

FAST TRACK

High-throughput protein expression analysis using tissue microarray technology of a large well-characterised series identifies biologically distinct classes of breast cancer confirming recent cDNA expression analyses

Dalia M. Abd El-Rehim¹, Graham Ball², Sarah E. Pinder¹, Emad Rakha¹, Claire Paish¹, John F.R. Robertson¹, Douglas Macmillan¹, Roger W. Blamey¹ and Ian O. Ellis^{1*}

¹Departments of Histopathology and Surgery, The Breast Unit, Nottingham City Hospital NHS Trust and University of Nottingham, Nottingham, United Kingdom

²Division of Life Sciences, Nottingham Trent University, Nottingham, United Kingdom

Recent studies on gene molecular profiling using cDNA microarray in a relatively small series of breast cancer have identified biologically distinct groups with apparent clinical and prognostic relevance. The validation of such new taxonomies should be confirmed on larger series of cases prior to acceptance in clinical practice. The development of tissue microarray (TMA) technology provides methodology for high-throughput concomitant analyses of multiple proteins on large numbers of archival tumour samples. In our study, we have used immunohistochemistry techniques applied to TMA preparations of 1,076 cases of invasive breast cancer to study the combined protein expression profiles of a large panel of well-characterized commercially available biomarkers related to epithelial cell lineage, differentiation, hormone and growth factor receptors and gene products known to be altered in some forms of breast cancer. Using hierarchical clustering methodology, 5 groups with distinct patterns of protein expression were identified. A sixth group of only 4 cases was also identified but deemed too small for further detailed assessment. Further analysis of these clusters was performed using multiple layer perceptron (MLP)-artificial neural network (ANN) with a back propagation algorithm to identify key biomarkers driving the membership of each group. We have identified 2 large groups by their expression of luminal epithelial cell phenotypic characteristics, hormone receptors positivity, absence of basal epithelial phenotype characteristics and lack of c-erbB-2 protein overexpression. Two additional groups were characterized by high c-erbB-2 positivity and negative or weak hormone receptors expression but showed differences in MUC1 and E-cadherin expression. The final group was characterized by strong basal epithelial characteristics, p53 positivity, absent hormone receptors and weak to low luminal epithelial cytokeratin expression. In addition, we have identified significant differences between clusters identified in this series with respect to established prognostic factors including tumour grade, size and histologic tumour type as well as differences in patient outcomes. The different protein expression profiles identified in our study confirm the biologic heterogeneity of breast cancer and demonstrate the clinical relevance of classification in this manner. These observations could form the basis of revision of existing traditional classification systems for breast cancer.

© 2005 Wiley-Liss, Inc.

Key words: breast cancer; classification; protein expression; tissue microarray

Routine clinical management of breast cancer relies on traditional histopathologic classification including tumour grade, histologic tumour type, carcinoma size and lymph node stage. Despite the overall association of these variables with prognosis and outcome,¹ these systems remain relatively weakly predictive of behaviour in some circumstances. Tumours of apparently homogeneous morphologic character vary in response to therapy and have divergent outcomes.²

In addition, the current classification methods are descriptive and relatively subjective with reliance on assessment by expert histopathologists. Furthermore, the histologic appearance of the tumours cannot fully reveal the underlying complex genetic alterations and the biologic events involved in their development and progression. Intuitively, this requires development of a new classi-

fication based on key molecular events involved in the process of carcinogenesis, providing a molecular explanation for the different morphologic phenotypes and behaviours. The cellular and molecular heterogeneity in breast cancer and the large number of molecular events involved in controlling cell growth, differentiation, proliferation, invasion and metastases³ emphasize the importance of studying multiple molecular alterations in concert. Recent high-throughput genomic studies have offered the opportunity to challenge the molecular complexity of breast cancer and provided evidence for classifying breast cancer into biologically and clinically distinct groups based on gene expression patterns.^{4–7} Such new molecular taxonomies have identified many genes, some of which are being proposed as candidate genes for subgrouping breast cancer. Such studies have been applied on a relatively small number of tumours and require validation in large series and comparison with traditional classification systems prior to acceptance in clinical practice. This can be achieved using high-throughput tissue screening tissue microarray (TMA) technology, which allows concomitant analyses of many proteins on a large number of tumour samples⁸ and provides new opportunities to examine combined protein expression profiles in breast cancer to determine their relevance and ability to challenge existing taxonomy.

Clustering is the grouping of a collection of objects into populations by calculating mathematical resemblances between individuals, where objects in the same cluster are more closely related to each other than those assigned to another cluster.⁹ Thus hierarchical clustering analysis is a powerful technique for class discovery; however, it does not provide information about the influence of each object in each cluster with respect to other groups. Further analyses of the cluster data can be provided by the analyses of data belonging to each cluster using artificial neural networks (ANNs). ANNs are a form of artificial intelligence that have found applications in different fields, including the medical field, and have given superior results to the standard statistical methods.^{10,11} One of the important advantages of ANN over other standard statistical approaches is its ability to model complicated data with nonlinear relationships. This type of ANN is a powerful tool for analyzing large complicated data containing a high level of background noise.¹²

Thus, in our study, IHC was applied to TMA sections of a large series of invasive breast cancer using a panel of the most relevant biomarkers. The IHC results were analyzed using hierarchical clustering and ANN to categorize cases into groups and to examine the driving biomarker in each group. Then, the association between these groups and the different clinicopathologic

*Correspondence to: Department of Histopathology, Nottingham City Hospital NHS Trust, Hucknall Road, Nottingham, NG5 1PB, United Kingdom. Fax: +44-115-962-7768. E-mail: ian.ellis@nottingham.ac.uk
Received 8 June 2004; Accepted after revision 26 November 2004
DOI 10.1002/ijc.21004
Published online 7 April 2005 in Wiley InterScience (www.interscience.wiley.com).

