Published in final edited form as:
Breast Cancer Res Treat. 2014 January ; 143(1): 181–187. doi:10.1007/s10549-013-2771-z.

Breast cancer susceptibility risk associations and heterogeneity by E-cadherin tumor tissue expression

Hisani N. Horne[1], **Mark E. Sherman**[2], **Montserrat Garcia-Closas**[3], **Paul D. Pharoah**[4], **Fiona M. Blows**[4], **Xiaohong R. Yang**[1], **Stephen M. Hewitt**[5], **Catherine Conway**[5], **Jolanta Lissowska**[6], **Louise A. Brinton**[1], **Ludmila Prokunina-Olsson**[1], **Sarah-Jane Dawson**[4], **Carlos Caldas**[4], **Douglas F. Easton**[4], **Stephen J. Chanock**[1], and **Jonine D. Figueroa**[1]

[1]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Rockville, MD, USA [2]Division of Cancer Prevention, National Cancer Institute, National Institutes of Health, Rockville, MD, USA [3]Section of Epidemiology and Genetics, The Institute of Cancer Research, Sutton, UK [4]Department of Oncology and Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK [5]Laboratory of Pathology, National Cancer Institute, National Institutes of Health, Bethesda, MD, USA [6]Department of Cancer Epidemiology and Prevention, M. Sklodowska-Curie Memorial Cancer Center and Institute of Oncology, Warsaw, Poland

Abstract

E-cadherin is involved in cell-cell adhesion and epithelial-to-mesenchymal transitions (EMT). In cancers, loss or inactivation of E-cadherin is associated with epithelial cell proliferation and invasion. Here, we sought to determine if risk associations for 18 breast cancer susceptibility single nucleotide polymorphisms (SNPs) differed by E-cadherin tumor tissue expression in the Polish Breast Cancer Study (PBCS), using data on 1,347 invasive breast cancer cases and 2,366 controls. E-cadherin expression (low/high) was assessed using immunohistochemical staining of tumor tissue microarrays. Replication data on 2,006 cases and 6,714 controls from the Study of Epidemiology and Risk Factors in Cancer Heredity (SEARCH) was used to follow-up promising findings from PBCS. In PBCS, we found the rs11249433 SNP at the 1p11.2 locus to be more strongly associated with risk of E-cadherin low tumors (OR = 1.30, 95% CI 1.08 – 1.56) than with E-cadherin high tumors (OR = 1.06, 95% CI 0.95 – 1.18; case-only *p*-heterogeneity (*p*-het) = 0.05). Findings in PBCS for rs11249433 were replicated in SEARCH. Combined analyses of the two datasets for SNP rs11249433 revealed significant heterogeneity by E-cadherin expression (combined case-only *p*-het = 0.004). Further, among carriers of rs11249433, the highest risk was seen for E-cadherin low tumors that were ER-positive and of lobular histology. Our results in two independent data sets suggest that rs11249433, which is located between the *NOTCH2* and *FCGR1B* genes within the 1p11.2 locus, is more strongly associated with risk of breast tumors

Correspondence to: Hisani N. Horne, National Institutes of Health/NCI/DCEG/HREB, 9609 Medical Center Drive, Rm 7E234, MSC 7234, Bethesda, MD 20892-7234.

Conflict of interest
The authors declare that they have no conflict of interest.

with low or absent E-cadherin expression, and suggest that evaluation of E-cadherin tumor tissue expression may be useful in clarifying breast cancer risk factor associations.

INTRODUCTION

Genome-wide association studies (GWAS) have identified numerous breast cancer susceptibility single nucleotide polymorphism (SNP) markers (1, 2). Several follow-up studies have reported that the risk for breast cancer associated with these markers can vary by clinical and pathological characteristics. For example, Broeks and colleagues noted the associations with breast cancer risk for multiple SNPs differed by hormone receptor expression in breast tumor tissue (3). Analysis of SNP risk associations by tumor characteristics could provide additional etiological insights and more precise relative risk estimates for risk prediction models. Prediction of subtype-specific risk might result in improved prevention and screening interventions offered to individuals at highest risk of tumors that are more likely to respond to specific interventions such as hormonal chemoprevention.

