

# Hierarchical Clustering Analysis of Tissue Microarray Immunostaining Data Identifies Prognostically Significant Groups of Breast Carcinoma

Nikita A. Makretsov,<sup>1</sup> David G. Huntsman,<sup>1</sup>  
Torsten O. Nielsen,<sup>1</sup> Erika Yorida,<sup>1</sup>  
Michael Peacock,<sup>1</sup> Maggie C. U. Cheang,<sup>1</sup>  
Sandra E. Dunn,<sup>2</sup> Malcolm Hayes,<sup>3</sup>  
Matt van de Rijn,<sup>5</sup> Chris Bajdik,<sup>4</sup> and  
C. Blake Gilks<sup>1</sup>

<sup>1</sup>Genetic Pathology Evaluation Centre of the Department of Pathology, and Prostate Research Centre of Vancouver General Hospital, British Columbia Cancer Agency and University of British Columbia; <sup>2</sup>Laboratory for Oncogenomic Research, Department of Pediatrics, British Columbia Research Institute for Children's and Women's Health, University of British Columbia; <sup>3</sup>Department of Pathology, and <sup>4</sup>Cancer Control Research Program, British Columbia Cancer Agency, Vancouver, Canada; and <sup>5</sup>Department of Pathology, Stanford University Medical Center, Stanford, California

## ABSTRACT

Prognostically relevant cluster groups, based on gene expression profiles, have been recently identified for breast cancers, lung cancers, and lymphoma. Our aim was to determine whether hierarchical clustering analysis of multiple immunomarkers (protein expression profiles) improves prognostication in patients with invasive breast cancer. A cohort of 438 sequential cases of invasive breast cancer with median follow-up of 15.4 years was selected for tissue microarray construction. A total of 31 biomarkers were tested by immunohistochemistry on these tissue arrays. The prognostic significance of individual markers was assessed by using Kaplan-Meier survival estimates and log-rank tests. Seventeen of 31 markers showed prognostic significance in univariate analysis ( $P \leq 0.05$ ) and 4 markers showed a trend

toward significance ( $P \leq 0.2$ ). Unsupervised hierarchical clustering analysis was done by using these 21 immunomarkers, and this resulted in identification of three cluster groups with significant differences in clinical outcome.  $\chi^2$  analysis showed that expression of 11 markers significantly correlated with membership in one of the three cluster groups. Unsupervised hierarchical clustering analysis with this set of 11 markers reproduced the same three prognostically significant cluster groups identified by using the larger set of markers. These cluster groups were of prognostic significance independent of lymph node metastasis, tumor size, and tumor grade in multivariate analysis ( $P = 0.0001$ ). The cluster groups were as powerful a prognostic indicator as lymph node status. This work demonstrates that hierarchical clustering of immunostaining data by using multiple markers can group breast cancers into classes with clinical relevance and is superior to the use of individual prognostic markers.

## INTRODUCTION

Breast cancer encompasses a highly heterogeneous group of tumors. Tumor size and grade and status of axillary lymph nodes have been established as prognostic indicators for breast cancer patients, with lymph node metastasis being the most powerful of these. Many other factors have been suggested as prognostic indicators, but none of them, individually, has consistently reached the prognostic significance of lymph node status in multivariate analysis (1).

Gene expression profiling has allowed stratification of breast tumors into clinically relevant groups through the use of expression levels of thousands of genes rather than single prognostic indicators. Novel prognostically relevant subsets of breast cancer that are not identifiable on routine light microscopy have been identified by this approach (2–7). The objective of this study was to apply analytical techniques, developed for gene expression profiling, to determine whether multiple immunohistochemical prognostic markers could be used to identify prognostically relevant groups of breast cancer patients, and to determine the optimal panel of immunomarkers necessary to define those groups.

## MATERIALS AND METHODS

**Case Series.** This study group comprised 438 women with primary invasive breast cancer who underwent surgery for breast cancer between 1974 and 1995 at Vancouver General Hospital. These were consecutive cases, and the presence of invasive breast carcinoma was the only selection criterion in this study. Outcome data were available for all of the patients, with median follow-up of 15.4 years (range, 6.3–26.6 years). The patients' ages ranged from 28 to 87 years with a mean age of 61 years. The patients were

Received 3/3/04; revised 4/22/04; accepted 4/28/04.

**Grant support:** Supported in part by an educational grant from Aventis Canada, by a National Cancer Institute of Canada (NCIC) Streams of Excellence grant, and by a grant from the Professional Development Fund of the Department of Pathology and Laboratory Medicine, Vancouver General Hospital. C. Bajdik, D. Huntsman, and T. Nielsen are scholars of the Michael Smith Foundation for Health Research. M. van de Rijn is supported by National Cancer Institute, NIH, grant CA85129. M. Peacock was a summer student, University of British Columbia Bachelor of Science Microbiology and Immunology program.

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.

**Requests for reprints:** Blake Gilks, Department of Pathology and Laboratory Medicine, Vancouver General Hospital, Room 1438 JPPN, 855 West 12th Avenue, Vancouver, BC, Canada V5M 1Z9. Phone: 604-875-4111, extension 63305; Fax: 604-875-4797; E-mail: bgilks@vanhosp.bc.ca.

©2004 American Association for Cancer Research.

