

Research article

Open Access

Comparison of molecular phenotypes of ductal carcinoma *in situ* and invasive breast cancer

Rulla M Tamimi^{1,2}, Heather J Baer^{1,2}, Jonathan Marotti³, Mark Galan³, Laurie Galaburda³, Yineng Fu³, Anne C Deitz⁴, James L Connolly³, Stuart J Schnitt³, Graham A Colditz⁵ and Laura C Collins³

¹Channing Laboratory, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, 181 Longwood Avenue, Boston, MA, 02115, USA

²Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA, 02115, USA

³Department of Pathology, Beth Israel Deaconess Medical Center and Harvard Medical School, 330 Brookline Avenue, Boston, MA 02115, USA

⁴Worldwide Epidemiology, GlaxoSmithKline, 1250 S Collegeville Rd, Collegeville, PA, 19426, USA

⁵Department of Surgery, Washington University School of Medicine, 660 S. Euclid Avenue, St. Louis, MO 63110, USA

Corresponding author: Rulla M Tamimi, Rulla.Tamimi@channing.harvard.edu

Received: 2 Apr 2008 Revisions requested: 2 Jun 2008 Revisions received: 11 Jul 2008 Accepted: 5 Aug 2008 Published: 5 Aug 2008

Breast Cancer Research 2008, **10**:R67 (doi:10.1186/bcr2128)

This article is online at: <http://breast-cancer-research.com/content/10/4/R67>

© 2008 Tamimi *et al.*; licensee BioMed Central Ltd.

This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Introduction At least four major categories of invasive breast cancer that are associated with different clinical outcomes have been identified by gene expression profiling: luminal A, luminal B, human epidermal growth factor receptor 2 (HER2) and basal-like. However, the prevalence of these phenotypes among cases of ductal carcinoma *in situ* (DCIS) has not been previously evaluated in detail. The purpose of this study was to compare the prevalence of these distinct molecular subtypes among cases of DCIS and invasive breast cancer.

Methods We constructed tissue microarrays (TMAs) from breast cancers that developed in 2897 women enrolled in the Nurses' Health Study (1976 to 1996). TMA slides were immunostained for oestrogen receptor (ER), progesterone receptor (PR), HER2, cytokeratin 5/6 (CK5/6) and epidermal growth factor receptor (EGFR). Using these immunostain results, cases were grouped into molecularly defined subtypes.

Results The prevalence of the distinct molecular phenotypes differed significantly between DCIS (n = 272) and invasive breast cancers (n = 2249). The luminal A phenotype was

significantly more frequent among invasive cancers (73.4%) than among DCIS lesions (62.5%) (p = 0.0002). In contrast, luminal B and HER2 molecular phenotypes were both more frequent among DCIS (13.2% and 13.6%, respectively) as compared with invasive tumours (5.2% and 5.7%, respectively) (p < 0.0001). The basal-like phenotype was more frequent among the invasive cancers (10.9%) than DCIS (7.7%), although this difference was not statistically significant (p = 0.15). High-grade DCIS and invasive tumours were more likely to be HER2 type and basal-like than low- or intermediate-grade lesions. Among invasive tumours, basal-like and HER2 type tumours were more likely to be more than 2 cm in size, high-grade and have nodal involvement compared with luminal A tumours.

Conclusion The major molecular phenotypes previously identified among invasive breast cancers were also identified among cases of DCIS. However, the prevalence of the luminal A, luminal B and HER2 phenotypes differed significantly between DCIS and invasive breast cancers.

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 subgroups that vary in their gene expression signatures and clinical course. The molecularly distinct breast cancer subgroups identified to date include luminal subtypes A and B (both of

BMI = body mass index; CBCS = Carolina Breast Cancer Study; CK5/6 = cytokeratin 5/6; CI = Confidence interval; DCIS = ductal carcinoma *in situ*; EGFR = epidermal growth factor receptor; ER = oestrogen receptor; HER2 = human epidermal growth factor receptor 2; OR = odds ratio; PMH = postmenopausal hormone; PR = progesterone receptor; TMA = tissue microarrays.

which are hormone receptor-positive), the HER2 subtype and a group known as basal-like cancers [1-4].

Immunohistochemical staining of paraffin sections using antibody panels has been shown to be a reliable surrogate for molecular classification of invasive breast cancers as categorised by gene expression profiling studies [4-8]. Antibodies against oestrogen 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 fact, this approach to molecular classification (that is, using immunostaining as a surrogate for expression profiling) is arguably the most practical approach to phenotyping large numbers of archived specimens for which fresh tissue is not available for expression array analysis [9-11]. In addition, application of this method to archival tissues from existing studies provides the opportunity to correlate tissue-marker data with long-term exposures and follow-up data.

