Heterogeneity of luminal breast cancer characterised by immunohistochemical expression of basal markers

Hyuna Sung*,1, Montserrat Garcia-Closas2, Jenny Chang-Claude3, Fiona M Blows4, H Raza Ali5, Jonine Figueroa1, Heli Nevanlinna6, Rainer Fagerholm6, Paivi Heikkila7, Carl Blomqvist8, Graham G Giles9,10, Roger L Milne9,10, Melissa C McLean12, Arto Mannermaa13,14, Veli-Matti Kosma13,14, Vesa Kataja15, Reijo Sironen13,14, Fergus J Couch16, Janet E Olson17, Emily Hallberg17, Curtis Olswold17, Angela Cox18, Simon S Cross19, Peter Kraft20, Rulla M Tamimi21,22, A Heather Eliassen21,22, Marjanka K Schmidt23, Manjeet K Bolla24, Qin Wang24, Douglas Easton24,25, William J Howat26, Penny Coulson27, Paul DP Pharoah24,25, Mark E Sherman1 and Xiaohong R Yang1

1Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, 9609 Medical Center Drive, Bethesda, 20850 MD, USA; 2Division of Breast Cancer Research, Breakthrough Breast Cancer Research Centre, The Institute of Cancer Research, 15 Cotswold Rd, Sutton, Surrey, SM2 5NG London, UK; 3Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany; 4Department of Oncology, University of Cambridge, Worts Causeway, CB1 8RN Cambridge, UK; 5Cancer Research UK Cambridge Institute, University of Cambridge, Robinson Way, CB2 0RE Cambridge, UK; 6Department of Obstetrics and Gynecology, Helsinki University Hospital and University of Helsinki, PO Box 700, 00029 HUS Helsinki, Finland; 7Department of Pathology, Helsinki University Hospital and University of Helsinki, PO Box 400, 00029 Helsinki, Finland; 8Department of Oncology, Helsinki University Hospital and University of Helsinki, PO Box 400, 00029 Helsinki, Finland; 9Cancer Epidemiology Centre, Cancer Council Victoria, 615 St Kilda Rd, Melbourne, 3004 Victoria, Australia; 10Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Parkville, 3010 Victoria, Australia; 11Genetic Epidemiology Laboratory, Department of Pathology, The University of Melbourne, Melbourne, 3010 Victoria, Australia; 12Anatomical Pathology, The Alfred Hospital, Commercial Rd, Prahran, 3181, 3053 Victoria, Australia; 13School of Medicine, Institute of Clinical Medicine, Pathology and Forensic Medicine, Cancer Center of Eastern Finland, University of Eastern Finland, Yliopistokatu 1, PO Box 1627, 70211 Kuopio, Finland; 14Department of Clinical Pathology, Imaging Center, Kuopio University Hospital, PO Box 100, 70029 KYS Kuopio, Finland; 15Jyväskylä Central Hospital, Central Finland Health Care District, Adm Bldg 6/2, Keskusraatteenkatu 19, 40620 Jyväskylä, Finland; 16Department of Laboratory Medicine and Pathology, Mayo Clinic, Stabile 2-42, 200 First Street SW, Rochester, 55905 MN, USA; 17Department of Health Sciences Research, Mayo Clinic, 200 First St SW, Rochester, 55905 MN, USA; 18Sheffield Cancer Research, Department of Oncology, University of Sheffield, Beech Hill Road, S10 2RX Sheffield, UK; 19Department of Neuroscience, University of Sheffield, Beech Hill Road, S10 2RX Sheffield, UK; 20Program in Genomic Epidemiology and Statistical Genetics, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, 02115 MA, USA; 21Department of Medicine, Channing Division of Network Medicine, Harvard Medical School and Brigham and Women’s Hospital, 181 Longwood Avenue, Boston, 02115 MA, USA; 22Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Ave, Boston, 02115 MA, USA; 23Netherlands Cancer Institute, Antoni van Leeuwenhoek Hospital, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands; 24Department of Public Health and Primary Care, Centre for Cancer Genetic Epidemiology, University of Cambridge, CB1 8RN Cambridge, UK; 25Department of Oncology, Centre for Cancer Genetic Epidemiology, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, CB1 8RN Cambridge, UK; 26Cancer Research UK Cambridge Institute, Li Ka Shing Centre, CB2 0RE Cambridge, UK and 27Division of Genetics and Epidemiology, The Institute of Cancer Research, 15 Cotswold Rd, Sutton, Surrey, SM2 5NG London, UK

*Correspondence: Dr H Sung; E-mail: hyuna.sung@nih.gov

Received 13 August 2015; revised 23 October 2015; accepted 21 November 2015; published online 17 December 2015

© 2016 Cancer Research UK. All rights reserved 0007 – 0920/16
Background: Luminal A breast cancer defined as hormone receptor positive and human epidermal growth factor receptor 2 (HER2) negative is known to be heterogeneous. Previous study showed that luminal A tumours with the expression of basal markers (cytokeratin (CK) 5 or CK5/6) or epidermal growth factor receptor (EGFR) were associated with poorer prognosis compared with those that stained negative for basal markers. Prompted by this study, we assessed whether tumour characteristics and risk factors differed by basal marker status within luminal A tumours.

