Featured Article

Molecular Cytogenetic Identification of Subgroups of Grade III Invasive Ductal Breast Carcinomas with Different Clinical Outcomes

Chris Jones,[1] Emily Ford,[1] Cheryl Gillett,[2] Ken Ryder,[2] Samantha Merrett,[1] Jorge S. Reis-Filho,[1] Laura G. Fulford,[1,3] Andrew Hanby,[4] and Sunil R. Lakhani[1,5]

[1]The Breakthrough Toby Robins Breast Cancer Research Centre, Institute of Cancer Research, London; [2]Hedley Atkins/Cancer Research UK Breast Pathology Laboratory, Guy's Hospital, London; [3]Ludwig Institute for Cancer Research, University College London; [4]Department of Histopathology, St. James' Hospital, Leeds; and [5]Department of Histopathology, Royal Marsden Hospital, London, United Kingdom

ABSTRACT

Tumor grade is an established indicator of breast cancer outcome, although considerable heterogeneity exists even within-grade. Around 25% of grade III invasive ductal breast carcinomas are associated with a "basal" phenotype, and these tumors are reported to be a distinct subgroup. We have investigated whether this group of breast cancers has a distinguishing pattern of genetic alterations and which of these may relate to the different clinical outcome of these patients. We performed comparative genomic hybridization (CGH) analysis on 43 grade III invasive ductal breast carcinomas positive for basal cytokeratin 14, as well as 43 grade- and age-matched CK14-negative controls, all with up to 25 years (median, 7 years) of clinical follow-up. Significant differences in CGH alterations were seen between the two groups in terms of mean number of changes (CK14+ve − 6.5, CK14−ve − 10.3; $P = 0.0012$) and types of alterations at chromosomes 4q, 7q, 8q, 9p, 13q, 16p, 17p, 17q, 19p, 19q, 20p, 20q and Xp. Supervised and unsupervised algorithms separated the two groups on CGH data alone with 76% and 74% accuracy, respectively. Hierarchical clustering revealed distinct subgroups, one of which contained 18 (42%) of the CK14+ve tumors. This subgroup had significantly shorter overall survival ($P = 0.0414$) than other grade III tumors, regardless of CK14 status, and was an independent prognostic marker ($P = 0.031$). These data provide evidence that the "basal" phenotype on its own does not convey a poor prognosis. Basal tumors are also heterogeneous with only a subset, identifiable by pattern of genetic alterations, exhibiting a shorter overall survival. Robust characterization of this basal group is necessary if it is to have a major impact on management of patients with breast cancer.

INTRODUCTION

Breast cancer is a heterogeneous disease with a disparate variety of histological types and a wide spectrum of responsiveness to different treatments, making clinical management difficult. The majority of breast carcinomas fall into the category of invasive ductal carcinoma, no special type, for which histological grade is one of the best predictors of behavior. Poorly differentiated grade III invasive ductal carcinomas are strongly associated with shorter recurrence-free and overall survival times than lower grade I and II tumors (1). Within these groups, however, considerable heterogeneity still exists, and delineation of the most aggressive subtypes within grades would be of considerable clinical benefit.

Invasive ductal carcinoma-no special type, as determined morphologically, is thought to arise exclusively from the luminal epithelial cells of the breast. It has been known for some time, however, that a proportion (2–18% of all invasive ductal carcinomas and up to 25% of grade III tumors) of these tumors have been demonstrated to show a basal/myoepithelial cell phenotype by immunohistochemical analysis using a range of markers including intermediate filaments cytokeratin (CK) 5 and 14 (2–17). Recent microarray studies have also identified a "basal-like" group of breast tumors based on their patterns of gene expression (18, 19).

Although a comprehensive characterization and definition of basal tumors is lacking, there are a number of features reported to be associated with this phenotype of invasive ductal carcinoma-no special type. Morphologically, they appear to be predominantly grade III (14, 16) and are reported to frequently contain large central acellular zones composed of necrosis, tissue infarction, collagen, and hyaline material on their cut surfaces (15). Immunohistochemically, as well as in the expression of a number of myoepithelial markers, these tumors seem to be predominantly estrogen receptor (ER), progesterone receptor (PR), and ERBB2 negative (16), an immunophenotype resembling BRCA1 tumors (20). Microarray analysis has also

Received 12/15/03; revised 2/24/04; accepted 3/4/04.
Grant support: J. Reis-Filho is the recipient of the Gordon Signy International Fellowship Award and is partially supported by a Ph.D. grant (Ref:SFRH/BD/5386/2001) from the Fundação para a Ciência e a Tecnologia, Portugal.
The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.
Note: Supplementary data for this article may be found at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org).
Requests for reprints: Sunil R. Lakhani, The Breakthrough Toby Robins Breast Cancer Research Centre, Institute of Cancer Research, Fulham Road, London SW3 6JB, United Kingdom. Phone: 20-7153-5525; Fax: 20-7153-5533; E-mail: Sunil.Lakhani@icr.ac.uk.

demonstrated a similarity between sporadic, basal-like tumors and those familial tumors harboring a BRCA1 mutation, based on their patterns of gene expression (21).

The pathogenesis of such lesions appear to indicate a poor prognosis. As well as the association with high histological grade and hormone receptor negativity, the myoepithelial phenotype has been reported to be associated with a high risk of brain and lung metastases and of death by cancer independent of nodal status and tumor size (22). The basal-like phenotype identified by expression profiling experiments conferred a shorter survival time than the other tumor groups described (19). Tissue microarray analysis of basal keratins 5 and 17 showed a poorer clinical outcome in node-negative tumors expressing one or both of these markers (17).

Preliminary analysis has indicated that the pattern and number of alterations, as detected by comparative genomic hybridization (CGH), more closely resembles pure myoepithelial carcinomas than other grade III invasive ductal carcinoma (16). In this study, we have chosen to use CK14 to identify grade III invasive ductal carcinomas as exhibiting a basal/myoepithelial phenotype and have carried out CGH analysis on 43 CK14 positive and 43 age- and grade-matched CK14 negative tumors, in an attempt to investigate the molecular events associated with the different subtypes and to correlate these data to clinical outcome.

MATERIALS AND METHODS

Tumor Specimens. Patients diagnosed as having a grade III mammary carcinoma at the Hedley Atkins/Cancer Research UK Breast Pathology Laboratory, Guy's Hospital between 1975 and 1992 were included in the study and have formed a cohort upon which studies have been published previously (23). Stained sections from the primary tumor were retrieved, and the tumor was regraded according to the modified Bloom and Richardson method (1). Any patient subsequently found to have a grade II tumor on review was excluded from the study. An additional group of patients was also excluded because insufficient tumor tissue remained in the diagnostic block for study. Therefore, a total of 418 patients were available. Of these, 89 were found to be either focally or diffusely positive for CK14 by immunohistochemistry, and 43 of these tumors proved amenable to microdissection, DNA extraction, and CGH analysis. Forty-three age-matched, CK14-negative tumors were chosen from the same cohort as controls and were subject to the same experimental difficulties. Matching for tumor size, stage, and nodal status were also carried out as closely as possible, with the two groups showing no significant differences in distribution of these markers.