While new information regarding the molecular heterogeneity of invasive cancers is rapidly emerging, far less is known about the spectrum of molecular phenotypes among cases of ductal carcinoma *in situ* (DCIS), the immediate precursor to invasive breast cancer. In particular, whether or not the same molecular subtypes identified among invasive breast cancers are also seen in DCIS has not been widely studied [12-15]. The purpose of the current study, therefore, was to determine in a large, well-characterised population of women with invasive breast cancer and DCIS which of the molecular phenotypes found in invasive cancers are also seen in DCIS and the frequency of the various molecular phenotypes in DCIS relative to that in invasive breast cancers.

Materials and methods

Study design and population

The Nurses' Health Study was initiated in 1976, with 121,700 US-registered nurses aged between 30 and 55 returning 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 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 more than 90% through 1996.

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 histological type, size, and presence or absence of metastases. To identify cases of cancer in non-respondents 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

In 1993, we began collecting archived formalin-fixed paraffin-embedded breast cancer blocks for participants with primary incident breast cancers over 20 years of follow-up (1976 to 1996). Cases who reported a prevalent cancer including breast cancer at baseline were excluded from collection. Of the 5610 patients with breast cancers that were eligible for block collection, we were unable to obtain any pathology material for 1858 cases. The primary reason was because they had been destroyed by the hospital (45%). The majority of hospitals archive tissue blocks for between five and 10 years, therefore we were more successful in obtaining more recent blocks. The year of diagnosis and age at diagnosis were highly correlated (Spearman correlation = 0.49; $p < 0.0001$) and the temporal effect on our collections is evident, not only in the differences in age at diagnosis, but also in the frequency of premenopausal breast cancers when comparing the women from whom we obtained specimens with those for whom we did not. However, these two groups of women were very similar with regards to a number of other breast cancer risk factors and tumour characteristics (Table 1). After taking into account age and year of diagnosis, the participants whose tumours were included in the tissue microarrays (TMAs) were very similar to those for whom we were unable to obtain tissue blocks.

We obtained pathology samples for 3752 participants. Of these, 390 specimens were only slides stained with haematoxylin and eosin (H&E) and 45 tissue blocks had to be returned to the lending hospital before construction of the TMAs and therefore could not be included. A single pathologist (YF) reviewed H&E sections from eligible cases to confirm the cancer diagnosis, classify the cancer according to histological type and grade (Nottingham), and circle the area from which the cores for the TMAs would be taken. Pathology review identified 420 tumour blocks as being unusable for TMA construction (eg, the block did not contain residual tumour or there was insufficient tumour in the block). TMAs were constructed in the Dana Farber Harvard Cancer Center Tissue Microarray Core Facility, Boston, MA.

Three cores 0.6 mm in diameter were obtained from each breast cancer sample and inserted into the recipient TMA blocks. In total, 23 TMA blocks were constructed from 3093 cancers and positive lymph nodes from 2897 participants. 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 ($n = 13$), and additional rare tumour types including malignant phyllodes tumours, neuroendocrine carcinoma and angiosarcoma ($n = 10$). *In situ* carcinomas with both ductal and lobular features were excluded because their ambiguous histological features precluded their definitive categorisation as either DCIS

Table 1

Comparison of breast cancer risk factors and breast tumour characteristics according to those who were eligible for the study, those for whom we received pathology specimens and those who were included in the tissue microarrays, Nurses' Health Study (1976 to 1996)