Methods: We pooled 5040 luminal A cases defined by immunohistochemistry (4490 basal-negative (CK5 (or CK5/6) – and EGFR – ) and 550 basal-positive ((CK5 (or CK5/6 + ) or EGFR + )) from eight studies participating in the Breast Cancer Association Consortium. Case–control comparison was performed using unconditional logistic regression.

Results: Tumour characteristics and risk factors did not vary significantly by the expression of basal markers, although results suggested that basal-positive luminal A tumours tended to be smaller and node negative, and were more common in women with a positive family history and lower body mass index.

Conclusions: Most established breast cancer risk factors were similar in basal-positive and basal-negative luminal A tumours. The non-significant but suggestive differences in tumour features and family history warrant further investigations.

BREITHE JOURNAL OF CANCER

Heterogeneity of luminal A breast cancer

BREITHE JOURNAL OF CANCER

Breast cancer can be classified into several molecular subtypes based on gene expression profiling analyses (Perou et al., 2000; Sorlie et al., 2001), which can be approximated with the use of key immunohistochemical (IHC) markers, including estrogen receptor (ER), progesterone receptor (PR), human epidermal growth factor receptor 2 (HER2), and basal markers such as epidermal growth factor receptor (EGFR), cytokeratin 5 (CK5) or cytokeratin 5/6 (CK5/6). In general, well-known breast cancer hormonal and lifestyle risk factors, such as early age at menarche, late age at first birth, nulliparity, prolonged interval between menarche and age at first birth, and postmenopausal obesity showed stronger associations with ER-positive (luminal) subtypes (Yang et al., 2011; Anderson et al., 2014). In contrast, these factors showed either a lack of association or associations in the opposite direction for ER-negative (non-luminal) tumours. For example, parity and premenopausal obesity were protective for luminal cancers but associated with increased risk for non-luminal tumours, particularly triple-negative breast cancer (TNBC: ER−/PR−/HER2−; Millikan et al., 2008; Phipps et al., 2011). We have previously shown that risk factor associations differed most strikingly between luminal A (ER+ or PR+/HER2−) and core-basal phenotype (CBP; TNBC expressing (CK5 or CK5/6) or EGFR), suggesting that these two subtypes may develop from etiologically different pathways (Yang et al., 2011).

Experimental and clinical studies suggest more complex layers of heterogeneity within major breast cancer subtypes (Perou et al., 2000; Sorlie et al., 2003; Colleoni et al., 2012; Ali et al., 2014). In particular, luminal cancers demonstrated substantial variability in molecular characteristics (Cancer Genome Atlas Network, 2012) and clinical behaviour, including responsiveness to endocrine treatment (Ciriello et al., 2013; Howell, 2013; Ignatiadis and Sorlie, 2013). In line with this, in a recent large pooled analysis including >10,000 invasive breast cancer cases, Blows et al., 2010 showed that luminal A tumours expressing basal markers (CK5 or CK5/6 or EGFR, luminal A basal-positive) had worse prognosis than luminal A tumours that were negative for basal markers (luminal A basal-negative). However, to our knowledge, there have been no reports on etiological heterogeneity within luminal A tumours so far.

To assess whether luminal A basal-positive tumours (ER+ or PR+/HER2−/basal markers + ) represent a distinct disease entity from an etiologic perspective, we performed a large pooled analysis for 5040 luminal A breast cancer cases contributed by eight studies participating in the Breast Cancer Association Consortium (BCAC), with risk factor information and expression status for ER, PR, HER2, and basal markers. The goal of this study was to examine whether tumour characteristics and risk factors of luminal A basal-positive tumours are different from those of luminal A basal-negative tumours (ER+ or PR+/HER2−/basal markers − ).

MATERIALS AND METHODS

Study participants. Among studies participating in the BCAC (Yang et al., 2011), eight studies that had IHC data on ER (and/or PR), HER2, and basal markers (CK5 (or CK5/6) and/or EGFR) as well as breast cancer risk factor information were eligible for inclusion. Study details are summarised in Supplementary Table 1. These include four population-based studies (Kuopio Breast Cancer Project (KBCP), Melbourne Collaborative Cohort Study ( MCCS), Nurses’ Health Study (NHS), and NCI’s Polish Breast Cancer Study (PBCS)) and four hospital-based case-control studies or studies of mixed design (Helsinki Breast Cancer Study (HEBCS), Mayo Clinic Breast Cancer Study (MCBCS), Sheffield Breast Cancer Study (SBCS), and Study of Epidemiology and Risk factors in Cancer Heredity (SEARCH)). As the goal of our analysis was to determine whether tumour characteristics and risk factors differed by basal marker status within luminal tumours, we restricted the analysis to 5040 luminal A cases (ER+ or PR+/HER2−) that were known to express or not to express basal markers (CK5 (or CK5/6) or EGFR) (Table 1). Study participants were recruited under protocols approved by the institutional review board at each institution and all subjects provided written informed consent.

