal breast cancer cases. It is worth noting that the frequency of receptor status positivity and molecular subtype frequency among invasive tumors in our study population was very similar to other populations suggesting that samples included in this study are representative of the overall US population.

In conclusion, in this study we found that traditional breast cancer risk factors demonstrated different relations with the molecular subtypes of breast cancer. In general, many of the reproductive factors were most strongly associated with the luminal A subtype. In addition, we confirmed a previously reported strong inverse association between lactation and basal-like tumors. It is unclear whether classifying breast tumors according to the molecular phenotypes (ie. subtypes that are known to have prognostic importance) permits identification of differences in risk factor profiles. Additional work to determine if the

differences that have been observed are due to single markers or to the molecular phenotype is necessary. It remains to be seen whether non-traditional breast cancer risk factors may exist that are associated with less common tumor subtypes. Identifying risk factors for these less common subtypes such as HER2 and basal-like tumors, which also have a poorer prognosis, has important implications for prevention of these tumor subtypes.

Acknowledgments

Funding/Support: Supported by GlaxoSmithKline (WE234 (EPI40307)); Public Health Service Grants CA087969, and SPORE in Breast Cancer CA089393, from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services and Breast Cancer Research Fund. Dr. Graham Colditz is supported in part by an American Cancer Society Cissy Hornung Clinical Research Professorship.

We thank the participants and staff of the Nurses' Health Study cohort, for their valuable contributions. We thank the following state cancer registries for their help: AL, AZ, AR, CA, CO, CT, DE, FL, GA, ID, IL, IN, IA, KY, LA, ME, MD, MA, MI, NE, NH, NJ, NY, NC, ND, OH, OK, OR, PA, RI, SC, TN, TX, VA, WA, WY. The study sponsors had no role in the design of the study; the collection, analysis, and interpretation of the data; the writing of the manuscript; or the decision to submit the manuscript for publication.

References

1. Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC, et al. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci U S A. 1999; 96(16):9212–9217. [PubMed: 10430922]
2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001; 98(19):10869–10874. [PubMed: 11553815]
3. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003; 100(14):8418–8423. [PubMed: 12829800]
4. Brenton JD, Carey LA, Ahmed AA, Caldas C. Molecular classification and molecular forecasting of breast cancer: ready for clinical application? J Clin Oncol. 2005; 23 (29):7350–7360. [PubMed: 16145060]
5. 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):218–228. [PubMed: 14759989]
6. Ma H, Bernstein L, Pike MC, Ursin G. Reproductive factors and breast cancer risk according to joint estrogen and progesterone receptor status: a meta-analysis of epidemiological studies. Breast Cancer Res. 2006; 8(4):R43. [PubMed: 16859501]
7. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res. 2004; 10(16):5367–5374. [PubMed: 15328174]
8. Abd El-Rehim DM, Ball G, Pinder SE, Rakha E, Paish C, Robertson JF, Macmillan D, Blamey RW, Ellis IO. High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses. Int J Cancer. 2005; 116(3):340–350. [PubMed: 15818618]
9. Abd El-Rehim DM, Pinder SE, Paish CE, Bell J, Blamey RW, Robertson JF, Nicholson RI, Ellis IO. Expression of luminal and basal cytokeratins in human breast carcinoma. J Pathol. 2004; 203(2): 661–671. [PubMed: 15141381]
10. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Kochli OR, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol. 2002; 161(6):1991–1996. [PubMed: 12466114]