Immunohistochemistry. Staining was performed on 3-μm-thick sections of formalin-fixed, paraffin-embedded tissues. Sections were cut and placed onto Vectabond-coated slides and allowed to dry overnight at 37°C. Before staining, the sections were melted onto the slide by placing them in an oven at 60°C for 16 h. Sections were dewaxed, and endogenous peroxidase activity was inhibited using methanol and hydrogen peroxide. After thorough washing, heat-mediated antigen retrieval was used to expose sites of immunoreactivity. This was achieved by placing the slides in a pressure cooker containing boiling 0.01 M citrate buffer (pH 6.0) where they were kept under pressure (15 psi) for 2 min. Once the sections were cool, the remaining immunohistochemical staining was carried out using an Optimax Automated Immunostainer (A. Menarini) with a standard peroxidase-conjugated streptavidin-biotin complex method. CK19 (clone RCK108; 1:50), ER (clone 1D5; 1:100), and PR (clone PgR636; 1:400) were obtained from Dako Ltd., CK14 (clone LL002; 1:50) was obtained from Biogenex, and ERBB^2 (clone 3B5; 1:1500) was obtained from Oncogene Science. Sites of peroxidase activity were detected using 3,3′-diaminobenzidine/hydrogen peroxide, which produced a brown reaction product. The slides were lightly stained with hematoxylin. A positive control, known to express the antigen, was included in every batch of staining.

ER and PR were assessed using the Quick score method (24) with a score of ≤ 3 considered negative. c-erbB2 was assessed in accordance with the Dako HercepTest guidelines with a score of ≤ 1 considered negative. Cytokeratin 14 and 19 were scored according to the presence or absence of expression in the invasive component.

CGH. Microdissection, DNA extraction, and CGH analysis were carried out as described previously (25). Briefly, microdissection was carried out using the PixCell II Laser Capture Microdissection system (Arcturus, Mountain View, CA) from formalin-fixed, paraffin-embedded tissue and the DNA extracted with 0.5 μg/μl proteinase K. Amplification and fluorescent labeling of the DNA from microdissected tumor and normal tissue was carried out by degenerate oligonucleotide primed-PCR in two rounds as published previously (26) and hybridized to normal male metaphase spreads (Vysis United Kingdom Ltd., Richmond, England) for 2–3 days at 37°C. Metaphase chromosome preparations were captured using a Zeiss Axioskop microscope, Photometrics (Munich, Germany) KAF1400 CCD camera and Vysis SmartCapture software. Image analysis was performed using Vysis Quips CGH software. Between 5 and 10 representative images of high quality hybridizations were analyzed and the results combined to produce an average fluorescence ratio for each chromosome. Control experiments were carried out using normal:normal (microdissected normal lymph node) cohybridizations, of which the average red:green ratio levels and 95% confidence intervals were used to set the lower and upper limits for scoring losses and gains of genetic material as 0.80–1.20.

Statistical Analysis. All of the changes in DNA copy number were reduced to "gains" or "losses" at the chromosomal arm resolution. The data were retabulated as $+2$ for gains and -2 for losses for supervised and unsupervised analysis. Unsupervised hierarchical clustering was carried out using "hclust" in R 1.7.0,[6] and plotted using the "heatmap" function in the mva package in R. Hierarchical clustering was based on a Euclidean distance measure using Ward's minimum variance method. Supervised analysis was carried out using the nearest shrunken centroid classifier (27) implemented in R 1.7.0. Survival analysis was carried out using the statistical platform S-Plus version 6.1 for Windows (Insightful) on our right-censored clinical

[6] Internet address: http://www.r-project.org.

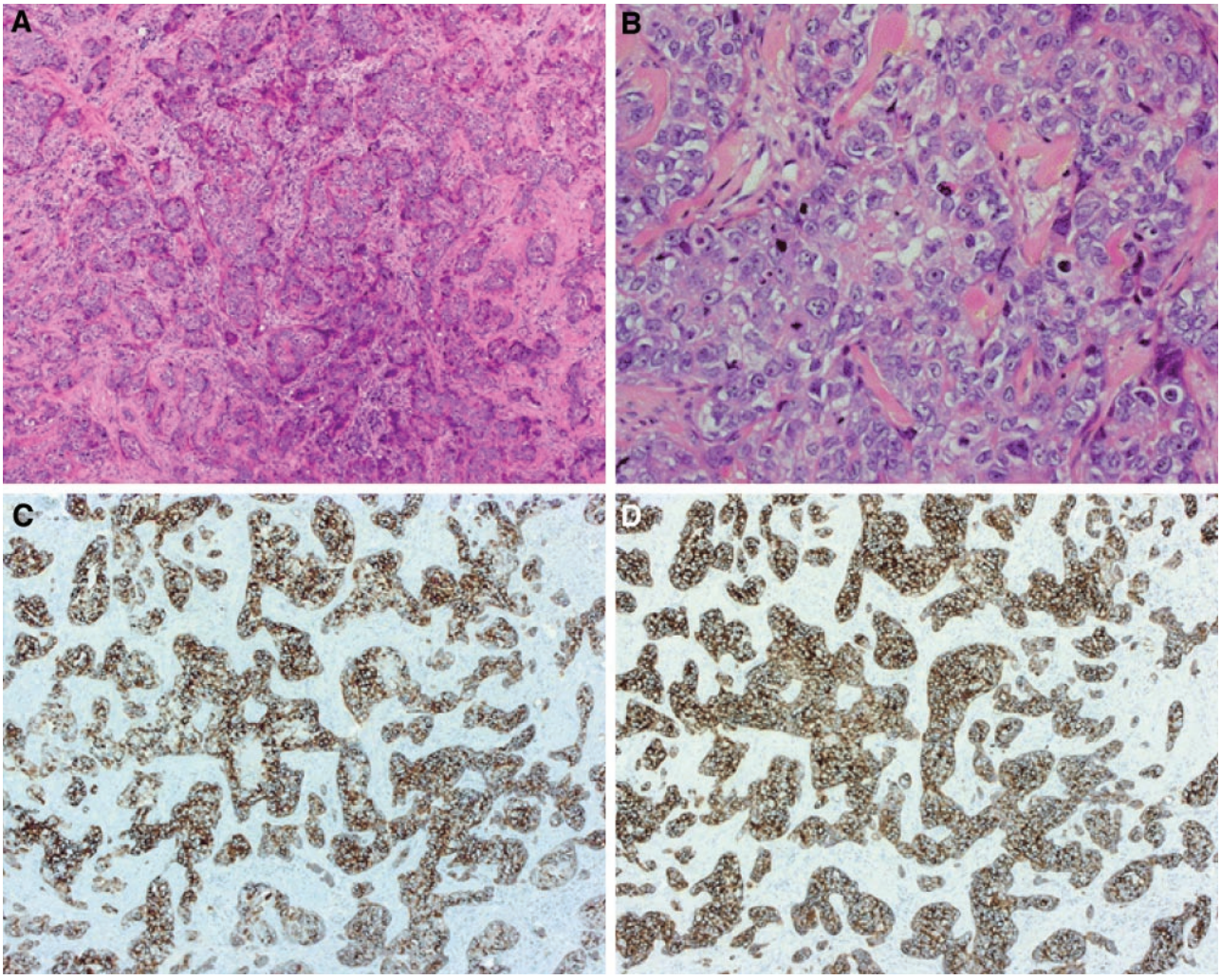

Fig. 1 Photomicrographs of a grade III invasive ductal breast carcinoma CK272. *A,* H&E, low power (×10). *B,* H&E, high power (×40). *C,* diffuse immunohistochemical staining of cytokeratin 14 (×10). *D,* diffuse staining of cytokeratin 19 (×10).

follow-up data. Kaplan-Meier plots were generated using the function "survfit," and the log-rank test was carried out to determine whether curves were significantly different from each other using the "survdiff" function. Multivariate analysis was carried out using the Cox proportional hazards model with "coxph" used to investigate the independence of the individual proteins on prognosis.

