Modern Pathology (2011) 24, 64–81

P-cadherin expression as a prognostic biomarker in a 3992 case tissue microarray series of breast cancer

Gulisa Turashvili[1], Steven E McKinney[1], Ozge Goktepe[1], Samuel C Leung[2], David G Huntsman[2], Karen A Gelmon[3], Gerrit Los[4], Paul A Rejto[4] and Samuel AJR Aparicio[1]

[1]*Molecular Oncology Department, BC Cancer Research Centre, Vancouver, BC, Canada;* [2]*Genetic Pathology Evaluation Centre, University of British Columbia, Vancouver, BC, Canada;* [3]*Medical Oncology, BC Cancer Agency, Vancouver, BC, Canada and* [4]*Oncology Research Unit, Pfizer Worldwide Research and Development, San Diego, CA, USA*

P-cadherin is a calcium-dependent cell–cell adhesion glycoprotein. P-cadherin expression is restricted to the myoepithelial cells in normal breast tissue, and aberrant staining has also been described in invasive tumors. Several small studies have reported P-cadherin as a marker of poor outcome in breast cancer patients but its prognostic significance in relation to other variables has not been established in a large series of breast cancers. A tissue microarray was constructed from 3992 cases of invasive breast carcinoma, and P-cadherin expression was evaluated using immunohistochemistry. Median follow-up was 12.5 years. The immunohistochemistry-based definitions of cancer subtypes were luminal (ER+ or PR+/HER2−), luminal/HER2+ (ER+ or PR+/HER2+), HER2+ (ER−/PR−/HER2+), and basal (ER−/PR−/HER2−/CK5/6+ or EGFR+). Clinical covariate and biomarker associations were assessed using contingency tables, and Pearson's χ^2 or Fisher's exact test. Survival associations were assessed using Kaplan–Meier plots, logrank and Breslow tests, and Cox proportional hazards regression analysis. P-cadherin was expressed in 34.8% (1290/3710, 50% cut point) of cases. P-cadherin staining was strongly associated with HER2+ and basal carcinoma subtypes ($P<0.0005$). P-cadherin-positive patients showed significantly poorer short-term (0–10 years) overall survival, disease-specific survival, distant relapse-free interval, and locoregional relapse-free interval in univariable models ($P<0.05$). In multivariable Cox models containing standard clinical covariates and cancer subtypes, P-cadherin did not show independent prognostic value. P-cadherin expression was positively associated with histological grade, chemotherapy, Ki-67, EGFR, CK5/6, p53, YB-1, and HER2 expression ($P<0.002$), and negatively associated with age at diagnosis, ER, PR, and Bcl-2 expression ($P<0.0005$). This study shows the value of P-cadherin as a marker of poor prognosis. The large sample size of this series clarifies contradictory findings of many smaller studies. P-cadherin positivity is associated with high-grade tumor subtypes and well-established markers of poor prognosis, and may represent a promising antibody therapeutic target.
Modern Pathology (2011) **24,** 64–81; doi:10.1038/modpathol.2010.189; published online 17 September 2010

Keywords: breast cancer; immunohistochemistry; P-cadherin; prognosis; tissue microarray

Cadherins are a family of transmembrane glycoproteins involved in calcium-dependent cell–cell adhesion in many tissues.[1,2] Although these proteins are similar in their domain structure, calcium and protease sensitivity, and molecular weight, they have distinct tissue expression patterns and immunological reactivity. P-cadherin is localized in placenta, whereas E-cadherin and N-cadherin are found in epithelial and neural tissues, respectively.[3–5] P-cadherin has been shown to be crucial for orderly progression of terminal differentiation of the epidermis,[6] and P-cadherin-mediated adhesion seems to determine mammary gland growth control and maintenance of an undifferentiated state in

Correspondence: Professor SAJR Aparicio, BM, BCh, PhD, FRCPath, Molecular Oncology Department, Nan and Lorraine Robertson Chair UBC/BC Cancer Agency, 675 West 10th Avenue, Vancouver, BC V5Z-1L3, Canada.
E-mail: saparicio@bccrc.ca or http://molonc.bccrc.ca
Received 28 June 2010; revised 2 August 2010; accepted 4 August 2010; published online 17 September 2010

embryogenesis.[7] E-cadherin and P-cadherin have important roles in the architecture, function, and differentiation of the normal adult non-lactating mammary gland. E-cadherin is expressed by both luminal and myoepithelial cells, whereas P-cadherin is restricted to myoepithelial/basal cells in normal breast tissue.[5,8,9] During late pregnancy and lactation, P-cadherin is secreted by epithelial cells, and its soluble fragment is frequently found in human milk.[10,11] P-cadherin can also be detected in nipple aspirate fluid,[12] and serum from breast cancer patients.[13] Aberrant expression measured by P-cadherin cDNA has been reported in about 30% of known mammary cancer cell lines.[5] Aberrant P-cadherin protein has also been found in a minority of ductal carcinoma *in situ*,[14] and 4–71% of invasive breast carcinomas.[15–18] Increased P-cadherin expression has been reported to be associated with reduced expression of E-cadherin and high histological grade,[15] estrogen and progesterone receptor negative (ER−/PR−) tumors,[9,15,17,19–22] increased expression of epidermal growth factor receptor (EGFR), oncoprotein p53, HER2, and proliferation marker Ki-67.[15,19,20] Several small studies have reported that P-cadherin is a marker of poor patient survival over short-term follow-up.[9,15,17,19,20,22]

Recently, microarray profiling of breast tumors has identified five distinct tumor subtypes (luminal A, luminal B, normal breast like, HER2 overexpressing, and basal like) that are associated with different clinical outcomes.[23] Basal-like carcinoma is characterized by cytokeratin 5 (CK5) and cytokeratin 17 (CK17) expression, and the basal epithelial cell-enriched gene cluster that also includes P-cadherin.[23] Immunohistochemistry can be used to identify basal-like carcinoma by staining for P-cadherin, in conjunction with p63 and CK5.[24–26] This subset of breast tumors usually lacks the expression of all three standard biomarkers (ER−/PR−/HER2−, ie triple-negative phenotype) and despite increased sensitivity to standard cytotoxic chemotherapy regimens, has a particularly unfavorable prognosis.[27] Efforts to identify targets for basal tumors, which may be used therapeutically or to further classify these tumors may have clinical utility.

In this study, we describe the immunohistochemical expression and prognostic value of P-cadherin protein in a large tissue microarray series constructed from 3992 primary breast carcinomas linked to treatment and outcome information. This large sample size allows us to provide stronger evidence for the role of P-cadherin compared with previous publications involving much smaller patient cohorts.

Materials and methods

This study cohort included 3992 female patients with primary invasive breast carcinoma diagnosed in 1986–1992 and referred to the British Columbia Cancer Agency for treatment. The clinicopathological characteristics and the treatment strategies of the patients included in this study have previously been reported.[28] The median follow-up was 12.5 years and the median age at diagnosis was 60 years. Abstracted clinical information included age, menstrual status, histological type and grade of tumor, clinical and pathological TNM stage, status of final surgical margin at diagnosis, tumor size, number of involved axillary lymph nodes, type of local and initial adjuvant systemic therapy, dates of diagnosis, and first locoregional or distant recurrence and death. HER2 fluorescent *in situ* hybridization data and immunohistochemistry scores were also available for the following biomarkers: ER, PR, Ki-67, Bcl-2, HER2, EGFR, CK5/6, keratin 5 (KRT5), p53, YB-1, and E-cadherin, as previously published.[29–35] Ethical approval for the study was obtained from the Clinical Research Ethics Board of the University of British Columbia and the British Columbia Cancer Agency.

Previously frozen breast cancer tissue samples were fixed in 10% neutral buffered formalin and embedded in paraffin. Representative areas of invasive carcinoma were selected and marked on the hematoxylin and eosin stained slides, and their corresponding tissue blocks to be sampled for tissue microarray. Seventeen single core (0.6 mm size) tissue microarray blocks were then assembled using a manual arrayer (Beecher Instruments, Silver Springs, MD, USA) as previously described.[36] From each tissue microarray block, 4 μm thick sections were cut and immunostained on a Ventana Discovery XT staining system (Ventana Medical Systems, AZ, USA). Sections were deparaffinized in xylene, dehydrated through three alcohol changes, and transferred to Ventana Wash solution. Endogenous peroxidase activity was blocked in 3% hydrogen peroxide. Antigen retrieval was performed in mild cell conditioner 1 and slides were incubated with anti-P-cadherin mouse monoclonal antibody (1:20 dilution, clone 56, BD Transduction, ON, Canada) for 32 min. Finally, sections were incubated with the pre-diluted Ventana Universal Secondary Antibody and DAB Map detection system, counterstained with hematoxylin, dehydrated, cleared, and mounted. Optimization of immunohistochemical protocol involved three different antigen-retrieval conditions and a serial dilution of the antibody to establish the optimal staining concentration. Although freezing of the tissue samples prior to formalin fixation could have potentially affected the immunoreactivity for P-cadherin, appropriate negative and positive controls were performed to ensure the quality and adequacy of staining. The negative control was performed by omission of the primary antibody, and myoepithelial/basal cells were used as an internal positive control.

P-cadherin expression was scored visually based on the determination of staining intensity

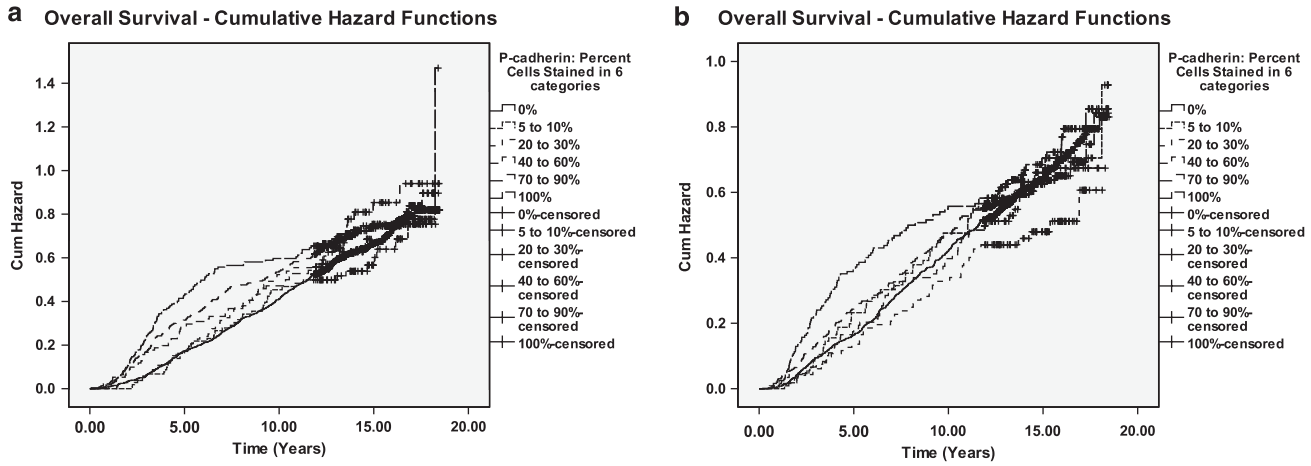

Figure 1 Changing hazard structure: hazard curves show marked change between 5 and 10 years in training (**a**) and validation (**b**) sets, 2004 censor date.