TABLE I – FREQUENCIES AND PERCENTAGE OF DIFFERENT HISTOLOGIC TUMOUR TYPES AND GRADES

	No.	%
Tumour type		
Invasive ductal/NST	650	60.5%
Tubular mixed	171	15.9%
Medullary	30	2.8
Typical	4	
Atypical	26	
Lobular	112	10.4%
Classical	66	
Alveolar	2	
Solid	4	
Tubulo-lobular	3	
Mixed	37	
Tubular	27	2.5%
Mucinous	11	1%
Invasive cribriform	5	0.5%
Invasive papillary	3	0.3%
Mixed NST & lobular	37	3.4%
Mixed NST & special type	24	2.2%
Other rare types	4	0.4%
Tumour grade		
1	160	14.9%
2	343	31.9%
3	572	53.2%

NST, no special type.

parameters was studied to examine their biologic and prognostic implications.

Material and methods

Patients

A consecutive series of 1,944 cases of primary operable invasive breast carcinoma from patients presenting from 1986–1998 and entered into the Nottingham Tenovus Primary Breast Carcinoma Series were used. Data on histologic grade,¹³ histologic tumour type,¹⁴ vascular invasion,¹⁵ tumour size and lymph node stage are routinely assessed and recorded in the database as well as patient information such as age and menopausal status. The mean survival for this subgroup of the Series was 57 months (range 1–192 months). Information on local, regional and distant recurrence and survival is maintained on a prospective basis. Patients were followed up at 3-month intervals initially, then every 6 months, then annually for a median period of 58 months. The disease-free interval was defined as the interval (in months) from the date of the primary treatment to the first locoregional or distant metastatic recurrence. The overall survival was taken as the time (in months) from the date of the primary treatment to the time of death.

At the time of the primary diagnosis, 654 (61%) of the patients had lymph node-negative disease and 419 (39%) had positive lymph nodes (332 cases with 1–3 positive nodes, 87 cases with 4 or more positive). A total of 1,076 cases were informative for all markers. Frequencies for different histologic tumour types and grades are shown in Table I.

Tissue microarray construction

Breast cancer TMA were prepared as described previously.^{8,16,17} Briefly, cores of 0.6 mm thickness were obtained from the most representative areas of the tumours then reembedded in microarray blocks. Each case was sampled twice; one core was obtained from the centre and the other from the periphery of the tumour. TMAs of 100 cases per block were made.

Immunohistochemistry

Immunohistochemical staining for the sections was performed using the streptavidin-Biotin Complex method using a large panel of well-characterized commercially available tumour markers

(Table 2). To avoid loss or decline of immunoreactivity of tissue sections with increasing storage time,¹⁸ sections from TMA blocks were cut shortly prior to staining of each antibody.

Immunohistochemistry scoring

The modified histochemical score (H-score)¹⁹ was used because it includes a semiquantitative assessment of both the intensity of staining and the percentage of positive cells. For the intensity, a score of 0 to 3, corresponding to negative, weak, moderate and strong positivity, was recorded. In addition, the percentage of positive cells at each intensity was estimated. The H-score is calculated as $0 \times \text{negative } \% + 1 \times \text{weak } \% + 2 \times \text{moderate } \% + 3 \times \text{strongly stained } \%$. The range of possible scores is thus 0 to 300. H-score and similar semiquantitative scoring systems have been successfully used for TMA evaluation.^{20–22} By using such a score, we were able to explore rationalization of our cases into biologically relevant groups depending on different levels of expression, which could not be obtained by using simpler scoring methods (e.g., positive vs. negative).

Two cores were evaluated from each tumour. Each core was scored individually, then the mean of the 2 readings was calculated. If one core was uninformative (either lost or contained no tumour tissues), the overall score applied was that of the remaining core. Previous studies have validated the use of one core to study the expression of tumour markers having heterogeneous distribution.^{16,17} One observer scored all cases, which were rechecked randomly by the same investigator after a period of time. A good correlation was found between the 2 estimations.

Selection of the biochemical markers

For clustering analyses, we used a large panel of tumour markers (Table II). Most of the proteins selected to study in our work have a well-established role in breast carcinogenesis.³ In addition, the gene transcripts of these proteins have been reported to be important candidate discriminator genes in stratifying breast cancer into distinct groups based on previous cDNA microarray studies.^{4,5,23} Furthermore, some of these markers have been reported to have prognostic and predictive power in some series of breast cancers,²⁴ and finally we used some markers to detect specific forms of differentiation.

The amount of information generated in our study was large and multidimensional (1,076 \times 26 data points), being based on 26 immunohistochemical markers studied in 1,076 invasive breast cancer cases. For further weighting analyses using ANN, we reduced the number of markers analyzed to avoid the background noise that may be produced. For this analysis, we used a subset of markers whose expression patterns showed marked variability among the tumour groups according to their mean level of expression and those known to have key biologic roles in breast cancer.

Cluster analyses (hierarchical clustering)

Hierarchical cluster analysis was conducted based on Euclidean distance measure. We used a complete linkage cluster algorithm implemented in the program Statistica (Statsoft, www.statsoft.com, Tulsa, OK). The closest tumours were merged in an agglomerative way by identifying the pairs of cases that were most similar to each other, as determined by their correlation coefficient across all markers, then a new case is added to the cluster by selecting the one with the greatest closeness to the cluster. By using this method, the closest tumours merged in an agglomerative way, and then we could cluster the breast cancers on the basis of similarity of expression of the markers used. Groups within the cluster analysis were identified by segregating the data by a Euclidian distance of 860. This value was selected because use of a lower distance would result in a much greater number of clusters.