RESULTS

Of 418 grade III invasive ductal breast carcinomas in our study, 89 (21.2%) were either focally of diffusely positive for CK14. Of these, 43 were amenable to CGH analysis and 43 age-matched CK14-negative grade III tumors were analyzed as controls. Of these 86 tumors analyzed by molecular cytogenetics, 33 were ER positive (38.4%), 25 were PR positive (34.7%), 15 were ERBB2 positive (18.3%), and 83 were positive for CK19 (96.5%). Forty-nine showed positive lymph node metastasis (57.6%). Follow-up data were available for all of the patients who had a mean age at diagnosis of 47.8 years (median, 46.5), mean tumor size of 3.4 cm (median, 3.5 cm), a mean disease-free time of 8.7 years (median, 4.2), and mean survival time of 9.9 years (median, 7.0). Full details of clinicopathological data are available in Supplementary Table S1.

Fig. 1 shows a representative CK14-positive tumor taken for CGH analysis. Positive CK14 tumors were shown to be significantly associated with negative ER status ($P < 0.00001$), negative PR ($P = 0.001$), and negative ERBB2 ($P = 0.0201$), but not with nodal status ($P = 0.191$) or tumor size ($P = 0.9479$).

CGH analysis of the 86 tumors showed a mean number of alterations of 8.2 (median, 6.0). Full details of all CGH data for each case are available in Supplementary Table S1. The most frequent alterations over all samples were gains at 1q (21%), 17q (26%), and 20q (41%); and losses at 1p (21%), 16q (20%), 17p (20%), 17q (24%), 19p (21%), and 19q (21%). Differences in CGH profiles between the CK14-positive and -negative tumors are shown in Table 1. Statistical significance was deter-

Table 1 Differences in CGH alterations between CK14-positive and CR14-negative tumors

Chromosome gain	Overall (%)	CK14 positive (%)	CK14 negative (%)	*P**	Chromosome loss	Overall (%)	CK14 positive (%)	CK14 negative (%)	*P**
1p	5 (6)	4 (9)	1 (2)	0.2077	1p	18 (21)	9 (21)	9 (21)	>0.9999
1q	18 (21)	7 (16)	11 (26)	0.4271	1q	1 (1)	0 (0)	1 (2)	>0.9999
2p	7 (8)	2 (5)	5 (12)	0.4331	2p	11 (13)	3 (7)	8 (19)	0.1951
2q	10 (12)	3 (7)	7 (16)	0.3134	2q	12 (14)	3 (7)	9 (21)	0.1171
3p	8 (9)	3 (7)	5 (12)	0.7130	3p	10 (12)	3 (7)	7 (16)	0.3134
3q	7 (8)	2 (5)	5 (12)	0.4331	3q	7 (8)	2 (5)	5 (12)	0.4331
4p	4 (5)	1 (2)	3 (7)	0.6160	4p	7 (8)	1 (2)	6 (14)	0.1096
4q	11 (13)	6 (14)	5 (12)	>0.9999	*4q*	*12 (14)*	*2 (5)*	*10 (23)*	*0.0261*
5p	3 (3)	0 (0)	3 (7)	0.2412	5p	6 (7)	1 (2)	5 (12)	0.2020
5q	9 (10)	2 (5)	7 (16)	0.1561	5q	12 (14)	3 (7)	9 (21)	0.1171
6p	7 (8)	1 (2)	6 (14)	0.1096	6p	5 (6)	2 (5)	3 (7)	>0.9999
6q	8 (9)	5 (12)	3 (7)	0.7130	6q	10 (12)	3 (7)	7 (16)	0.3134
7p	6 (7)	4 (9)	2 (5)	0.7383	7p	6 (7)	1 (2)	5 (12)	0.2020
7q	10 (12)	6 (14)	4 (9)	0.7383	*7q*	*6 (7)*	*0 (0)*	*6 (14)*	*0.0259*
8p	7 (8)	3 (7)	4 (9)	>0.9999	8p	7 (8)	2 (5)	5 (12)	0.4331
8q	11 (13)	5 (12)	6 (14)	>0.9999	*8q*	*12 (14)*	*2 (5)*	*10 (23)*	*0.0261*
9p	4 (5)	2 (5)	2 (5)	>0.9999	*9p*	*6 (7)*	*0 (0)*	*6 (14)*	*0.0259*
9q	8 (9)	5 (12)	3 (7)	0.7130	9q	14 (16)	5 (12)	9 (21)	0.3816
10p	2 (2)	1 (2)	1 (2)	>0.9999	10p	2 (2)	1 (2)	1 (2)	>0.9999
10q	3 (3)	1 (2)	2 (5)	>0.9999	10q	3 (3)	1 (2)	2 (5)	>0.9999
11p	2 (2)	0 (0)	2 (5)	0.4941	11p	7 (8)	3 (7)	4 (9)	>0.9999
11q	6 (7)	3 (7)	3 (7)	>0.9999	11q	7 (8)	3 (7)	4 (9)	>0.9999
12p	6 (7)	2 (5)	4 (9)	0.7383	12p	6 (7)	1 (2)	5 (12)	0.2020
12q	2 (2)	1 (2)	1 (2)	>0.9999	12q	10 (12)	2 (5)	8 (19)	0.0887
13q	5 (6)	2 (5)	3 (7)	>0.9999	*13q*	*9 (10)*	*0 (0)*	*9 (21)*	*0.0025*
14q	6 (7)	2 (5)	4 (9)	0.7383	*14q*	*6 (7)*	*0 (0)*	*6 (14)*	*0.0259*
15q	4 (5)	1 (2)	3 (7)	0.6160	15q	9 (10)	2 (5)	7 (16)	0.1561
16p	2 (2)	1 (2)	1 (2)	>0.9999	*16p*	*16 (19)*	*14 (33)*	*2 (5)*	*0.0016*
16q	1 (1)	0 (0)	1 (2)	>0.9999	16q	17 (20)	11 (26)	6 (14)	0.2787
17p	*11 (13)*	*2 (5)*	*9 (21)*	*0.0488*	17p	17 (20)	12 (28)	5 (12)	0.1025
17q	*22 (26)*	*4 (9)*	*18 (42)*	*0.0010*	*17q*	*21 (24)*	*16 (37)*	*5 (12)*	*0.0110*
18p	6 (7)	2 (5)	4 (9)	0.7383	18p	3 (3)	2 (5)	1 (2)	>0.9999
18q	2 (2)	1 (2)	1 (2)	>0.9999	18q	14 (16)	4 (9)	10 (23)	0.1424
19p	5 (6)	1 (2)	4 (9)	0.3600	19p	18 (21)	11 (26)	7 (16)	0.4271
19q	5 (6)	0 (0)	5 (12)	0.0553	*19q*	*18 (21)*	*14 (33)*	*4 (9)*	*0.0155*
20p	*9 (10)*	*0 (0)*	*9 (21)*	*0.0025*	20p	9 (10)	6 (14)	3 (7)	0.4833
20q	*35 (41)*	*9 (21)*	*26 (60)*	*0.0004*	20q	4 (5)	0 (0)	4 (9)	0.1162
21q	6 (7)	3 (7)	3 (7)	>0.9999	21q	6 (7)	5 (12)	1 (9)	0.2020
22q	4 (5)	0 (0)	4 (9)	0.1162	22q	14 (16)	8 (19)	6 (14)	0.7712
Xp	1 (1)	0 (0)	1 (2)	>0.9999	*Xp*	*14 (16)*	*11 (26)*	*3 (7)*	*0.0381*
Xq	3 (3)	0 (0)	3 (7)	0.2412	Xq	13 (15)	7 (16)	6 (14)	>0.9999