Table 1 Distribution of the percentage of P-cadherin-positive cells and P-cadherin-staining intensity

P-cadherin measure	*Training set no. (%)*	*Validation set no. (%)*	*Whole series no. (%)*
P-cadherin-staining intensity, categorized			
0	953 (50.7)	905 (49.4)	1858 (50.1)
1+	246 (13.1)	252 (13.8)	498 (13.4)
2+	285 (15.2)	290 (15.8)	575 (15.5)
3+	395 (21.0)	384 (21.0)	779 (21.0)
P-cadherin-staining intensity, binarized			
Negative (0)	953 (50.7)	905 (49.0)	1858 (50.1)
Positive (1+, 2+, 3+)	926 (49.3)	926 (50.6)	1852 (49.9)
Percentage of P-cadherin-positive cells, binarized (≤50% vs >50%)			
Negative (0–50%)	1222 (65.0)	1198 (65.4)	2420 (65.2)
Positive (51–100%)	657 (35.0)	633 (34.6)	1290 (34.8)
Percentage of P-cadherin-positive cells, binarized (0% vs any percent positive)			
Negative (0%)	953 (50.7)	905 (49.4)	1858 (50.1)
Positive (1–100%)	926 (49.3)	926 (50.6)	1852 (49.9)
Missing	124	158	282
Total	2003	1989	3992

(0 = negative, 1 = weak, 2 = moderate, 3 = strong) and percentage of cells with membranous and/or cytoplasmic staining (0–100%), as previously reported.[37–39] Scores were entered into a standardized Excel worksheet with a sector map matching each tissue microarray section. Biomarker information was considered uninterpretable if there were no tumor cells in the cores or the cores were missing. Original scoring grids were converted to tables using Deconvoluter 1.10[40] and combined in a single text file with TMA-Combiner 1.00.[41] The resulting text files were imported into SPSS 15.0 and R2.4.0 for Windows.[42]

The hematoxylin and eosin and P-cadherin immunohistochemistry images and scores of all cores used in this study are publicly available at the companion site (http://www.gpecimage.ubc.ca; username: pcad; password: abc123). This site was constructed using the Genetic Pathology Evaluation Centre database and image viewers provided by Olympus America. All slides were scanned with the BLISS scanner (Olympus America, Center Valley, PA, USA; Bacus Laboratories, Lombard, IL, USA).

The expanded surrogate immunopanel of ER, PR, HER2, Ki-67, EGFR, and CK5/6 was used to define five major biologically distinct immunohistochemical subtypes of breast cancer: (1) Luminal A—ER+ or PR+ and HER2− and Ki-67−; (2) Luminal B—ER+ or PR+ and HER2− and Ki-67+. Ki-67 positivity was defined based on a 14% cut point; (3) Luminal/HER2+—HER2+ and ER+ or PR+; (4) HER2+—HER2+ and ER− and PR−; and (5) Basal—this subtype was defined differently by two classification schemes: the triple-negative phenotype and the five-biomarker method as published previously.[28] Using the triple-negative phenotype method, basal-like carcinoma is triple-negative (ER−/PR−/HER2−). Using the five-biomarker method, triple-

Table 2 Clinicopathological characteristics of breast cancer patients

Characteristics	*Training set no. cases (%)*	*Validation set no. cases (%)*	*Whole series no. cases (%)*
Age at diagnosis (years)			
<40	151 (7.5%)	143 (7.2%)	294 (7.4%)
40–49	412 (20.6%)	432 (21.7%)	844 (21.1%)
50–65	697 (34.8%)	728 (36.6%)	1425 (35.7%)
>65	743 (37.1%)	686 (34.5%)	1429 (35.8%)
Menstrual status			
Premenopausal	594 (29.7%)	586 (29.5%)	1180 (29.5%)
Postmenopausal	1366 (68.2%)	1355 (68.1%)	2721 (68.2%)
Unknown	43 (2.1%)	48 (2.4%)	91 (2.3%)
Histological type of tumor			
Ductal	1807 (90.2%)	1806 (90.8%)	3613 (90.5%)
Lobular	160 (8.0%)	143 (7.2%)	303 (7.6%)
Other	36 (1.8%)	40 (2.0%)	76 (1.9%)
Histological grade			
Grade 1	119 (5.9%)	90 (4.5%)	209 (5.2%)
Grade 2	798 (39.8%)	765 (38.5%)	1563 (39.2%)
Grade 3	991 (49.5%)	1049 (52.7%)	2040 (51.1%)
Unknown	95 (4.7%)	85 (4.3%)	180 (4.5%)
Tumor size, cm			
≤2	1040 (51.9%)	1038 (52.2%)	2078 (52.1%)
>2–5	850 (42.4%)	817 (41.1%)	1667 (41.8%)
>5	103 (5.1%)	118 (5.9%)	221 (5.5%)
Unknown	10 (0.5%)	16 (0.8%)	26 (0.7%)
Percentage of positive/total number of examined axillary lymph nodes			
0%	1051 (52.5%)	1037 (52.1%)	2088 (52.3%)
1–25	378 (18.9%)	399 (20.1%)	777 (19.5%)
>25%	443 (22.1%)	436 (21.9%)	879 (22.0%)
Unknown	131 (6.5%)	117 (5.9%)	248 (6.2%)
Lymphovascular invasion			
Positive	814 (40.6%)	896 (45.0%)	1710 (42.0%)
Negative	1090 (54.4%)	1016 (51.1%)	2106 (52.8%)
Unknown	99 (4.9%)	77 (3.9%)	176 (4.4%)
Local therapy			
No breast surgery	30 (1.5%)	30 (1.5%)	60 (1.5%)
Mastectomy ± any radiotherapy	1086 (54.2%)	1058 (53.2%)	2144 (53.7%)
Breast-conserving surgery+any radiotherapy	817 (40.8%)	837 (42.1%)	1654 (41.4%)
Breast-conserving surgery alone	70 (3.5%)	64 (3.2%)	134 (3.4%)
Initial systemic therapy			
No systemic therapy	846 (42.2%)	830 (41.7%)	1676 (42.0%)
Tamoxifen, no chemotherapy	643 (32.1%)	633 (31.8%)	1276 (32.0%)
Chemotherapy, no hormonal therapy	357 (17.8%)	370 (18.6%)	727 (18.2%)
Chemotherapy+tamoxifen	151 (7.5%)	146 (7.3%)	297 (7.4%)
Other	6 (0.3%)	10 (0.5%)	16 (0.4%)
Unknown	1 (50.0%)	1 (50.0%)	2 (100.0%)
Clinical T stage			
T0	7 (0.4%)	8 (0.4%)	15 (0.4%)
T1	946 (48.5%)	977 (50.5%)	1923 (49.5%)
T2	877 (45.0%)	821 (42.4%)	1698 (43.7%)
T3	68 (3.5%)	61 (3.2%)	129 (3.3%)
T4	53 (2.7%)	68 (3.5%)	121 (3.1%)
Clinical N stage			
N0	1707 (87.3%)	1718 (88.7%)	3425 (88.0%)
N1	231 (11.8%)	194 (10.0%)	425 (10.9%)
N2	18 (0.9%)	24 (1.2%)	42 (1.1%)
N3	0 (0.0%)	1 (0.1%)	1 (0.0%)

Table 2 Continued

Characteristics	*Training set no. cases (%)*	*Validation set no. cases (%)*	*Whole series no. cases (%)*
Clinical M stage			
M0	1931 (96.4%)	1919 (96.5%)	3850 (96.4%)
M1	0 (0.0%)	0 (0.0%)	0 (0.0%)
Missing	72 (3.6%)	70 (3.5%)	142 (3.6%)
Pathological T stage			
T0	0 (0.0%)	2 (0.1%)	2 (0.1%)
T1	957 (53.4%)	942 (53.5%)	1899 (53.4%)
T2	736 (41.0%)	693 (39.4%)	1429 (40.2%)
T3	69 (3.8%)	81 (4.6%)	150 (4.2%)
T4	31 (1.7%)	43 (2.4%)	74 (2.1%)
Pathological N stage			
N0	1076 (56.8%)	1049 (55.5%)	2125 (56.2%)
N1	780 (41.2%)	808 (42.8%)	1588 (42.0%)
N2	37 (2.0%)	30 (1.6%)	67 (1.8%)
N3	1 (0.1%)	2 (0.1%)	3 (0.1%)
Pathological M stage			
M0	417 (20.8%)	399 (20.1%)	816 (20.4%)
M1	0 (0.0%)	0 (0.0%)	0 (0.0%)
Missing	1586 (79.2%)	1590 (79.9%)	3176 (79.6%)
Total	**2003**	**1989**	**3992**

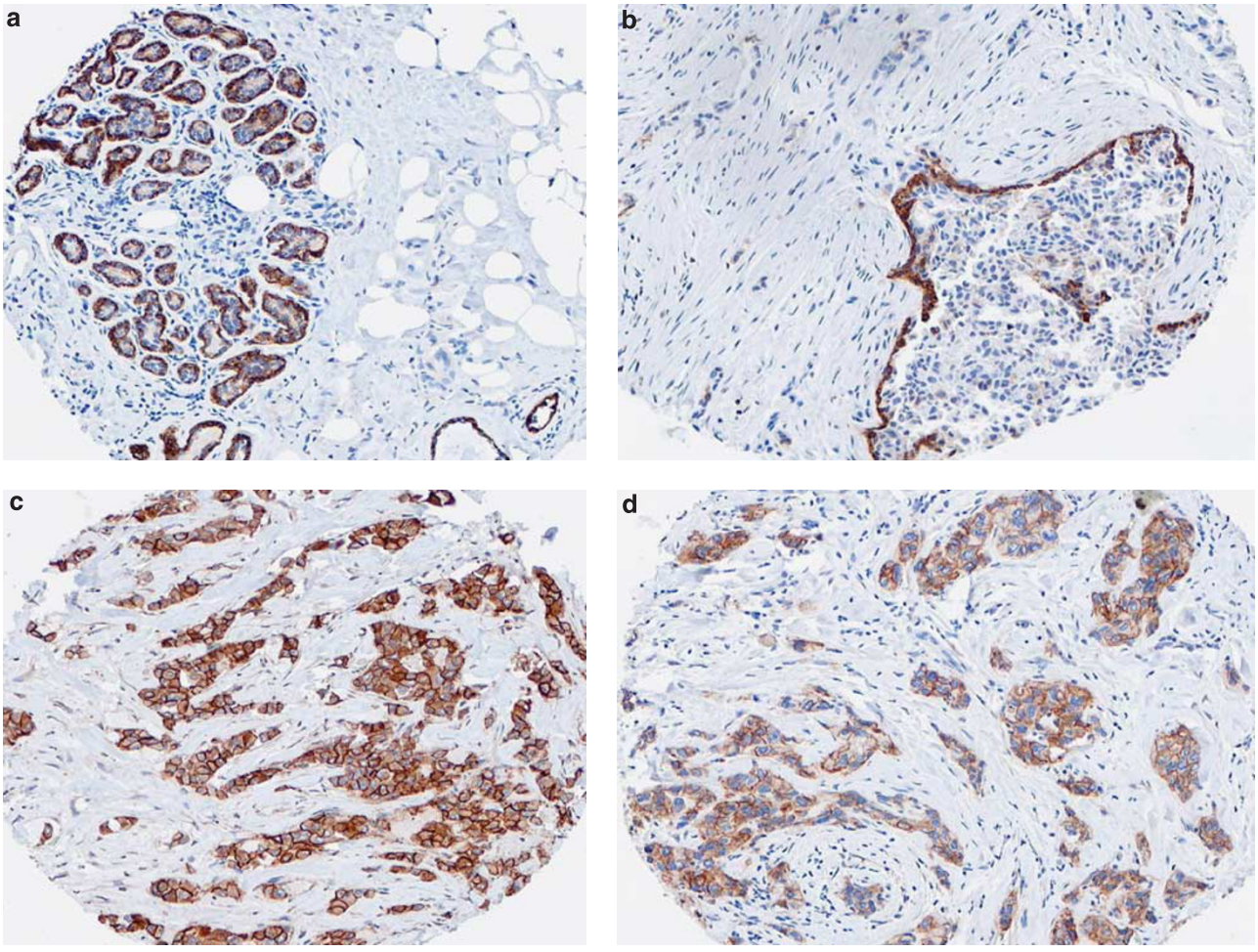

Figure 2 P-cadherin expression in myoepithelial/basal cells in normal terminal duct lobular units (**a**), and ductal carcinoma *in situ* (**b**), and in tumor cells in invasive ductal carcinoma (**c** and **d**). X200.