ANN model development and parameterisation

We used a three-layer MLP-ANN with a back-propagation algorithm and a sigmoid activation function, an approach that has

TABLE II – SOURCE, DILUTION AND PRETREATMENT OF ANTIBODIES USED

Antibody, clone	Dilution	Source	Pretreatment
Luminal phenotype			
CK 7/8 [clone CAM 5.2]	1:2	Becton Dickinson	Microwave
CK 18 [clone DC 10]	1:50	DakoCytomation	Microwave
CK 19 [clone BCK 108]	1:100	DakoCytomation	Microwave
Basal phenotype			
CK 5/6 [cloneD5/16134]	1:100	Boehringer Biochemica	Microwave
CK 14 [clone LL002]	1:100	Novocastra	Microwave
SMA [clone 1A4]	1:2000	DakoCytomation	No
p63 ab-1 [clone 4A4]	1:200	Neomarkers	No
Hormone receptors			
ER [clone 1D5]	1:80	DakoCytomation	Microwave
PgR [clone PgR 636]	1:100	DakoCytomation	Microwave
AR [clone F39.4.1]	1:30	Biogenex	Microwave
EGFR family members			
EGFR [clone EGFR.113]	1:10	Novocastra	Microwave
c-erbB-2	1:250	DakoCytomation	No
c-erbB-3 [clone RTJ1]	1:20	Novocastra	Microwave
c-erbB-4 [clone HFR1]	6:4	Neomarkers	No
Tumour suppressor genes			
p53 [clone DO7]	1:50	Novocastra	Microwave
BRCA1 Ab-1 [clone MS110]	1:150	Oncogene Research Products	Microwave
Anti-FHIT [clone ZR44]	1:600	Zymed Laboratories	Microwave
Cell adhesion molecules			
Anti E-cad [clone HECD-1]	1:10 then 1:20	Zymed Laboratories	Microwave
Anti P-cad [clone 56]	1:200	BD Biosciences	Microwave
Mucins			
NCL-Muc-1 [clone Ma695] NCL-	1:300	Novocastra	Microwave
Muc-1 core [clone Ma552]	1:250	Novocastra	Microwave
NCL muc2 [clone Ccp58]	1:250	Novocastra	Microwave
Apocrine differentiation			
Anti-GCDFP-15	1:30	Novocastra	No
Neuroendocrine differentiation			
Chromogranin A [clone DAK-A3]	1:100	DakoCytomation	Microwave
Synaptophysin [clone SY38]	1:30	DakoCytomation	Microwave

received wide application in the medical field.^{12,25} Prior to training, the data were scaled linearly between 0 and 1 using minimums and maximums. The MLP architecture consisted of 13 inputs (representing the 13 markers used in our study), 60 hidden nodes (optimized for performance on blind data results not presented) and 6 outputs each representing an individual group. The initial weights of the network were randomized with a low standard deviation of 0.01, a constrained approach designed to maximize the accuracy of the analysis of weights. In the data set, each of these outputs were encoded with Boolean representation. The ANN model was trained until the performance for predictions on blind data failed to improve for 20,000 epochs. Once training and optimization were complete across a range of architectures based on the convergence of error, weighting of interconnections in the ANN model were analyzed to determine the relative importance of each input variable. This approach has some inherent sources of error, as it is a linearization of the nonlinear ANN model. To minimize these errors, the initial weights of the model were randomized with a low standard deviation (a constrained approach designed to amplify the importance of features within the weighting analysis). Further confidence can be gained in the parameterisation results because the data were scaled as described earlier and the models used in the analysis had a good predictive performance.

Weighting analyses were conducted by multiplying the weight of the link between input and hidden nodes by the link between the corresponding hidden and output nodes, then summing the values for each input to determine the weighting for that input. The importance value obtained for each input provides a measure of its influence when absolute/unsigned weight values were used and indicated whether that influence is positive or negative when signed values were used indicating if these inputs drive into or away from a certain group.²⁶ By parameterization of ANN models, inputs that have little or no influence on the output can be excluded.

TABLE III – MARKERS OF HIGH TO MODERATE INFLUENCE ON THE GROUPS BASED ON ABSOLUTE WEIGHTINGS

Group	Markers of high to moderate absolute weightings
Group 1	AR, nBRCA1, GCDFF, c-erbB-2 ¹ Muc1, CK 18, E-cad, ER ²
Group 2	AR, nBRCA1, c-erbB-2, CK18 ¹ E-cad, muc1, GCDFF, p53 ²
Group 3	c-erbB-2, CK18, AR, GCDFF ¹ Muc1, ER, E-cad, p53 ²
Group 4	c-erbB-2, E-cad, nBRCA1, p53 ¹ EGFR, muc1 ²
Group 5	CK 18, c-erbB-2, AR, p53 ¹ ER, CK 5/6, nBRCA1, chromogranin ²
Group 6	c-erbB-2, AR, CK 18, ER ¹ p53, nBRCA1, Muc 1, CK5/6 ²

¹Markers with high absolute weighting values.–²Markers with moderate absolute weighting values.

Associations with clinicopathologic features

To investigate the relationship of the classified groups with tumour and patient characteristics and patient outcome, statistical analyses were carried out using the χ^2 test with the SPSS statistical software, version 10.0.