	Eligible for TMA (n = 5610)	Received pathology specimens (n = 3752)	Included in TMA (n = 2897)
<i>Means</i>			
Age at diagnosis, y	56.4	57.4	57.5
Age at menarche, y	12.4	12.4	12.4
Age at first birth, y	25.4	25.4	25.3
Age at menopause, y	46.9	47.1	47.1
BMI at age 18, kg/m ²	21.0	21.0	21.0
Parity	3.2	3.2	3.2
Alcohol g/week	6.6	6.1	6.2
<i>Frequencies, n (%)</i>			
Family history of BC	805 (14.4)	559 (14.9)	409 (14.1)
Nulliparous	430 (7.7)	281 (7.5)	211 (7.3)
Premenopausal	1345 (24.0)	782 (20.8)	600 (20.7)
Prior BBD	2662 (47.5)	1852(49.4)	1360 (47.0)
Current PMH user	1239 (33.8)	884 (34.2)	688 (34.6)
<i>Year of diagnosis</i>			
1976 to 1980	517 (9.4)	223 (6.1)	158 (5.6)
1980 to 1985	1066 (19.4)	565 (15.4)	431 (15.2)
1985 to 1990	1678 (30.6)	1152 (31.3)	899 (31.6)
1990 to 1996	2230 (40.6)	1739 (47.3)	1357 (47.7)
<i>Tumour characteristics^a</i>			
DCIS	654 (11.7)	443 (11.8)	270 (9.3)
Invasive	4956 (88.3)	3309 (88.2)	2627 (90.7)
<i>Tumour size</i>			
Less than 2 cm	3162 (65.6)	2197 (66.7)	1685 (64.9)
More than 2 cm	1656 (34.4)	1098 (33.3)	911 (35.1)
<i>Grade</i>			
Well differentiated	319 (15.2)	240 (16.2)	176 (14.7)
Moderately	860 (41.0)	630 (42.4)	515 (43.1)
Poorly	919 (43.8)	615 (41.4)	505 (42.2)
<i>Nodal involvement</i>			
No nodes	3931 (70.1)	2684 (71.5)	2006 (69.2)

Table 1 (Continued)

Comparison of breast cancer risk factors and breast tumour characteristics according to those who were eligible for the study, those for whom we received pathology specimens and those who were included in the tissue microarrays, Nurses' Health Study (1976 to 1996)

1 to 3 nodes	904 (16.1)	585 (15.6)	485 (16.7)
4 to 9 nodes	428 (7.6)	268 (7.1)	213 (7.4)
More than 10 nodes	228 (4.1)	140 (3.7)	129 (4.5)
Metastatic	119 (2.1)	75 (2.0)	64 (2.2)
Receptor status			
ER+	2738 (74.5)	1953 (75.2)	1572 (75.3)
PR+	2083 (61.2)	1518 (61.9)	1235 (62.3)

^aTumour characteristics abstracted from medical pathology reports. BC, breast cancer; BBD, benign breast disease; BMI, body mass index; DCIS, ductal carcinoma *in situ*; ER, oestrogen receptor; PMH, postmenopausal hormone; PR, progesterone receptor; TMA, tissue microarrays.

or lobular carcinoma *in situ*. For participants with both tumour and lymph node tissue, only the tumour tissue was evaluated in the current study. If the invasive case was present with DCIS, the tumour was considered invasive and scored as such.

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). These particular biomarkers were selected for analysis because they are commonly used as a surrogate to classify invasive breast cancers according to their molecular phenotypes [4-8]. Sources and dilutions of the primary antibodies used in this study are listed in Table 2. Immunostaining was conducted according to established protocols. Appropriate positive and negative controls were included in all staining runs.

Immunostained TMA slides were evaluated for ER and PR expression, HER2 protein overexpression and expression of CK5/6 and EGFR in each core. Tumour cells that showed

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 tumour cell staining. Of note, low ER-positive or PR-positive (1 to 10% of tumour cell nuclei staining) and ER-positive or PR-positive (>10% of tumour cell nuclei staining) were categorised as a single "positive" category for the purposes of this analysis. Tumour cells were considered positive for HER2 protein overexpression 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 tumour cells, even if focal. These latter criteria are similar to those previously used for scoring these markers in invasive basal-like cancers [4-6].

Classification of molecular phenotype

Immunostained TMA sections were reviewed under a microscope and visually scored for each individual tissue core as previously described. We classified a case as positive if there was staining in any of the three cores from that patient and

Table 2

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
CK5/6	D5/16B4	Dako	1:50
EGFR	2-18C9	Dako	pre-diluted (pharmDX kit)

CK5/6, cytokeratin 5/6; EGFR, epidermal growth factor receptor; ER, oestrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2.

negative if there was no immunostaining present. Cases that were ER-positive and/or PR-positive and HER2-negative were classified as luminal A cancers; cases that were ER-positive and/or PR-positive and HER2-positive were classified as luminal B cancers; 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 categorised as basal-like. Cases that lacked expression of all five markers were considered "unclassified" or "null".

Statistical analysis

Information on breast cancer risk factors was obtained from questionnaires completed biennially. Covariate data at the time of diagnosis were obtained from the questionnaire before the report of breast cancer diagnosis. Chi-squared tests were used to evaluate the independence of selected variables under the null hypothesis. All statistical tests were two-sided and $p < 0.05$ was considered statistically significant. Informed consent was obtained from each participant. This study was approved by the Committee on the Use of Human Subjects in Research at Brigham and Women's Hospital.