Table 1 Antibodies and TMA scoring criteria

	Marker	Source	Dilution	Pretreatment
1	ER	DAKO	1:100	Steam 20 min EDTA
2	PR	DAKO	1:100	Steam 20 min EDTA
3	Her 2	DAKO	1:500	Steam 20 min, TRS
4	Her 3	Neomarker	1:400	Steam 30 min, citrate buffer (pH 6.0)
5	Her 4	Neomarker	1:160	Steam 30 min, citrate buffer (pH 6.0)
6	p53	DAKO	1:400	Pressure cooker in MV, 2 min
7	p63	Neomarker	1:1000	Steam 30 min, citrate buffer (pH 6.0)
8	Relaxin	Immune Diagnostics	1:100	Steam 30 min, citrate buffer (pH 6.0)
9	E2F1	Neomarker	1:200	Microwave 15 min, citrate buffer (pH 6.0)
10	Ki67	Novocastra	1:100	Pressure cooker in MV, 2 min
11	Cyclin E	Neomarker	1:320	Microwave, 30 min, citrate buffer (pH 6.0)
12	PTEN	*	1:300	Microwave, 30 min, citrate buffer (pH 6.0)
13	CA IX	[31]†	1:50	Steam 30 min, citrate buffer (pH 6.0)
14	HSP27	Novocastra	1:800	Steam 30 min, citrate buffer (pH 6.0)
15	IGFBP2	Santa Cruz	1:200	Steam 30 min, citrate buffer (pH 6.0)
16	IGFBP5	Santa Cruz	1:300	Steam 30 min, citrate buffer (pH 6.0)
17	Cox2	Cayman Chemical	1:100	Steam 20 min, citrate buffer (pH 6.0)
18	CK5/6	Chemicon	1:100	Pressure cooker in MV, 2 min
19	YB1	‡	1:2000	Steam 30 min, citrate buffer (pH 6.0)
20	E-cadherin	Intermedico-Zymed	1:150	Decloak 5 min TRS
21	TIP1	[32]§	1:80	Citrate buffer (pH 6.0), 5 min
22	CD3	Novocastra	1:600	Decloak 5 min, EDTA
23	CD10	Novocastra	1:50	Decloak 5 min TRS
24	CD20	DAKO	1:500	0.05% pronase
25	CD43	Quartett	1:10	Decloak 5 min TRS
26	CD68	DAKO	1:800	Decloak 5 min TRS
27	CD117	DAKO	1:50	None
28	Osteo-protegerin		1:2500	None
29	NSE	DAKO	1:100	None
30	Chromogranin A	Biogenex	1:1000	None
31	Synaptophysin	DAKO	1:200	None

treated by mastectomy or lumpectomy with (392 patients) or without (46) axillary lymph node dissection. One hundred twenty-six patients had histologically confirmed lymph node metastasis, and 266 had no lymph node metastasis. Tumor size information, measured in millimeters, was available for 366 patients; 217 patients had tumors  $\leq 20$  mm, and 149 had tumors  $> 20$  mm. Adjuvant therapy varied substantially during the period 1974 through 1995, and no information on individual adjuvant treatment was available. Ethical approval was obtained from the institutional ethical review board to perform this study.

**Tissue Microarrays and Immunostaining.** Tissue microarrays (TMAs) were designed as described previously (8, 9) by using two 0.6-mm tissue cores per case, taken from formalin-fixed, paraffin-embedded archival tumor blocks. All of the immunostains were done with standardized protocols.<sup>6</sup> The panel of antibodies consisted of those available in the Genetic Pathology Evaluation Centre that showed case-to-case variation within the series of breast carcinomas (*i.e.*, none of the tumors showed identical or near-identical staining patterns). The antibody choice was empirical, based on availability, suitability for paraffin-embedded archival material, and biological and clinical relevance to breast cancer, and was also partially driven by an attempt to reproduce the classification based on recent investigations in gene expression profiles in

breast cancer (Table 1; Refs. 2–7). Once new antibodies were selected, each antibody was titrated with four to five different dilutions (with at least a 2-fold difference between each dilution) on the whole-mount tissue sections, according to the manufacturer's recommendation (Table 1). If there were no well-established positive-control tissues, we used our in-house multitumor array to find model positive-control tissue. If signal-to-background ratio was not acceptable for the dilution tested, the pretreatment/incubation time/concentration were readjusted. Immunostains were scored semi-quantitatively by two pathologists considering either cytoplasmic or membranous staining intensity, or percentage of positive nuclei (Table 1). Any discrepancies were resolved with a multihead microscope. The higher score was considered as a final score in case of a difference between duplicate tissue cores. Scoring was done without knowledge of patient outcomes.

**Data Analysis.** For each patient, the date of breast cancer diagnosis, date of last follow-up, and vital status at last follow-up (*i.e.*, living or deceased) were recorded. For patients who died, the death due to breast cancer or death due to other causes was recorded, and the date of death was recorded as the last follow-up. For patients who were still alive, the date of last follow-up was the one recorded in their medical chart. We considered two outcomes: overall survival and disease-specific survival. Both outcomes were measured from a patient's date of the breast cancer diagnosis until her date of death. If a patient was alive at the date of last follow-up, the outcomes were censored. For patients who

<sup>6</sup> Internet address for standardized protocols: [www.gpec.ubc.ca](http://www.gpec.ubc.ca); "Research" and "Research protocols."

Table 1 Continued.

Staining pattern	Score 0	Score 1	Score 2	Score 3
Nuclear	10%	10 to <30%	30 to <70%	≥70%
Nuclear	<10%	10 to <30%	30 to <70%	≥70%
Membranous	Negative	Herceptest score 1	N/a	Herceptest scores 2 and 3
Cytoplasmic	Negative	Weak	Intermediate	Strong
Cytoplasmic	Negative	N/a	N/a	Positive
Nuclear	<10%	10 to <50%	N/a	≥50%
Nuclear	≤2 cells	>3 cells to <10%	N/a	≥10%
Extracellular matrix	Negative	Weak	Intermediate	Strong
Nuclear	<1%	1 to <10%	≥10 to 30%	>30%
Nuclear	<10%	10 to 30%	N/a	>30%
Nuclear	<10%	10 to 50%	N/a	>50%
Cytoplasmic	Negative	Weak	Intermediate	Strong
Membranous	≤1 cell	>1 cell to <30%	N/a	≥30%
Membranous	≤1 cell	>1 cell to <30%	N/a	≥30%
Cytoplasmic	Negative	Weak	Intermediate	Strong
Cytoplasmic	Negative	Weak	Intermediate	Strong
Cytoplasmic	Negative	Weak	Intermediate	Strong
Cytoplasmic	Negative	Weak	Intermediate	Strong
Cytoplasmic	Negative	Weak	Intermediate	Strong
Cytoplasmic	Negative	Weak	Intermediate	Strong
Cytoplasmic	Negative	Weak	Intermediate	Strong
Stromal T cell	<10 cell	N/a	N/a	≥10 cell
Stromal fibroblasts	Negative	Weak	Intermediate	Strong
Stromal B cell	<10 cell	N/a	N/a	≥10 cell
Stromal T cell	<10 cell	N/a	N/a	≥10 cell
Stromal macrophages	<10 cell	N/a	N/a	≥10 cell
Stromal mast cell	Negative	N/a	N/a	≥1 cell
Nuclear	Negative	10 to 50%	N/a	>50%
Cytoplasmic	Negative	Weak	Intermediate	Strong
Cytoplasmic	Negative	Weak/focal	Intermediate	Strong
Cytoplasmic	Negative	Weak/focal	Intermediate	Strong

Note. Manufacturer's locations are, as follows: DAKO (Glostrup, Denmark); Neumarker (Fremont, CA); Immune Diagnostics (Buffalo, NY); Novocastra (Newcastle, UK); Santa Cruz (Santa Cruz, CA); Cayman Chemical (Ann Arbor, MI); Chemicon (Temecula, CA); Intermedico-Zymed (San Francisco, CA), Quartett (Berlin, Germany), and Biogenex (San Ramon, CA).