NOTE: Numbers of tumors (and percentages) exhibiting a change in DNA copy number in all cases, as well as stratified by CK14 status are given along with the *P* value as determined by Fisher's exact test to probe differences between the two groups. Significant ($P < 0.05$) differences in gains are in italics.

* Determined by Fisher's exact test.

mined by Fisher's exact test. The CK14-positive tumors showed an increased number of losses at 16p (33% CK14 positive *versus* 5% CK14 negative; $P = 0.016$), 17q (37% *versus* 12%; $P = 0.0110$), 19q (33% *versus* 9%; $P = 0.0155$) and Xp (26% *versus* 7%, $P = 0.0381$). CK14-negative tumors displayed an increased number of gains at 17p (5% *versus* 21%; $P = 0.0488$), 17q (9% *versus* 42%; $P < 0.0010$), 20p (0% *versus* 21%; $P = 0.0025$), and 20q (21% *versus* 60%; $P < 0.0004$); and losses at 4q (5% *versus* 23%; $P = 0.0261$), 9p (0% *versus* 14%; $P = 0.0259$), and 13q (0% *versus* 21%; $P = 0.025$). Summary karyograms displaying all of the changes in DNA copy number are shown in Fig. 2.

Clinical follow-up data were available for all of the cases, and univariate analysis was carried out to determine which clinicopathological, immunohistochemical, and molecular variables were associated with prognosis (Table 2). In this grade III invasive ductal carcinoma cohort, the most significant prognostic indicator was nodal status, with positive lymph nodes predicting for shorter disease-free time ($P = 0.00269$) and overall survival ($P = 0.00275$). Trends were observed linking poor prognosis with ER negativity, ERBB2 positivity, and number of CGH alterations; however, these did not reach statistical significance. A number of chromosomal loci showed a significant association with disease-free and overall survival, including gain of 1q, loss of 3p, loss of 4p, gain of 6p, and loss of 8p. CK14 positivity showed no significant influence on prognosis in this cohort (disease-free survival $P = 0.193$; overall survival $P = 0.385$). CK14-positive tumors showed a significantly lower overall mean number of CGH alterations than the CK14-negative group (6.5 *versus* 10.3; $P = 0.0012$). Interestingly, the few CK19-negative tumors (3 of 83) showed a significant

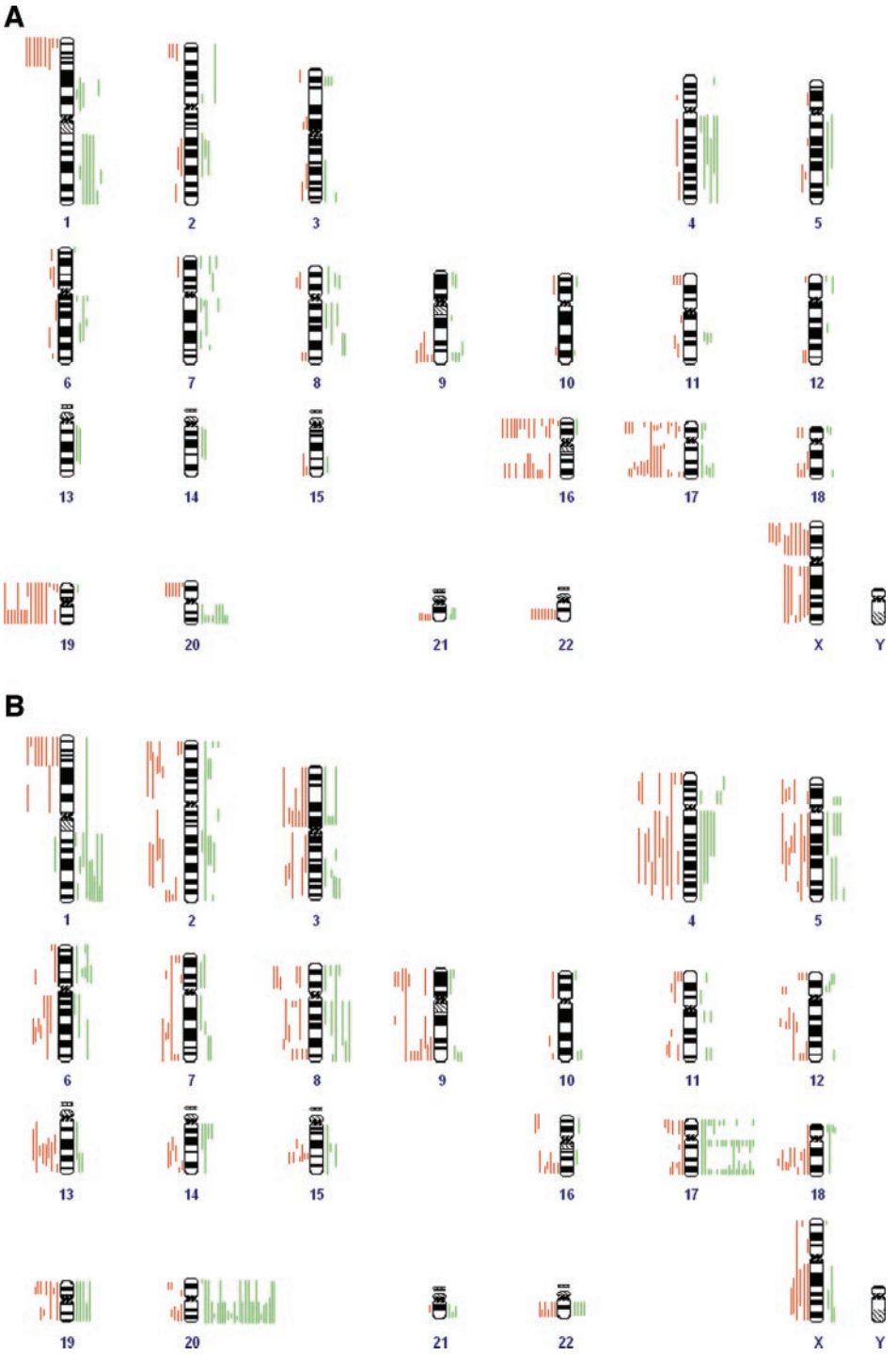

Fig. 2 Summary karyograms of comparative genomic hybridization alterations in grade III invasive ductal breast carcinomas. *A*, CK14-positive tumors. *B*, CK14-negative tumors. *Red bars* to the *left* of the chromosomal ideogram represent a loss at that locus in a single case; *green bars* to the *right* of the ideogram represent a gain in DNA copy number.

association with shorter disease-free time ($P = 0.0209$) and overall survival ($P = 0.0402$), despite these small numbers.