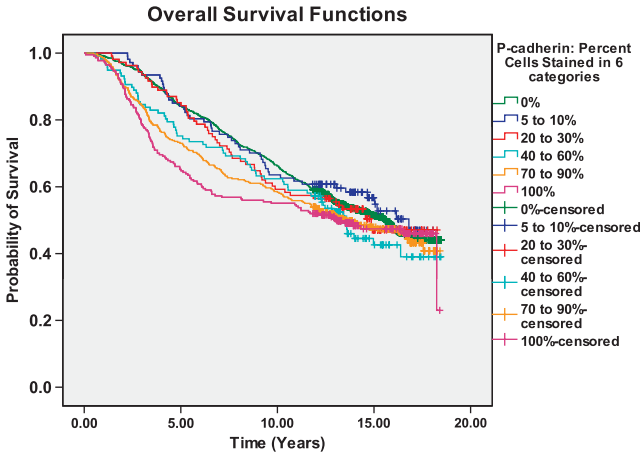

Figure 3 Kaplan–Meier overall survival curves for the percentage of P-cadherin-positive cells in the training set, 2004 censor date.

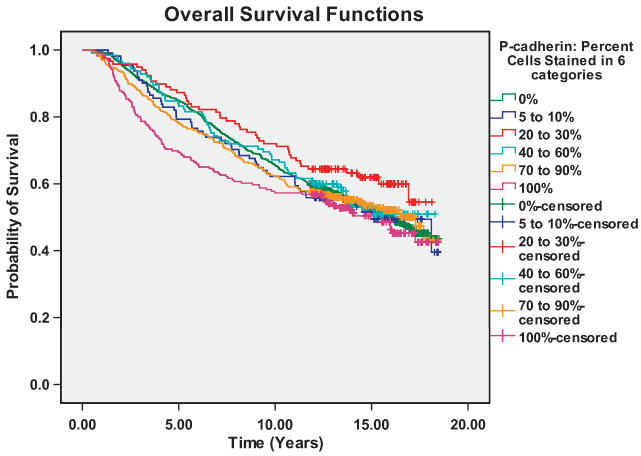

Figure 4 Kaplan–Meier overall survival curves for the percentage of P-cadherin-positive cells in the validation set, 2004 censor date.

negative phenotype can be divided into two groups: (1) Core Basal—triple-negative subset, which also expresses EGFR or CK5/6 and (2) five-marker negative phenotype—ER−/PR−/HER2−/EGFR−/CK5/6−. Tumors missing any of ER, PR, or HER2 data were categorized as unassigned.

Statistical analysis was performed using SPSS 17.0 and R-2.10.0. Clinical covariate and biomarker associations were assessed using contingency tables, and significance of associations was determined using Pearson's χ^2 or Fisher's exact test. Survival and relapse interval associations were graphically assessed using Kaplan–Meier plots. Significance of time-to-event associations was assessed using log-rank and Breslow tests and Cox proportional hazards regression analysis. Survival and relapse interval end point groups include overall survival, disease-specific survival, distant relapse-free interval, and locoregional relapse-free interval. Cox proportional hazards models were used to calculate adjusted hazard ratios to account for covariates of known clinical relevance.

Complete survival data for this tissue microarray series was obtained as of 30 June 2004. One of the complexities of working with such a large series with long-term follow-up data, noted in previous Cox modeling exercises using this series, was the changing structure in the hazard functions for breast cancer subtypes and subsets of other variables, between 5 and 10 years after diagnosis. This is readily seen in the hazard function plots in Figure 1. Hazards are proportional over the first 5–8 years, then exhibit a shift and tend to converge by 15–20 years. This long-term shift in hazard structure yields a decrease in power in testing for associations with a Cox model, as its basic proportional hazards requirement is violated. To mitigate this violation of the proportional hazards assumption, we formed an additional survival record, with data followed completely through 30 June 1999. Censoring data at an earlier time point to restrict analysis to a period in which hazards exhibit proportional structure represents a simple and effective strategy for handling this issue of changing hazard structure across long periods of time.[43] Survival analysis results thus reveal short-term and long-term associations that can differ, and both should be considered when assessing associations. Shorter-term (0–10 years) associations will be better assessed with the Breslow statistic in Kaplan–Meier analysis and with statistics from Cox models fitted to the data censored in 1999. Longer-term associations are better assessed via the logrank statistic in Kaplan–Meier analysis, and with Cox models fitted to the data censored in 2004. This strategy was proposed and implemented prior to any data analysis. We used a split-sample validation technique for statistical analysis, as described previously.[44,45] In brief, a large data collection ($n=3992$) was randomly split into a 'training' set ($n=2003$) and a 'validation' set ($n=1989$). Although it is common practice in studies of prognostic markers to conduct exploratory analyses with the training set, and take only a selected subset of analyses onward to the validation set, such strategies tend to result in the overreporting of positive findings and an underreporting of negative findings. To avoid the resultant reporting bias associated with this analysis strategy,[46] we prespecified a set of hypotheses, all of which were evaluated in the training and validation sets. Concordant findings in the training and validation sets, whether negative or positive, are reported here.

Results

Association of P-cadherin Expression with Survival

Out of 3710 interpretable cases on the tissue microarrays, P-cadherin was positive (50% cut point) in 1290 patients (34.8%), including 657/1879 (35%) cases in the training set and 633/1831 (34.6%) cases in the validation set (Table 1).

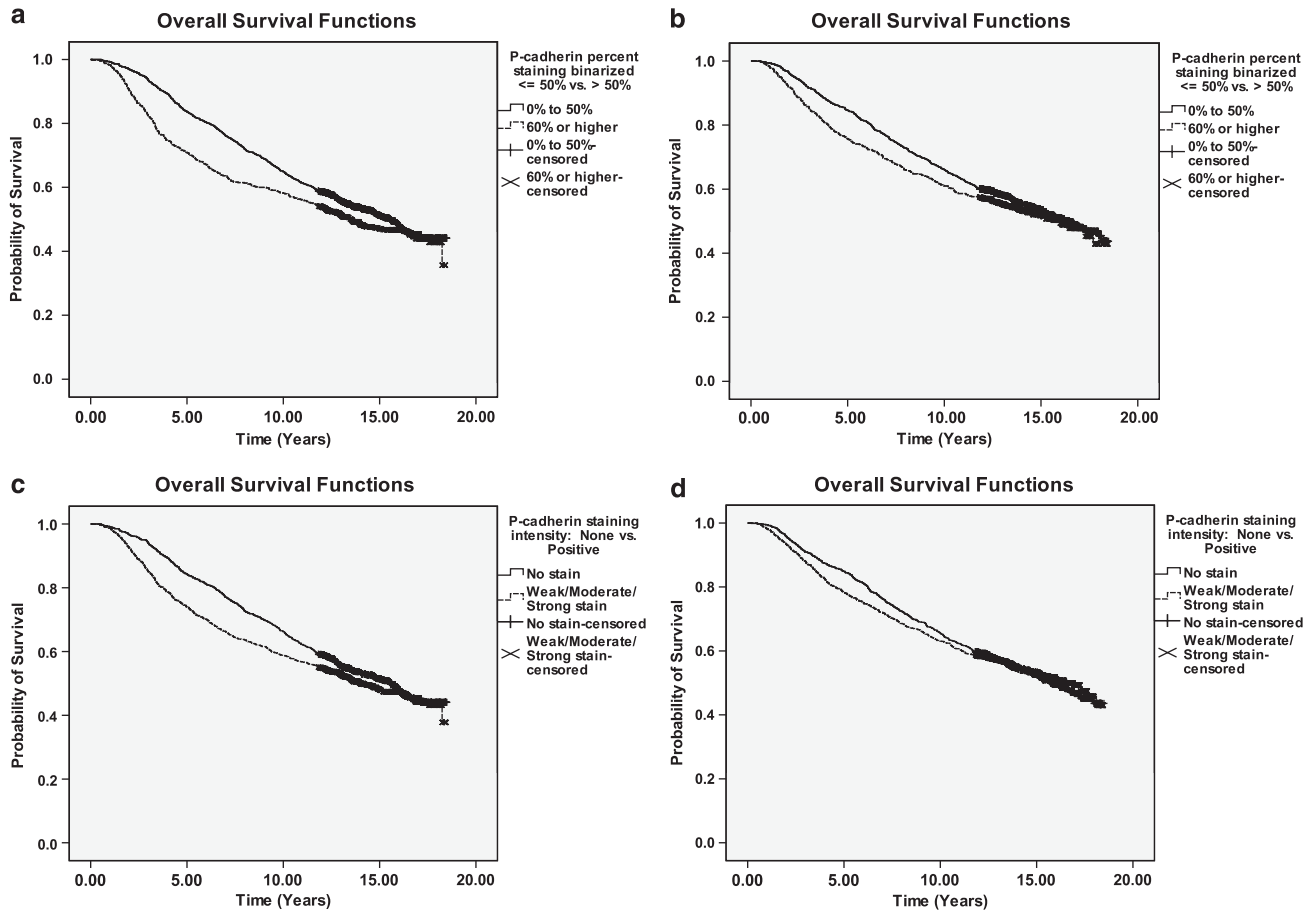

Figure 5 Univariable overall survival curves for (1) Percentage of P-cadherin-positive cells, binarized to groups 0–50% (negative) *vs* 60–100% (positive) in training (**a**) and validation (**b**) sets; (2) P-cadherin-staining intensity, binarized to groups {0} (negative) *vs* {1+,2+,3+} (positive), in training (**c**) and validation (**d**) sets, 2004 censor date.

Clinicopathological characteristics of breast cancer patients are summarized in Table 2, and distribution of P-cadherin expression according to these clinical covariates is shown in Supplementary Table 1. Representative photomicrographs of P-cadherin immunostaining are shown in Figure 2.

Binarization

The percentage of P-cadherin-positive cells and P-cadherin-staining intensity appeared linearly related to survival in the training and validation sets, so binarizing using a single cut point was reasonable (Figures 3 and 4). Maximum prognostic value as assessed by model fit Akaike information criteria was achieved in the training set by binarizing the percentage of P-cadherin-positive cells at a cut point of $\leq 30\%$, with $>30\%$ P-cadherin expressing cells defining P-cadherin positivity (Supplementary Table 2). In the validation set, this cut point did not validate. The lowest Akaike information criteria was achieved at 70%; however, the difference in Akaike information criteria values across a range of cut points was small: Akaike information criteria values from 30 to 90% ranged from 12 402.52 to 12 404.49. Thus, although no single cut point value was strongly suggested by these data, for validation and publication purposes, the cut point value of 50% ($\leq 50\%$ *vs* $>50\%$) was chosen to facilitate comparisons between the training and validation analysis results and publications in the literature,[9,37-39,47] many of which use a 50% cut point value. Staining intensity did not appear to improve this rule in either set (Figure 5c and d). Cox models with various combinations of the percentage of P-cadherin-positive cells and P-cadherin-staining intensity, recapitulating 'h-scores' or 'q-scores' commonly used in other P-cadherin studies,[15,38,39,47,48] showed no improvement over the percentage of P-cadherin-positive cells binarization rule as measured by Akaike information criteria or likelihood ratio χ^2 tests (Supplementary Table 3).