Analysis of molecular markers in breast tumors has clarified risk factor associations that may be obscured when considering breast cancer as a single homogenous disease (1, 3, 4). In this study we explored E-cadherin as a potential protein marker for clarifying breast cancer risk associations. Expression of E-cadherin protein (encoded by the *CDH1* gene) is critical for maintaining epithelial cell-cell adhesion and epithelial-to-mesenchymal transitions (EMT) (5). E-cadherin is considered a tumor-suppressor protein because its loss or inactivation by mutations is frequently seen in invasive epithelial cell cancers and is thought to be an essential step in both tumorigenesis and progression (5, 6). Decreased cellular adhesion due to loss of E-cadherin leads to enhanced invasion of tumor cells and metastases. Ductal and lobular carcinomas comprise the two major histologic subtypes of invasive breast cancers. Loss of E-cadherin expression has been noted more frequently in invasive lobular carcinomas compared to invasive ductal carcinomas (7).

Here we explored whether common breast cancer susceptibility loci were differentially associated with tumors classified by either low or high E-cadherin expression. These analyses were done in two independent breast cancer case-control studies, the Polish Breast Cancer Study (PBCS) and the Study of Epidemiology and Risk Factors in Cancer Heredity (SEARCH).

MATERIAL AND METHODS

Study population

Polish Breast Cancer Study (PBCS)—The study population has previously been described in detail (8, 9). In brief, eligible cases included all women between the ages of 20 and 75 years who were residents of Warsaw or ŁódŸ, Poland from 2000 to 2003 and who were diagnosed with incident *in situ* or invasive breast cancer. These cases were confirmed and reviewed centrally to provide standardized classification. Approximately 2,386 cases (79% of eligible) and 2,502 (69% of eligible) age and study site frequency matched

population controls agreed to participate in the study and provided informed consent required by the National Cancer Institute (USA) and local institutional review boards. This analysis is based on 1,347 invasive cases with available E-cadherin tumor tissue expression data.

Study of Epidemiology and Risk Factors in Cancer Heredity (SEARCH)—We used data from the SEARCH Breast Cancer Study (10) an ongoing population-based study of women diagnosed with breast cancer in the region of England included in the Eastern Cancer Registration and Information Centre (ECRIC, formerly East Anglia Cancer Registry). Eligible participants included women diagnosed with invasive breast cancer who were either under 70 years of age since the beginning of the study on July 1, 1996 (incident cases) or age 55 or younger since January 1, 1991 and were alive at the start of the study (prevalent cases). Controls, frequency matched to cases by age and geographic region, were selected from the EPIC-Norfolk cohort study recruited between 1992 and 1994 and from general practitioner practices from March 2003 to present. Approximately 64% of eligible cases and 41% of controls agreed to participate in the study and provided blood samples. All participants in the study provided informed consent, and the study was approved by the Cambridgeshire 4 Research Ethics Committee.

Pathology and tumor markers

For PBCS, histopathologic features including diagnosis, grade, tumor size, ER status of the tumors and axillary lymph node metastases were assessed using clinical reports and independent evaluation by the study pathologist (M.E.S.). Routinely prepared formalin-fixed paraffin-embedded (FFPE) blocks of 1,347 invasive breast tumors were used to construct tissue microarray (TMA) blocks with 2-fold representation of 1.0 mm diameter cores per tumor (Pathology Devices, Westminister, MD). TMA blocks were sectioned with a microtome at 5-μm thickness onto charged slides. Due to the limited availability of tissue, cases with small tumors were underrepresented in the TMAs (52% of cases with tissue samples in the TMAs had tumors ≤2 cm compared with 66% of cases not included in the TMAs, $P < 0.0001$; data not shown). This might contribute to the fact that tumors in the TMAs were more frequently lobular (18 vs. 12%), node positive (41 vs. 29%), ER-positive (69 vs. 59%) and low grade (75 vs. 69%) compared to patients with available FFPE blocks but not included in the TMAs due to limited tissue.

For immunohistochemistry (IHC) analysis: TMA sections were deparaffinized with zylene and graded alcohols, antigen retrieval was mediated with citrate buffer pH 9 (Dako) for 20 minutes in a pressure cooker. Primary mouse monoclonal antibody, anti-E-cadherin (clone NCH-38, 1:500; Dako, Carpinteria, CA) was applied at room temperature for 2 hours. Detection of the antigen-antibody complex was done with Envision+ (Dako) and DAB was applied for 20 minutes. Slides were counterstained with hematoxylin, dehydrated and coverslipped. Slides were imaged with a Hamamatsu Nanozoomer (Bridgewater, NJ), at 20X magnification. A cytotechnologist assessed whole digital images of stained TMAs using the SlidePath Digital Image Hub (Leica Biosystems, Wetzlar, Germany), blinded to any clinical phenotypes or genetic data. Manual readings for each TMA spot was performed and the percentage of cells positively stained for E-cadherin (0–100%) and the intensity of