Because there were a number of alterations that were more or less prevalent in the different groups of grade III tumors, supervised cluster analysis was carried out to determine whether the CGH data would be able to predict for CK14 status on its own. Using the nearest shrunken centroid classifier, 76% of breast tumors were correctly assigned into the relevant group by CGH data alone using leave-one-out cross-validation (cross-validated error-rate, 0.24) after first training the classifier (Fig. 3*A*). We next attempted to determine whether we could identify subgroups of these tumors by their CGH data, which may have different biological behaviors. Hierarchical clustering of the CGH data split the tumors into two large branches, containing

Table 2 Univariate analysis of prognostic indicators

Factor	No. of cases	Disease-free time			Death from breast cancer		
		Mean survival (y)	SE	*P*	Mean survival (y)	SE	*P*
Size of tumor				0.8			0.825
<2.0 cm	21	10.11	2.31		13.6	2.41	
2.0–5.0 cm	54	11.76	1.49		13.2	1.47	
>5.0 cm	10	9.39	3.2		10	3.04	
Nodal status				*0.00269*			*0.00275*
Positive	49	8.23	1.44		10.4	1.55	
Negative	36	16.51	1.85		17.7	1.77	
ER[a]				0.151			0.139
Positive	33	12.5	1.73		13.6	1.59	
Negative	53	10.5	1.57		12.5	1.59	
PR				0.471			0.588
Positive	25	9.68	2.08		11.1	1.99	
Negative	47	12.03	1.63		13.7	1.59	
ERBB2				0.318			0.412
Positive	15	9.25	2.64		10.6	2.6	
Negative	67	12.45	1.46		14	1.41	
CK14				0.193			0.385
Positive	43	12.7	1.73		14.2	1.67	
Negative	43	10.5	1.72		12.5	1.71	
CK19				*0.0209*			*0.0402*
Positive	83	12.38	1.29		14.1	1.261	
Negative	3	1.17	0.395		2.72	0.784	
CGH alterations				0.888			0.931
0	8	13.35	4.35		13.9	4.17	
1–10	52	11.98	1.54		13.9	1.52	
11–20	20	9.53	2.26		11.9	2.43	
21 or more	6	9.53	3.54		10.5	3.21	
Gain of 1q				*0.0469*			*0.043*
Present	18	6.63	2.16		7.91	1.96	
Absent	67	13.42	1.44		15.32	1.41	
Loss of 3p				*0.00339*			*0.00244*
Present	10	2.54	1.01		3.5	0.954	
Absent	68	13.26	1.43		15.2	1.392	
Loss of 4p				*0.00397*			*0.0423*
Present	7	2.57	1.31		5.76	3.01	
Absent	75	12.82	1.36		14.43	1.32	
Gain of 6p				*0.0183*			*0.0408*
Present	7	2.44	1.33		3.64	1.21	
Absent	74	12.34	1.37		14.19	1.34	
Loss of 8p				0.0032			*0.0186*
Present	7	2.34	1.51		3.55	1.37	
Absent	72	12.68	1.39		14.22	1.31	
CGH profile				0.175			*0.0414*
Cluster 1	23	8.19	2.06		8.83	1.98	
All others	63	13.08	1.48		15.2	1.44	

NOTE. Statistical significance was calculated for disease-free and overall survival between different stratifications of clinicopathological, immunohistochemical, and CGH data by the log-rank test. Significant factors ($P < 0.05$) are highlighted in italics.

[a] ER, estrogen receptor; PR, progesterone receptor; CGH, comparative genomic hybridization.

38 of 55 (69%) CK14-positive and 26 of 31(84%) CK14-negative tumors, respectively, giving an overall error rate of 0.256 (Fig. 3*B*).

These branches could be additionally refined into 6 smaller clusters. When multivariate analysis was carried out using the Cox proportional hazards model, ER negativity, CK14 positivity, positive lymph nodes, and 1 cluster of tumors were found to be independent predictors of poor survival (Table 3). Within the large, predominantly CK14-positive branch, "cluster 1" was composed of 18 of 23 (78%) CK14 positive tumors and was found to indicate shorter overall survival ($P = 0.0270$) and disease-free time ($P = 0.0503$).

Univariate analyses showed that cluster 1 tumors predicted shorter overall survival for all of the tumors ($P = 0.0414$) and when the CK14-positive tumors were considered ($P = 0.0475$). Shorter disease-free times were also observed, but the differences were not statistically significant ($P = 0.175$, all tumors; $P = 0.127$, CK14-positive tumors). The Kaplan-Meier survival curves are shown in Fig. 4.

The cluster 1 tumors, as well as being predominantly CK14 positive (18 of 23; 79%; $P = 0.003$), were found to be associated with negative ER ($P = 0.0231$) and negative PR ($P = 0.039$) but not with ERBB2 ($P = 0.7458$), nodal status ($P > 0.9999$), or tumor size ($P = 0.6237$). The CGH alterations,