Abbreviations: TRS, Tris; N/a, not applicable; MV, microwave; NSE, neuron-specific enolase.

\* The Anti-PTEN is a polyclonal antibody provided as a gift from Dr. Heung-Chin Cheng, University of Melbourne, Melbourne, Victoria, Australia.

† A gift from Dr. Stephen Chia, British Columbia Cancer Agency, Vancouver, BC, Canada.

‡ A gift from Dr. Colin Nelson, Prostate Centre of Vancouver General Hospital, Vancouver, BC, Canada.

§ A gift from Dr. Calvin Roskelley, University of British Columbia, Vancouver, BC, Canada.

|| Gift from the University of Sheffield, Sheffield, United Kingdom.

died of breast cancer, disease-specific survival was from the date of breast cancer diagnosis until the date of death. If a patient died of some other cause, disease-specific survival was censored at the date of death.

Univariate analyses were performed by using Kaplan-Meier estimates and log-rank tests, with raw score data obtained for each individual biomarker (according to the empirical score system described in Table 1). The binarization of score data for the purpose of correlational and multivariate analysis (*i.e.*, determining "positive" versus "negative" staining results) was done with historically established cutoffs for 27 of 31 markers. For four biomarkers (IGFBP2, IGFBP5, TIP1, and HSP27) binary cutoff points were found by testing multiple combinations of ordinal score groups (based on a separate test of equality of all factor levels for each stratum), *e.g.*, 0 versus 1, 2, 3 or 0, 1, 2 versus 3, but not 0, 2 versus 1, 3, and so forth) to measure the clinical effect of expression levels of these proteins to disease-specific survival. Clustering analysis was based on the complete dataset and not on the binary results.

Multivariate analyses were performed with a proportional

hazard model (*i.e.*, Cox regression) and a backward stepwise method to remove variables from the model. A significant difference was declared if the *P* value from a two-sided test was less than 0.05 and a near-significant result was declared if the *P* value was less than 0.2.

We used unsupervised hierarchical clustering analysis to organize TMA score data (*i.e.*, the results of immunostaining) into meaningful structures, applying the same approach that has previously been adopted for cDNA microarrays (10, 11) and has also been applied to TMA data (9). Clustering analysis organizes cases according to the similarity or dissimilarity of immunostaining profiles, placing the cases with similar immunoprofiles together as neighboring rows in the clustergram. The relationship between cases and immunomarkers is depicted graphically as a dendrogram in which branch length is determined by the correlation between immunostaining results. All raw score data for each biomarker (as shown in Table 1) were used for clustering analysis. Only cases with immunostaining data for 80% or more of the markers under consideration were entered into clustering analysis.  $\chi^2$  tests were used to determine

Table 2 Prognostic markers, clinical variables, and cluster groups in univariate analysis of disease-specific survival

	Marker or Factor (cutoff)*	No. of interpretable cases	Significance ( <i>P</i> value)
1	PR (score 0 vs. 1, 2, and 3)	357	<i>P</i> = 0.007, positive factor
2	ER (score 0 vs. 1, 2, and 3)	324	<i>P</i> = 0.005, positive factor
3	HER2 (score 0 and 1 vs. 3)	337	<i>P</i> = 0.03, negative factor
4	p53 (score 0 vs. 1 and 3)	309	<i>P</i> = 0.001, negative factor
5	Her3 (scores 0 and 1 vs. 2 and 3)	278	<i>P</i> = 0.01, negative factor
6	Cox2 (scores 0 and 1 vs. 2 and 3)	356	<i>P</i> = 0.003, negative factor
7	NSE (score 0 vs. 1, 2, and 3)	335	<i>P</i> = 0.04, positive factor
8	Relaxin (score 0 vs. 1, 2, and 3)	361	<i>P</i> = 0.02, † positive factor
9	Ki67 (score 0 vs. 1, 2, and 3)	319	<i>P</i> = 0.02, negative factor
10	CA IX (score 0 vs. 1 and 2)	322	<i>P</i> = 0.05, negative factor
11	TIP1 (scores 0, 1, and 2 vs. 3)	272	<i>P</i> = 0.01, negative factor
12	PTEN (scores 0 vs. 1, 2, and 3)	344	<i>P</i> = 0.003, † positive factor
13	IGFBP2 (scores 0 and 1 vs. 2 and 3)	301	<i>P</i> = 0.04, negative factor
14	IGFBP5 (scores 0 and 1 vs. 2 and 3)	301	<i>P</i> = 0.05, negative factor
15	Mast cell (CD117) (score 0 vs. 1)	346	<i>P</i> = 0.003, positive factor
16	CD10 (scores 0, 1, and 2 vs. 3)	300	<i>P</i> = 0.008, negative factor
17	YB1 (scores 0, 1, and 2 vs. 3)	305	<i>P</i> = 0.01, negative factor
18	CD68 (scores 0 and 1 vs. 2 and 3)	331	NS ( <i>P</i> = 0.074)
19	HSP27 (score 0 vs. 1, 2, and 3)	300	NS ( <i>P</i> = 0.14)
20	p63 (score 0 vs. 1, 2, and 3)	376	NS ( <i>P</i> = 0.2)
21	CK5/6 (score 0 vs. 1, 2, and 3)	350	NS ( <i>P</i> = 0.06)
22	Lymph node metastasis	392	<i>P</i> < 0.0001, negative factor
23	Tumor size	366	<i>P</i> < 0.0001, negative factor
24	Tumor grade	438	NS ( <i>P</i> = 0.08), negative factor
25	Cluster groups based on 21 markers ‡	281	NS ( <i>P</i> = 0.15)
26	Cluster groups based on 20 markers ‡	277	<i>P</i> = 0.0025
27	Cluster groups based on 19 markers	267	<i>P</i> = 0.0003 ‡ ( <i>P</i> = 0.0014) §
28	Cluster groups based on 18 markers ‡	272	<i>P</i> = 0.0012
29	Cluster groups based on 17 markers ‡	273	NS ( <i>P</i> = 0.18)
30	Cluster groups based on 11 markers ‡	294	<i>P</i> < 0.0001 ( <i>P</i> = 0.0001) §

Abbreviations: NS, nonsignificant (*P* > 0.05); NSE, neuron-specific enolase.