11. Missmer SA, Eliassen AH, Barbieri RL, Hankinson SE. Endogenous estrogen, androgen, and progesterone concentrations and breast cancer risk among postmenopausal women. J Natl Cancer Inst. 2004; 96(24):1856–1865. [PubMed: 15601642]
12. Colditz GA, Hankinson SE. The Nurses' Health Study: lifestyle and health among women. Nat Rev Cancer. 2005; 5(5):388–396. [PubMed: 15864280]
13. Tamimi RM, Baer HJ, Marotti J, Galan M, Galaburda L, Fu Y, Deitz AC, Connolly JL, Schnitt SJ, Colditz GA, et al. Comparison of molecular phenotypes of ductal carcinoma in situ and invasive breast cancer. Breast Cancer Res. 2008; 10(4):R67. [PubMed: 18681955]
14. Collins LC, Marotti JD, Baer HJ, Tamimi RM. Comparison of estrogen receptor results from pathology reports with results from central laboratory testing. J Natl Cancer Inst. 2008; 100(3): 218–221. [PubMed: 18230800]
15. Collins LC, Botero ML, Schnitt SJ. Bimodal frequency distribution of estrogen receptor immunohistochemical staining results in breast cancer: An analysis of 825 cases. Am J Clin Pathol. 2005; 123:16–20. [PubMed: 15762275]
16. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev. 2007; 16(3):439–443. [PubMed: 17372238]
17. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Smith LV, Labbok MH, Geradts J, Bensen JT, Jackson S, et al. Epidemiology of basal-like breast cancer. Breast Cancer Res Treat. 2008; 109(1):123–139. [PubMed: 17578664]
18. Hankinson, S.; Hunter, DJ. Breast Cancer. In: Adami, HOHD.; Trichopoulos, D., editors. Textbook of Cancer Epidemiology. New York: Oxford University Press; 2002. p. 301-339.
19. Glynn RJ, Rosner B. Methods to evaluate risks for composite end points and their individual components. J Clin Epidemiol. 2004; 57(2):113–122. [PubMed: 15125620]
20. Glynn RJ, Rosner B. Comparison of risk factors for the competing risks of coronary heart disease, stroke, and venous thromboembolism. Am J Epidemiol. 2005; 162(10):975–982. [PubMed: 16207808]
21. Lunn M, McNeil D. Applying Cox regression to competing risks. Biometrics. 1995; 51 (2):524–532. [PubMed: 7662841]
22. Kwan ML, Kushi LH, Weltzien E, Maring B, Kutner SE, Fulton RS, Lee MM, Ambrosone CB, Caan BJ. Epidemiology of breast cancer subtypes in two prospective cohort studies of breast cancer survivors. Breast Cancer Res. 2009; 11(3):R31. [PubMed: 19463150]
23. Phipps AI, Malone KE, Porter PL, Daling JR, Li CI. Reproductive and hormonal risk factors for postmenopausal luminal, HER-2-overexpressing, and triple-negative breast cancer. Cancer. 2008; 113(7):1521–1526. [PubMed: 18726992]
24. Gaudet MM, Press MF, Haile RW, Lynch CF, Glaser SL, Schildkraut J, Gammon MD, Douglas Thompson W, Bernstein JL. Risk factors by molecular subtypes of breast cancer across a population-based study of women 56 years or younger. Breast Cancer Res Treat. 2011 Jun 11. [Epub ahead of print].
25. Cheang MC, Chia SK, Voduc D, Gao D, Leung S, Snider J, Watson M, Davies S, Bernard PS, Parker JS, et al. Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst. 2009; 101(10):736–750. [PubMed: 19436038]

Table 1

Tumor characteristics according to breast cancer phenotypes among women with invasive breast cancer, Nurses' Health Study (1976–1996).