Results

Based on the immunohistochemical data of all markers used, 6 main clusters could be identified at Euclidean distance 860 of the dendrogram (Fig. 1). Group 1 and 2 merge to form a single cluster then also merge with group 3 to form one main cluster at a higher level. On the other side of the dendrogram, clusters 4, 5 and 6 are derived from a common branch, which splits to give a separate branch for group 4 and a common branch that then divides further into clusters 5 and 6. The ANN models developed based on these

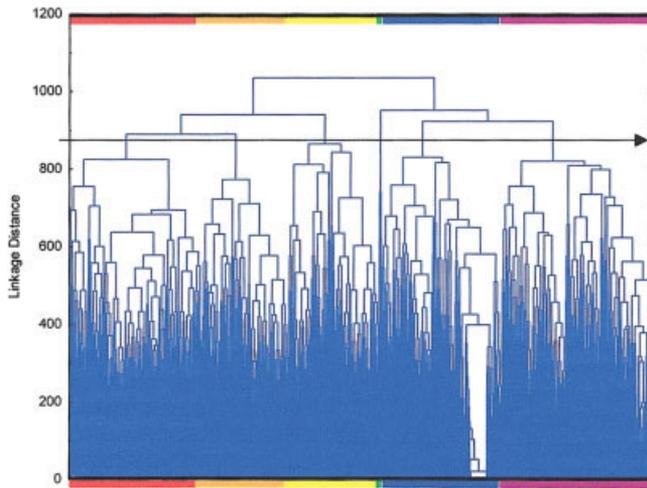


FIGURE 1 – Cluster tree diagram of all tumours clustered into 6 groups at Euclidean distance of 860 (arrow). Clusters are arranged from left to right, starting from cluster 1 and ending at cluster 6.

clusters were able to predict classes 1, 2, 4, 5 and 6 with 100% correct prediction rates. Group 3 was predicted with 99.57% accuracy (1 misclassification out of 234, misclassified as group 2).

Groups' description according to weighting analyses

Analysis and description of each group was performed based on the means and the standard deviation (SD) of the mean of expression of the markers in each group (Fig. 2) and ANN, as described above. Figures 3 and 4 show the absolute and the signed weighting values from ANN among the 6 groups.

Analysis of the absolute weightings indicated that the inputs could be ranked in a descending order (Fig. 3). Markers of high to moderate absolute weighting values are summarized in Table III. Inputs of little influence (low absolute weighting values) were ignored because they had relatively little or no effect on driving their corresponding groups.

Group 1, n = 336 (31.2%), and group 2, n = 180 (16.7%)

Both groups showed broadly similar expression patterns being derived from a common branch in the cluster tree (Fig. 1). They were characterized in general by moderate to strong expression of the luminal markers as well as moderate to strong MUC1 expression. Both groups contained breast cancers that were mainly hormone receptor positive with occasional expression of the classical basal markers. The low means of expression of c-erbB-2 and p53 indicated the predominant negative phenotype of both markers in both groups. Few cases were found to express c-erbB-2 at low levels. Both groups were broadly similar; however, we found a distinct difference regarding c-erbB-3 and c-erbB-4 expression. Group 1 showed relatively stronger combined expression of c-erbB-3 and c-erbB-4 compared to group 2.

In addition, the mean of expression of nuclear BRCA1 was lower in group 1 than in group 2. These findings were confirmed by ANN. Nuclear BRCA1 expression was an important discriminating feature in both groups, as reflected by its high absolute weighting values. Unexpectedly, we also found from signed weightings analysis that smooth muscle actin had a high positive weight value, driving cases into group 1; the signed weighting analysis sometimes misrepresents general trends by cancellation of positive and negative elements if there are 1 or 2 strong negative or positive values in a predominance of opposite but weaker values. Furthermore, because weighting analysis is a linearization of the aggregate of 2 levels in the ANN, the levels may be strongly

negative and thus by multiplication the overall results are strongly positive.

Group 3, n = 234 (21.7%)

This group showed common features with groups 1 and 2, namely strong luminal differentiation and MUC1 overexpression. Group 3, however, was characterized by overexpression of c-erbB-2. In addition, group 3 mainly, but not exclusively, consisted of tumours that were hormone receptor negative or weakly expressing, reflected by low mean H-score levels of ER, PgR and AR. The basal phenotype was detected in a higher proportion of cases than seen in groups 1 and 2. ANN confirmed c-erbB-2 as a key driver for membership of this group, as shown by relatively high mean and absolute weighting values. It also confirmed the relevance of negative or weak hormone receptor expression.

Group 4, n = 4 (0.4%)

This cluster contained only 4 tumours, thus making it difficult to determine which features were driving membership of the group. However, this group showed high mean levels of expression of nuclear BRCA1, p53 and the basal markers. A basal phenotype was detected in 2 cases by expression of CK5/6, whereas all 4 cases were P-cadherin positive, also recognized as a basal marker. The luminal markers means were of low to moderate levels. Two important features of this group were the highest EGFR expression and the absolutely negative ER and PgR phenotype.

Group 5, n = 183 (17%)

In group 5, luminal differentiation was low with predominantly absent hormone receptor expression. High p53 protein expression was a characteristic feature, which distinguished this group from the others. Expression of c-erbB-2 in this group was uncommon, with a low mean compared to other groups. In addition, nuclear BRCA1 was markedly different from other groups, with either negative or reduced levels of BRCA1 expression. The dominant expression of the basal epithelial markers was mainly confined to this group, although rarely, individual cases showed expression in the other groups.