Tumour marker assessment and subtype classification. ER, PR, and HER2 status were primarily extracted from medical records. Accordingly, the source of tumour marker data and definition of positivity for each marker varied across studies (Supplementary Table 2). Among 5040 luminal A cases defined based on medical records for ER and PR, centralised quantitative scores for ER or PR status obtained through automated imaging analysis of tissue microarrays were available for 3702 participants (Supplementary Table 3). More than 99% of luminal A cases (n = 3670/3702) had tumours with ≥1% cells and 97% (n = 3592/3702) with ≥10% tumour cells stained positive for either ER or PR, respectively. Given the high concordance of clinical data and centralised measurements for ER and PR, we used clinical data for these markers in the main analyses because they were available for more cases. Data for CK5 (or CK5/6) and EGFR status were obtained from centralised visual scoring of tissue microarray slides by pathologists. Expression was determined to be positive if >10% tumour cells were stained. When the proportion of positive cells
Table 1. Description of participating studies and distribution of tumour subtypes by study

| Study (abbreviation)      | Study design                      | Total   | Luminal A basal-negative | Luminal A basal-positive | HR+/HER2+ | HR−/HER2+ | CBP | 5-NP | Missing |
|---------------------------|-----------------------------------|---------|--------------------------|--------------------------|-----------|-----------|-----|------|---------|
| Study of Epidemiology and Risk factors in Cancer Heredity (SEARCH) | Case-control study                | 3096    | 1791 (67.7)              | 196 (7.4)                | 198 (7.5) | 111 (4.2) | 176 (6.7) | 173 (5.9) | 451      |
| NCI Polish Breast Cancer Study (PBCS) | Case-control study                | 1342    | 768 (58.9)               | 57 (4.4)                 | 115 (8.8) | 119 (9.1) | 116 (8.9) | 128 (9.8) | 39       |
| Helsinki Breast Cancer Study (HEBCS) | Hospital-based case-control study & additional familial cases | 1067    | 625 (63.2)               | 87 (8.8)                 | 95 (9.6)  | 53 (5.4)  | 77 (7.8)  | 52 (5.3)  | 78       |
| Melbourne Collaborative Cohort Study (MCCS) | Prospective cohort study           | 796     | 539 (67.8)               | 23 (2.9)                 | 86 (10.8) | 40 (5.0)  | 75 (9.4)  | 32 (4.0)  | 1        |
| Kuopio Breast Cancer Project (KBCP) | Prospective clinical cohort        | 427     | 232 (62.7)               | 42 (11.4)                | 21 (5.7)  | 26 (7.0)  | 19 (5.1)  | 30 (8.1)  | 57       |
| Mayo Clinic Breast Cancer Study (MCBCS) | Hospital-based case-control study  | 422     | 189 (58.0)               | 53 (16.3)                | 27 (8.3)  | 10 (3.1)  | 42 (12.9) | 5 (1.5)   | 96       |
| Sheffield Breast Cancer Study (SSBCS) | Hospital-based case-control study  | 358     | 176 (53.8)               | 61 (18.7)                | 18 (5.5)  | 11 (3.4)  | 41 (12.5) | 20 (6.1)  | 31       |
| Nurses’ Health Study (NHS) | Prospective cohort study: nested case-control study | 349     | 170 (50.0)               | 31 (9.1)                 | 72 (21.2) | 14 (4.1)  | 32 (9.4)  | 21 (6.2)  | 9        |
| Total                     |                                    | 7857    | 4490 (63.3)              | 550 (7.8)                | 632 (8.9) | 384 (5.4) | 578 (8.2) | 461 (6.5) | 742      |

Luminal A basal-negative = ER+ or PR+/HER2− / (CK5 or CK5/6) / EGFR−; Luminal A basal-positive = ER+ or PR+/HER2− / (CK5 or CK5/6) / EGFR− and basal-positive (ER+ or PR+/HER2− / (CK5 or CK5/6) / EGFR+).

Breast cancer risk factors. The collection of information on tumour characteristics and risk factors for BCAC studies has been previously described (Yang et al, 2011). Briefly, each study collected information on one or more of the following factors: family history of breast cancer in first-degree relatives, age at menarche, age at menopause, age at first full-term pregnancy, parity (never/ever), number of children, breast feeding (never/ever), and body mass index (BMI) at baseline (MCCS and KBCP) or at the time of diagnosis (all others). NHS was not included in risk factor analysis owing to the lack of data.

Statistical analyses. We compared the distribution of tumour characteristics and risk factors between luminal A basal-negative and basal-positive subtypes using unconditional logistic regression with luminal A basal-negative subtype as the reference group. Tumour characteristics included histology (ductal, lobular, other), grade (well, moderately, poorly differentiated), size (<2 cm, >2 cm), and axillary node status (negative/positive). Breast cancer risk factors included family history of breast cancer among first-degree relatives (present, absent), age at menarche (<12, 13–14, >14 years), parity (parous, nulliparous), and BMI (<25, 25–30, ≥30 kg m<sup>2</sup> or per 5 unit of increase); and in analyses restricted to parous women included age at first full-term birth (<20, 20–24, 25–29, ≥30 years), number of full-term pregnancies (1, 2, ≥3), and breast feeding (ever, never). Multivariable models were used in all analyses to control for age (10-year frequency), study, other tumour characteristics and risk factors. Given that risk associated with BMI is known to vary by menopausal status, we stratified the BMI analysis by menopausal status. We used age groups (<50 and ≥50 years) as a surrogate for menopausal status to maximise power. A sensitivity analysis using known menopausal status yielded similar results. Between-study heterogeneity was assessed with $I^2$ statistics using study-specific odds ratio (ORs) and 95% confidence intervals (CIs). Analyses were conducted using SAS (version 9.3; SAS Institute, Cary, NC, USA) or Stata/SE (version 11.2; StataCorp LP, College Station, TX, USA).