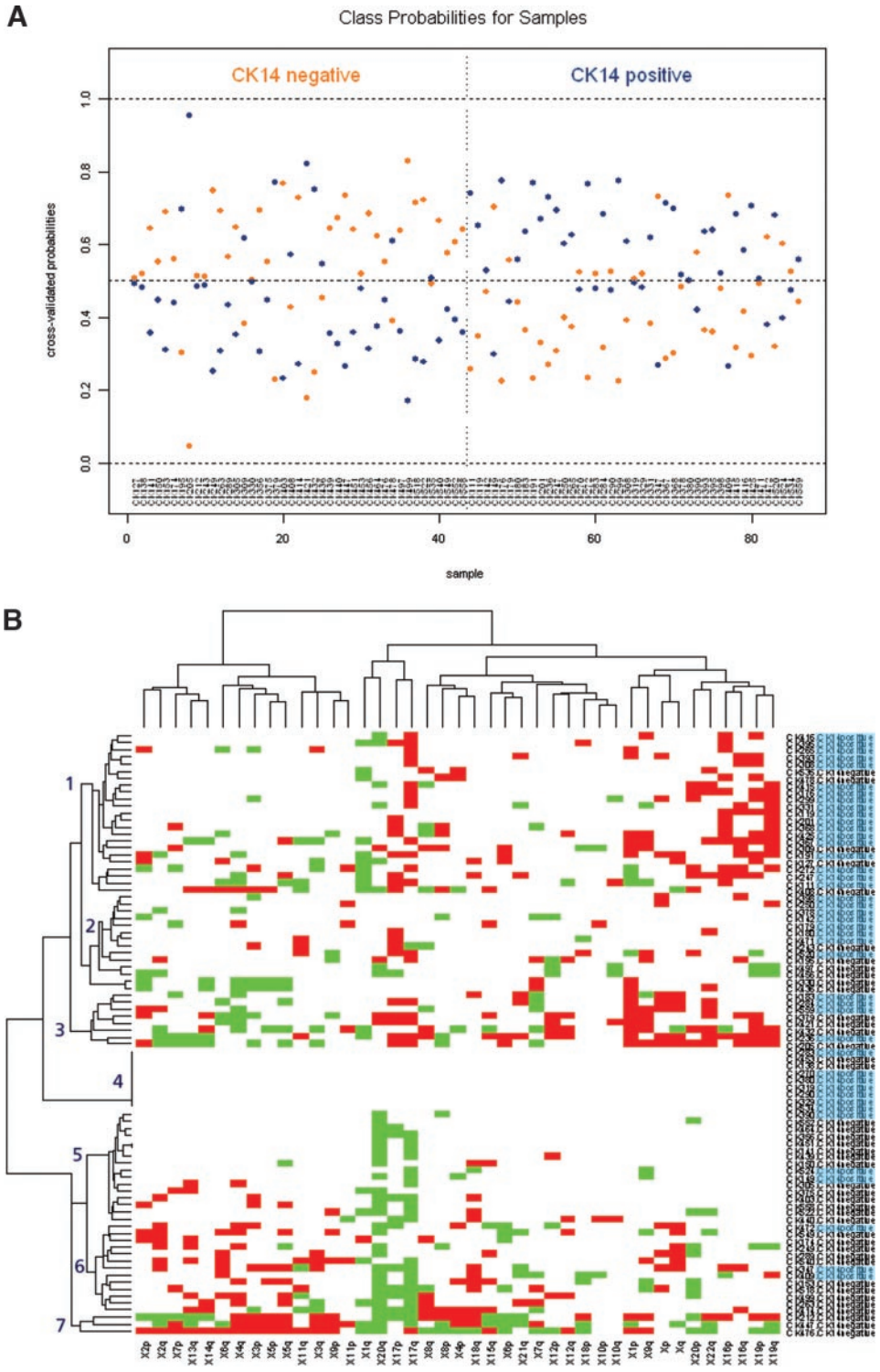

Fig. 3 Statistical analysis of comparative genomic hybridization (CGH) profiles across all tumors. *A*, supervised analysis using the nearest shrunken centroid classifier. The cross-validated probabilities of class assignment are shown for all tumors separated by their CK14 status. The overall cross-validated error rate was 0.24. *B*, unsupervised hierarchical clustering of the CGH data. Agglomerative clustering identifies two major branches of tumors (*horizontal*) according to patterns of gain and loss on individual chromosomal arms (*vertical*), separating the cases according to CK14 status in 74% of samples. Seven additional subdivisions are identified, with cluster 1 composed of 18 of 23 CK14-positive cases. *Green squares* indicate a gain by CGH; *red square* a loss. CK14-positive tumors are highlighted in *blue*.

which were found to be significantly associated with the cluster 1 tumors *versus* the rest, were gain of 1q ($P = 0.004$), and losses at 8p ($P = 0.0199$), 16p ($P < 0.0001$), 16q ($P = 0.0065$), 17p ($P = 0.0488$), 17q ($P = 0.0001$), 19p ($P = 0.001$), and 19q ($P = 0.0001$), and are shown in Table 4.

DISCUSSION

There is accumulating evidence to suggest that different histological grades of invasive ductal breast carcinomas may have distinct molecular origins and pathogenesis and do not

Table 3 Multivariate analysis of prognostic indicators

	Disease-free time		Death from breast cancer	
Factor	Hazard ratio (95% CI)[a]	*P**	Hazard ratio (95% CI)	*P**
ER negative	1.598 (1.417–1.853)	*0.0072*	1.682 (1.498–1.950)	*0.0060*
Positive nodes	1.661 (1.213–2.276)	*0.0016*	1.789 (1.280–2.501)	*0.0007*
CK14 negative	1.608 (1.426–1.865)	*0.0071*	1.570 (1.377–1.846)	*0.0140*
Cluster 1	1.980 (0.991–3.957)	*0.0530*	2.231 (1.097–4.537)	*0.0270*

NOTE. Independent factors that predict for disease-free and overall survival were calculated by the Cox proportional hazards model with simultaneous inclusion of all factors shown. Significant factors ($P < 0.05$) are highlighted in italics.

[a] CI, confidence interval; ER, estrogen receptor.

* Determined by Cox proportional hazards model.

typically progress from one grade group to another (28–31). The different grades have different clinical behaviors, and within-grade studies to identify the more aggressive subgroups of these classes of breast tumors would be of great assistance in clinical management. The expression of basal/myoepithelial markers has been observed in a proportion of grade III invasive breast tumors, and the spectrum of basal-like tumors, also recognized by morphology (15, 32), molecular cytogenetics (16, 33), and expression profiling (18, 19), has been associated with poor prognosis (17). CGH has the advantage of being applied to archival pathology specimens with long-term follow-up as well as being amenable to microdissection strategies to profile the molecular genetic change occurring in a pure population of tumor cells.