Group 6, n = 139 (12.9%)

Although this group was derived from a separate branch in the dendrogram, it had some homology to group 3, as reflected by high mean levels of expression of c-erbB-2, a negative or weak hormone receptor phenotype and moderate to strong luminal markers expression. Two distinct differences were identified: weak/negative MUC1 and strong positive E-cadherin were seen in group 6 compared to strong positive expression of MUC1 and negative or weak E-cadherin in group 3. The negative or the very weak expression of MUC1 in group 6 was the characteristic feature that caused group 6 to be derived from a separate branch in the dendrogram. The ANN signed weighting results reflected these data, demonstrating high positive MUC1 and high negative E-cadherin weights in group 3 with high positive E-cadherin and high negative MUC1 in group 6.

Associations of the groups with clinicopathologic parameters

Significant differences between the groups with respect to tumour grade, size and lymph node stage were identified (Table IV). The distribution of tumour grades among the different clusters was highly significantly different ($\chi^2 = 260.552, p < 0.001$); grade 3 carcinomas were observed in 37%, 18.3%, 59.8%, 100%, 88.5% and 78.4% of groups 1, 2, 3, 4, 5 and 6, respectively. The classification of tumours was also significantly related to tumour size ($\chi^2 = 33.593, p < 0.001$); large tumour sizes were more frequently noticed in the last 4 groups compared to groups 1 and 2. Some differences were noticed regarding nodal stage among the groups ($\chi^2 = 21.198, p = 0.020$); node-negative disease was more frequently seen in group 2 and metastases to 4 or more lymph nodes was more prevalent in group 5. In addition, significant

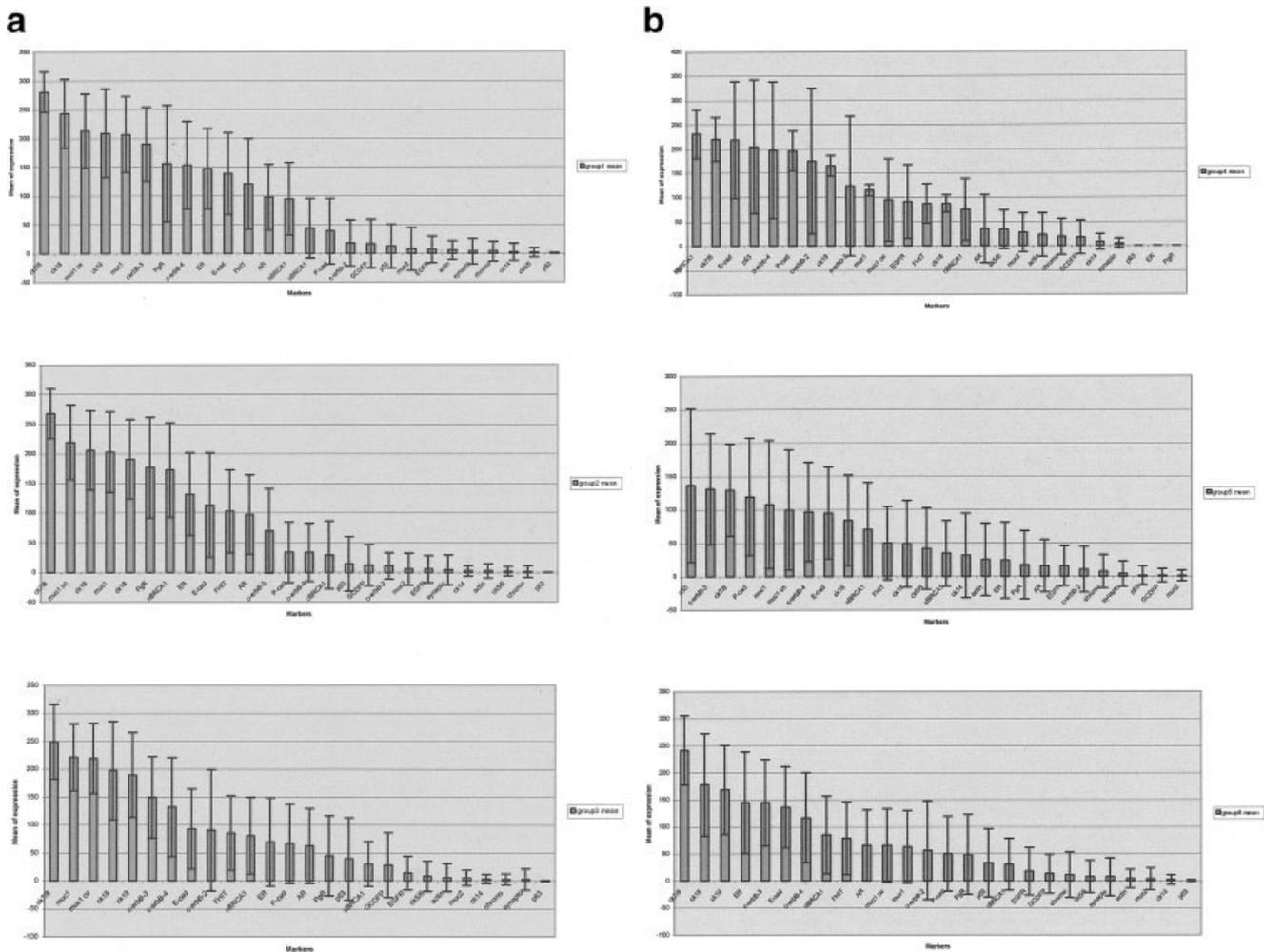


FIGURE 2 – Results of mean of expression among different groups.

differences were noticed among clusters in relation to patient age ($\chi^2 = 33.553$, $p = 0.004$; Table V).