Survival analysis

Univariable survival analysis showed that P-cadherin-positive patients had significantly poorer short-term (0–10 years) overall survival (logrank for 1999 and Breslow for 2004 and 1999 censorings, all $P<0.05$), whereas long-term overall survival (15+ years) was not significantly different (logrank for 2004 censoring) (Figure 5). P-cadherin expression

Table 3 Distribution of P-cadherin expression (50% cut point) within breast cancer subtypes

Tumor subtype	*Percentage of P-cadherin-positive cells*	*Training set no. cases (%)*	*Validation set no. cases (%)*	*Whole series no. cases (%)*
Luminal NOS (ER+ or PR+, HER2+)	P-cadherin negative	87 (74.4%)	80 (76.2%)	167 (75.2%)
	P-cadherin positive	30 (25.6%)	25 (23.8%)	55 (24.8%)
	Uninterpretable	9	13	22
Luminal A (ER+ or PR+, HER2−, Ki-67−)	P-cadherin negative	570 (78.2%)	560 (76.4%)	1130 (77.3%)
	P-cadherin positive	159 (21.8%)	173 (23.6%)	332 (22.7%)
	Uninterpretable	23	33	56
Luminal B (ER+ or PR+, HER2−, Ki-67+)	P-cadherin negative	294 (71.2%)	290 (74.9%)	584 (73.0%)
	P-cadherin positive	119 (28.8%)	97 (25.1%)	216 (27.0%)
	Uninterpretable	11	18	29
HER2+ (HER2+/ER−/PR−)	P-cadherin negative	36 (30.8%)	53 (41.4%)	89 (36.3%)
	P-cadherin positive	81 (69.2%)	75 (58.6%)	156 (63.7%)
	Uninterpretable	3	2	5
Luminal/HER2+ (ER+ or PR+, HER2+)	P-cadherin negative	65 (55.6%)	58 (58.6%)	123 (56.9%)
	P-cadherin positive	52 (44.4%)	41 (41.4%)	93 (43.1%)
	Uninterpretable	4	4	8
Core Basal (ER−/PR−/HER2−, EGFR+ or CK5/6+)	P-cadherin negative	30 (18.6%)	28 (17.5%)	58 (18.1%)
	P-cadherin positive	131 (81.4%)	132 (82.5%)	263 (81.9%)
	Uninterpretable	3	6	9
Five-marker negative phenotype (ER−/PR−/HER2−/EGFR−/CK5/6−)	P-cadherin negative	78 (53.4%)	71 (51.1%)	149 (52.3%)
	P-cadherin positive	68 (46.6%)	68 (48.9)	136 (47.7%)
	Uninterpretable	8	7	15
Unassigned (missing ER, PR or HER2 data)	P-cadherin negative	62 (78.5%)	58 (72.5%)	120 (75.5%)
	P-cadherin positive	17 (21.5%)	22 (27.5%)	39 (24.5%)
	Uninterpretable	63	75	138
Total	P-cadherin negative	1222 (65.0%)	1198 (65.4%)	2420 (65.2%)
	P-cadherin positive	657 (35.0%)	633 (34.6%)	1290 (34.8%)
	Uninterpretable	124	158	282
Total no. cases		**2003**	**1989**	**3992**

Percentages include scorable P-cadherin cases only.

was strongly associated with breast cancer subtypes ($P<0.0005$), two poor prognostic subtypes in particular, namely HER2+ and Core Basal, showed markedly higher overall rates of P-cadherin expression (Table 3). This association yielded an occurrence of 'Simpson's Paradox', wherein P-cadherin-positive cases showed significantly poorer overall survival in the whole cohort, yet showed less evidence of poorer overall survival within breast cancer subtypes, although some of this effect may also be due to smaller sample sizes within subtype groups.

Within breast cancer subtypes, P-cadherin-positive cases showed similar overall survival to P-cadherin-negative cases ($P>0.05$; Figure 6). P-cadherin did not show independent prognostic value in a multivariable Cox model containing: (1) breast cancer subtypes and P-cadherin scores (adjusting for breast cancer subtypes); (2) age, grade, tumor size, node status, and P-cadherin scores (adjusting for clinical covariates); and (3) breast cancer subtypes, clinical covariates, and P-cadherin scores (adjusting for clinical covariates and breast cancer subtypes) (Supplementary Tables 4–7). Parameter estimates and corresponding relative risk estimates from the multivariable Cox models for overall survival and disease-specific survival are shown in Table 4 and Supplementary Table 8. Univariable and multivariable models included all histological subtypes of breast cancer. No change in results was noted upon excluding the special type carcinomas.

Binarization of the percentage of P-cadherin-positive cells showed similar survival curve patterns for the other time-to-event end points. Disease-specific survival (Figure 7), distant relapse-free interval, and locoregional relapse-free interval, all validated for the short term as measured using the 1999 censor date data (logrank and Breslow statistics), and the 2004 censor date data (Breslow statistic) (Table 5). Univariable association of the percentage of P-cadherin-positive cells with time-to-event over the long term (15+ years) validated only for disease-specific survival (logrank statistic for 2004 censoring). Similarly, P-cadherin-positive cases as defined by univariable binarized P-cadherin-staining intensity data showed poorer short-term event rates for all time-to-event end points (Breslow

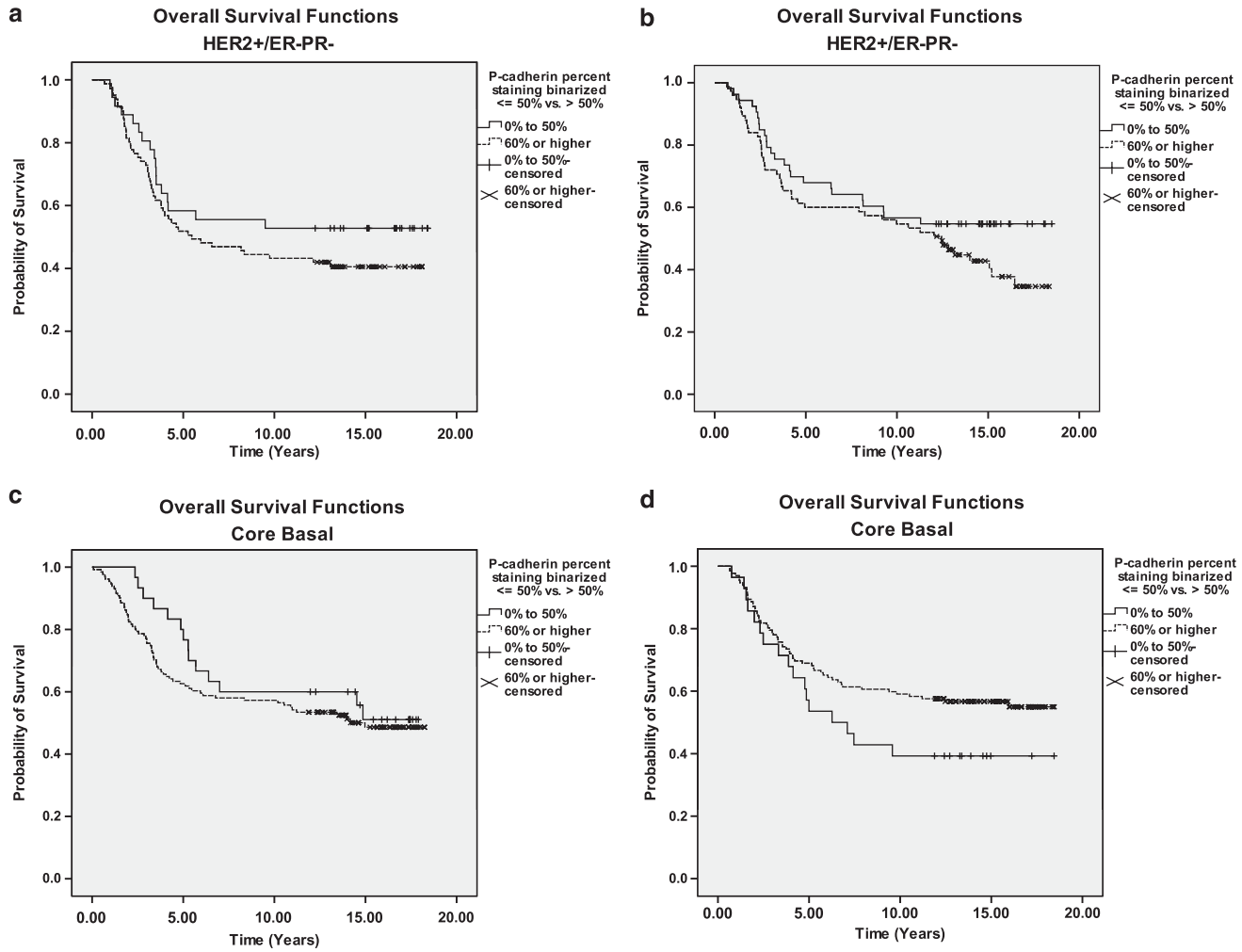

Figure 6 Univariable overall survival curves for (1) HER2+ breast carcinomas in the training (**a**) and validation (**b**) sets; (2) Core Basal carcinomas in the training (**c**) and validation (**d**) sets, 2004 censor date.

statistic for 1999 censoring, Table 5), though univariable association of P-cadherin-staining intensity with time-to-event over the long term did not validate for any time-to-event end point (logrank statistic for 2004 censoring).

Association of P-cadherin Expression with Clinicopathological and Immunohistochemical Variables

Herein, we describe only the findings observed in the training set that were validated in the validation set (ie showed a concordant statistical test of association outcome, statistically significant in both sets, or statistically not significant). P-cadherin expression differed according to the histological type of breast cancer. Invasive ductal carcinoma cases showed P-cadherin-positive rates equal to the overall P-cadherin-positive rate (35.5% (1195/3364) in the whole series), invasive lobular carcinoma showed lower P-cadherin-positive rates (16.2% (45/278) in the whole series) and other special type tumors showed higher P-cadherin-positive rates (73.5% (50/68) in the whole series). P-cadherin was positive in 44 out of 66 medullary carcinomas, 1 of 1 acinar cell carcinoma, 1 of 1 oat cell carcinoma, 1 of 1 adenoid cystic carcinoma, 1 of 2 adenosquamous carcinoma, and 2 of 2 adenocarcinomas with squamous metaplasia.