Results

This analysis population consisted of breast cancers that developed in women participating in the Nurses' Health Study after the baseline questionnaire (in 1976) through to the 1996 follow-up cycle and that could be classified into one of five molecular phenotypes ($n = 2521$; 272 DCIS and 2249 invasive). Based on immunostaining data from the five markers used, overall 1820 tumours were classified as luminal A; 152 were luminal B; 165 were HER2; 266 were basal-like; and 118 tumours were unclassifiable (ER-negative/PR-/HER2-negative/EGFR-negative/CK5/6-negative). An additional 297 cases were excluded because of staining that could not be evaluated or insufficient tumour tissue in the core of the sample.

In general, breast cancer risk factors in women with DCIS and invasive breast cancer were found to be similar (Table 3). Women with DCIS were slightly younger when they first gave birth (24.9 vs. 25.4, $p = 0.05$), were more likely to report a family history of breast cancer (19.0 vs 13.3%; $p = 0.02$) and a previous benign breast disease (57.0 vs 44.8%; $p = 0.0001$) compared with women with invasive breast cancer. As expected, women with DCIS were more likely to report their tumour being detected by screening mammography (81.6%) compared with women with invasive tumours (38.2%).

Compared with invasive tumours, DCIS lesions were more likely to be HER2-positive ($p < 0.0001$). The prevalence of the molecular phenotypes differed significantly between DCIS and invasive breast cancers (Table 4). Invasive tumours were significantly more likely than DCIS to be luminal A ($p = 0.0002$). In contrast, luminal B and HER2 molecular pheno-

Table 3

Age and age-standardised characteristics of breast cancer cases, Nurses' Health Study (1976 to 1996)

Risk factors	DCIS (<i>n</i> = 272)	Invasive (<i>n</i> = 2249)
<i>Means</i>		
Age at diagnosis, y	57.9	57.3
Age at menopause, y	45.2	45.5
Age at menarche, y	12.3	12.4
Body mass index at age 18, kg/m ²	20.9	21.0
Parity ^a	3.1	3.2
Age at first birth ^a , y	24.9	25.4
Alcohol, g/week	5.7	6.3
<i>Frequency, %</i>		
Family history of breast cancer	19.0	13.3
Nulliparous	6.9	6.9
Premenopausal at diagnosis	22.4	20.8
Postmenopausal at diagnosis	76.2	77.2
Prior benign breast disease	57.0	44.8
Current postmenopausal hormone use ^b	37.3	31.9
<i>Molecular Markers, %</i>		
ER positive	74.0	77.0
PR positive	59.9	64.4
HER2 positive	27.2	10.9

^aAmong parous women only. ^bAmong postmenopausal women only. DCIS, ductal carcinoma *in situ*; ER, oestrogen receptor; HER2, human epidermal growth factor receptor 2; PR, progesterone receptor.

types were more frequent among DCIS than among invasive tumours ($p < 0.0001$). The basal-like phenotype was more frequent among the invasive tumours than among DCIS, although this difference was not statistically significant ($p = 0.15$). However, when the analysis of invasive tumours was restricted to infiltrating ductal carcinomas ($n = 1550$), the frequency of the basal-like phenotype (14.4%) was significantly higher than among DCIS ($p = 0.005$). In an effort to determine if a single marker was responsible for distinguishing invasive tumours from DCIS, we examined ER, PR and HER2 status in multivariate analyses. HER2 was the only marker that significantly distinguished invasive tumours from DCIS ($p < 0.0001$). However, in distinguishing infiltrating ductal tumours from DCIS, ER was also a strong marginally significant predictor ($p = 0.08$).

Molecular phenotypes of DCIS and invasive tumours varied according to the grade of the lesions (Table 5). High-nuclear-

Table 4

Frequency of molecular phenotypes among DCIS and invasive breast cancers, Nurses' Health Study (1976 to 1996)

Immunophenotype	DCIS (n = 272) n (%)	All invasive (n = 2249) n (%)	Infiltrating ductal, NOS only (n = 1550) N (%)	p value ^a	p value ^b
Luminal A	170 (62.5)	1650 (73.4)	1053 (67.9)	0.0002	0.08
Luminal B	36 (13.2)	116 (5.2)	90 (5.8)	<0.0001	<0.0001
HER2+	37 (13.6)	128 (5.7)	107 (6.9)	<0.0001	<0.0001
Basal-like	21 (7.7)	245 (10.9)	223 (14.4)	0.15	0.005
Unclassified	8 (2.9)	110 (4.9)	77 (5.0)	0.15	0.14