Highly significant differences were observed among the groups in relation to histologic tumour types ($\chi^2 = 231.479$, $p < 0.001$). Ductal/no special type (NST) was more or less distributed equally in all groups, with the exception of group 2, which showed the lowest frequency of invasive NST carcinomas. Tubular mixed carcinomas, lobular and other tumours of special morphology were predominantly seen in groups 1 and 2 and to a lesser degree in group 3. Medullary tumours (both of typical and atypical patterns) were predominantly confined to groups 5 and 3 (Fig. 5).

Survival analyses

As shown by Kaplan-Meier survival analyses, molecular clustering analysis was highly significantly associated with overall survival (OS) and disease-free survival (DFS). The highest frequency of breast cancer mortality was seen in patients whose tumours belonged to group 5. A lower, but still high, frequency was seen in patients with tumours clustered in groups 3 and 6 ($\chi^2 = 33.107$, $p < 0.001$; Table IV). Significant associations between the groups and outcome were observed on Kaplan-Meier survival analyses. The OS and DFS were the poorest for group 5, and the longest survivals were seen in groups 1 and 2 (Figs. 6 and 7). No reported deaths due to breast cancer were seen in patients with tumours in group 4, and neither recurrence nor distant metastases were noted during the period of follow-up in this group.

Cox regression analysis was performed to evaluate the independent prognostic effect of the clustering. This was performed with inclusion of the 3 most important and well-recognised parameters related to patient outcome, namely histologic grade, tumour size and lymph node stage. Multivariate analysis demonstrated the independent prognostic effect of the molecular clustering in predicting patient outcomes, independent of tumour grade, size and nodal stage (Tables VI and VII).

Discussion

Our findings, using a large panel of markers, have identified distinct protein expression patterns, which defined 6 classes of breast cancer. This diversity in protein expression patterns reflects the complex heterogeneous molecular nature of this disease.

Analyses of the means and weightings were carried out to determine the relative importance of each variable to each group. We demonstrated that AR, c-erbB-2, CK18, MUC1, CK5/6, p53, nuclear BRCA1, ER and E-cadherin were the driving key markers and the most important discriminators among different clusters. We will focus the discussion on these markers.

To summarise our results, groups 1 and 2 as identified by cluster analysis contain invasive breast cancers that are luminal epithelial cell and hormone receptor positive; group 3 is c-erbB-2-driven, hormone receptor weak/negative and strongly MUC1 positive with altered E-cad expression; group 5 is p53-driven, hormone