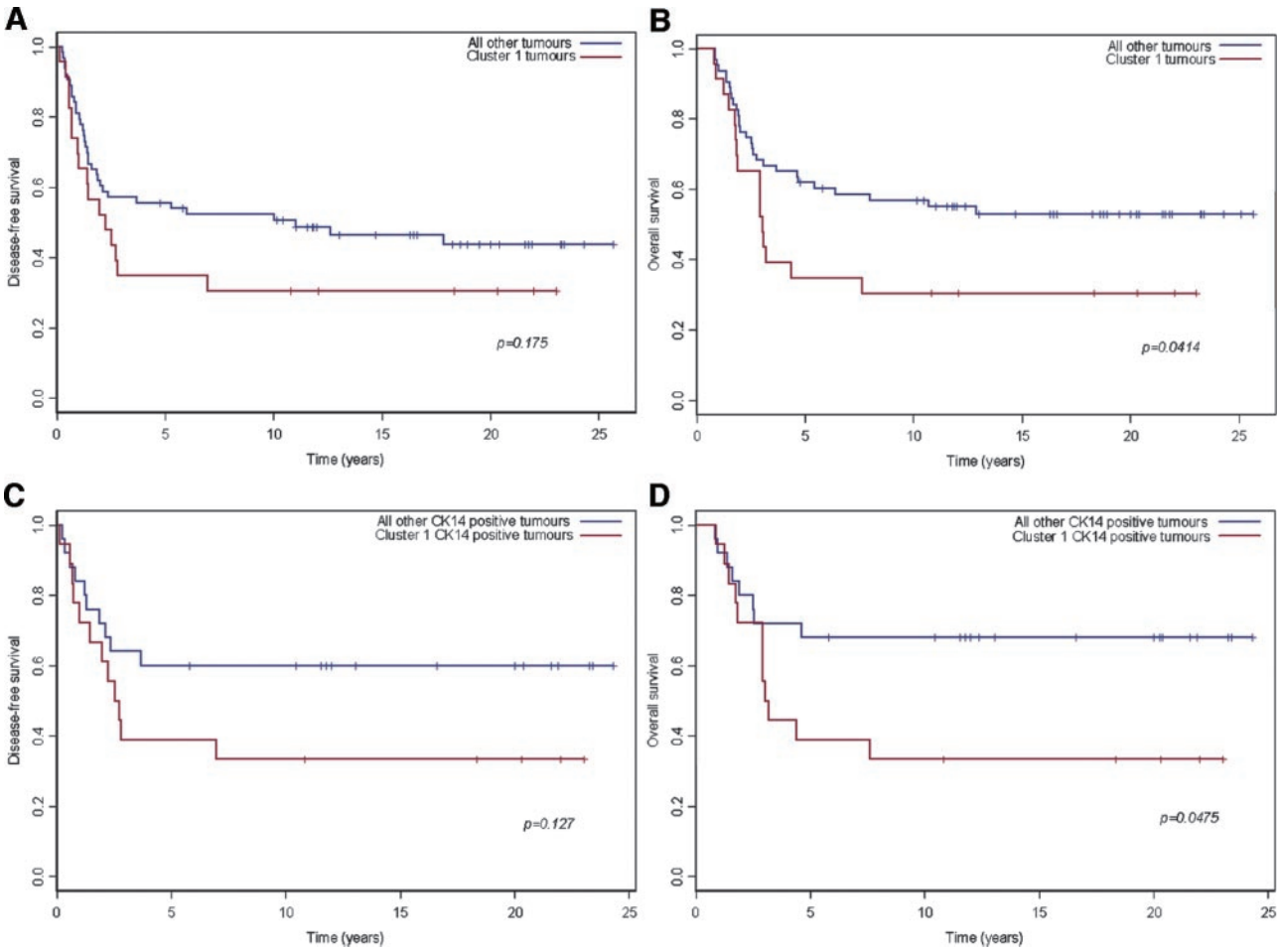

Fig. 4 Kaplan-Meier survival curves for subgroups of grade III breast tumors identified by cluster analysis of their comparative genomic hybridization profiles. *A* and *C*, disease-free survival; *B* and *D*, overall survival. *A* and *B*, cluster 1 tumors plotted against all others; *C* and *D*, CK14-positive cluster 1 tumors plotted against all other CK14-positive tumors. Statistical significance was calculated by the log-rank test.

Table 4 Association of specific immunohistochemical and CGH data with cluster 1 tumors

Factor	Cluster 1 tumors (%)	All other tumors (%)	*P* (Fisher's exact)
CK14 positive	18 (78.3)	25 (39.7)	*0.003*
ER[a] negative	19 (82.6)	34 (54.0)	*0.0231*
PR negative	15 (65.2)	32 (50.8)	*0.039*
1q gain	10 (43.5)	8 (12.7)	*0.004*
8p loss	5 (21.7)	2 (3.2)	*0.0199*
16p loss	13 (56.5)	3 (4.8)	*<0.0001*
16q loss	9 (39.1)	8 (12.7)	*0.0116*
17p loss	8 (34.8)	9 (14.3)	*0.0488*
17q loss	13 (56.5)	8 (12.7)	*0.0001*
19p loss	12 (52.2)	6 (9.5)	*0.0001*
19q loss	12 (52.2)	6 (9.5)	*0.0001*

NOTE. Numbers of tumors (and percentage) exhibiting a change in DNA copy number in cluster 1 and other tumors are given along with the *P* value as determined by Fisher's exact test to probe differences between the two groups. Significant ($P < 0.05$) differences in gains are highlighted in italics.

[a] ER, estrogen receptor; PR, progesterone receptor.

* Determined by Fisher's exact test.

In the cohort of grade III tumors in the present study, CK14 positivity was not associated significantly with prognosis. At the CGH level, the CK14-positive tumors showed fewer overall changes in DNA copy number than the CK14-negative group. This may in part explain the conflict with a previous study (33), which found a higher number of CGH alterations in the basal-like group of tumors as determined by CK5/6 positivity. This incongruity with published data suggesting that basal keratin expression confers a poorer prognosis must be interpreted in the light of the fact that we have focused only on poorly differentiated, high-grade malignancies. Within this cohort, ER negativity is not a significant prognostic indicator by univariate analysis, nor is tumor size or ERBB2 status. Interestingly, the few (3 of 83) tumors that were negative for the luminal epithelial keratin CK19 showed a very poor prognosis, statistically significant despite the small numbers.

The CK14-positive and -negative groups were clearly different in terms of their CGH profiles. In particular, the CK14-positive tumors showed an increased prevalence for losses at 16p, 17q, and 19q, all alterations associated with pure myoepithelial carcinomas (34). None of these alterations on their own conferred any prognostic information, although gain of 1q and 6p, as well as losses of 3p, 4p, and 8p did indicate shorter disease-free and overall survival times in the whole cohort. Taken as a whole, the CGH profiles of the tumors alone were able to predict the CK14 status in approximately three-quarters of cases by supervised analysis, using leave-one-out cross-validation. This demonstrates the inherent differences in the molecular evolution of the tumor groups.

Unsupervised hierarchical clustering has been applied to gene expression profiling data to identify subgroups of breast tumors with different clinical outcomes (19). Such a statistical approach may also be applied to CGH data, with the advantage in this instance that we are profiling pure populations of microdissected tumor cells. In our study, hierarchical clustering revealed two large branches, which predicted for CK status again with approximately three-quarters accuracy. Within the predominantly CK14-positive group, 4 separate clusters were identified, with the cluster 1 tumors accounting for 18 of 43 CK14-positive tumors. This subgroup of tumors had a worse prognosis than the rest of the tumors, both in terms of the whole data set, as well as that stratified purely by CK14 positivity. In multivariate analysis, this cluster of tumors was found to be an independent indicator of shorter overall survival in a model including ER, CK14, and nodal status, identified by the Cox proportional hazards test.

Substratification of CK14-positive breast tumors into two groups, based on their CGH profiles, which is reflected in their biological behavior, is of considerable clinical interest. This poor prognosis group was associated with negative hormone receptor status and exhibited a number of CGH alterations with higher prevalence than the better prognosis tumors, including gain at 1q, and losses at 8p, 16p, 16q, 17p, 17q, 19p, and 19q. The association of these loci with a poor clinical outcome group of tumors will provide clues for targeted studies hoping to unravel the underlying molecular events associated with the pathogenesis of these lesions.

The immunophenotype of these tumors, as has been pointed out in smaller studies (16), exhibits a resemblance to tumors with germ-line BRCA1 mutation (20), and this possible association has been postulated recently by gene expression analysis (21). Investigations into epigenetic mechanisms of BRCA1 inactivation in basal-like breast tumors seem warranted.

The clinical heterogeneity of grade III invasive ductal breast carcinomas is well-known. The idea that we could identify, at diagnosis, subgroups of these patients that will do badly (and thus require aggressive therapy) or relatively well (where the patient may be spared such treatment) is an attractive one. Clearly some form of basal/myoepithelial differentiation is apparent in these lesions, as determined by their expression of some basal markers, and a pattern of CGH alterations at loci associated with pure myoepithelial carcinomas. Accurate characterization of these lesions at the morphological level, along with an additional immunohistochemical refinement using an extensive panel of basal/myoepithelial markers will be necessary to produce a set of criteria that will allow the accurate diagnosis of these tumors, with the implications that will have on patient management. Unraveling the molecular pathways that drive the divergent groups of good- and poor-prognosis tumors will be required to identify potential targets for novel therapeutic strategies.