staining (0 = negative, 1 = weak, 2 = intermediate, and 3 = strong) were recorded. Quality assurance and reproducibility of IHC scoring done on 200 images, by cytotechnologist and a clinical pathologist (M.E.S.) showed good inter- and intra-observer agreement (≥90%) calculated by weighted kappa analyses. An overall E-cadherin score was generated using the product of percent positive tumor cells and intensity, resulting in a product score range of 0–300. Tumors having a score of <100 were classified as E-cadherin low and those with a score ≥100 as E-cadherin high (Figure 1). This dichotomous cut-point was determined prior to genetic association analyses and was based on the observed distribution of the composite scores and supported by evidence in the literature (11, 12).

We obtained SEARCH data on clinicopathologic characteristics, specifically, ER status, histology, grade, tumor size and axillary node involvement. Clinical characteristics for SEARCH participants were obtained through Eastern Cancer Registration and Information Centre (formerly the East Anglian Cancer Registry). For this report we used genotyping and clinicopathologic data from 2006 incident breast cancer cases with E-cadherin expression data and 6,714 controls. In this set, IHC staining for E-cadherin (M3612, 1:25, Dako, Carpineteria, CA) was done on TMAs constructed from paraffin-embedded TMA blocks as previously described (13). Scoring for E-cadherin was done manually by a study pathologist who assessed the percent of cells staining positive (0–100%). Tumors with ≤10 percent of cells stained were classified as E-cadherin low while tumors with >10 percent of cells stained were classified as E-cadherin high.

As described above, the IHC in the two studies was performed with different anti-E-cadherin antibodies and different scoring criteria; in PBCS E-cadherin scores were based on both intensity of staining and percentage of positive cells, while SEARCH scores were based solely on percentage of positive cells. Of the tumors stained for E-cadherin expression, 10% (N=138) were categorized as E-cadherin low in PBCS and 23% (N=471) in SEARCH.

Genetic analyses

Genotyping for PBCS and SEARCH was performed using pre-designed TaqMan assays (Applied Biosystems, Foster City, CA, USA) according to standard protocols described previously in the framework of the Breast Cancer Association Consortium (BCAC) of which both PBCS and SEARCH are participating studies (3, 14–16). For PBCS, we used genotyping data for 18 breast cancer SNPs available at the time of analyses (see Supplementary Table 1). Importantly, the SNPs in this study were not associated with missingess of the samples in the PBCS TMA (OR = 1.06, 95% CI (0.90 – 1.25). For SEARCH we received existing genotyping data to replicate our two significant findings of risk differences observed with E-cadherin expression in PBCS: SNPs rs2046210 and rs11249433 (17).

Statistical analyses

Polytomous logistic regression models adjusted for age and study site (PBCS or SEARCH) were used to estimate per-allele odds ratios (OR) and 95% confidence intervals (95% CI) for genetic association between SNPs and E-cadherin low/high tumor expression, compared to

controls. We used fixed effects meta-analysis to generate the combined ORs and 95% CI for PBCS and SEARCH. *P*-values to test for heterogeneity (*P*-het) for tumor characteristics were obtained using logistic regression models restricted to cases (case-only analyses), adjusted for age and study site (PBCS) or age only (SEARCH). Differences between E-cadherin low/high expressing tumors and clinicopathologic characteristics were examined using chi-square test for all tumors and then stratified by ductal and lobular histology. Results with $p < 0.05$ were considered statistically significant; no adjustment for multiple testing was implemented. All analyses were conducted using Stata version 11.2 for Windows (College Station, TX).

RESULTS

Among the 18 susceptibility loci evaluated, 10 were significantly associated with breast cancer risk at a $p < 0.05$ significance level in PBCS (Supplementary Table 1). We next determined if these 18 SNPs exhibited risk differences by E-cadherin tumor tissue expression. Two SNPs, rs2046210 (*ESR1)* and rs11249433 (*NOTCH2/FCGR1B*), showed stronger associations with breast cancer risk among patients with E-cadherin low compared to E-cadherin high tumors: rs2046210 SNP at 6q25.1 with OR = 1.43, 95% CI 1.18 – 1.75 for E-cadherin low tumors, compared to OR = 1.06, 95% CI 0.94 – 1.19 for E-cadherin high tumors (case-only *p*-het = 0.007; Table 1 & Supplementary Table 2). Similarly, rs11249433 at 1p11.2 was more strongly associated with E-cadherin low tumors (OR = 1.30, 95% CI 1.08 – 1.56) than with E-cadherin high tumors (OR = 1.06, 95% CI 0.95 – 1.18; case-only *p*-het = 0.05).