\* Survival analysis was done on binary data using the cutoff points identified.

† Reached significance in node-negative subset of patients only; *P* < 0.2 for all patients.

‡ Cluster group 1 versus 2 and 3.

§ Cluster group 1 versus 2 versus 3.

which markers contributed to the formation of cluster groups. The agreement in classification of cases based on different sets of immunomarkers was assessed with the kappa statistic. A kappa value of 0.41 to 0.6 indicates moderate agreement, 0.61 to 0.8 substantial agreement, and more than 0.8 near-perfect agreement (12).

Software for TMA data management was used [Deconvoluter version 1.04, EXCEL for Windows; Liu *et al.* (11)]. Unsupervised hierarchical clustering analysis with average and complete linkage algorithms was done with GeneCluster, and graphical representation of the results of clustering was done with TreeView.<sup>7</sup> SPSS for Windows version 11.0 (Chicago, IL) was used for statistical analysis of the data.

## RESULTS

**Univariate Analysis of Basic Clinicopathologic Variables (Tumor Size, Nodal Status, and Tumor Grade) and Immunostaining Data.** Positive nodal status and larger tumor size (>20 mm) showed a highly significant correlation with poor outcome; Nottingham grade did not show a statistically

significant correlation with outcome in our series, although tumors with higher grade showed a trend toward a worse outcome (Table 2).

A set of 31 biomarkers was tested by immunohistochemistry on the TMAs. A total of 10,041 (74.5%) of a maximum possible number of 13,454 paired tissue core sections produced interpretable immunostains and were scored semiquantitatively, as shown in Table 1. Missing immunostain data were usually caused by loss of a core from the section, less commonly by exhaustion of tumor material in the core. The list of 21 markers that were prognostically significant (*P* ≤ 0.05) or approached significance (*P* ≤ 0.2) in univariate analyses is shown in Table 2. The remaining 10 markers were not prognostically significant (*P* > 0.2). Examples of representative immunostaining results for two cases, with four immunomarkers per case, are shown as Fig. 1. Results of immunostaining with 11 immunomarkers per case, for all cases, can be viewed online at <http://gpec.bliss.ubc.ca> (under "view/Clustering Breast Cancer").<sup>8</sup>

<sup>7</sup> Both GeneCluster and TreeView are available at <http://rana.lbl.gov/EisenSoftware>.

<sup>8</sup> Results of immunostainings, with 11 immunomarkers per case, for all cases can be viewed online at <http://gpec.bliss.ubc.ca> (under "view/Clustering Breast Cancer").

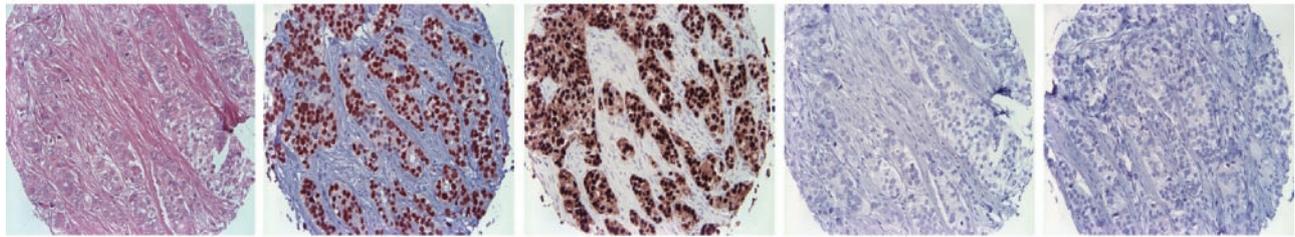
**Case 1**

Fig. 1 Representative immunostaining of two cases with hematoxylin and eosin and four markers per case.