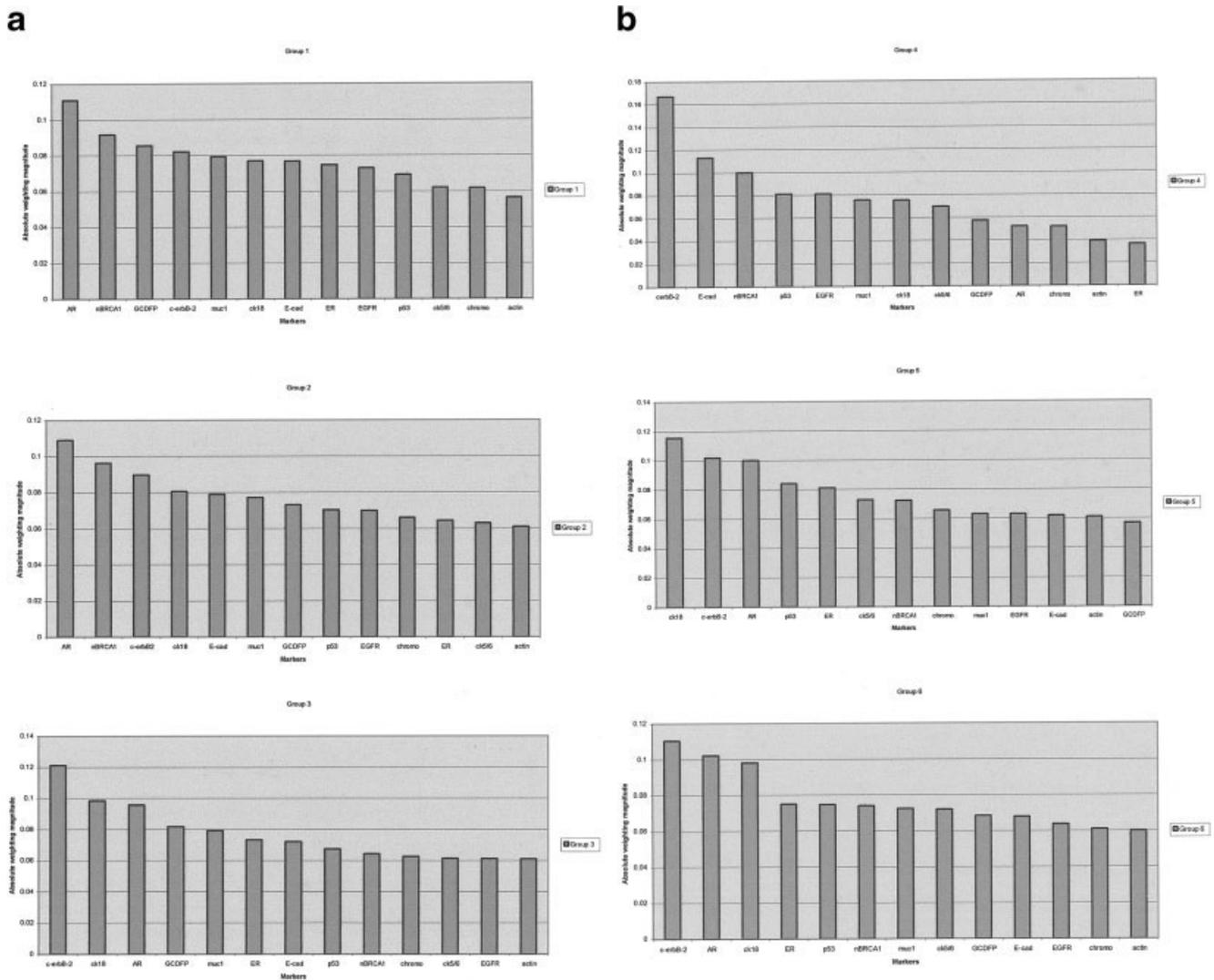


FIGURE 3 – Results of absolute weightings among different groups.

receptor negative but consists of tumours that are in general basal marker positive; and finally, group 6 bears c-erbB-2 positive, hormone receptor weak/negative and MUC1 weak/negative and E-cad strong positive breast cancers. Group 4 consists of only 4 tumours, making it difficult to give a general description of the features of that cluster; however, it appears to be characterized by a basal phenotype with negative hormone receptor expression and strong expression of c-erbB2, p53 and nuclear BRCA1. Our findings are broadly comparable to clusters described in previous cDNA-based clustering studies,^{4,5} where breast cancers have been stratified into 4 groups: (i) luminal and ER positive; (ii) basal and ER negative; (iii) c-erbB-2 positive and ER negative or low; and (iv) normal breast cluster.

It has been postulated that any tumour biology reflects to some extent the biology of the cell of origin at the time of initiation and that it follows distinct pathways related to the cell of origin. Tumours originating from more undifferentiated epithelial cells have a rapid growth pattern and more aggressive behaviour and outcome compared with those originating in more differentiated epithelial cells.²⁷ Two of the important discriminator proteins, the luminal epithelial cytokeratin CK18 and the basal epithelial cytokeratin CK5/6 identified in our study, relate to the mammary gland anatomy and cellular structure of its parenchymal tissue. Two pre-

vious studies using gene expression⁴ and protein expression patterns²⁸ have demonstrated distinctive patterns of luminal and basal cytokeratin expression in breast cancer. Our findings are consistent with these reports, with groups 1 and 2 representing a more differentiated luminal phenotype and group 5 a less differentiated basal phenotype, the luminal expression-negative group.

We have examined whether there were associations between different protein expression patterns within groups and the cellular phenotype identified by these luminal and the basal markers. Tumours expressing luminal and basal markers displayed remarkably different patterns of other proteins in addition to cytokeratins, which may be attributable to their evolution through distinct cell lineage or differentiation related to gene expression characteristics. For example, p53 protein expression was mainly confined to group 5, whose tumours were predominantly of the basal phenotype. These findings can be explained by the postulate that subsets of breast cancer exist that are either derived from different normal cell populations or differentiate along different pathways as a consequence of having different alterations in the control mechanisms of cell proliferation and hence neoplastic progression; this may be p53-dependent in the tumours with a less differentiated, basal phenotype and p53-independent in those tumours with a more differentiated, pure luminal form. The existence of p53 mutations by