We next evaluated the two significant findings from PBCS in the SEARCH dataset. In SEARCH, evaluation of rs2046210 and rs11249433 replicated only rs11249433 association, with effect sizes being in the same direction as in PBCS. For rs11249433 we found stronger associations with E-cadherin low tumors (OR = 1.30, 95% CI 1.14 – 1.49) than with E-cadherin high tumors (OR = 1.11, 95% CI 1.03 – 1.21; *p*-het = 0.04; Table 1). However, the association for rs2046210 for SEARCH participants was in the opposite direction of PBCS, with stronger risk associations observed for E-cadherin high tumors compared with E-cadherin low tumors, and combined analysis of rs2046210 was not significant (combined case-only *p*-het = 0.84; Table 1). Combined analysis of the PBCS and SEARCH studies (3,231 case and 9,054 controls) for rs11249433 showed differing estimates of breast cancer risk based on E-cadherin expression: E-cadherin low tumors (combined OR = 1.30, 95% CI 1.16 – 1.45) and E-cadherin high tumors (combined OR = 1.09, 95% CI 1.02 – 1.16); combined case-only *p*-het = 0.004 (Table 1).

To further refine the risk associations for rs11249433, we next assessed relationships according to combined parameters of E-cadherin and both ER status and histologic subtype (ductal or lobular; Table 2). Risk associated with rs11249433 was stronger in patients with E-cadherin low tumors that were ER-positive (combined OR = 1.38, 95% CI 1.21 – 1.56) as well as those of lobular histology (OR = 1.41, 95% CI 1.20 – 1.61; Table 2). Further stratification of the ER-positive E-cadherin low tumors by histology did not reveal significantly different risk estimates from what was observed in ER-positive tumors alone (Table 2). There was no evidence of an association for rs11249433 among ER-negative

tumors, irrespective of histologic subtype (data not shown). When assessing differences in risk associations by E-cadherin tissue expression among the stratified tumor cases, we only observed significant heterogeneity among ER-positive tumors (combined case-only *p*-het = 0.001; Table 2).

Given the known association of E-cadherin with tumor histology and its probable association with ER status, we evaluated the differences in the distribution of tumor characteristics among E-cadherin low and high breast cancer cases. Supplementary Table 3 summarizes the distribution of select clinicopathologic features stratified by E-cadherin tumor tissue expression for both PBCS and SEARCH. As expected, E-cadherin low tumors were largely composed of lobular breast tumors (38–49% lobular for E-cadherin low vs. 8–10% for E-cadherin high; Supplementary Table 3). Further, compared to E-cadherin high tumors, E-cadherin low tumors were more frequently larger in size both in PBCS and SEARCH (40–55% >2 cm for E-cadherin low vs. 33–46% for E-cadherin high; Supplementary Table 3). There were observed differences between PBCS and SEARCH for associations between E-cadherin expression status and tumor characteristics. In PBCS only, E-cadherin expression was associated with tumor grade, whereas in SEARCH, E-cadherin expression was associated with ER status ($p < 0.05$; Supplementary Table 3). Comparing PBCS and SEARCH cancer cases with available E-cadherin IHC data revealed significant differences by histology, grade, tumor size and ER status. Specifically, tumors in SEARCH were more frequently of ductal histology, poorly differentiated, smaller in size (≤2 cm) and ER-positive ($p < 0.001$; data not shown), compared to PBCS.

DISCUSSION

In this report from two large breast cancer studies totaling over 3300 cases, we provide evidence that E-cadherin breast tumor expression can be used as a protein marker to further refine SNP breast cancer risk associations. Specifically, we show that the SNP rs11249433 on 1p11.2 was more strongly associated with breast tumors with low or absent levels of E-cadherin. We also demonstrate a consistent association with low E-cadherin expression and both lobular histology and large (≥2 cm) tumor size. To our knowledge, E-cadherin has not been included as a molecular marker in previous analyses of risk differences for breast cancer susceptibility loci.