RESULTS

Among all 7857 invasive breast cancer cases in the 8 studies, 63.3% (n = 4490) and 7.8% (n = 550) were classified as luminal A basal-negative and luminal A basal-positive subtype, respectively (Table 1). Mean age at diagnosis was not significantly different between the two subtypes, although women with luminal A basal-positive tumours were diagnosed less frequently after 60 years compared with the women with luminal A basal-negative tumours (Table 2). Compared with the luminal A basal-negative tumours, basal-positive tumours were more likely to be smaller (OR > 2 cm vs ≤ 2 cm = 0.83; 95% CI = 0.67–1.03; P = 0.09; $I^2 = 0$%) and negative for axillary nodes (OR = 0.83; 95% CI = 0.67–1.02; P = 0.08; $I^2 = 0$%), however, the differences were not statistically significant. The association with tumour grade did not follow a logical trend, with luminal A basal-positive tumours showing a lower frequency of moderately differentiated tumours (OR = 0.75; 95% CI = 0.60–0.94; P = 0.01), but a higher frequency of poorly differentiated tumours (OR = 1.13; 95% CI = 0.85–1.50; P = 0.42) compared with luminal A basal-negative tumours (Table 2). Study heterogeneity was not significant in the former ($I^2 = 10.2%; P = 0.35$) but was significant in the latter association ($I^2 = 63.2%; P = 0.01$).

Compared with basal-negative cases, cases with luminal A basal-positive tumours were more likely to have a positive family history (OR = 1.27; 95% CI = 0.99–1.63; P = 0.06; $I^2 = 1.1%$; Table 3) particularly among younger (<50 years) women (OR = 1.81; 95% CI = 1.16–2.82; P = 0.009). In addition, basal-positive cases tended to have lower BMI (OR<sub>per 5 units</sub> = 0.90; 95% CI = 0.81–1.01, $P = 0.07$; $I^2 = 0.0%$) especially among older (≥50 years) women.
In a previous BCAC analysis (Blows et al, 2010), we showed that all-cause mortality among cases with luminal A basal-negative tumours was slightly but significantly higher than that of cases with luminal A basal-negative tumours, and the difference was persistent up to 15 years after diagnosis. Similar but non-significant difference in survival by basal marker expression (adjusted hazard ratio = 1.20
_\text{basal-positive} vs _\text{basal-negative}; 95\% CI = 0.69–2.08, P = 0.51) was observed in our study when we analysed a subset of cases (1245 luminal A cases; 1124 basal-negative cases, and 121 basal-positive cases) with the follow-up data available. Interestingly, luminal A basal-positive tumours were not associated with more aggressive features, rather, they tended to be less aggressive (smaller, lower grade, and node negative) compared with basal-negative tumours.

The apparent discrepancy between less aggressive tumour features and poorer prognosis in basal-positive cases might be explained by different responses to endocrine therapy among cases with luminal tumours. Previous studies using luminal tumour xenografts identified a subpopulation of ER-PR-CK5
_\text{+} cells that were resistant to endocrine therapies (Horwitz et al, 2008); when ER
_\text{+} tumours with ER-PR-CK5
_\text{+} cells were treated with 17\beta-
estradiol plus anti-estrogens tamoxifen or fulvestrant, the number of CK5
_\text{+} cells in post compared with pre-treatment tumours coupled with decreased expression of ER and increased expression of CK5 (Kabos et al, 2011). Studies with detailed pathology data incorporating cellular subpopulation, as well as treatment regimens with long-term follow-up are needed to definitively address this question.

Known breast cancer risk factors did not appear to vary significantly by basal marker expression within luminal A tumours, although we observed weak associations between basal-positive tumours and higher frequency of positive family history especially among younger women and lower prevalence of obesity. The higher frequency of slim women with luminal A basal-positive tumours might be also related to smaller tumour size of luminal A basal-positive tumours as we observed a significant correlation between tumour size and BMI among our study subjects. Indeed, when we adjusted for BMI, the association between luminal A basal-positive subtype and smaller tumour size was attenuated (OR = 0.87; 95\% CI = 0.69–1.09; P = 0.22). This finding is consistent with previous reports that obese breast cancer patients had larger tumours and higher rates of lymph node metastases (Ewertz et al, 2011; Haakinson et al, 2012).