Characteristic	Luminal A	Luminal B	HER2	Basal-like	Unclassified
N (%)	**1267 (62.7)**	**321 (15.9)**	**113 (5.6)**	**226 (11.2)**	**95 (4.7)**
Mean age at diagnosis, yrs	58.0	56.5	56.1	55.3	55.6
Median age at diagnosis, yrs	59.0	57.0	56.0	56.0	55.0
Tumor size[1]					
0.1 to 1.0 cm	323 (26.7)	47 (15.3)	18 (17.0)	27 (12.9)	19 (21.6)
1.1 to 2.0 cm	503 (41.6)	116 (37.8)	27 (25.5)	74 (35.2)	31 (35.2)
2.1 to 4.0 cm	281 (23.2)	100 (32.6)	44 (41.5)	87 (41.4)	24 (27.3)
4.1+ cm	103 (8.5)	44 (14.3)	17 (16.0)	22 (10.5)	14 (15.9)
Missing	57	14	7	16	7
Lymph node status[1]					
No Nodes	787 (66.3)	165 (55.4)	47 (43.9)	119 (56.9)	50 (57.5)
1–3 Nodes	234 (19.7)	65 (21.8)	28 (26.2)	54 (25.8)	14 (16.1)
4–9 Nodes	86 (7.2)	38 (12.8)	14 (13.1)	22 (10.5)	9 (10.3)
10+ Nodes	55 (4.6)	23 (7.7)	13 (12.2)	7 (3.4)	7 (8.1)
Metastatic at diagnosis	26 (2.2)	7 (2.4)	5 (4.7)	7 (3.4)	7 (8.1)
Missing	79	23	6	17	8
Stage at diagnosis[1]					
I/II	1026 (84.6)	230 (74.7)	71 (67.0)	173 (80.8)	66 (73.3)
III/IV	187 (15.4)	78 (25.3)	35 (33.0)	41 (19.2)	24 (26.7)
Missing	54	13	7	12	5
Grade[2]					
Well differentiated	366 (28.9)	8 (2.7)	3 (2.8)	7 (3.1)	16 (16.8)
Int. differentiated	901 (71.1)	57 (18.9)	54 (49.5)	57 (25.3)	37 (39.0)
Poorly differentiated	0	237 (78.5)	52 (47.7)	161 (71.6)	42 (44.2)
Missing	0	19	4	1	0
Histology[1]					
Invasive ductal	1002(79.4)	300 (94.0)	111 (99.1)	210 (94.6)	79 (85.9)
Invasive lobular	180 (14.3)	10 (3.1)	0	3 (1.4)	8 (8.7)

Characteristic	Luminal A	Luminal B	HER2	Basal-like	Unclassified
N (%)	**1267 (62.7)**	**321 (15.9)**	**113 (5.6)**	**226 (11.2)**	**95 (4.7)**
Invasive ductal and lobular	70 (5.6)	5 (1.6)	0	0	2 (2.2)
Invasive not specified	10 (0.8)	4 (1.3)	1 (0.9)	9 (4.1)	3 (3.3)
Missing	5	2	1	4	3
Race					
White	1209 (95.4)	304 (94.7)	109 (96.5)	210 (92.9)	90 (94.7)
Black	11 (0.9)	7 (2.2)	1 (0.9)	8 (3.5)	2 (2.1)
American Indian	2 (0.2)	0	0	0	0
Asian	4 (0.3)	0	1 (0.9)	2 (0.9)	0
Hawaiian	0	0	0	1 (0.4)	0
Other	31 (2.5)	8 (2.5)	0	4 (1.8)	3 (3.2)
Multiracial	10 (0.8)	2 (0.6)	2 (1.8)	1 (0.4)	0

[1] Information obtained from pathology records

[2] Information obtained from centralized pathology review of slides

Table 2

Relative risk[1] of invasive breast cancer molecular subtypes breast cancer according to breast cancer risk factors