^ap value from Chi-Squared test comparing DCIS to all invasive tumours. ^bp value from Chi-Squared test comparing DCIS to infiltrating ductal tumours. DCIS, ductal carcinoma *in situ*.

grade DCIS was significantly more likely than low-grade or intermediate-grade lesions to be HER2 type ($p < 0.0001$) and basal-like ($p = 0.009$). Similarly, high-grade invasive cancers were more likely to be HER2 type ($p < 0.0001$) and basal-like ($p < 0.0001$) than low-grade tumours. Results were similar when the analysis was limited to invasive ductal tumours only.

Among invasive tumours, molecular phenotypes were differentially associated with prognostic factors (Table 6). Compared with the luminal A subtype, HER2-type tumours were 2.6 (OR = 2.6, 95%CI 1.8 to 3.9) times as likely to be more than 2 cm in size, 3.6 times (OR = 3.6, 95%CI 2.1 to 6.3) as likely to be high grade and twice as likely to have nodal involvement (OR = 2.1, 95%CI 1.5 to 3.1). Similarly, basal-like tumours were more likely to be more than 2 cm in size (OR = 2.0, 95% 1.5 to 2.7), high grade (OR = 5.3, 95%CI 3.5 to 8.1) and have nodal involvement (OR = 1.5, 95%CI 1.3–1.7) when compared with luminal A tumours.

Discussion

In this large case series, we have shown that by using a panel of five immunostains, DCIS can be classified into five molecularly defined phenotypes that have been described for invasive breast carcinomas. Furthermore, we have shown that the prevalence of the molecularly defined phenotypes differed significantly between DCIS and invasive breast cancers. DCIS were more likely to be of the luminal B and HER2 phenotypes than invasive tumours. HER2 and basal-like phenotypes were common among both high-grade DCIS and high-grade invasive lesions.

These data provide evidence that DCIS and invasive tumours are both molecularly heterogeneous. Our finding of an increased prevalence of luminal B and HER2 molecular subtypes (ie, HER2-positive) in DCIS is consistent with earlier studies demonstrating a higher prevalence of HER2 protein overexpression and gene amplification among DCIS than invasive breast cancers [16-19]. The explanation for the higher prevalence of HER2 overexpression in DCIS compared with

Table 5

Frequency of molecular phenotypes among DCIS and invasive breast cancers according to tumour grade^a, Nurses' Health Study (1976 to 1996)

Tumour type	Luminal A	Luminal B	Her2 type	Basal	Unclassified
DCIS					
DCIS, low nuclear grade	26 (92.9)	1 (3.6)	0	1 (3.6)	0
DCIS, intermediate grade	109 (79.0)	15 (10.9)	6 (4.4)	6 (4.4)	2 (1.5)
DCIS, high nuclear grade	35 (33.0)	20 (18.9)	31 (29.3)	14 (13.2)	6 (5.7)
Invasive tumours					
Well-differentiated	134 (95.7)	2 (1.4)	0	2 (1.4)	2 (1.4)
Moderately differentiated	344 (79.3)	24 (5.5)	21 (4.8)	31 (7.1)	14 (3.2)
Poorly differentiated	252 (56.8)	20 (4.5)	43 (9.7)	99 (22.3)	30(6.8)

^aGrade of DCIS and invasive tumours as determined by centralised pathology review. DCIS, ductal carcinoma *in situ*.

Table 6**Odds ratios (95% confidence interval) of prognostic factors among invasive breast cancers according to molecular phenotype, Nurses' Health Study (1976 to 1996)**

Tumour type	Luminal A		Luminal B		Her2 type		Basal		Unclassified	
	<i>n</i>	OR	<i>n</i>	OR (95%CI)	<i>n</i>	OR (95%CI)	<i>n</i>	OR (95%CI)	<i>n</i>	OR (95%CI)
Tumour size^{a,b}										
≤ 2.0 cm	1025		54		48		107		57	
>2.0 cm	537	1.0 (Ref)	55	1.9 (1.3 to 2.8)	68	2.6 (1.8 to 3.9)	119	2.0 (1.5 to 2.7)	44	1.4 (0.9 to 2.1)
Tumour grade^a										
Low/Intermediate	478		26		21		33		16	
High	252	1.0 (Ref)	20	1.4 (0.8 to 2.6)	43	3.6 (2.1 to 6.3)	99	5.3 (3.5 to 8.1)	30	3.4 (1.8 to 6.4)
Nodal status^{a,b}										
No nodes	1103		72		61		145		70	
Any nodes	547	1.0 (Ref)	44	1.2 (0.8 to 1.7)	67	2.1 (1.5 to 3.1)	100	1.3 (1.0 to 1.7)	40	1.1 (0.7 to 1.6)
Menopausal status										
Postmenopausal	1289		76		97		187		83	
Premenopausal	334	1.0 (Ref)	35	1.8 (1.2 to 2.7)	27	1.1 (0.7 to 1.7)	51	1.1 (0.8 to 1.5)	25	1.2 (0.7 to 1.8)