REFERENCES

1. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up.“ Histopathology 1991;19:403–10.

2. Gusterson BA, Warburton MJ, Mitchell D, Ellison M, Neville AM, Rudland PS. Distribution of myoepithelial cells and basement membrane proteins in the normal breast and in benign and malignant breast diseases Cancer Res 1982;42:4763–70.

3. Dairkee SH, Blayney C, Smith HS, Hackett AJ. Monoclonal antibody that defines human myoepithelium Proc Natl Acad Sci USA 1985;82:7409–13.

4. Altmannsberger M, Dirk T, Droese M, Weber K, Osborn M. Keratin polypeptide distribution in benign and malignant breast tumors: subdivision of ductal carcinomas using monoclonal antibodies.“ Virchows Arch B Cell Pathol Incl Mol Pathol 1986;51:265–75.

5. Nagle RB, Bocker W, Davis JR, Heid HW, Kaufmann M, Lucas DO, Jarasch ED. Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells J Histochem Cytochem 1986;34:869–81.

6. Dairkee SH, Ljung BM, Smith H, Hackett A. Immunolocalization of a human basal epithelium specific keratin in benign and malignant breast disease Breast Cancer Res Treat 1987;10:11–20.

7. Jarasch ED, Nagle RB, Kaufmann M, Maurer C, Bocker WJ. Differential diagnosis of benign epithelial proliferations and carcinomas of the breast using antibodies to cytokeratins Hum Pathol 1988;19:276–89.

8. Wetzels RH, Holland R, van Haelst UJ, Lane EB, Leigh IM, Ramaekers FC. Detection of basement membrane components and basal cell keratin 14 in noninvasive and invasive carcinomas of the breast Am J Pathol 1989;134:571–9.

9. Gould VE, Koukoulis GK, Jansson DS, Nagle RB, Franke WW, Moll R. Coexpression patterns of vimentin and glial filament protein with cytokeratins in the normal, hyperplastic, and neoplastic breast Am J Pathol 1990;137:1143–55.

10. Wetzels RH, Kuijpers HJ, Lane EB, et al. C. Basal cell-specific and hyperproliferation-related keratins in human breast cancer Am J Pathol 1991;138:751–63.

11. Heatley M, Maxwell P, Whiteside C, Toner P. Cytokeratin intermediate filament expression in benign and malignant breast disease J Clin Pathol 1995;48:26–32.

12. Santini D, Ceccarelli C, Taffurelli M, Pileri S, Marrano D. Differentiation pathways in primary invasive breast carcinoma as suggested by intermediate filament and biopathological marker expression J Pathol 1996;179:386–91.

13. Rejthar A, Nenutil R. The intermediate filaments and prognostically oriented morphological classification in ductal breast carcinoma. Neoplasma 1997;44:370–3.

14. Malzahn K, Mitze M, Thoenes M, Moll R. Biological and prognostic significance of stratified epithelial cytokeratins in infiltrating ductal breast carcinomas Virchows Arch 1998;433:119–29.

15. Tsuda H, Takarabe T, Hasegawa T, Murata T, Hirohashi S. Myoepithelial differentiation in high-grade invasive ductal carcinomas with large central acellular zones Hum Pathol 1999;30:1134–9.

16. Jones C, Nonni AV, Fulford L, et al. CGH analysis of ductal carcinoma of the breast with basaloid/myoepithelial cell differentiation Br J Cancer 2001;85:422–7.

17. van de Rijn M, Perou CM, Tibshirani R, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome Am J Pathol 2002;161:1991–6.

18. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours Nature 2000;406:747–52.

19. Sorlie T, Perou CM, Tibshirani R, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications Proc Natl Acad Sci USA 2001;98:10869–74.

20. Lakhani SR, Van De Vijver MJ, Jacquemier J, et al. The pathology of familial breast cancer: predictive value of immunohistochemical markers estrogen receptor, progesterone receptor, HER-2, and p53 in patients with mutations in BRCA1 and BRCA2. J Clin Oncol 2002;20:2310–8.

21. Sorlie T, Tibshirani R, Parker J, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 2003;10:8418–23.

22. Tsuda H, Takarabe T, Hasegawa F, Fukutomi T, Hirohashi S. Large, central acellular zones indicating myoepithelial tumor differentiation in high-grade invasive ductal carcinomas as markers of predisposition to lung and brain metastases. Am J Surg Pathol 2000;24:197–202.

23. Gillett CE, Miles DW, Ryder K, et al. Retention of the expression of E-cadherin and catenins is associated with shorter survival in grade III ductal carcinoma of the breast J Pathol 2001;193:433–41.

24. Barnes DM, Harris WH, Smith P, Millis RR, Rubens RD. Immunohistochemical determination of oestrogen receptor: comparison of different methods of assessment of staining and correlation with clinical outcome of breast cancer patients Br J Cancer 1996;74:1445–51.

25. Jones C, Damiani S, Wells D, Chaggar R, Lakhani SR, Eusebi V. Molecular cytogenetic comparison of apocrine hyperplasia and apocrine carcinoma of the breast Am J Pathol 2001;158:207–14.

26. Wells D, Sherlock JK, Handyside AH, Delhanty JD. Detailed chromosomal and molecular genetic analysis of single cells by whole genome amplification and comparative genomic hybridisation Nucleic Acids Res 1999;27:1214–8.

27. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression Proc Natl Acad Sci USA 2002;99:6567–72.

28. Buerger H, Otterbach F, Simon R, et al. Different genetic pathways in the evolution of invasive breast cancer are associated with distinct morphological subtypes J Pathol 1999;189:521–6.

29. Roylance R, Gorman P, Harris W, et al. Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer Cancer Res 1999;59:1433–6.

30. Buerger H, Mommers EC, Littmann R, et al. Ductal invasive G2 and G3 carcinomas of the breast are the end stages of at least two different lines of genetic evolution J Pathol 2001;194:165–70.

31. Roylance R, Gorman P, Hanby A, Tomlinson I. Allelic imbalance analysis of chromosome 16q shows that grade I and grade III invasive ductal breast cancers follow different genetic pathways J Pathol 2002;196:32–6.

32. Jimenez RE, Wallis T, Visscher DW. Centrally necrotizing carcinomas of the breast: a distinct histologic subtype with aggressive clinical behavior Am J Surg Pathol 2001;25:331–7.

33. Korsching E, Packeisen J, Agelopoulos K, et al. Cytogenetic alterations and cytokeratin expression patterns in breast cancer: integrating a new model of breast differentiation into cytogenetic pathways of breast carcinogenesis. Lab Invest 2002;82:1525–33.

34. Jones C, Foschini MP, Chaggar R, et al. Comparative genomic hybridization analysis of myoepithelial carcinoma of the breast. Lab Invest 80: 831–6, 2000.