	Luminal A	Luminal B	HER2	Basal	Unclassified
Cases	1267	321	113	226	95
Person-years	2008898	2009767	2009968	2009865	2009986
Age at menarche					
<12	1.0 (ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)
12	1.0 (0.8–1.1)	1.1 (0.8–1.5)	1.4 (0.8–2.3)	0.8 (0.5–1.1)	0.8 (0.4–1.3)
13	0.8 (0.7–1.0)	1.1 (0.8–1.5)	0.9 (0.5–1.5)	0.7 (0.5–1.0)	0.5 (0.3–0.9)
14	0.7 (0.6–0.8)	1.0 (0.7–1.5)	0.9 (0.4–1.8)	0.8 (0.5–1.3)	0.6 (0.3–1.2)
14+	0.7 (0.5–0.9)	1.0 (0.6–1.6)	1.1 (0.5–2.3)	0.8 (0.5–1.3)	0.8 (0.4–1.7)
P for Trend	0.002	0.14	0.31	0.72	0.11
P-Heterogeneity	0.92				
BMI at age 18					
<20	1.0 (ref)	1.0 (ref)	1.0 (ref)	1.0 (Ref)	1.0 (Ref)
20–21.9	0.9 (0.7–1.0)	1.0 (0.8–1.3)	0.6 (0.4–1.0)	0.8 (0.5–1.1)	0.5 (0.3–0.9)
22.0–23.9	0.8 (0.7–1.0)	0.8 (0.6–1.2)	0.7 (0.4–1.3)	0.9 (0.6–1.3)	0.6 (0.3–1.2)
24.0–26.9	0.6 (0.5–0.8)	1.0 (0.6–1.5)	0.5 (0.2–1.1)	0.7 (0.4–1.3)	0.8 (0.4–1.8)
27-	0.5 (0.4–0.8)	0.7 (0.3–1.4)	0.6 (0.2–1.7)	0.4 (0.1–1.1)	0.2 (0.0–1.2)
P for Trend	<0.0001	0.56	0.27	0.04	0.02
P-Heterogeneity	0.49				
Weight Gain since 18					
Loss <2 kg	0.9 (0.6–1.1)	0.9 (0.5–1.7)	1.1 (0.4–2.9)	0.8 (0.4–1.7)	3.1 (0.8–11.9)
Stable	1.0 (ref)	1.0 (Ref)	1.0 (ref)	1.0 (ref)	1.0 (Ref)
Gain 2.1–5kg	0.9 (0.7–1.1)	0.9 (0.5–1.5)	0.6 (0.2–1.6)	1.4 (0.8–2.5)	3.1 (0.9–10.8)
Gain 5.1–10kg	0.9 (0.7–1.1)	1.1 (0.7–1.8)	1.4 (0.6–3.0)	1.1 (0.6–2.0)	2.6 (0.7–8.8)
Gain 10.1–20kg	0.9 (0.8–1.2)	1.3 (0.8–2.1)	1.3 (0.6–2.7)	1.2 (0.7–2.1)	3.5 (1.1–11.5)
Gain 20.1–25kg	0.9 (0.7–1.3)	1.2 (0.7–2.1)	1.3 (0.5–3.3)	1.0 (0.5–2.2)	2.4 (0.6–9.6)
Gain 25+ kg	1.1 (0.8–1.4)	1.5 (0.9–2.5)	1.2 (0.5–3.2)	1.5 (0.8–2.8)	2.5 (0.6–9.6)
P for trend	0.05	0.001	0.31	0.11	0.35