^aAdjusted for age (five-year categories). ^bTumour characteristics abstracted from medical pathology reports. CI, confidence interval; OR, odds ratio.

invasive carcinoma remains unresolved and several possible explanations have been offered. These include: some HER2-positive DCIS lose HER2 expression when they progress to invasive cancers [19]; invasive carcinomas may arise more frequently from HER2-negative DCIS than from HER2-positive DCIS; there is a bias in mammographically screened populations toward the detection of DCIS lesions that are HER2-positive since these lesions are more frequently associated with comedo necrosis and, in turn, suspicious mammographic microcalcifications that prompt biopsy than non-high-grade DCIS lesions. In support of this explanation, previous studies have shown that screen-detected DCIS is more often due to the presence of linear branching and coarse granular calcifications, as well as DCIS of high nuclear grade and HER2 overexpression than interval DCIS [20,21].

Consistent with previous studies our results indicate that low-grade invasive cancers are more likely to have the luminal A phenotype, whereas high-grade invasive carcinomas are more likely to be HER2-type and basal-like [6]. In addition, both the HER2 and basal-like tumours were significantly more likely to be associated with poorer prognostic factors including larger tumour size, higher grade and nodal involvement. The Carolina Breast Cancer Study (CBCS) also examined molecular phenotypes of invasive tumours classified using the same immunohistochemical markers and categories as used in the current study [22]. Consistent with our study, Carey and col-

leagues [22] reported that HER2-type tumours ($n = 33$) were more likely to be high grade and have nodal involvement and that basal-like tumours ($n = 100$) were more likely to be high grade relative to luminal A tumours. We observed a strong association between basal-like tumours and nodal involvement, which was not found in the CBCS study. Although information on prognostic and tumour characteristics was not identical between our studies, both support the hypothesis that immunohistochemical classification of invasive tumours is associated with breast cancer prognosis.

We demonstrated a similar relationship between grade and molecular phenotype for DCIS lesions as was observed for invasive tumours. In particular, HER2-type (84%) and basal-like (67%) DCIS lesions were significantly more likely to be high grade than low-grade or intermediate-grade lesions. A similar distribution of molecular phenotypes of DCIS according to grade was reported in the CBCS, with 92% of HER2 DCIS and 84% of basal-like DCIS being high-grade [13]. These data suggest that similar to invasive tumours, molecular classification of DCIS lesions may be important in identifying more aggressive lesions. Additional support for this hypothesis comes from a recent case-control study of women with DCIS, in which 32 women went on to develop subsequent cancer (cases) and 38 did not develop subsequent disease (controls) [23]. This study found that eight of eight DCIS

cases expressing high levels of both markers of basal-like subtype, p16 and Ki67, developed a subsequent tumour [23].

The majority of epidemiological studies have reported very similar risk factors for DCIS and invasive breast cancer [16,24]. However, in these epidemiological studies, DCIS and invasive tumours are often considered together as one endpoint or are considered as two distinct groups without further classification according to receptor status. Studies considering invasive breast cancer according to receptor status have identified risk factors specific for certain phenotypes and not others [25,26]. It is likely, however, that the impact of several established risk factors for DCIS and invasive tumours will differ according to molecularly defined phenotypes. Further identification of risk factors for the different molecular phenotype could be of value in risk assessment and prevention.

A limitation of the current study is that we were unable to obtain tissue blocks from all breast cancers arising in this cohort. Our success in doing so was highly correlated with time between diagnosis and initiation of our tissue block collection. Many hospitals destroy paraffin-embedded tissue blocks after five to 10 years. After taking into account the effect of age and year of diagnosis, the women for whom we were able to obtain tumour specimens were very similar to those for whom we were unable to obtain specimens. The primary differences were due to the temporal consequences of the timing of tumour block collection. In addition, the frequency of receptor status positivity among invasive tumours was very similar to other populations suggesting that samples included in this study are representative of the overall US population.