Our study has limitations. Although it is one of the largest consortium studies with breast tumour subtype information and risk factor data collected, statistical power was limited to assess risk factors in uncommon subtypes especially when controlling for potential confounding factors such as breastfeeding, menopausal

### Table 2. Comparison of tumour characteristics between luminal A basal-negative and basal-positive subtypes

| Tumour characteristic | Basal-negative (n = 4490) | Basal-positive (n = 550) |
|-----------------------|--------------------------|-------------------------|
|                       | n (%)                    | n (%)                   | OR (95% CI) |
| Age, years (mean, s.d.) | 56.6 (10.8)              | 56.3 (11.4)             | 5.6         |
| <40                   | 207 (4.6)                | 31 (5.6)                | 1 (reference) |
| 40–49                 | 1024 (22.8)              | 127 (23.1)              | 0.90 (0.59–1.38) |
| 50–59                 | 1504 (33.5)              | 187 (34.0)              | 0.91 (0.60–1.38) |
| ≥60                   | 1755 (39.1)              | 205 (37.3)              | 0.79 (0.52–1.21) |
| P_{trend}             | —                        | —                       | 0.91 (0.82–1.02) |
| Tumour histology      |                          |                         |             |
| Ductal                | 2994 (72.6)              | 376 (74.6)              | 1 (reference) |
| Lobular               | 807 (19.6)               | 94 (18.7)               | 1.08 (0.84–1.39) |
| Other                 | 323 (7.8)                | 34 (6.8)                | 0.89 (0.61–1.30) |
| Missing               | 366 (–)                  | 46 (–)                  | —            |
| Tumour grade          |                          |                         |             |
| Well differentiated    | 1113 (27.1)              | 160 (32.0)              | 1 (reference) |
| Moderately differentiated | 2318 (56.4)          | 240 (48.0)              | 0.75 (0.60–0.94) |
| Poorly differentiated  | 676 (16.5)               | 100 (20.0)              | 1.13 (0.85–1.50) |
| Missing               | 383 (–)                  | 50 (–)                  | —            |
| P_{trend}             | —                        | —                       | —            |
| Tumour size           |                          |                         |             |
| ≤2 cm                 | 2798 (65.3)              | 362 (69.8)              | 1 (reference) |
| >2 cm                 | 1488 (34.7)              | 157 (30.3)              | 0.83 (0.67–1.03) |
| Missing               | 204 (–)                  | 31 (–)                  | —            |
| Axillary node status  |                          |                         |             |
| Negative              | 2523 (61.7)              | 332 (66.5)              | 1 (reference) |
| Positive              | 1565 (38.3)              | 167 (33.5)              | 0.83 (0.67–1.02) |
| Missing               | 402 (–)                  | 51 (–)                  | —            |

Abbreviations: CI = confidence interval; OR = odds ratio.

*Odds ratios (95% CI) for being basal-positive cases were estimated with adjustment for age (10-year category), tumour grade, histology, tumour size, axillary node status and study.

b Student’s t-test was used to compare the distribution of continuous variable.
hormone therapy usage, and tumour size. As expected for any analysis pooling data from multiple studies, there were variations in study populations, study designs, data collection methods, and marker measurement, which may cause study heterogeneity and subtype misclassification. However, we found no significant heterogeneity across studies at least for the associations in risk factors we analysed. In addition, the proportions of CBP (8.5%) and 5-NP (6.3%) subtypes were also comparable to those reported previously (Cheang et al, 2008; Blows et al, 2010; Yang et al, 2011; Liu et al, 2012). Of note, although we used centralised