	Luminal A	Luminal B	HER2	Basal	Unclassified
P-Heterogeneity			0.05		
Parity					
Nulliparous	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)
1 child	0.8 (0.6–1.3)	2.2 (0.8–5.9)	1.1 (0.3–4.2)	0.6 (0.2–1.8)	1.4 (0.3–8.2)
2 children	0.7 (0.5–1.1)	1.2 (0.5–3.0)	1.0 (0.3–3.5)	0.9 (0.3–2.6)	1.1 (0.2–5.4)
3+ children	0.7 (0.5–1.0)	1.1 (0.5–2.6)	0.7 (0.2–2.3)	1.1 (0.4–2.9)	1.4 (0.3–6.0)
P for trend	0.01	0.01	0.12	0.04 (+)	0.65
P-Heterogeneity			0.60		
Age at first birth					
Per 1 year increase	1.018 (1.007–1.030)	0.994 (0.964–1.024)	1.024 (0.991–1.059)	1.002 (0.968–1.037)	1.006 (0.959–1.056)
P-Heterogeneity			0.36		
Age at Menopause					
Per year increase	1.039 (1.021–1.058)	1.067(1.024–1.111)	1.075 (1.006–1.150)	1.012 (0.972–1.053)	1.089 (1.006–1.179)
P-Heterogeneity			0.10		
Previous BBD					
No	1.0 (Ref)	1.0 (ref)	1.0 (ref)	1.0 (Ref)	1.0 (Ref)
Yes	1.4 (1.2–1.5)	1.7 (1.3–2.2)	1.2 (0.8–1.9)	1.5 (1.1–2.0)	1.5 (0.9–2.4)
P-Heterogeneity			0.69		
Menopausal status/PMH Use					
Premenopausal	1.2 (1.0–1.5)	1.5 (1.0–2.2)	1.1 (0.6–2.1)	1.1 (0.7–1.9)	1.7 (0.8–3.7)
Post Never Use	1.0(REF)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	(Ref)
Post Past Use	0.9 (0.8–1.1)	1.2 (0.8–1.7)	1.3 (0.7–2.4)	0.8 (0.5–1.4)	1.5 (0.7–3.2)
Post Current E only	1.4 (1.1–1.7)	1.0 (0.7–1.7)	1.1 (0.6–2.3)	1.4 (0.8–2.3)	2.0 (0.9–4.4)
Post Current E+P	1.5 (1.2–2.0)	1.3 (0.8–2.4)	0.3 (0.0–2.2)	1.8 (1.0–3.4)	2.9 (1.1–7.6)
P-Heterogeneity			0.08		
Alcohol Consumption					
None	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)
<5 g/week	1.1 (1.0–1.3)	0.8 (0.6–1.1)	1.1 (0.6–1.8)	0.8 (0.6–1.2)	0.5 (0.3–1.0)
5–10	1.0 (0.8–1.3)	1.1 (0.7–1.7)	1.3 (0.7–2.7)	0.7 (0.4–1.3)	0.9 (0.5–1.9)

	Luminal A	Luminal B	HER2	Basal	Unclassified
10–15	1.2(1.0–1.5)	0.8 (0.5–1.3)	1.4 (0.7–2.8)	0.7 (0.4–1.2)	0.6 (0.3–1.3)
15+	1.3 (1.0–1.6)	1.0 (0.7–1.5)	1.4 (0.7–2.8)	0.8 (0.5–1.4)	0.8 (0.4–1.6)
P-for trend	0.04	0.91	0.13	0.31	0.58
P-Heterogeneity	0.32				
Lactation					
Never	1.0 (Ref)	1.0 (Ref)	1.0 (ref)	1.0 (Ref)	1.0 (Ref)
0–3 months	0.8 (0.7–0.9)	1.0 (0.7–1.4)	0.8 (0.4–1.3)	0.8 (0.6–1.2)	1.0 (0.6–1.8)
4+ months	0.8 (0.7–1.0)	0.8 (0.6–1.1)	0.9 (0.6–1.5)	0.6 (0.4–0.9)	0.6 (0.4–1.1)
P-Heterogeneity	0.88				
Family History of breast cancer					
None	1.0 (Ref)	1.0 (ref)	1.0 (Ref)	1.0 (Ref)	1.0 (Ref)
1 1st degree relative	1.6 (1.4–1.9)	1.5 (1.0–2.1)	1.5 (0.8–2.6)	1.0 (0.6–1.6)	0.9 (0.4–1.9)
2+ 1st degree relat	2.3 (1.3–4.2)	--	2.5 (0.3–18.1)	2.9 (0.7–11.7)	--
P-Heterogeneity	0.01				

[1] Mutually adjusted for all variables in the table

Appendix 1

Sources and dilutions of primary antibodies used in this study.

Antibody	Clone	Manufacturer	Dilution
ER	1D5	Dako	1:200
PR	PgR 636	Dako	1:50
HER2	A0485 (rabbit polyclonal)	Dako	1:400
CK 5/6	D5/16B4	Dako	1:50
EGFR	2-18C9	Dako	pre-diluted (pharmDX kit)

ER=estrogen receptor; PR=progesterone receptor; HER2=human epidermal growth factor receptor 2; CK5/6=cytokeratin 5/6; EGFR=epidermal growth factor receptor