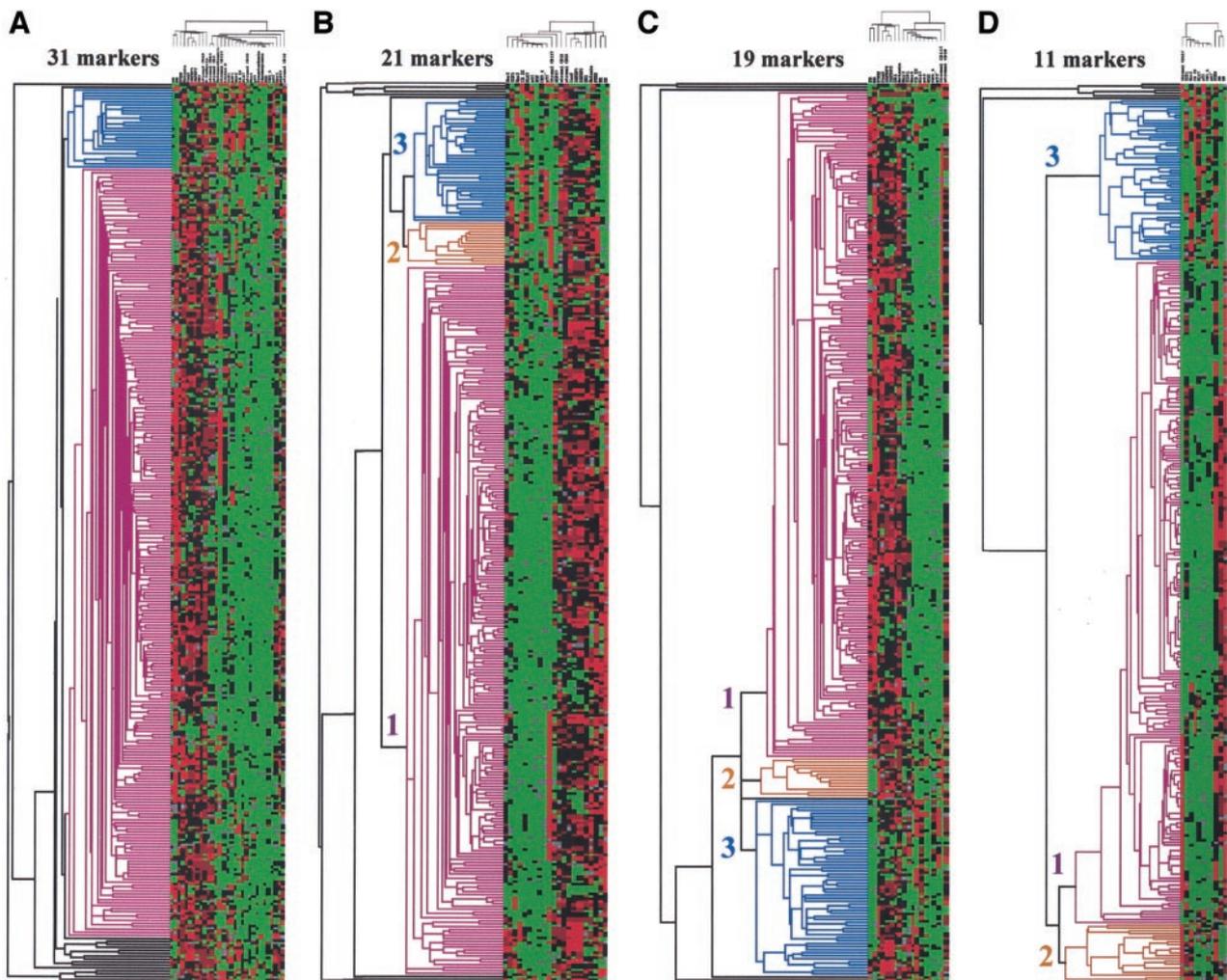
**Unsupervised Hierarchical Clustering Analysis of Immunostaining Data.** Unsupervised hierarchical clustering analysis (average linkage method) was applied to the dataset of 31 immunomarkers. This did not produce a dendrogram with well-defined cluster groups of cases (Fig. 2A). The ten markers, which did not show significance or trend toward significance in univariate analysis were then excluded from the clustering model. Clear separation of cases into three distinct groups with large linkage distances was apparent when clustering was done with the 21 markers listed in Table 2 (Fig. 2B). To determine the stability of the cluster groups and to possibly reduce the number of markers needed to define cluster groups, we clustered the data by using, in turn, 20, 19 (Fig. 2C), 18, and 17 markers by stepwise exclusion of CD68, HSP27, p63, and CK5/6 [markers that showed a trend toward significance ( $P \leq 0.2$ ) in univariate analysis]. Prognostically significant cluster groups were identified by clustering that was based on 20, 19, and 18 markers. The most significant survival differences between cluster groups were seen with the set of 19 markers (Table 2).

It is noteworthy that the removal of the basal-like breast cancer marker CK5/6 (13, 14) from the clustering analysis when the analysis went from 18 to 17 markers resulted in the identification of cluster groups with no significant differences in survival ( $P = 0.18$ ). CK5/6 did not quite reach prognostic significance in univariate analysis ( $P = 0.06$ ) but contributed significantly to accurate classification of cases into “good” and “bad” prognosis cluster groups. CK5/6-positive cases accumulated in cluster group 3, associated with the worst prognosis, as shown in Table 3.

**Determination of a Minimal Immunomarker Set to Define Cluster Groups.** As has been shown, the cluster groups produced by 19 markers were of highest prognostic significance, when compared with the cluster groups formed by 21, 20, 18, and 17 markers. To define the minimal set of markers necessary to identify prognostically significant cluster groups, we performed  $\chi^2$  tests, where each marker was measured across the

defined three cluster groups formed by the set of 19 markers. We found that Her3, Cox2, NSE, Relaxin, CD10, YB1, IGFBP2, and IGFBP5 showed no significant differences in staining results among the three cluster groups, whereas the remaining 11 markers (ER, PR, Her2, p53, Ki67, CA IX, TIP1, stromal CD117, PTEN, p63, and CK5/6) showed statistically significant difference in distribution between the three cluster groups (Table 3). Hierarchical clustering was performed with the set of these 11 markers (Fig. 2D). The overall concordance between assignment of individual cases to one of the three cluster groups formed when 19 *versus* 11 markers were used, respectively, was near perfect ( $\kappa = 0.89$ ). There were highly significant differences in disease-specific survival and overall survival among the 3 cluster groups defined with 11 markers (Fig. 3A and B). Cluster groups 2 and 3 were characterized by predominance of ER- and PR-negative breast cancers, whereas cluster group 1 consisted of tumors that were uniformly ER positive and PR positive in a majority of cases (Table 3). Accordingly, comparison was made between survival differences based on cluster group designation and ER/PR status (Fig. 3C and D). The differences in survival of patients in different cluster groups was greater than was seen with ER and/or PR alone. Also, there was a highly significant positive correlation between cluster groups and tumor grade ( $P < 0.001$ ) but only a trend to correlation with tumor size ( $P = 0.07$ ). No correlation was observed with lymph node status or patient age.

**Assessment of Reproducibility of Cluster Groups by Application of Complete Linkage Clustering Algorithm.** To assess variation in cluster group membership designation when using different hierarchical clustering methods, we applied complete linkage clustering algorithm to the data set of 11 markers. The concordance between designation of individual cases to one of three cluster groups using average linkage *versus* complete linkage clustering algorithms was near perfect ( $\kappa = 0.92$ ), with only 9 mismatches of 269 paired cases.



**Fig. 2** Clustergrams graphically depicting unsupervised hierarchical clustering analysis with 31(A), 21(B), 19(C), and 11(D) immunomarkers. Each row, a single case; each column, a single immunomarker. Green, negative immunostaining (score 0); black, weak immunostaining (score 1); brown, moderately intense immunostaining (score 2); red, strong immunoreactivity (score 3); gray, missing data. The dendrograms on the left of each clustergram, the relatedness of the immunoprofiles of individual cases; the longer the horizontal dendrogram arm, the greater the difference in immunoprofiles between individual cases inside a cluster group. The small dendrograms on the top of each clustergram, the relatedness of the immunoprofiles of individual biomarkers; the longer the horizontal dendrogram arm, the greater the difference in immunoprofiles between individual biomarkers. Numbers and different colors, cluster groups 1 (purple), 2 (brown), and 3 (blue). Black long horizontal arms, unclassifiable cases.