Both the percentage of P-cadherin-positive cells and P-cadherin-staining intensity were positively associated with histological grade, and negatively associated with age at diagnosis (Table 6). The percentage of P-cadherin-positive cells and P-cadherin-staining intensity showed association with breast cancer subtypes ($P<0.0005$), with HER2+ and basal subtypes of breast cancer showing elevated P-cadherin-positive rates and luminal subtypes showing reduced rates. In Core Basal carcinomas, P-cadherin was expressed (50% cut point) in 81.4% (131/161) in the training set and in 82.5% (132/160) in the validation set. In HER2+ carcinomas, P-cadherin was expressed in 69.2%

Table 4 Complete case analysis model fit results with relative risk estimates and 95% confidence intervals from the full multivariable Cox model for overall survival using 1999 censored data

Clinical variables	*Estimate, B*	*Standard error*	*Wald statistic*	*Degrees of freedom*	*Significance*	*Relative risk, Exp(B)*	*95.0% confidence interval for Exp(B)*	
							Lower	*Upper*
Age, years								
Age 40–49 *vs* ≤40	−0.225	0.122	3.431	1	0.064	0.798	0.629	1.013
Age 50–65 *vs* ≤40	0.036	0.111	0.102	1	0.749	1.036	0.833	1.289
Age >65 *vs* ≤40	0.548	0.110	24.665	1	0.000	1.729	1.393	2.146
Histological grade								
3 *vs* 1,2	0.386	0.062	38.451	1	0.000	1.471	1.302	1.661
Lymph node status								
Positive *vs* negative	0.638	0.058	120.551	1	0.000	1.892	1.688	2.120
Tumor size, cm								
>2–5 *vs* ≤2	0.413	0.060	47.290	1	0.000	1.511	1.343	1.699
>5 *vs* ≤2	0.674	0.112	36.430	1	0.000	1.961	1.576	2.441
Breast cancer subtype. luminal as reference								
Luminal/HER2+	0.512	0.136	14.220	1	0.000	1.668	1.279	2.176
HER2+/ER−/PR−	0.349	0.169	4.286	1	0.038	1.417	1.019	1.972
Core Basal	0.412	0.195	4.478	1	0.034	1.510	1.031	2.211
Five-marker negative phenotype	0.128	0.151	0.726	1	0.394	1.137	0.846	1.528
Unassigned	−0.269	0.188	2.055	1	0.152	0.764	0.529	1.104
Percentage of P-cadherin-positive cells								
>50% *vs* ≤50%	0.100	0.081	1.531	1	0.216	1.106	0.943	1.296
Percentage of P-cadherin-positive cells by breast cancer subtype interaction								
Luminal/HER2+	−0.245	0.214	1.308	1	0.253	0.783	0.515	1.191
HER2+/ER−/PR−	0.158	0.215	0.538	1	0.463	1.171	0.768	1.785
Core Basal	−0.044	0.228	0.037	1	0.847	0.957	0.612	1.497
Five-marker negative phenotype	0.000	0.220	0.000	1	0.999	1.000	0.650	1.540
Unassigned	0.580	0.327	3.147	1	0.076	1.786	0.941	3.391

Model fitted to 3523 cases from the whole series (3992 cases total) having complete data for the variables included.

(81/117) in the training set and 58.6% (75/128) in the validation set. The percentage of P-cadherin-positive cells and P-cadherin-staining intensity showed no association with tumor size, clinical and pathological T stage, and pathological N stage of the TNM system as well as the status of final surgical margin at diagnosis (Table 6). As the P-cadherin-positive cases were associated with a higher risk cancer, it was not surprising to find an association with receiving initial or subsequent chemotherapy ($P<0.0005$). However, no clinically useful conclusions can be drawn about the interaction with systemic therapy as the indications for treatment and the protocols used are not consistent with contemporary therapies.

The expression of other immunohistochemical biomarkers in relation to P-cadherin expression is summarized in Table 7. The percentage of P-cadherin-positive cells and P-cadherin-staining intensity were negatively associated with ER, PR, and Bcl-2 expression ($P<0.0005$), and positively associated with HER2 overexpression based on both immunohistochemistry and fluorescent *in situ* hybridization data ($P<0.002$), as well as with CK5/6, KRT5, EGFR, Ki-67, p53, and YB-1 ($P<0.0005$). P-cadherin expression showed no association with E-cadherin (data binarized at {0} *vs* {1,2}) (Table 8).

Discussion

P-cadherin is a calcium-dependent transmembrane glycoprotein in adherens-type junctions, mainly promoting homotypic interactions in epithelium.[1,2] P-cadherin expression is restricted to normal myoepithelial/basal cells, having an important role in the architecture, function, and differentiation of the normal adult breast.[5,7] P-cadherin has been detected in breast cancer cell lines, high-grade or basal-like ductal carcinoma *in situ*,[14,25] papillary lesions,[49] and 4–71% invasive breast carcinomas.[15–18] In this study, we report the immunohistochemical expression of P-cadherin protein in 34.8% of cases (1290/3710 at a cut point of 50%) in a large tissue microarray series consisting of 3992 primary breast carcinomas. For statistical analysis, our data were split into 'training'

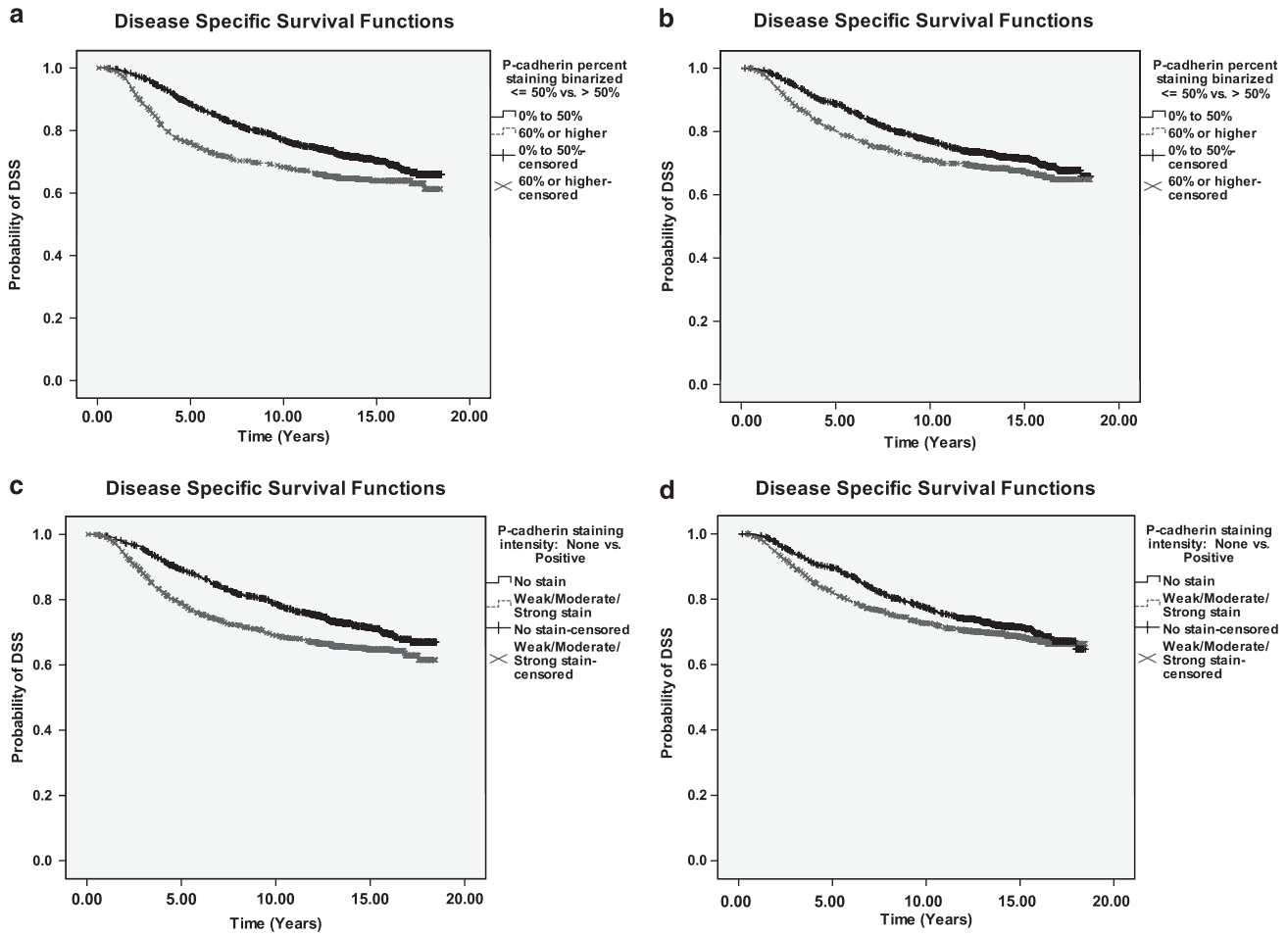

Figure 7 Univariable disease-specific survival curves for (1) Percentage of P-cadherin-positive cells, binarized to groups 0–50% (negative) *vs* 60–100% (positive), in training (**a**) and validation (**b**) sets; (2) P-cadherin-staining intensity, binarized to groups {0} (negative) *vs* {1+,2+,3+} (positive), in training (**c**) and validation (**d**) sets, 2004 censor date.

and 'validation' sets, as described previously.[44,45,50] Although maximum prognostic value was achieved in the training set by binarizing the percentage of P-cadherin-positive cells at a cut point of 30%, the validation data did not strongly support this cut point, indicating that analysis results were not heavily dependent on choice of cut point for any cut point greater than 0. Assessment of several binarization schemes using both the percentage of P-cadherin-positive cells and P-cadherin-staining intensity showed no prognostic improvement. This suggests that P-cadherin positivity is best defined using the percentage of positive cells without further consideration of the staining intensity, as suggested previously.[14,19] As <10%, 10–50%, and >50% cut points have been described in the literature,[37–39,47,51] the commonly used cut point of 50% has been used in this study to yield results more readily comparable to other studies.

Univariable survival analysis in our study showed that P-cadherin expression was associated with all time-to-event end points. P-cadherin-positive patients had significantly poorer short-term (0–10 years) overall survival, disease-specific survival, distant relapse-free interval, and locoregional relapse-free interval ($P<0.0005$). Univariable association of the percentage of P-cadherin-positive cells with long-term survival (15+ years) validated only for disease-specific survival (logrank statistic for 2004 censoring). Our study also showed that P-cadherin does not have independent prognostic value in multivariable Cox models when adjusting for clinical covariates, or breast cancer subtypes and clinical covariates. In contrast to our findings, P-cadherin has been reported as an independent prognostic factor, as it was associated with poorer survival in a Cox model including histological grade and lymph node status.[17] Peralta Soler and coworkers selected cases based on a minimum of a 5-year patient follow-up after surgery for living patients and breast carcinoma as the cause of death for those who died, thus introducing length-bias sampling (cases with at least 5 years follow-up will tend to be healthier)[52] and unknown bias in survival rate estimates.[53] They report association of P-cadherin-positive cases with ER−/PR− status,

Table 5 Survival analysis using univariable logrank and Breslow statistics on binarized percentage of P-cadherin-positive cells (a) and P-cadherin-staining intensity (b) data