Our finding for rs11249433 is consistent with previous data showing that the 1p11.2 locus was most strongly associated with ER-positive tumors that were of low grade and lobular histology (15). Our current analyses extends these previous findings by showing a stronger association for rs11249433 among patients with either ER-positive or lobular tumors that are also E-cadherin low, indicating that E-cadherin tumor tissue expression can further refine this association. Though the SNP we identified appears to modify the risk of developing a more favorable breast cancer subtype, the utility of SNPs for refining breast cancer risk profiles is relevant for both more and less aggressive tumor types. Specifically, prediction of subtype-specific risk may benefit women at highest risk of tumors that are more likely to respond to specific therapies. Notably, invasive lobular tumors comprise a special subtype of breast cancer displaying a low response to preoperative chemotherapy indicating that a more tailored approach to treatment is warranted (18). Also, given the recent identification of 41

new susceptibility loci associated with breast cancer risk (19), it will be interesting to determine if any of these SNPs display heterogeneity in risk by E-cadherin tissue expression. Furthermore, combining multiple SNPs into breast cancer risk prediction models could increase the discriminatory ability of individual genetic risk factors (20).

The rs11249433 SNP resides in a pericentromeric, nongenic region on chromosome 1p11.2, in a large linkage disequilibrium block neighboring the *NOTCH2* and *FCGR1B* genes (21). Using HaploReg (22) we found 3 motif changes linked to SNP rs11249433, including Mef2, Pax-2 and Pou1f1, which could potentially influence transcriptional regulation. Although the rs11249433 SNP is about 600 kb away from the *NOTCH2* gene, a study examining the association between rs11249433 and all surrounding genes reported that *NOTCH2* expression in breast tumors was increased in carriers of the risk allele of rs11249433 (23). Interestingly, this study also noted that the association of *NOTCH2* expression and rs11249433 was only found in *TP53* wild-type/ER-positive tumors. NOTCH overexpression leads to preferential differentiation of mammary stem cells into luminal type epithelium (24). Furthermore, it is important to note that the NOTCH signaling pathway has been shown to promote the EMT process during oncogenesis via the transcriptional induction of the Snail repressor leading to loss of E-cadherin expression (25). These findings provide a feasible link between carriers of the SNP rs11249433, reduction in *NOTCH2* expression and development of E-cadherin low tumors, particularly those that are ER-positive.

Our analyses in the PBCS found that the SNP rs2046210 at 6q25.1 (*ESR1)* also displayed risk differences by E-cadherin tumor expression levels. However, we were unable to replicate this finding in the SEARCH population, where we observed stronger risk associations of the SNP with E-cadherin high tumors. It is possible that the difference seen in the PBCS and SEARCH for rs2046210 is due to population differences: PBCS is unscreened whereas SEARCH is a highly screened population. Also, there may be differences between PBCS and SEARCH in the prevalence of menopausal hormone therapy (MHT) use among participants. MHT may be an important factor to consider given the results of a case-only analysis by O'Connor et al. (26) which found a trend towards a higher incidence of E-cadherin negative tumors among women on MHT compared to those who did not receive MHT. Further, the IHC analysis of E-cadherin expression in PBCS and SEARCH was performed with different antibodies and utilizing different scoring protocols, thus, some technical differences between these sets could contribute to non-replication of the results. There is also the possibility that the original finding in PBCS was in fact a false-positive, hence the inability to replicate the finding in SEARCH. Future studies containing larger sample sizes will be needed to determine the validity of the rs2046210 association.

The strengths of this study include the novel use of E-cadherin tissue expression as a biomarker to assess differences in genetic associations with breast cancer risk and the ability to replicate our findings for the rs11249433 (1p11.2) locus in two large, independent datasets. A limitation of our study was the technical differences in E-cadherin IHC analysis and scoring used in the two studies. However, in both studies, the cut-points for low and high E-cadherin expression were chosen *a priori* to avoid data-driven inference. Also, the percentage of lobular tumors that we observed to be E-cadherin low (57%) was lower than what has been previously reported in the literature which typically ranges from 80 – 90% (7,