### Table 3. Comparison of risk factors between luminal A basal-negative and basal-positive subtypes

|                      | Basal-negative (n = 4490) | Basal-positive (n = 550) | OR (95% CI)* | P-value |
|----------------------|--------------------------|--------------------------|--------------|---------|
| **Age, years (mean, s.d.)** | 56.6 (10.8) | 56.3 (11.4) | — | 0.66* |
| **Family history of breast cancer** | | | | |
| Absent               | 2859 | 78.7 | 345 | 73.4 | 1 (reference) | — |
| Present             | 776  | 21.4 | 125 | 26.6 | 1.27 (0.99–1.63) | 0.06 |
| Missing            | 855  | —    | 80  | —    | —              | —    |
| **Age at menarche** | | | | |
| ≤12                 | 1421 | 36.6 | 168 | 36.7 | 1 (reference) | — |
| 13–14               | 1751 | 45.1 | 200 | 43.7 | 0.97 (0.78–1.22) | 0.82 |
| >14                | 711  | 18.3 | 90  | 19.7 | 1.05 (0.79–1.39) | 0.75 |
| Missing            | 607  | —    | 92  | —    | —              | —    |
| P<sub>τ</sub>       | —    | —    | —   | —    | 1.02 (0.88–1.17) | 0.82 |
| **Parity** | | | | |
| Nulliparous         | 659  | 15.9 | 69  | 13.9 | 1 (reference) | — |
| Primiparous         | 3490 | 84.1 | 429 | 86.1 | 1.17 (0.89–1.54) | 0.25 |
| Missing             | 341  | —    | 52  | —    | —              | —    |
| **Age at first full-term pregnancy** | | | | |
| <20                | 362  | 11.2 | 47  | 12.2 | 1 (reference) | — |
| 20–24               | 1403 | 43.2 | 159 | 41.4 | 0.82 (0.57–1.17) | 0.27 |
| 25–29               | 969  | 29.9 | 115 | 30.0 | 0.85 (0.59–1.24) | 0.41 |
| ≥30                | 512  | 15.8 | 63  | 16.4 | 0.90 (0.59–1.38) | 0.63 |
| Missing             | 244  | —    | 45  | —    | —              | —    |
| P<sub>τ</sub>       | —    | —    | —   | —    | 0.99 (0.87–1.12) | 0.85 |
| **Number of live births** | | | | |
| 1                   | 727  | 20.8 | 76  | 17.7 | 1 (reference) | — |
| 2                   | 1583 | 45.4 | 209 | 48.7 | 1.28 (0.96–1.72) | 0.09 |
| ≥3                  | 1180 | 33.8 | 144 | 33.6 | 1.09 (0.79–1.52) | 0.59 |
| P<sub>τ</sub>       | —    | —    | —   | —    | 1.00 (0.84–1.18) | 0.97 |
| **Breast feeding**  | | | | |
| Never               | 475  | 17.9 | 53  | 17.6 | 1 (reference) | — |
| Ever                | 2181 | 82.1 | 248 | 82.4 | 1.11 (0.79–1.55) | 0.54 |
| P<sub>τ</sub>       | 834  | —    | 128 | —    | —              | —    |
| **Body mass index (cm m<sup>−2</sup>)** | | | | |
| Overall             | 26.6 (5.1) | 26.0 (4.9) | — | 0.02<sup>b</sup> | — |
| <25                 | 1701 | 43.6 | 215 | 47.4 | 1 (reference) | — |
| 25–30               | 1355 | 34.7 | 157 | 34.6 | 0.92 (0.74–1.16) | 0.50 |
| ≥30                 | 847  | 21.7 | 82  | 18.1 | 0.78 (0.59–1.03) | 0.08 |
| Missing             | 587  | —    | 96  | —    | —              | —    |
| Per 5 unit          | —    | —    | —   | —    | 0.90 (0.81–1.01) | 0.07 |
| **Premenopausal (age <50)** | | | | |
| Mean, s.d.          | 25.1 (4.80) | 24.9 (4.66) | — | 0.63<sup>b</sup> | — |
| <25                 | 631  | 57.7 | 79  | 57.3 | 1 (reference) | — |
| 25–30               | 323  | 29.6 | 44  | 31.9 | 1.12 (0.75–1.69) | 0.57 |
| ≥30                 | 139  | 12.7 | 15  | 10.9 | 0.78 (0.42–1.44) | 0.43 |
| Missing             | 138  | —    | 20  | —    | —              | —    |
| Per 5 unit          | —    | —    | —   | —    | 0.94 (0.77–1.15) | 0.56 |
| **Postmenopausal (age ≥50)** | | | | |
| Mean, s.d.          | 27.2 (5.10) | 26.5 (4.89) | — | 0.03<sup>b</sup> | — |
| <25                 | 1070 | 38.1 | 136 | 43.0 | 1 (reference) | — |
| 25–30               | 1032 | 36.7 | 113 | 35.8 | 0.89 (0.67–1.18) | 0.41 |
| ≥30                 | 708  | 25.2 | 67  | 21.2 | 0.81 (0.58–1.12) | 0.20 |
| Missing             | 449  | —    | 76  | —    | —              | —    |
| Per 5 unit          | —    | —    | —   | —    | 0.89 (0.79–1.01) | 0.07 |

Abbreviations: CI = confidence interval; OR = Odds ratio.

*Odds ratios (95% CI) for being basal-positive cases were estimated with adjustment for age (10-year categories), family history of breast cancer, age at menarche, parity, and BMI.

<sup>b</sup> Student’s t-test was used to compare the distribution of continuous variables.

<sup>c</sup> Parous women only; adjusted for age (10-year categories), family history of breast cancer, age at menarche, parity, age at first full-term pregnancy, number of live births, breastfeeding, BMI, and study.
measurement for CK5/6 and EGFR expression, we used ER, PR, and HER2 status retrieved from clinical records in each study instead of centralised data to maximise the power of our study. Accordingly, IHC methods and cut-point for positivity varied substantially among studies. However, we observed high concordance for ER and PR status between clinical records and centralised quantitative measurements among a subset of study subjects with both data available. Further, the overall proportions of positivity for these five markers (ER, 78.0%; PR, 64.2%; HER2, 14.5%; CK5/6, 13.5%; EGFR, 12.8%) were generally consistent with what was reported in previous studies (El-Rehim et al, 2004; Carey et al, 2006; Rakha et al, 2006; Cheang et al, 2008; Liu et al, 2012). Finally, information on proliferation marker (such as Ki-67) was not available for most studies, which made the accurate classification of real luminal A tumours a challenge. However, results from the sensitivity analysis restricting to cases with high ER and PR expression levels using centralised data did not change results significantly, suggesting that the potential subtype misclassification caused by study heterogeneity or marker measurement and scoring did not significantly influence our conclusion.