**Multivariate Analysis of Cluster Groups and Clinicopathologic Variables.** Multivariate analysis showed independence of cluster groups defined by 19 (as well as by a reduced set of 11) markers from nodal status and tumor size for both disease-specific survival and overall survival. The relative risk of death from breast cancer for patients in cluster group 1 versus groups 2 and 3, combined, was comparable with the relative risk associated with positive lymph node status (Table 4). Another multivariate model included cluster groups of 11 markers, ER, and Her2, and the clinicopathologic variables of lymph node status, tumor size, tumor grade, and patient age. After six cycles of iteration, when the least significant variables were eliminated, only cluster groups and lymph node status ( $P = 0.001$ ) remained of independent prognostic significance with respect to disease-specific survival. When overall survival

was considered, cluster groups ( $P = 0.002$ ), lymph node status ( $P = 0.039$ ), and tumor size ( $P = 0.008$ ) remained of independent significance. None of the immunomarkers, used in isolation, achieved the prognostic significance of cluster groups in terms of disease-specific or overall survival.

## DISCUSSION

Invasive breast cancer is the most common carcinoma in North American women (1), and the prognosis is favorable if it is detected at an early stage. Significant improvements in survival have been realized through the combined effects of mammographic screening and adjuvant treatment (15). Continuing improvement will require development of more refined and effective treatment strategies, and these will probably be based

Table 3 Impact of individual immunomarkers on cluster group determination ( $\chi^2$  test results)

Marker	n (within the tissue arrays)	Distribution of the results by cluster groups			P
		Cluster group 1	Cluster group 2	Cluster group 3	
<b>ER, positive</b>	202/253	194/194 (100%)	2/8 (25%)	6/51 (11.8%)	< <b>0.001</b>
<b>PR, positive</b>	162/265	151/200 (75.5%)	2/10 (20%)	9/55 (16.3%)	< <b>0.001</b>
<b>CK5/6, positive</b>	44/254	14/191 (7.3%)	4/9 (44.4%)	26/54 (48.2%)	< <b>0.001</b>
<b>p63, positive</b>	34/260	17/195 (8.7%)	4/9 (44.4%)	13/56 (23.3%)	< <b>0.001</b>
YB1, strongly positive	116/262	82/196 (41.9%)	5/10 (50%)	29/56 (51.7%)	0.8
CD10, strongly positive	84/239	56/140 (40%)	3/9 (33.3%)	25/51 (49%)	0.4
IGFBP5, strongly positive	94/265	68/199 (34.2%)	6/10 (60%)	20/56 (35.7%)	0.7
IGFBP2, strongly positive	129/263	99/263 (50.2%)	6/10 (60%)	24/54 (42.8%)	0.5
<b>CD117 (mast cell), positive</b>	57/261	46/199 (23.6%)	9/10 (90%)	2/56 (3.6%)	< <b>0.001</b>
<b>PTEN, negative</b>	35/260	23/197 (11.7%)	0/9 (0%)	12/54 (22.2%)	<b>0.02</b>
<b>TIP1, strongly positive</b>	12/232	5/176 (2.8%)	0/10 (0%)	7/47 (14.9%)	< <b>0.001</b>
<b>Ki67, positive</b>	82/263	35/197 (17.7%)	4/10 (40%)	45/56 (82.3%)	< <b>0.001</b>
<b>CA IX, positive</b>	49/265	14/200 (7%)	1/10 (10%)	34/55 (61.8%)	< <b>0.001</b>
NSE, positive	12/261	11/195 (5.6%)	1/10 (10%)	0/56 (0%)	0.5
Relaxin, positive	218/259	163/194 (84%)	9/10 (90%)	46/55 (83.7%)	0.3
Cox2, strongly positive	103/261	77/198 (38.9%)	6/8 (75%)	20/55 (36.4%)	0.1
<b>p53, positive</b>	38/248	12/186 (6.4%)	4/10 (40%)	22/52 (42.3%)	< <b>0.001</b>
Her3, strongly positive	37/250	28/189 (14.8%)	5/10 (50%)	27/51 (42.9%)	0.6
<b>Her2, strongly positive</b>	30/257	12/192 (6.3%)	4/10 (40%)	14/55 (21.4%)	< <b>0.001</b>

NOTE. Bold font has been used for markers that contributed significantly to cluster group designation, and their respective *P* values (*P* < 0.05).

on advances in molecular biology and pathology (16, 17). Numerous biological markers have been proposed to be prognostic markers in invasive breast carcinoma; however, stratification of tumors into prognostic groups to guide therapeutic decisions is based mainly on tumor stage and grade and on assessment of ER, PR, and Her2 (18). Lymph node status remains the most powerful prognostic indicator, and prognostication is especially problematic for lymph node-negative invasive breast cancer, in which a decision about the advisability of adjuvant chemotherapy requires accurate risk assessment (18).

The potential for combinations of prognostic markers to be superior to any single marker has been observed previously (19, 20). Clinical oncologists have been using Tumor-Node-Metastasis (TNM) classification for decades (21), which is a form of supervised clustering analysis combining tumor size, lymph node status, and distant metastasis. Unsupervised hierarchical clustering analysis has been used to classify tumors based on mRNA expression levels of thousands of genes. Prognostically relevant cluster groups have been identified for breast cancer, lymphoma, and lung cancer with this approach (2, 3, 6, 16). There have been attempts to classify breast cancers based on hierarchical clustering analysis of immunomarker data, but the prognostic significance of the cluster groups identified is unclear because of limited available outcome data or none (22, 23).

We applied unsupervised (*i.e.*, without consideration of other histologic and clinical variables) hierarchical clustering analysis to immunostaining data and identified prognostically significant cluster groups. We were able to identify a minimal set of 11 biomarkers necessary to define these cluster groups. ER and PR status of tumors were the major determinants of cluster group designation in hierarchical clustering analysis; however, the prognostic value of these two markers in either univariate or multivariate analyses was much less significant than for the cluster groups formed by multiple immunomarkers. Although this establishes that consideration of multiple markers

is superior to single markers, the basis for this superiority is important to consider. A small proportion of ER- and/or PR-positive tumors clustered along with ER-/PR-negative cases, which suggests that expression of multiple negative prognostic markers associated with tumor aggressiveness in these tumors overrides the effect of positive prognostic factors such as ER. All of the patients were treated at the British Columbia Cancer Agency according to provincial treatment guidelines; however, management varied, based on the year of diagnosis, and detailed treatment information is not available. This is a limitation of this study.