We used immunohistochemical markers as a surrogate to classify breast cancers into the molecular phenotypes defined by expression profiles. While the antibody panel we used in this study has been shown to be a reliable proxy for classification of invasive breast cancers categorised by gene expression [4-8], the correlation is not perfect and there will be some misclassification of these phenotypes. The molecular phenotype categories as defined by the immunohistochemical markers have been shown to be associated with prognostic markers and survival consistent with what has been seen with classification based on RNA expression assays, suggesting that both methods are capturing distinct subgroups [5,22]. Misclassification of phenotypes may underestimate true differences between the subtypes.

Because lesions had to be large enough to yield multiple cores, it is possible that we may have biased our sample towards larger DCIS lesions. In addition, the frequency of DCIS lesions that we were unable to evaluate in the TMA for ER and HER2 staining was 22% and 18%, respectively. These are higher than what was observed for invasive lesions (6% and 7%, respectively), suggesting that the ability to assess DCIS lesions with three replicate cores is less than for

invasive tumours. However, the frequency of molecular phenotypes observed in our study is very similar to those reported in the CBCS among DCIS cases ($n = 245$) [13], suggesting that the DCIS cases included in our study are representative of DCIS cases in the US population.

Conclusion

The introduction of widespread mammographic screening has resulted in a dramatic increase in the diagnosis of DCIS [27]. A better understanding of the progression from DCIS to invasive cancer to avoid over-treatment is an important public health concern. The HER2 and basal-like molecular phenotypes are more common among high-grade DCIS lesions. Because these phenotypes are also associated with poor prognosis among women with invasive tumours, molecular characterisation of in situ tumours may help predict which lesions will progress to invasive tumours. More aggressive treatments could be targeted to the subset of women with DCIS lesions at highest risk of progressing to invasive cancer.

Competing interests

ACD is a full-time employee of GlaxoSmithKline and has received shares of company stock. The Nurses' Health Study investigators (GAC, RMT) received funding from GlaxoSmithKline to conduct biomarker studies on tissue microarrays. However, study investigators acted independently in generating research hypotheses, results and interpreting the data and deciding to submit the paper for publication.

Authors' contributions

RMT was responsible for data analyses, manuscript preparation and editing. RMT, HJB, JLC, SJS, GAC and LCC made substantial contributions to the study design and to the interpretation of data. LCC, JM, MG, LG and YF were involved in central pathology review, scoring of stains and contributed substantially to manuscript editing. ACD contributed to manuscript editing. All authors read and approved the final manuscript.

Acknowledgements

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 are grateful to the participants of the Nurses' Health Study for their outstanding dedication and commitment to the study. 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, Rijn M van de, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC, Lashkari D, Shalon D, Brown PO, Botstein D: **Distinctive gene expression**

- patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci USA* 1999, **96**:9212-9217.
2. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, Rijn M van de, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein Lønning P, Børresen-Dale AL: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci USA* 2001, **98**:10869-10874.
 3. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, Lønning PE, Brown PO, Børresen-Dale AL, Botstein D: **Repeated observation of breast tumor subtypes in independent gene expression data sets.** *Proc Natl Acad Sci USA* 2003, **100**:8418-8423.
 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**:7350-7360.
 5. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, Hernandez-Boussard T, Livasy C, Cowan D, Dressler L, Aksent LA, Ragaz J, Gown AM, Gilks CB, van de Rijn M, Perou CM: **Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma.** *Clin Cancer Res* 2004, **10**:5367-5374.
 6. 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**:340-350.
 7. 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**:661-671.
 8. Rijn M van de, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Kochli OR, Mross F, Dieterich H, Seitz R, Ross D, Botstein D, Brown P: **Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome.** *Am J Pathol* 2002, **161**:1991-1996.
 9. Cheang MC, Voduc D, Bajdik C, Leung S, McKinney S, Chia SK, Perou CM, Nielsen TO: **Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype.** *Clin Cancer Res* 2008, **14**:1368-1376.
 10. Yang XR, Sherman ME, Rimm DL, Lissowska J, Brinton LA, Peplonska B, Hewitt SM, Anderson WF, Szeszenia-Dabrowska N, Bardin-Mikolajczak A, Zatonski W, Cartun R, Mandich D, Rymkiewicz G, Ligaj M, Lukaszek S, Kordek R, Garcia-Closas M: **Differences in risk factors for breast cancer molecular subtypes in a population-based study.** *Cancer Epidemiol Biomarkers Prev* 2007, **16**:439-443.
 11. Millikan RC, Newman B, Tse CK, Moorman PG, Conway K, Smith LV, Labbok MH, Geradts J, Bensen JT, Jackson S, Nyante S, Livasy C, Carey L, Earp HS, Perou CM: **Epidemiology of basal-like breast cancer.** *Breast Cancer Res Treat* 2008, **109**:123-139.
 12. Bryan BB, Schnitt SJ, Collins LC: **Ductal carcinoma in situ with basal-like phenotype: a possible precursor to invasive basal-like breast cancer.** *Mod Pathol* 2006, **19**:617-621.
 13. Livasy CA, Perou CM, Karaca G, Cowan DW, Maia D, Jackson S, Tse CK, Nyante S, Millikan RC: **Identification of a basal-like subtype of breast ductal carcinoma in situ.** *Hum Pathol* 2007, **38**:197-204.
 14. Steinman S, Wang J, Bourne P, Yang Q, Tang P: **Expression of cytokeratin markers, ER-alpha, PR, HER-2/neu, and EGFR in pure ductal carcinoma in situ (DCIS) and DCIS with co-existing invasive ductal carcinoma (IDC) of the breast.** *Ann Clin Lab Sci* 2007, **37**:127-134.
 15. Paredes J, Lopes N, Milanezi F, Schmitt FC: **P-cadherin and cytokeratin 5: useful adjunct markers to distinguish basal-like ductal carcinomas in situ.** *Virchows Arch* 2007, **450**:73-80.
 16. Claus EB, Stowe M, Carter D: **Breast carcinoma in situ: risk factors and screening patterns.** *J Natl Cancer Inst* 2001, **93**:1811-1817.
 17. Zafrani B, Leroyer A, Fourquet A, Laurent M, Trophilme D, Validire P, Sastre-Garau X: **Mammographically-detected ductal in situ carcinoma of the breast analyzed with a new classification. A study of 127 cases: correlation with estrogen and progesterone receptors, p53 and c-erbB-2 proteins, and proliferative activity.** *Semin Diagn Pathol* 1994, **11**:208-214.
 18. Evans AJ, Pinder SE, Ellis IO, Sibbering DM, Elston CW, Poller DN, Wilson AR: **Correlations between the mammographic features of ductal carcinoma in situ (DCIS) and C-erbB-2 onco-gene expression.** Nottingham Breast Team. *Clin Radiol* 1994, **49**:559-562.
 19. Allred DC, Clark GM, Molina R, Tandon AK, Schnitt SJ, Gilchrist KW, Osborne CK, Tormey DC, McGuire WL: **Overexpression of HER-2/neu and its relationship with other prognostic factors change during the progression of in situ to invasive breast cancer.** *Hum Pathol* 1992, **23**:974-979.
 20. de Roos MA, Vegt B van der, de Vries J, Wesseling J, de Bock GH: **Pathological and biological differences between screen-detected and interval ductal carcinoma in situ of the breast.** *Ann Surg Oncol* 2007, **14**:2097-2104.
 21. Evans AJ, Pinder S, Ellis IO, Sibbering M, Elston CW, Poller DN, Wilson R: **Screening-detected and symptomatic ductal carcinoma in situ: mammographic features with pathologic correlation.** *Radiology* 1994, **191**:237-240.
 22. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC: **Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study.** *JAMA* 2006, **295**:2492-2502.
 23. Gauthier ML, Berman HK, Miller C, Kozakeiwicz K, Chew K, Moore D, Rabban J, Chen YY, Kerlikowske K, Tlsty TD: **Abrogated response to cellular stress identifies DCIS associated with subsequent tumor events and defines basal-like breast tumors.** *Cancer Cell* 2007, **12**:479-491.
 24. Longnecker MP, Bernstein L, Paganini-Hill A, Enger SM, Ross RK: **Risk factors for in situ breast cancer.** *Cancer Epidemiol Biomarkers Prev* 1996, **5**:961-965.
 25. 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**:218-228.
 26. Powles TJ, Ashley S, Tidy A, Smith IE, Dowsett M: **Twenty-year follow-up of the Royal Marsden randomized, double-blinded tamoxifen breast cancer prevention trial.** *J Natl Cancer Inst* 2007, **99**:283-290.
 27. Ernster VL, Barclay J, Kerlikowske K, Grady D, Henderson C: **Incidence of and treatment for ductal carcinoma in situ of the breast.** *JAMA* 1996, **275**:913-918.