Event	*Censoring date*	*Univariable $X^2_{(1)}$ Statistic*	*Training set*	*Validation set*	*Whole Series*
Percentage of P-cadherin-positive cells					
Overall survival	2004	Logrank[a]	$P=0.006$	$P=0.110$	$P=0.002$
		Breslow[b]	$P<0.0005$	$P=0.013$	$P<0.0005$
	1999	Logrank[b]	$P<0.0005$	$P=0.017$	$P<0.0005$
		Breslow[b]	$P<0.0005$	$P=0.001$	$P<0.0005$
Disease-specific survival	2004	Logrank[b]	$P<0.0005$	$P=0.017$	$P<0.0005$
		Breslow[b]	$P<0.0005$	$P=0.001$	$P<0.0005$
	1999	Logrank[b]	$P<0.0005$	$P=0.001$	$P<0.0005$
		Breslow[b]	$P<0.0005$	$P<0.0005$	$P<0.0005$
Distant relapse-free interval	2004	Logrank[a]	$P<0.0005$	$P=0.132$	$P<0.0005$
		Breslow[b]	$P<0.0005$	$P=0.023$	$P<0.0005$
	1999	Logrank[b]	$P<0.0005$	$P=0.038$	$P<0.0005$
		Breslow[b]	$P<0.0005$	$P=0.011$	$P<0.0005$
Locoregional relapse-free interval	2004	Logrank[a]	$P=0.098$	$P<0.0005$	$P<0.0005$
		Breslow[b]	$P=0.004$	$P<0.0005$	$P<0.0005$
	1999	Logrank[b]	$P=0.027$	$P<0.0005$	$P<0.0005$
		Breslow[b]	$P=0.002$	$P<0.0005$	$P<0.0005$
P-cadherin-staining intensity					
Overall survival	2004	Logrank[a]	$P=0.025$	$P=0.743$	$P=0.073$
		Breslow[a]	$P<0.0005$	$P=0.250$	$P=0.001$
	1999	Logrank[a]	$P<0.0005$	$P=0.122$	$P<0.0005$
		Breslow[b]	$P<0.0005$	$P=0.027$	$P<0.0005$
Disease-specific survival	2004	Logrank[a]	$P<0.0005$	$P=0.072$	$P<0.0005$
		Breslow[b]	$P<0.0005$	$P=0.008$	$P<0.0005$
	1999	Logrank[b]	$P<0.0005$	$P=0.003$	$P<0.0005$
		Breslow[b]	$P<0.0005$	$P<0.0005$	$P<0.0005$
Distant relapse-free interval	2004	Logrank[a]	$P<0.0005$	$P=0.193$	$P<0.0005$
		Breslow[b]	$P<0.0005$	$P=0.040$	$P<0.0005$
	1999	Logrank[a]	$P<0.0005$	$P=0.053$	$P<0.0005$
		Breslow[b]	$P<0.0005$	$P=0.014$	$P<0.0005$
Locoregional relapse-free interval	2004	Logrank[a]	$P=0.063$	$P<0.0005$	$P<0.0005$
		Breslow[b]	$P=0.003$	$P<0.0005$	$P<0.0005$
	1999	Logrank[b]	$P=0.007$	$P<0.0005$	$P<0.0005$
		Breslow[b]	$P=0.001$	$P<0.0005$	$P<0.0005$

[a]Discordant between training and validation sets.
[b]Validated findings showing association.

suggesting that in their study, P-cadherin status is an indicator of breast cancer subtype, though they report no interaction effect of P-cadherin with ER/PR status in a survival model. P-cadherin has also been reported to be an independent predictor of nodal positivity in a multivariable logistic model containing stage, grade, tumor type, menstrual status, and proliferation rate.[18] Our study strongly suggests that P-cadherin is not an independent prognostic marker, and is a surrogate marker for the basal-like profile.

Survival analysis in our study showed no difference between P-cadherin-positive and P-cadherin-negative cases within luminal A, luminal B, luminal/HER2+, HER2+, and basal subtypes of breast cancer ($P>0.05$). However, P-cadherin expression was strongly associated with HER2+ and basal carcinomas ($P<0.0005$), two of the five distinct molecular subtypes of breast cancer identified by gene expression profiling.[23] The HER2+ tumor subtype lacks the expression of hormonal receptors and has poor prognosis but can benefit from targeted therapy. Basal-like tumors have the triple-negative phenotype (ER−/PR−/HER2−), express basal cytokeratins, and do not benefit from hormonal therapy or trastuzumab. Although this histologically heterogeneous group of cancers includes poorly differentiated carcinomas and rare tumor types ranging from those with an excellent prognosis to aggressive metaplastic carcinomas,[54] it is widely accepted that these tumors are associated with particularly poor survival.[27] Immunohistochemistry surrogate panels have been proposed to identify basal-like breast cancer, including triple-negative phenotype alone,[55] triple-negative phenotype but EGFR+ or CK5/6+,[56,57] and a panel of P-cadherin, p63 and CK5 antibodies.[24–26] It has been argued that identification of basal-like breast cancers on the basis of gene expression profiling data has been misleading in some respects due to a lack of standardized technology and terminology.[54]

In our study, higher P-cadherin expression was observed in invasive ductal carcinoma and special type tumors other than lobular carcinoma, including 66.7% of medullary carcinomas ($P<0.0005$). Frequent expression of P-cadherin in metaplastic and medullary breast carcinomas[20,58–63] supports a

Table 6 Association of P-cadherin expression with clinicopathological variables

Variable	*P-cadherin-staining intensity*		*Percentage of P-cadherin-positive cells*	
	Training set	*Validation set*	*Training set*	*Validation set*
Clinical variables that show concordance between training and validation sets				
Age at diagnosis[a]	$P<0.0005$	$P=0.021$	$P<0.0005$	$P=0.015$
Age at diagnosis binarized at 50[a]	$P=0.001$	$P=0.039$	$P=0.001$	$P=0.009$
Histological type[a]	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$
Histological grade[a]	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$
Tumor size[b]	$P=0.306$	$P=0.246$	$P=0.528$	$P=0.064$
Clinical T stage[b]	$P=0.124$	$P=0.245$	$P=0.352$	$P=0.081$
Pathological T stage[b]	$P=0.417$	$P=0.140$	$P=0.605$	$P=0.126$
Pathological N stage[b]	$P=0.548$	$P=0.315$	$P=0.697$	$P=0.634$
Margin at initial diagnosis[b]	(Did not validate)		$P=0.220$	$P=0.140$
Intent of initial radiotherapy treatment plan established at the time of diagnosis[b]	$P=0.131$	$P=0.588$	$P=0.415$	$P=0.836$
Whether or not radiotherapy to the breast/chest wall and/or regional nodes was done as part of the initial treatment plan established at the time of diagnosis[b]	$P=0.070$	$P=0.633$	$P=0.075$	$P=0.459$
Whether or not the patient received chemotherapy treatment[a]	$P<0.0005$	$P=0.018$	$P<0.0005$	$P<0.0005$
The type of adjuvant chemotherapy treatment received (if any) by the patient[a]	(Did not validate)		$P<0.0005$	$P=0.026$
Whether or not the patient received hormonal treatment[b]	$P=0.199$	$P=0.518$	$P=0.060$	$P=0.327$
Whether or not an axillary node dissection was done as part of the initial treatment plan established at the time of diagnosis[b]	$P=0.083$	$P=0.700$	$P=0.795$	$P=1.000$
Most definitive breast surgery done as part of the initial treatment plan established at the time of diagnosis[b]	$P=0.124$	$P=0.685$	$P=0.761$	$P=0.773$
Breast cancer subtypes[a]	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$
Clinical variables that show discordance between training and validation sets				
Menstrual status	$P<0.0005$	$P=0.354$	$P=0.003$	$P=0.083$
Clinical N stage	$P=0.021$	$P=0.076$	$P=0.008$	$P=0.149$
Systemic therapy	$P<0.0005$	$P=0.235$	$P<0.0005$	$P=0.059$
The type of adjuvant hormonal treatment received (if any) by the patient	$P=0.005$	$P=0.390$	$P=0.004$	$P=0.492$

P-values reported are from the likelihood ratio χ^2 test for association in a two-way contingency table, unless otherwise noted.
[a]Validated findings showing association.
[b]Validated findings showing no association.

myoepithelial/basal-like transcriptomic program for these tumor subtypes. Some basal-like carcinomas are associated with BRCA1 mutations,[64,65] and BRCA1 mutation has been found to be associated with P-cadherin expression.[39,51,65–70] P-cadherin expression was also positively associated with histological grade in agreement with previous findings.[15] P-cadherin expression has been reported to be associated with the lack of hormonal receptors,[9,15,17,19–22] expression of EGFR,[71,72] HER2 and Ki-67.[15,19,20] In our study, P-cadherin expression was negatively associated with ER, PR, and Bcl-2 expression ($P<0.0005$), and positively associated with HER2 overexpression based on both immunohistochemistry and fluorescent *in situ* hybridization data ($P<0.002$), as well as with CK5/6, KRT5, EGFR, Ki-67, p53, and YB-1 ($P<0.0005$).

Many hypotheses have been proposed to explain the aberrant expression of P-cadherin in breast cancer cells including the oncofetal properties of P-cadherin protein,[15] its histogenetic origin in cap cells, acquisition of a stem cell like phenotype,[17,20] and misexpression caused by epithelial transformation.[5] *In vitro* manipulation of P-cadherin expression modulates the motility and invasive potential of breast cancer cell lines. In luminal MCF-7 breast cancer cells with wild-type E-cadherin, P-cadherin overexpression promoted cell invasion, motility and migration, and siRNA knockdown of P-cadherin significantly reduced the invasive potential of BT-20 cells.[73] By contrast, induction of P-cadherin in mesenchymal MDA-MB-231 cells increased aggregation and induced a partial switch from fibroblastic to epithelial morphology.[74] Loss of E-cadherin-mediated cell adhesion is thought to account for the characteristic dyscohesion of tumor cells and the single file infiltration pattern of invasive lobular carcinoma.[75] Most lobular carcinomas show no immunohistochemical reactivity for E-cadherin as a consequence of mutation and loss of heterozygosity of the E-cadherin gene, or methylation of the E-cadherin promoter[76,77] and E-cadherin immunostaining has been of value in the differential diagnosis of ductal and lobular carcinomas.[78] E-cadherin downregulation is a key molecular change occurring in epithelial mesenchymal transition, a process by which epithelial cells modulate their phenotype and acquire mesenchymal-like properties.[79,80]