Sorlie *et al.* (2, 24), van 't Veer *et al.* (3), and van de Vijver *et al.* (25) independently identified subtypes of invasive breast cancer based on independent sets of gene expression data. Sorlie *et al.* (2, 24) were able to consistently identify basal-like, Her2-overexpressing, luminal type and normal-breast-like tumor subgroups with, respectively, worst, poor, intermediate, and good prognosis with hierarchical clustering analysis. Nevertheless, it was not possible to classify a significant proportion (6–36%) of breast cancers into these categories (2), which raised questions about further validation of this newly proposed classification system and its application to practice. Although cDNA microarray technology is widely used in cancer research, it is still far from clinical implementation because of the cost of the assay, the necessity of validation of initial findings, and the lack of standardization of protocols. Immunohistochemistry is routinely available in clinical laboratories and has proved to be a reliable and reproducible ancillary method in anatomic pathology. Application of TMAs has dramatically reduced the cost and time required for the testing of multiple biomarkers on large series of cases (26, 27). TMAs have been validated as useful tools for the study of prognostic markers in breast cancer (28); the concordance between TMAs and whole-section immunostains had been consistently high (22, 28, 29), and the prognostic value of

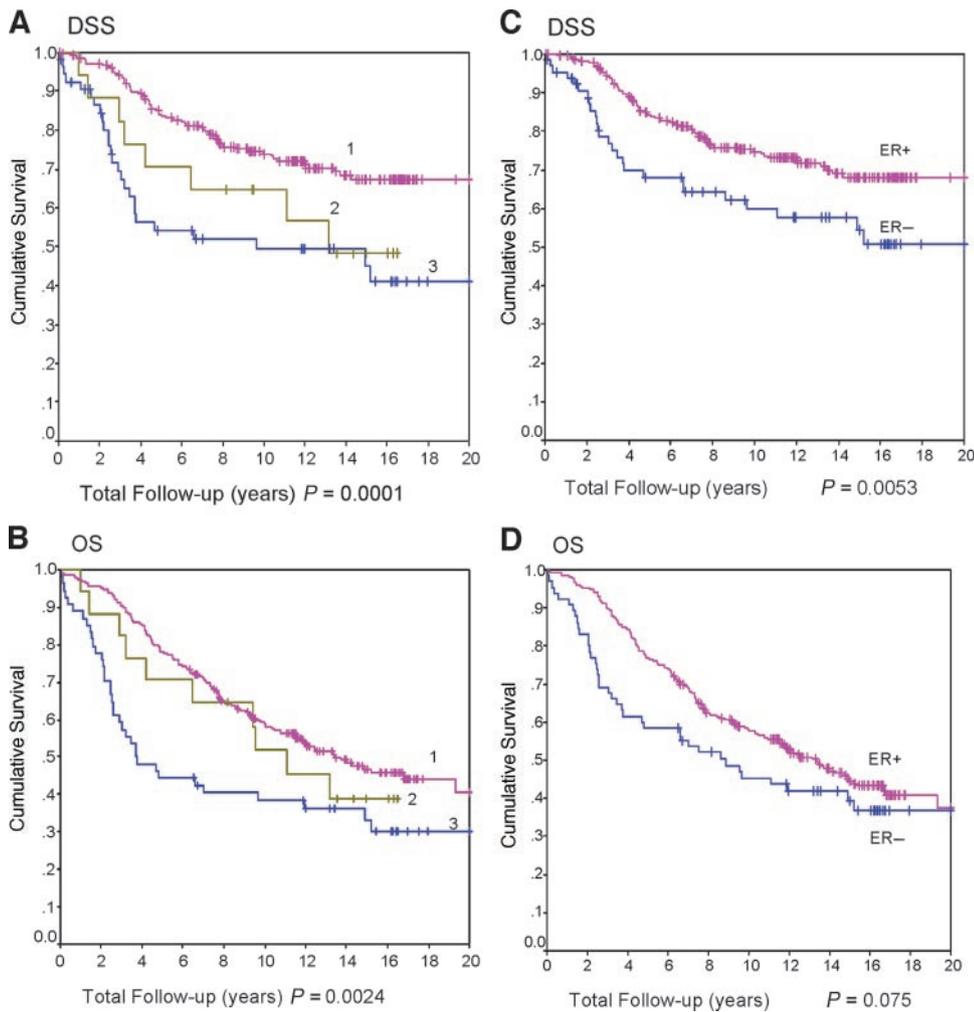


Fig. 3 Disease-specific survival (DSS; A) and overall survival (OS; B) for patients in cluster groups 1, 2, and 3, based on clustergram with 11 immunomarkers. DSS (C) and OS (D) for patients with positive (ER+) versus negative (ER-) estrogen receptor protein status.

markers when assessed with TMAs has been as good as, or superior to, that when whole sections were used (30).

Clinical application of clustering analysis of immunomarker data in breast cancer could allow combinations of prognostic markers to be used simultaneously, rather than the traditional

approach of assessment of prognosis based on single clinical facts. This could be done, for less cost than is currently incurred in the assessment of ER, PR, and Her-2, by batching clinical cases onto small TMAs, run once weekly. Each clinical center could have its own "training set" of cases with monitored

Table 4 Multivariate analysis of cluster group designation, tumor size, and lymph node status

Model and variables	Disease-specific survival		Overall survival	
	P	Relative risk (99% CI)	P	Relative risk (99% CI)
Cluster groups (19 markers) and clinical parameters				
Cluster groups*	0.001	2.5 (1.2–5)	0.002	1.9 (1.1–3.4)
Lymph node status	0.005	2.2 (1.1–4.5)	0.007	1.8 (1–3.1)
Tumor size	0.4	1.3 (0.6–2.6)	0.052	1.5 (0.9–2.6)
Cluster groups (11 markers) and clinical parameters				
Cluster groups*	<0.001	2.9 (1.5–5.7)	<0.001	2.2 (1.3–3.8)
Lymph node status	0.001	2.4 (1.2–4.6)	0.005	1.8 (1–3)
Tumor size	0.11	1.5 (0.8–3)	0.021	1.6 (0.9–2.8)

Abbreviations: CI, confidence interval.