27, 28). Consequently, some caution should be used when interpreting the results for lobular breast cancers. However, these limitations would not likely affect our overall conclusions of the paper but instead bias measures of association towards the null. The fact that the association for rs11249433 was detected in both datasets, regardless of their technical differences, strengthens our conclusions about the association of this variant with development of E-cadherin low tumors. In summary, we have demonstrated heterogeneity in risk associations for breast cancer susceptibility SNPs by expression of the tumor suppressor protein E-cadherin. Our data provide support for inclusion of E-cadherin as a novel molecular subtype marker in future molecular epidemiologic studies of breast cancer to improve our understanding of breast cancer etiology and risk prediction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

1. Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M. Genetic susceptibility to breast cancer. Mol Oncol. 2010; 4:174–91. [PubMed: 20542480]
2. Zhang B, Beeghly-Fadiel A, Long J, Zheng W. Genetic variants associated with breast-cancer risk: comprehensive research synopsis, meta-analysis, and epidemiological evidence. Lancet Oncol. 2011; 12:477–88. [PubMed: 21514219]
3. Broeks A, Schmidt MK, Sherman ME, et al. Low penetrance breast cancer susceptibility loci are associated with specific breast tumor subtypes: findings from the Breast Cancer Association Consortium. Human molecular genetics. 2011; 20:3289–303. [PubMed: 21596841]
4. Yang XR, Sherman ME, Rimm DL, et al. Differences in risk factors for breast cancer molecular subtypes in a population-based study. Cancer Epidemiol Biomarkers Prev. 2007; 16:439–43. [PubMed: 17372238]
5. Perl AK, Wilgenbus P, Dahl U, Semb H, Christofori G. A causal role for E-cadherin in the transition from adenoma to carcinoma. Nature. 1998; 392:190–3. [PubMed: 9515965]
6. Meiners S, Brinkmann V, Naundorf H, Birchmeier W. Role of morphogenetic factors in metastasis of mammary carcinoma cells. Oncogene. 1998; 16:9–20. [PubMed: 9467938]
7. Acs G, Lawton TJ, Rebbeck TR, LiVolsi VA, Zhang PJ. Differential expression of E-cadherin in lobular and ductal neoplasms of the breast and its biologic and diagnostic implications. Am J Clin Pathol. 2001; 115:85–98. [PubMed: 11190811]
8. Garcia-Closas M, Brinton LA, Lissowska J, et al. Established breast cancer risk factors by clinically important tumour characteristics. Br J Cancer. 2006; 95:123–9. [PubMed: 16755295]
9. Yang XR, Pfeiffer RM, Garcia-Closas M, et al. Hormonal markers in breast cancer: coexpression, relationship with pathologic characteristics, and risk factor associations in a population-based study. Cancer Res. 2007; 67:10608–17. [PubMed: 17968031]
10. Udler MS, Azzato EM, Healey CS, et al. Common germline polymorphisms in COMT, CYP19A1, ESR1, PGR, SULT1E1 and STS and survival after a diagnosis of breast cancer. International journal of cancer. Journal international du cancer. 2009; 125:2687–96. [PubMed: 19551860]
11. Szasz AM, Nemeth Z, Gyorffy B, et al. Identification of a claudin-4 and E-cadherin score to predict prognosis in breast cancer. Cancer science. 2011; 102:2248–54. [PubMed: 21883696]
12. Guler G, Balci S, Costinean S, et al. Stem cell-related markers in primary breast cancers and associated metastatic lesions. Mod Pathol. 2012; 25:949–55. [PubMed: 22388757]
13. Ali HR, Dawson SJ, Blows FM, Provenzano E, Pharoah PD, Caldas C. Cancer stem cell markers in breast cancer: pathological, clinical and prognostic significance. Breast Cancer Res. 2011; 13:R118. [PubMed: 22112299]