In conclusion, we found that tumour characteristics and known risk factors were generally similar in basal-positive and basal-negative luminal A tumours. The small differences in tumour features and family history between the two luminal A subtypes warrant further investigations in future studies with larger number of subjects and detailed annotation of subtype and risk factor information.

ACKNOWLEDGEMENTS

The Helsinki Breast Cancer Study (HEBCS) thanks Kristiina Attomäki, Kirsimari Aaltonen, Taru A Muranen, Karl von Smitten, and Irja Erkkilä for their kind help with the patient samples and data. The Study of Epidemiology and Risk factors in Cancer Heredity (SEARCH) thanks Elena Provenzano and Marie Mack. The HEBCS was supported by The Helsinki University Central Hospital Research Fund, Academy of Finland (266528), the Finnish Cancer Society, and The Nordic Cancer Union and the Sigrid Juselius Foundation. The Mayo Clinic Breast Cancer Study (MCBCS) was supported by The Breast Cancer Research Foundation, the National Institutes of Health Specialized Program of Research Excellence (SPORE) in Breast Cancer (CA116201), R01 grants (CA128978 and CA176785), and the Grohne Family. The Melbourne Collaborative Cohort Study (MCCS) was supported by The Cancer Council Victoria, VicHealth, NHMRC (209057, 251533, 396414, and 504711), Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index (NDI). The Nurses’ Health Study (NHS) was supported by National Institutes of Health/ National Cancer Institute (UM1 CA186107 and P01 CA 79569). The NHS thanks 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 Polish Breast Cancer Study (PBCS) was supported by Intramural Research Program of US NIH, NCI, Division of Cancer Epidemiology and Genetics (DCEG). The Sheffield Breast Cancer Study (SBCS) was supported by Yorkshire Cancer Research (S295, S299, and S305PA) and Sheffield Experimental Cancer Medicine Centre. The SEARCH was supported by Cancer Research UK (C490/A16561), the Biomedical Research Centre at the University of Cambridge. The Kuopio Breast Cancer Project (KBCP) was supported by The special Government Funding of Kuopio University Hospital grants, Cancer Fund of North Savo, the Finnish Cancer Organizations, the strategic funding of the University of Eastern Finland.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

Ali HR, Rueda OM, Chin SF, Curtis L, Dunning MJ, Aparicio SA, Caldas C (2014) Genome-driven integrated classification of breast cancer validated in over 7500 samples. Genome Biol 15: 431.

Anderson KN, Schwab RB, Martinez ME (2014) Reproductive risk factors and breast cancer subtypes: a review of the literature. Breast Cancer Res Treat 144: 1–10.

Blows FM, Driver KE, Schmidt MK, Brooks A, Van Leeuwen FE, Wesseling J, Cheang MC, Gelmon K, Nielsen TO, Blomqvist C, Heikilä P, Heikinen T, Nevanlinna H, Akslen LA, Begin LR, Foulkes WD, Couch FJ, Wang X, Caofoiek V, Olson JE, Baglietto L, Giles GG, Severi G, Mclean CA, Southey MC, Rakha E, Green AR, Ellis IO, Sherman ME, Lissowska J, Anderson WF, Cox A, Cross SS, Reed MWR, Provenzano E, Dawson SJ, Dunning AM, Humphreys M, Easton DF, Garcia-Closas M, Caldas C, Pharoah PD, Huntsman D (2010) Subtyping of breast cancer by immunohistochemistry to investigate a relationship between subtype and short and long term survival: a collaborative analysis of data for 10 159 cases from 12 studies. Plos Med 7: e1000279.

Cancer Genome Atlas Network (2012) Comprehensive molecular portraits of human breast tumours. Nature 490: 61–70.

Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Denning SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Mullikan RC (2006) Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 295: 2492–2502.

Cheang MC, Voduc D, Bajdik C, Leung S, Mckinney S, Chia SK, Perou CM, Nielsen TO (2008) Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res 14: 1368–1376.

Ciriello G, Sinha R, Hoadley KA, Jacobsen AS, Reva B, Perou CM, Sander C, Schultz N (2013) The molecular diversity of Luminal A breast tumors. Breast Cancer Res Treat 141: 409–420.

Colleoni M, Rotmensz N, Maisonneuve P, Mantgazou MG, Lunin A, Veronesi P, Intra M, Montagna E, Cancello G, Cardillo A, Mazza M, Perri G, Iorfida M, Pruneri G, Goldhirsch A, Viale G (2012) Outcome of special types of luminal breast cancer. Ann Oncol 23: 1428–1436.

El-Rehim DMA, Pinder SE, Paish CE, Bell J, Blamey R, Robertson JFR, Nicholson RI, Ellis IO (2004) Expression of luminal and basal cytokeratins in human breast carcinoma. J Pathol 203: 661–671.

Ewertz M, Jensen MB, Gunnarsdottir KA, Hojris I, Jakobsen EH, Nielsen D, Stenbygaard LE, Tange UB, Gold S (2011) Effect of obesity on prognosis after early-stage breast cancer. J Clin Oncol 29: 25–31.