\* Group 1 versus 2 and 3 (as shown on Fig. 2C and D).

long-term outcome. New cases, then, could be added one at a time to this training set and clustered, allowing assignment to one of the previously defined prognostic groups. Another advantage of this approach is that, if better immunohistochemical markers appear, they can be easily added to this model [for instance, more specific basal markers than are currently available (13)] to potentially improve the training set.

We conclude that it is possible to identify prognostic cluster groups of invasive breast cancer patients with maximal differences in survival, by the clustering of immunostaining data, based on a panel of 11 prognostic markers. Our data show that the application of a small panel of markers identifies prognostic profiles for individual cancers and is independent of the major clinical indicators of prognosis, *i.e.*, tumor size and lymph node status. From a clinical perspective, it is of critical importance to determine whether there are specific patient populations for whom it is reasonable to avoid the administration of cytotoxic chemotherapy. Limited information is available to answer this important question (18). Classification of breast cancer cases into prognostic cluster groups may aid in the individualization of adjuvant therapy, especially for a group of patients with node-negative status. Validation of this approach will require testing a sufficiently large (population-based) series, to allow analysis of subsets of patients with unclear prognosis. Quantification and further standardization of immunohistochemical analysis will also be an asset for future studies.

## ACKNOWLEDGMENTS

We are grateful to Shahriar Dabiri, Bev Dupuis, Dmitry Turbin, Jill Kucab, and Hamid Masoudi for the technical assistance with immunostains and data collection.

## REFERENCES

- World Health Organization classification of tumors. In: Tavassoli FA and Devilee P, editors. Pathology and genetics of tumors of the breast and female genital organs. Lyon: IACR Press; 2003.
- 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;100:8418–23.
- van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature (Lond)* 2002;415:530–6.
- Martin KJ, Kritzman BM, Price LM, et al. Linking gene expression patterns to therapeutic groups in breast cancer. *Cancer Res* 2000;60:2232–8.
- Iwao K, Matoba R, Ueno N, et al. Molecular classification of primary breast tumors possessing distinct prognostic properties. *Hum Mol Genet* 2002;11:199–206.
- Ahr A, Karn T, Solbach C, et al. Identification of high risk breast-cancer patients by gene expression profiling. *Lancet* 2002;359:131–2.
- Ahr A, Holtrich U, Solbach C, et al. Molecular classification of breast cancer patients by gene expression profiling. *J Pathol* 2001;195:312–20.
- Makretsov N, Gilks CB, Coldman A, Hayes M, Huntsman D. Tissue microarray analysis of neuroendocrine differentiation and its prognostic significance in breast cancer. *Hum Pathol* 2003; 34:1001–8.
- Hsu FD, Nielsen TO, Alkushi A, et al. TMAs are an effective quality assurance tool for diagnostic immunohistochemistry. *Mod Pathol* 2002; 15:1374–80.
- Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 1998;95:14863–8.
- Liu CL, Praopong W, Natkunam Y, et al. Software tools for high-throughput analysis and archiving of immunohistochemistry staining data obtained with TMAs. *Am J Pathol* 2002;161:1557–65.
- Landis JR, Koch GC. The measurements of observer agreement for categorical data. *Biometrics* 1977;33:159–74.
- 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.
- 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.
- Schwartz GF, Solin LJ, Olivetto IA, Ernster VL, Pressman P. Consensus Conference Committee. The Consensus Conference on the Treatment of in situ Ductal Carcinoma of the Breast, 22–25 April 1999. *Breast* 2000;9:177–86.
- Alizadeh AA, Ross DT, Perou CM, van de Rijn M. Towards a novel classification of human malignancies based on gene expression patterns. *J Pathol* 2001;195:41–52.
- Buchholz TA, Stivers DN, Stec J, et al. Global gene expression changes during neoadjuvant chemotherapy for human breast cancer. *Cancer J* 2002;8:461–8.
- Adjuvant therapy for breast cancer: NIH Consensus Statement. 2000; 17:1–35.
- Beenken SW, Grizzle WE, Crowe DR, et al. Molecular biomarkers for breast cancer prognosis: coexpression of c-erbB-2 and p53. *Ann Surg* 2001;233:630–8.
- Tsutsui S, Ohno S, Murakami S, Kataoka A, Kinoshita J, Hachitanda Y. Prognostic significance of the coexpression of p53 protein and c-erbB2 in breast cancer. *Am J Surg* 2003;185:165–7.
- Gospodarowicz MK, Miller D, Groome PA, Greene FL, Logan PA, Sobin LH. The process for continuous improvement of the TNM classification. *Cancer (Phila)* 2004;100:1–5.
- Callagy G, Cattaneo E, Daigo Y, et al. Molecular classification of breast carcinomas using TMAs. *Diagn Mol Pathol* 2003;12:27–34.
- Zhang DH, Salto-Tellez M, Chiu LL, Shen L, Koay ES. Tissue microarray study for classification of breast tumors. *Life Sci* 2003;73:3189–99.
- 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.
- van de Vijver MJ, He YD, van't Veer LJ, et al. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* 2002;347:1999–2009.
- Kallioniemi OP, Wagner U, Kononen J, Sauter G. Tissue microarray technology for high-throughput molecular profiling of cancer. *Hum Mol Genet* 2001;10:657–62.
- Milanes-Yearsley M, Hammond ME, Pajak TF, et al. Tissue microarray: a cost and time-effective method for correlative studies by regional and national cancer study groups. *Mod Pathol* 2002;15:1366–73.
- Camp RL, Charette LA, Rimm DL. Validation of tissue microarray technology in breast carcinoma. *Lab Invest* 2000;80:1943–9.
- Zhang D, Salto-Tellez M, Putti TC, Do E, Koay ES. Reliability of tissue microarrays in detecting protein expression and gene amplification in breast cancer. *Mod Pathol* 2003;16:79–84.
- Torhorst J, Bucher C, Kononen J, et al. Tissue microarrays for rapid linking of molecular changes to clinical endpoints. *Am J Pathol* 2001; 159:2249–56.
- Chia SK, Wykoff CC, Watson PH, et al. Prognostic significance of a novel hypoxia-regulated marker, carbonic anhydrase IX, in invasive breast carcinoma. *J Clin Oncol* 2001;16:3660–8.
- Kershaw DB, Beck SG, Wharram BL, et al. Molecular cloning and characterization of human podocalyxinlike protein. *J Biol Chem* 1997; 272:15708–14.