14. Azzato EM, Pharoah PD, Harrington P, et al. A genome-wide association study of prognosis in breast cancer. Cancer Epidemiol Biomarkers Prev. 2010; 19:1140–3. [PubMed: 20332263]
15. Figueroa JD, Garcia-Closas M, Humphreys M, et al. Associations of common variants at 1p11.2 and 14q24.1 (RAD51L1) with breast cancer risk and heterogeneity by tumor subtype: findings from the Breast Cancer Association Consortium. Human molecular genetics. 2011; 20:4693–706. [PubMed: 21852249]
16. Hein R, Maranian M, Hopper JL, et al. Comparison of 6q25 breast cancer hits from Asian and European Genome Wide Association Studies in the Breast Cancer Association Consortium (BCAC). PLoS One. 2012; 7:e42380. [PubMed: 22879957]
17. Easton DF, Pooley KA, Dunning AM, et al. Genome-wide association study identifies novel breast cancer susceptibility loci. Nature. 2007; 447:1087–93. [PubMed: 17529967]
18. Colleoni M, Russo L, Dellapasqua S. Adjuvant therapies for special types of breast cancer. Breast. 2011; 20(Suppl 3):S153–7. [PubMed: 22015285]
19. Michailidou K, Hall P, Gonzalez-Neira A, et al. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. Nat Genet. 2013; 45:353–61. 61e1–2. [PubMed: 23535729]
20. Husing A, Canzian F, Beckmann L, et al. Prediction of breast cancer risk by genetic risk factors, overall and by hormone receptor status. J Med Genet. 2012; 49:601–8. [PubMed: 22972951]
21. Thomas G, Jacobs KB, Kraft P, et al. A multistage genome-wide association study in breast cancer identifies two new risk alleles at 1p11.2 and 14q24.1 (RAD51L1). Nat Genet. 2009; 41:579–84. [PubMed: 19330030]
22. Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. 2012; 40:D930–4. [PubMed: 22064851]
23. Fu YP, Edvardsen H, Kaushiva A, et al. NOTCH2 in breast cancer: association of SNP rs11249433 with gene expression in ER-positive breast tumors without TP53 mutations. Molecular cancer. 2010; 9:113. [PubMed: 20482849]
24. Bouras T, Pal B, Vaillant F, et al. Notch signaling regulates mammary stem cell function and luminal cell-fate commitment. Cell stem cell. 2008; 3:429–41. [PubMed: 18940734]
25. Timmerman LA, Grego-Bessa J, Raya A, et al. Notch promotes epithelial-mesenchymal transition during cardiac development and oncogenic transformation. Genes & development. 2004; 18:99–115. [PubMed: 14701881]
26. O'Connor IF, Shembekar MV, Shousha S. Breast carcinoma developing in patients on hormone replacement therapy: a histological and immunohistological study. J Clin Pathol. 1998; 51:935–8. [PubMed: 10070338]
27. Goldstein NS, Kestin LL, Vicini FA. Clinicopathologic implications of E-cadherin reactivity in patients with lobular carcinoma in situ of the breast. Cancer. 2001; 92:738–47. [PubMed: 11550142]
28. Rakha EA, Abd El Rehim D, Pinder SE, Lewis SA, Ellis IO. E-cadherin expression in invasive non-lobular carcinoma of the breast and its prognostic significance. Histopathology. 2005; 46:685–93. [PubMed: 15910600]

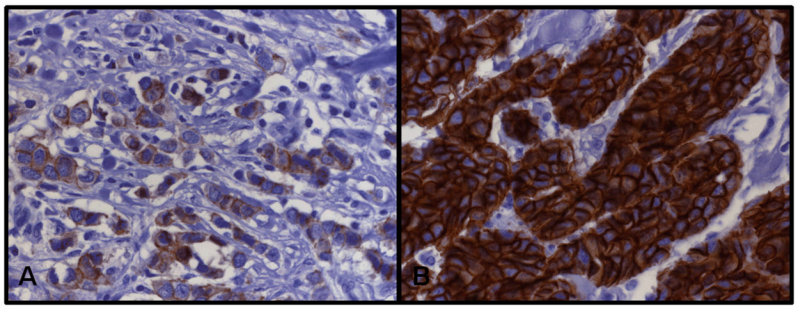

Figure 1. Detection of E-cadherin protein expression on tissue microarray (TMA) slides in PBCS

Tissue expression of E-cadherin in invasive breast tumors was assessed using immunohistochemical staining of TMAs. Representative sections of **(A)** a tumor classified as low for E-cadherin and **(B)** high for E-cadherin expression based on the intensity of the staining and percentage of cells stained positive. Magnification x500

Table 1

Per-allele OR and 95% CI for the association of established cancer susceptibility loci and breast cancer risk by E-cadherin tumor expression for two selected SNPs