Table 7 P-cadherin expression rates within biomarker subgroups

Biomarker	*No. P-cadherin positive/total no. (P-cadherin positive %)*		
	Training set	*Validation set*	*Whole series*
ER			
No nuclei stained or <1%	326/550 (59.3%)	313/542 (57.7%)	639/1092 (58.5%)
1–25% nuclei stained	50/164 (30.5%)	41/152 (27%)	91/316 (28.8%)
25–75% nuclei stained	142/579 (24.5%)	159/573 (27.7%)	301/1152 (26.1%)
>75% nuclei stained	135/571 (23.6%)	114/552 (20.7%)	249/1123 (22.2%)
PR			
No nuclei stained or <1%	388/828 (46.9%)	374/839 (44.6%)	762/1667 (45.7%)
1–25% nuclei stained	135/448 (30.1%)	102/389 (26.2%)	237/837 (28.3%)
25–75 % nuclei stained	62/228 (27.2%)	68/226 (30.1%)	130/454 (28.6%)
>75% nuclei stained	45/222 (20.3%)	59/241 (24.5%)	104/463 (22.5%)
HER2 immunohistochemistry			
0 (negative)	445/1256 (35.4%)	432/1226 (35.2%)	877/2482 (35.3%)
1+ (negative)	36/103 (35%)	44/112 (39.3%)	80/215 (37.2%)
2+ (equivocal)	29/59 (49.2%)	32/58 (55.2%)	61/117 (52.1%)
3+ (positive)	88/146 (60.3%)	169/140 (49.3%)	157/286 (54.9%)
Bcl-2 (60% cut point)			
Positive	324/1215 (26.7%)	297/1145 (25.9%)	621/2360 (26.3%)
Negative	299/551(54.3%)	302/580 (52.1%)	601/1131 (53.1%)
Ki-67 (14% cut point)			
Positive	355/743 (47.8%)	320/711 (45%)	675/1454 (46.4%)
Negative	234/905 (25.9%)	245/899 (27.3%)	479/1804 (26.6%)
p53			
<10% positive nuclei	422/1461 (28.9%)	413/1427 (28.9%)	835/2888 (28.9%)
>10–50% positive nuclei	128/248 (51.6%)	121/235 (51.5%)	249/483 (51.6%)
>50 % positive nuclei	100/133 (75.2%)	91/138 (65.9%)	191/271 (70.5%)
CK5/6			
Strong	23/29 (79.3%)	33/49 (67.3%)	56/78 (71.8%)
Weak	60/95 (63.2%)	57/101 (56.4%)	117/196 (59.7%)
Negative	498/1475 (33.8%)	466/1417 (32.9%)	964/2892 (33.3%)
KRT5			
Strongly positive (20%)	59/65 (90.8%)	73/80 (91.3%)	132/145 (91%)
Weakly positive (any staining)	60/78 (76.9%)	60/75 (80%)	120/153 (78.4%)
Negative	463/1443 (32.1%)	441/1408 (31.3%)	904/2851 (31.7%)
EGFR			
Strongly positive (20%)	52/66 (78.8%)	54/77 (70.1%)	106/143 (74.1%)
Weakly positive (any staining)	110/147 (74.8%)	110/143 (76.9%)	220/290 (75.9%)
Negative	415/1418 (29.3%)	400/1381 (29%)	815/2799 (29.1%)
YB-1			
Negative	261/956 (27.3%)	252/936 (26.9%)	513/1892 (27.1%)
Weakly positive ≥50%	211/420 (50.2%)	208/435 (47.8%)	419/855 (49.0%)
Moderately positive ≥50%	52/81 (64.2%)	52/80 (65%)	104/161 (64.6%)
Strongly positive	5/7 (71.4%)	4/5 (80%)	9/12 (75%)
HER2 fluorescent in situ hybridization			
Unamplified (<1.8)	175/531 (33%)	161/502 (32.1%)	336/1033 (32.5%)
Equivocal (1.8–2.2)	10/36 (27.8%)	14/29 (48.3%)	24/65 (36.9%)
Amplified (>2.2)	60/114 (52.6%)	59/126 (46.8%)	119/240 (49.6%)
Total	2003	1989	3992

Table shows the number of P-cadherin-positive cases/the total number of scorable cases in the biomarker category, with the percentage of P-cadherin-positive cases in parentheses.

P-cadherin expression has been reported to associate with reduced expression of E-cadherin in breast cancer patients[15] but no association was found in another study.[21] The majority of breast carcinomas coexpressing E-cadherin and P-cadherin had poor patient survival along with cytoplasmic expression of p120-catenin,[81] a master regulator of cadherin activity. In our study, p120-catenin was not

Table 8 Association of P-cadherin expression with other immunohistochemical biomarkers

Biomarker	*P-cadherin-staining intensity*		*Percentage of P-cadherin-positive cells*		
	Training set	*Validation set*	*Training set*	*Validation set*	*Type of association with P-cadherin*
Bcl-2[a]	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$	Negative
CK5/6	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$	Positive
E-cadherin[a]	$P=0.426$	$P=0.083$	$P=0.594$	$P=0.432$	None
EGFR	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$	Positive
ER	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$	Negative
PR	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$	Negative
HER2 immunohistochemistry scores, binarized {0,1} *vs* {3}[a]	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P=0.002$	Positive
HER2 immunohistochemistry scores, categorized {0} *vs* {1} *vs* {2} *vs* {3}	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$	Positive
HER2 fluorescent *in situ* hybridization ratios, binarized as negative (<1.8) and positive (>2.2)[a]	$P=0.001$	$P<0.0005$	$P<0.0005$	$P=0.002$	Positive
Ki-67[a]	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$	Positive
KRT5	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$	Positive
p53	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$	Positive
YB1	$P<0.0005$	$P<0.0005$	$P<0.0005$	$P<0.0005$	Positive

P-values reported are from the likelihood ratio χ^2 test for association in a two-way contingency table, unless otherwise noted.
[a]Fisher's exact test.

assessed, and P-cadherin expression did not show association with E-cadherin expression (whole series, Fisher's exact test $P=0.87$). P-cadherin and E-cadherin also failed to show association when the whole series was stratified by histological subtype (ductal carcinoma *vs* lobular carcinoma *vs* other special type tumors) (whole series, Mantel–Haenszel χ^2 $P=0.077$).

Scientific evidence suggests the concept of cancer immunosurveillance and immunoediting based on protection against development of spontaneous and chemically induced tumors in animal systems and identification of targets for immune recognition of human cancers.[82] Anti-tumor immune response, including the recognition of tumor-specific or tumor-associated antigens can be used to develop new vaccines and monoclonal antibody therapies.[83] It has been shown that apoptotic cell death is poorly immunogenic, whereas necrotic cell death is truly immunogenic.[84] Cell death during chemotherapy can potentially determine the immune response. Anthracyclines produce a beneficial immunogenic environment as these agents activate antigen-presenting dendritic cells, thus allowing a cytotoxic T lymphocyte response.[85] P-cadherin has been suggested as a possible target for immunotherapy of pancreatic, gastric, and colorectal cancers based on the identification of HLA-A2-restricted cytotoxic T lymphocyte epitopes of P-cadherin in HLA-A2.1 transgenic mice, and the *in vitro* and *in vivo* cytotoxicity against tumor cells of cytotoxic T lymphocyte specific to CDH3 induced from HLA-A2-positive healthy donors and cancer patients.[86]

In conclusion, this study shows the value of P-cadherin expression as a marker of poor prognosis in a large breast cancer series. P-cadherin positivity is associated with high-grade tumor subtypes (HER2 + and basal carcinomas), and well-established markers of poor prognosis (ER, PR, Bcl-2), and may represent a promising antibody therapeutic target, either by exploiting its association with poor prognosis tumor, or by modulating its role in cell adhesion and migration.

Acknowledgements

This research was supported in part by Pfizer Worldwide Research and Development, San Diego, CA, USA. The Genetic Pathology Evaluation Centre is supported in part by an unrestricted educational grant from Sanofi-Aventis Canada. S. Aparicio is supported by Canada Research Chair in Molecular Oncology. G Turashvili was supported by the Canadian Institute of Health Research Training Program for Clinician Scientists in Molecular Oncologic Pathology (STP-53912). We thank Cath Ennis for comments on the manuscript.

Disclosure/conflict of interest

The authors declare no conflict of interest.

References

1 Furukawa F, Fujii K, Horiguchi Y, *et al.* Roles of E- and P-cadherin in the human skin. Microsc Res Tech 1997;38:343–352.

2 Furukawa F, Takigawa M, Matsuyoshi N, *et al.* Cadherins in cutaneous biology. J Dermatol 1994;21:802–813.
3 Nose A, Nagafuchi A, Takeichi M. Isolation of placental cadherin cDNA: identification of a novel gene family of cell-cell adhesion molecules. EMBO J 1987;6:3655–3661.
4 Derycke L, De Wever O, Stove V, *et al.* Soluble N-cadherin in human biological fluids. Int J Cancer 2006;119:2895–2900.
5 Paredes J, Correia AL, Ribeiro AS, *et al.* P-cadherin expression in breast cancer: a review. Breast Cancer Res 2007;9:214.
6 Hines MD, Jin HC, Wheelock MJ, *et al.* Inhibition of cadherin function differentially affects markers of terminal differentiation in cultured human keratinocytes. J Cell Sci 1999;112:4569–4579.
7 Radice GL, Ferreira-Cornwell MC, Robinson SD, *et al.* Precocious mammary gland development in P-cadherin-deficient mice. J Cell Biol 1997;139:1025–1032.
8 Knudsen KA, Wheelock MJ. Cadherins and the mammary gland. J Cell Biochem 2005;95:488–496.
9 Kovacs A, Walker RA. P-cadherin as a marker in the differential diagnosis of breast lesions. J Clin Pathol 2003;56:139–141.
10 Soler AP, Russo J, Russo IH, *et al.* Soluble fragment of P-cadherin adhesion protein found in human milk. J Cell Biochem 2002;85:180–184.
11 De Paul AL, Bonaterra M, Soler AP, *et al.* Soluble p-cadherin found in human semen. J Androl 2005; 26:44–47.
12 Mannello F, Tonti GA, Medda V, *et al.* Increased shedding of soluble fragments of P-cadherin in nipple aspirate fluids from women with breast cancer. Cancer Sci 2008;99:2160–2169.
13 Knudsen KA, Lin CY, Johnson KR, *et al.* Lack of correlation between serum levels of E- and P-cadherin fragments and the presence of breast cancer. Hum Pathol 2000;31:961–965.
14 Paredes J, Milanezi F, Viegas L, *et al.* P-cadherin expression is associated with high-grade ductal carcinoma *in situ* of the breast. Virchows Arch 2002;440:16–21.
15 Palacios J, Benito N, Pizarro A, *et al.* Anomalous expression of P-cadherin in breast carcinoma. Correlation with E-cadherin expression and pathological features. Am J Pathol 1995;146:605–612.
16 Rasbridge SA, Gillett CE, Sampson SA, *et al.* Epithelial (E-) and placental (P-) cadherin cell adhesion molecule expression in breast carcinoma. J Pathol 1993;169: 245–250.
17 Peralta Soler A, Knudsen KA, Salazar H, *et al.* P-cadherin expression in breast carcinoma indicates poor survival. Cancer 1999;86:1263–1272.
18 Madhavan M, Srinivas P, Abraham E, *et al.* Cadherins as predictive markers of nodal metastasis in breast cancer. Mod Pathol 2001;14:423–427.
19 Paredes J, Albergaria A, Oliveira JT, *et al.* P-cadherin overexpression is an indicator of clinical outcome in invasive breast carcinomas and is associated with CDH3 promoter hypomethylation. Clin Cancer Res 2005;11:5869–5877.
20 Gamallo C, Moreno-Bueno G, Sarrio D, *et al.* The prognostic significance of P-cadherin in infiltrating ductal breast carcinoma. Mod Pathol 2001;14:650–654.
21 Kovacs A, Dhillon J, Walker RA. Expression of P-cadherin, but not E-cadherin or N-cadherin, relates to pathological and functional differentiation of breast carcinomas. Mol Pathol 2003;56:318–322.
22 Paredes J, Milanezi F, Reis-Filho JS, *et al.* Aberrant P-cadherin expression: is it associated with estrogen-independent growth in breast cancer? Pathol Res Pract 2002;198:795–801.
23 Perou CM, Sorlie T, Eisen MB, *et al.* Molecular portraits of human breast tumours. Nature 2000;406: 747–752.
24 Matos I, Dufloth R, Alvarenga M, *et al.* p63, cytokeratin 5, and P-cadherin: three molecular markers to distinguish basal phenotype in breast carcinomas. Virchows Arch 2005;447:688–694.
25 Paredes J, Lopes N, Milanezi F, *et al.* P-cadherin and cytokeratin 5: useful adjunct markers to distinguish basal-like ductal carcinomas *in situ*. Virchows Arch 2007;450:73–80.
26 Rakha EA, El-Sayed ME, Green AR, *et al.* Prognostic markers in triple-negative breast cancer. Cancer 2007;109:25–32.
27 Gluz O, Liedtke C, Gottschalk N, *et al.* Triple-negative breast cancer–current status and future directions. Ann Oncol 2009;20:1913–1927.
28 Cheang MC, Voduc D, Bajdik C, *et al.* Basal-like breast cancer defined by five biomarkers has superior prognostic value than triple-negative phenotype. Clin Cancer Res 2008;14:1368–1376.
29 Cheang MC, Treaba DO, Speers CH, *et al.* Immunohistochemical detection using the new rabbit monoclonal antibody SP1 of estrogen receptor in breast cancer is superior to mouse monoclonal antibody 1D5 in predicting survival. J Clin Oncol 2006;24:5637–5644.
30 Chia S, Norris B, Speers C, *et al.* Human epidermal growth factor receptor 2 overexpression as a prognostic factor in a large tissue microarray series of node-negative breast cancers. J Clin Oncol 2008;26:5697–5704.
31 Habibi G, Leung S, Law JH, *et al.* Redefining prognostic factors for breast cancer: YB-1 is a stronger predictor of relapse and disease-specific survival than estrogen receptor or HER-2 across all tumor subtypes. Breast Cancer Res 2008;10:R86.
32 Jensen KC, Turbin DA, Leung S, *et al.* New cutpoints to identify increased HER2 copy number: analysis of a large, population-based cohort with long-term follow-up. Breast Cancer Res Treat 2008;112:453–459.
33 Cheang MC, Chia SK, Voduc D, *et al.* Ki67 index, HER2 status, and prognosis of patients with luminal B breast cancer. J Natl Cancer Inst 2009;101:736–750.
34 Turashvili G, Leung S, Turbin D, *et al.* Inter-observer reproducibility of HER2 immunohistochemical assessment and concordance with fluorescent *in situ* hybridization (FISH): pathologist assessment compared to quantitative image analysis. BMC Cancer 2009;9:165.
35 Liu S, Chia SK, Mehl E, *et al.* Progesterone receptor is a significant factor associated with clinical outcomes and effect of adjuvant tamoxifen therapy in breast cancer patients. Breast Cancer Res Treat 2010;119:53–61.
36 Voduc D, Kenney C, Nielsen TO. Tissue microarrays in clinical oncology. Semin Radiat Oncol 2008;18:89–97.
37 Stefansson IM, Salvesen HB, Akslen LA. Prognostic impact of alterations in P-cadherin expression and related cell adhesion markers in endometrial cancer. J Clin Oncol 2004;22:1242–1252.
38 Nalwoga H, Arnes JB, Wabinga H, *et al.* Frequency of the basal-like phenotype in African breast cancer. APMIS 2007;115:1391–1399.