Haakinson DJ, Lees SG, Deutic AK, Gray RJ, Wasif N, Stucky CCH, Northfelt DW, Apsley HA, Pockaj B (2012) The impact of obesity on breast cancer: a retrospective review. Ann Surg Oncol 19: 3012–3018.

Horwitz KB, Dye WW, Harrell JC, Kaposi P, Stoffl CA (2008) Rare steroid receptor-negative basal-like tumorigenic cells in luminal subtype human breast cancer xenografts. Proc Nail Acad Sci USA 105: 5774–5779.

Howell SJ (2013) Advances in the treatment of luminal breast cancer. Curr Opin Obstet Gynecol 25: 49–54.

Ignatiadis M, Sotiropou C (2013) Luminal breast cancer: from biology to treatment. Nat Rev Clin Oncol 10: 494–506.

Kabos P, Haughian JM, Wang X, Dye WW, Finlayson C, Elias A, Horwitz KB, Dye WW, Harrell JC, Kabos P, Sartorius CA (2008) Rare steroid receptor-negative basal-like tumorigenic cells in luminal subtype human breast cancer xenografts. Proc Nail Acad Sci USA 105: 5774–5779.

Livasy C, Carey L, Earp HS, Perou CM (2008) Epidemiology of basal-like breast cancer: a retrospective review. Ann Surg Oncol 15: 3012–3018.

Milikan RC, Newman B, Tse CK, Moorman PG, Conway K, Dressler LG, Smith LV, Labbok MH, Geradts J, Jensen JT, Jackson S, Nyanta S, Livasy C, Carey L, Earp HS, Perou CM (2008) Epidemiology of basal-like breast cancer. Breast Cancer Res Treat 110: 121–139.

Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein
D (2000) Molecular portraits of human breast tumours. Nature 406: 747–752.

Phipps AI, Chlebowski RT, Prentice R, Mctiernan A, Wactawski-Wende J, Kuller LH, Adams-Campbell LL, Lane D, Stefanick ML, Vitolins M, Kabat GC, Rohan TE, Li CI (2011) Reproductive history and oral contraceptive use in relation to risk of triple-negative breast cancer. J Natl Cancer Inst 103: 470–477.

Rakha EA, El-Rehim DA, Paish C, Green AR, Lee AHS, Robertson JF, Blamey RW, Macmillan D, Ellis IO (2006) Basal phenotype identifies a poor prognostic subgroup of breast cancer of clinical importance. Eur J Cancer 42: 3149–3156.

Sotiriou C, Neo SY, Mcshane LM, Korn EL, Long PM, Jazaeri A, Martiat P, Fox SB, Harris AL, Liu ET (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. Proc Natl Acad Sci USA 100: 10393–10398.

Yang XR, Chang-Claude J, Goode EL, Couch FJ, Nevanlinna H, Milne RL, Gaudet M, Schmidt MK, Broeks A, Cox A, Fasching PA, Hein R, Spurdle AB, Blows F, Driver K, Flesh-Janys D, Heinz J, Sinn P, Vrieling A, Heikkinen T, Aittomaki K, Heikkila P, Blomqvist C, Lissowska J, Peplonska B, Chanock S, Figueroa J, Brinton L, Hall P, Czene K, Humphreys K, Darabi H, Liu J, van’t Veer LJ, van Leeuwen FE, Andrulis IL, Glendon G, Knight JA, Mulligan AM, O’Malley FP, Weerasooriya N, John EM, Beckmann MW, Hartmann A, Weihbrecht SB, Wachter DL, Jud SM, Loehberg CR, Baglietto L, English DR, Giles GG, Mclean CA, Severi G, Lambrechts D, Vandorpe T, Weltens C, Paridaens R, Smeets A, Neven P, Wilders H, Wang X, Olson JE, Cafourek V, Frederiksen Z, Kosel M, Vachon C, Cramp HE, Connelly D, Cross SS, Balasubramanian SP, Reed MW, Dork T, Bremer M, Meyer A, Karstens JH, Ay A, Park-Simon TW, Hillemanns P, Arias Perez JJ, Menendez Rodriguez P, Zamora P, Benitez J, Ko YD, Fischer HP, Hamann U, Pesch B, Bruning T, Justenhoven C, Brauch H, Eccles DM, Tapper WJ, Gerty SM, Sawyer EJ, Tomlinson IP, Jones A, Kerin M, Miller N, Mincerney N, Anton-Culver H, Ziegas A, Shen CY, Huising CN, Wu PE, Yang SL, Yu JC, Chen ST, Hsu GC, Haiman CA, Henderson BE, Le Marchand L, Colonel LN, Lindblom A, Margolin S, Jakubowska A, Lubiński J, Huzarski T, Byrski T, Görski B, Gronwald J, Hooing MJ, Hollestelle A, van den Ouweland AM (2011) Associations of breast cancer risk factors with tumor subtypes: a pooled analysis from the Breast Cancer Association Consortium studies. J Natl Cancer Inst 103: 250–263.

This work is published under the standard license to publish agreement. After 12 months the work will become freely available and the license terms will switch to a Creative Commons Attribution-NonCommercial-Share Alike 4.0 Unported License.

Supplementary Information accompanies this paper on British Journal of Cancer website (http://www.nature.com/bjc)