					Case-control analyses				Case-only
Locus (Gene)	**SNP**	**MAF[a]**	**Study**	**Controls, N**	**N**	**OR (95% CI)**	**N**	**OR (95% CI)**	***P*-het[b]**
					E-cad low tumors vs. controls		E-cad high tumors vs. controls		
1p11.2 (NOTCH2/FCGR1B)	rs11249433 (A/G)[c]	0.369	PBCS	2340	250	1.30 (1.08–1.56)	989	1.06 (0.95–1.18)	0.05
		0.409	SEARCH	6714	468	1.30 (1.14–1.49)	1524	1.11 (1.03–1.21)	0.04
			Combined	**9054**	**718**	**1.30 (1.16–1.44)**	**2513**	**1.09 (1.02–1.16)**	**0.004**
6q25.1 (ESR1)	rs2046210 (G/A)	0.294	PBCS	2291	239	1.43 (1.18–1.75)	949	1.06 (0.94–1.19)	0.007
		0.354	SEARCH	6716	465	0.96 (0.83–1.10)	1517	1.16 (1.07–1.26)	0.01
			Combined	**9007**	**704**	**1.18 (0.71–1.64)**	**2466**	**1.12 (1.02–1.21)**	**0.84**

[a] Minor allele frequency over all controls in PBCS.

[b] Case-only P-value was used to test for heterogeneity (Phet) and was estimated using a polytomous logistic regression model with E-cadherin (E-cad) status as the outcome adjusted for age in 5-year categories.

[c] Major/minor allele. Fixed effects meta-analysis by study (PBCS and SEARCH) was used to calculate the combined OR, 95% CI and Phet.

Table 2

Per-allele OR and 95% CI for the association of 1p11.2 (rs11249433 allele G) and breast cancer risk by E-cadherin tumor expression in PBCS and SEARCH studies

		Case-control analyses						Case Only
Tumor Subtype	**Study**	**N**	**OR (95% CI)**	***P*-value**	**N**	**OR (95% CI)**	***P*-value**	***P*-het[a]**
		E-cad low tumors vs. controls			E-cad high tumors vs. controls			
ER positive	PBCS	153	1.48 (1.18–1.85)	0.001	660	1.06 (0.94–1.20)	0.35	0.008
ER positive	SEARCH	339	1.35 (1.17–1.57)	<0.0001	1015	1.14 (1.05–1.25)	0.003	0.05
ER positive	**Combined**	**492**	**1.38 (1.21–1.56)**	**<0.0001**	**1675**	**1.11 (1.03–1.19)**	**0.003**	**0.001**
Lobular	PBCS	123	1.40 (1.13–1.74)	0.002	100	0.99 (0.84–1.16)	0.90	0.01
Lobular	SEARCH	177	1.41 (1.17–1.70)	<0.0001	124	1.12 (0.93–1.35)	0.23	0.09
Lobular	**Combined**	**300**	**1.41 (1.20–1.61)**	**<0.0001**	**224**	**1.04 (0.91–1.33)**	**0.39**	**0.09**
Ductal	PBCS	71	1.17 (0.91–1.51)	0.22	616	1.07 (0.96–1.20)	0.21	0.50
Ductal	SEARCH	240	1.22 (1.03–1.45)	0.02	1282	1.10 (1.01–1.20)	0.03	0.31
Ductal	**Combined**	**311**	**1.21 (1.03–1.38)**	**0.07**	**1898**	**1.09 (1.02–1.16)**	**0.01**	**0.63**
ER positive/lobular	PBCS	92	1.30 (0.92–1.83)	0.001	78	1.07 (0.94–1.22)	0.70	0.25
ER positive/lobular	SEARCH	147	1.30 (1.07–1.57)	0.001	101	1.13 (1.03–1.25)	0.14	0.23
ER positive/lobular	**Combined**	**239**	**1.45 (1.22–1.67)**	**<0.0001**	**179**	**1.08 (0.94–1.22)**	**0.29**	**0.17**
ER positive/ductal	PBCS	25	1.54 (1.20–1.97)	0.14	371	1.03 (0.87–1.23)	0.29	0.007
ER positive/ductal	SEARCH	156	1.40 (1.15–1.70)	0.01	839	1.16 (1.00–1.42)	0.01	0.22
ER positive/ductal	**Combined**	**181**	**1.30 (1.08–1.52)**	**0.01**	**1210**	**1.11 (1.03–1.19)**	**0.01**	**0.10**

Case-only P-value was used to test for heterogeneity (P-het) and was estimated using a polytomous logistic regression model with E-cadherin status as the outcome adjusted for age in 5-year categories and study stratum (Warsaw, ŁódŸ, SEARCH). Fixed effects meta-analysis by study (PBCS and SEARCH) was used to calculate the combined OR, 95% CI and P-het.