39 Arnes JB, Brunet JS, Stefansson I, *et al.* Placental cadherin and the basal epithelial phenotype of BRCA1-related breast cancer. Clin Cancer Res 2005;11: 4003–4011.
40 Liu CL, Prapong W, Natkunam Y, *et al.* Software tools for high-throughput analysis and archiving of immunohistochemistry staining data obtained with tissue microarrays. Am J Pathol 2002;161:1557–1565.
41 Liu CL, Montgomery KD, Natkunam Y, *et al.* TMA-Combiner, a simple software tool to permit analysis of replicate cores on tissue microarrays. Mod Pathol 2005;18:1641–1648.
42 The R Project for Statistical Computing Available from at http://www.r-project.org.
43 Therneau TM, Grambsch PM. Modeling Survival Data: Extending the Cox Model, Vol. XIII, Springer: New York, 2001, p 350.
44 Rajput AB, Turbin DA, Cheang MC, *et al.* Stromal mast cells in invasive breast cancer are a marker of favourable prognosis: a study of 4,444 cases. Breast Cancer Res Treat 2008;107:249–257.
45 Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer: New York, 2001, p 533.
46 Rifai N, Altman DG, Bossuyt PM. Reporting bias in diagnostic and prognostic studies: time for action. Clin Chem 2008;54:1101–1103.
47 Collett K, Stefansson IM, Eide J, *et al.* A basal epithelial phenotype is more frequent in interval breast cancers compared with screen detected tumors. Cancer Epidemiol Biomarkers Prev 2005;14: 1108–1112.
48 Ben Hamida A, Labidi IS, Mrad K, *et al.* Markers of subtypes in inflammatory breast cancer studied by immunohistochemistry: prominent expression of P-cadherin. BMC Cancer 2008;8:28.
49 Troxell ML, Masek M, Sibley RK. Immunohistochemical staining of papillary breast lesions. Appl Immunohistochem Mol Morphol 2007;15:145–153.
50 Crabb SJ, Bajdik CD, Leung S, *et al.* Can clinically relevant prognostic subsets of breast cancer patients with four or more involved axillary lymph nodes be identified through immunohistochemical biomarkers? A tissue microarray feasibility study. Breast Cancer Res 2008;10:R6.
51 Foulkes WD, Stefansson IM, Chappuis PO, *et al.* Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. J Natl Cancer Inst 2003;95:1482–1485.
52 van Belle G, Fisher LD, Heagerty PJ, *et al.* Biostatistics: A Methodology for the Health Sciences. John Wiley and Sons: New Jersey, 2004, p 896.
53 Nout RA, Fiets WE, Struikmans H, *et al.* The in- or exclusion of non-breast cancer related death and contralateral breast cancer significantly affects estimated outcome probability in early breast cancer. Breast Cancer Res Treat 2008;109:567–572.
54 Gusterson B. Do 'basal-like' breast cancers really exist? Nat Rev Cancer 2009;9:128–134.
55 Haffty BG, Yang Q, Reiss M, *et al.* Locoregional relapse and distant metastasis in conservatively managed triple negative early-stage breast cancer. J Clin Oncol 2006;24:5652–5657.
56 Nielsen TO, Hsu FD, Jensen K, *et al.* Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clin Cancer Res 2004;10:5367–5374.
57 Carey LA, Perou CM, Livasy CA, *et al.* Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA 2006;295:2492–2502.
58 Han AC, Soler AP, Knudsen KA, *et al.* Distinct cadherin profiles in special variant carcinomas and other tumors of the breast. Hum Pathol 1999;30:1035–1039.
59 Reis-Filho JS, Milanezi F, Paredes J, *et al.* Novel and classic myoepithelial/stem cell markers in metaplastic carcinomas of the breast. Appl Immunohistochem Mol Morphol 2003;11:1–8.
60 Jacquemier J, Padovani L, Rabayrol L, *et al.* Typical medullary breast carcinomas have a basal/myoepithelial phenotype. J Pathol 2005;207:260–268.
61 Livasy CA, Karaca G, Nanda R, *et al.* Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol 2006;19:264–271.
62 Fulford LG, Easton DF, Reis-Filho JS, *et al.* Specific morphological features predictive for the basal phenotype in grade 3 invasive ductal carcinoma of breast. Histopathology 2006;49:22–34.
63 Reis-Filho JS, Milanezi F, Steele D, *et al.* Metaplastic breast carcinomas are basal-like tumours. Histopathology 2006;49:10–21.
64 Reis-Filho JS, Tutt AN. Triple negative tumours: a critical review. Histopathology 2008;52:108–118.
65 McCarthy A, Savage K, Gabriel A, *et al.* A mouse model of basal-like breast carcinoma with metaplastic elements. J Pathol 2007;211:389–398.
66 Rakha EA, El-Sheikh SE, Kandil MA, *et al.* Expression of BRCA1 protein in breast cancer and its prognostic significance. Hum Pathol 2008;39:857–865.
67 Lakhani SR, Reis-Filho JS, Fulford L, *et al.* Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. Clin Cancer Res 2005;11:5175–5180.
68 Honrado E, Benitez J, Palacios J. The molecular pathology of hereditary breast cancer: genetic testing and therapeutic implications. Mod Pathol 2005;18: 1305–1320.
69 Palacios J, Honrado E, Osorio A, *et al.* Immunohistochemical characteristics defined by tissue microarray of hereditary breast cancer not attributable to BRCA1 or BRCA2 mutations: differences from breast carcinomas arising in BRCA1 and BRCA2 mutation carriers. Clin Cancer Res 2003;9:3606–3614.
70 Palacios J, Honrado E, Osorio A, *et al.* Phenotypic characterization of BRCA1 and BRCA2 tumors based in a tissue microarray study with 37 immunohistochemical markers. Breast Cancer Res Treat 2005; 90:5–14.
71 Arnes JB, Begin LR, Stefansson I, *et al.* Expression of epidermal growth factor receptor in relation to BRCA1 status, basal-like markers and prognosis in breast cancer. J Clin Pathol 2009;62:139–146.
72 Arnes JB, Collett K, Akslen LA. Independent prognostic value of the basal-like phenotype of breast cancer and associations with EGFR and candidate stem cell marker BMI-1. Histopathology 2008;52:370–380.
73 Ribeiro AS, Albergaria A, Sousa B, *et al.* Extracellular cleavage and shedding of P-cadherin: a mechanism underlying the invasive behaviour of breast cancer cells. Oncogene 2010;29:392–402.
74 Sarrio D, Palacios J, Hergueta-Redondo M, *et al.* Functional characterization of E- and P-cadherin in invasive breast cancer cells. BMC Cancer 2009;9:74.
75 Cleton-Jansen AM. E-cadherin and loss of heterozygosity at chromosome 16 in breast carcinogenesis:

different genetic pathways in ductal and lobular breast cancer? Breast Cancer Res 2002;4:5–8.
76 Berx G, Becker KF, Hofler H, *et al.* Mutations of the human E-cadherin (CDH1) gene. Hum Mutat 1998;12: 226–237.
77 Cowin P, Rowlands TM, Hatsell SJ. Cadherins and catenins in breast cancer. Curr Opin Cell Biol 2005; 17:499–508.
78 Yoder BJ, Wilkinson EJ, Massoll NA. Molecular and morphologic distinctions between infiltrating ductal and lobular carcinoma of the breast. Breast J 2007;13:172–179.
79 Guarino M, Rubino B, Ballabio G. The role of epithelial-mesenchymal transition in cancer pathology. Pathology 2007;39:305–318.
80 Boyer B, Valles AM, Edme N. Induction and regulation of epithelial-mesenchymal transitions. Biochem Pharmacol 2000;60:1091–1099.
81 Paredes J, Correia AL, Ribeiro AS, *et al.* Breast carcinomas that co-express E- and P-cadherin are associated with p120-catenin cytoplasmic localisation and poor patient survival. J Clin Pathol 2008;61: 856–862.
82 Dunn GP, Bruce AT, Ikeda H, *et al.* Cancer immunoediting: from immunosurveillance to tumor escape. Nat Immunol 2002;3:991–998.
83 Storkus WJ, Falo Jr LD. A 'good death' for tumor immunology. Nat Med 2007;13:28–30.
84 Igney FH, Krammer PH. Death and anti-death: tumour resistance to apoptosis. Nat Rev Cancer 2002;2: 277–288.
85 Obeid M, Tesniere A, Ghiringhelli F, *et al.* Calreticulin exposure dictates the immunogenicity of cancer cell death. Nat Med 2007;13:54–61.
86 Imai K, Hirata S, Irie A, *et al.* Identification of a novel tumor-associated antigen, cadherin 3/P-cadherin, as a possible target for immunotherapy of pancreatic, gastric, and colorectal cancers. Clin Cancer Res 2008; 14:6487–6495.

Supplementary Information accompanies the paper on Modern Pathology website (http://www.nature.com/modpathol)