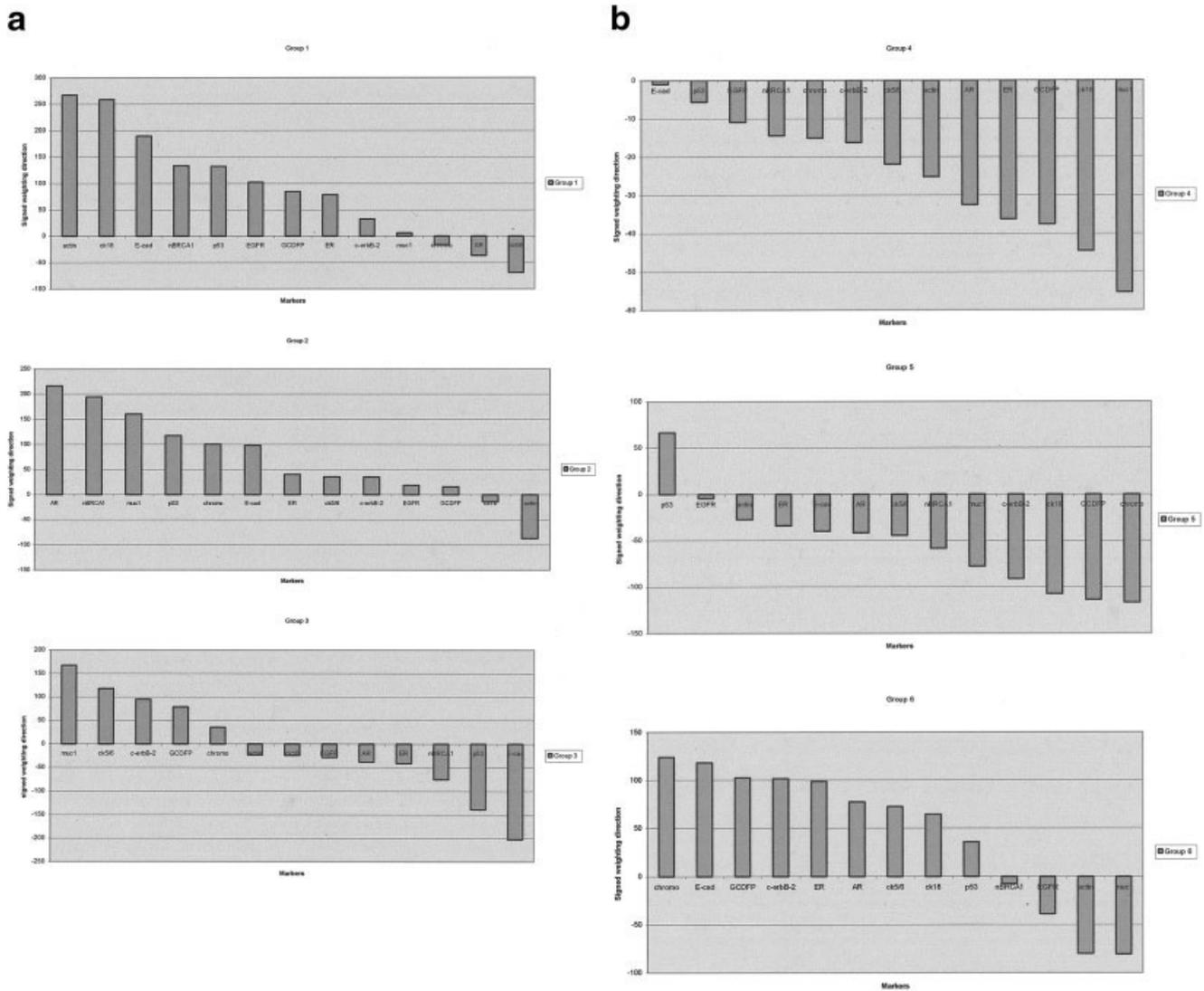


FIGURE 4 – Results of signed weighting among different groups.

chance in the later subset cannot, however, be excluded,^{29,30} as shown by our finding that occasional tumours in groups 1, 2, 3 and 6 exhibited p53 expression. This supports the suggestion that cell type-specific genetic pathways are maintained in distinct genetic evolutionary pathways in mammary carcinogenesis.

Known associations between the markers used in these studies support the biologic significance of these clustering analysis results. For example, the known associations between luminal epithelial differentiation and positive ER status, between basal differentiation and negative ER status are established,^{4,31–33} as are those between positive p53 expression and BRCA1 mutation.³⁴ Another interesting example is the inverse association between MUC1 expression and E-cadherin immunopositivity, which was typically noticed in groups 3 and 6 and which has been reported in previous *in vitro* studies.^{35–37}

Some associations would not be expected; for example, the combined strong expression of p53 and c-erbB-2 proteins occurred rarely in this large series. The same findings have been reported in 2 breast cancer studies: by examination of protein expression of p53 and c-erbB-2³⁸ and by determination of p53 mutation and c-erbB-2 amplification.³⁹ These findings imply that breast carcinogenesis

evolves in parallel pathways with different underlying, partially independent, partially overlapping mechanisms of carcinogenesis.

It is evident that the observed molecular heterogeneity of the tumours studied is not fully reflected in their morphologic appearance; tumours of apparently similar morphology have biologically different behaviours. We have identified by this clustering analysis of immunohistochemical expression of a large number of markers that there are biologically and clinically distinct groups within apparently homogenous tumours. For example, grade 3 carcinomas were found in all groups. Another example is the hormone receptor-negative or weakly expressing tumours, which were mainly clustered in the last 4 groups. Each group had a relatively distinct protein expression pattern and clinically distinct behaviour, in spite of being hormone receptor negative or weak. Such discrimination could have therapeutic implications; patients belonging to a certain group could be candidates for specific therapy, which was not optimal for patients from other groups. Clearly, this requires further study.

Despite the above observations, we have identified some broad relationships between molecular phenotypes and morphologic characteristics. Highly significant but not absolute associations

TABLE IV – GROUP DISTRIBUTION IN RELATION TO DIFFERENT CLINICOPATHOLOGIC PARAMETERS

Feature	Group					
	G1	G2	G3	G4	G5	G6
Grade						
1	73 (21.8%)	47 (26.1%)	29 (12.4%)	0	6 (3.3%)	5 (3.6%)
2	138 (41.2%)	100 (55.6%)	65 (27.8%)	0	15 (8.2%)	25 (18%)
3	124 (37%)	33 (18.3%)	140 (59.8%)	4 (100%)	162 (88.5%)	109 (78.4%)
Total	335	180	234	4	183	139
<i>p</i> - value < 0.001						
Size						
≤ 1.5 cm	126 (37.5%)	76 (42.2%)	69 (29.5%)	0	35 (19.1%)	34 (24.5%)
> 1.5 cm	210 (62.5%)	104 (57.8%)	165 (70.5%)	4 (100%)	148 (80.9%)	105 (75.5%)
Total	336	180	234	4	183	139
<i>p</i> - value < 0.001						
Lymph node stage						
1	201 (60%)	125 (69.8%)	133 (57.1%)	2 (50%)	121 (66.1%)	72 (51.8%)
2	108 (32.2%)	45 (25.1%)	79 (33.9%)	2 (50%)	42 (23%)	56 (40.3%)
3	26 (7.8%)	9 (5%)	21 (9%)	0	20 (10.9%)	11 (7.9%)
Total	335	179	233	4	183	139
<i>p</i> - value = 0.020						
Death						
No	317 (96.6%)	169 (93.9%)	201 (87%)	4 (100%)	147 (83.1%)	124 (89.9%)
Yes	11 (3.4%)	11 (6.1%)	30 (13%)	0	30 (16.9%)	14 (10.1%)
Total	328	180	231	4	177	138
<i>p</i> - value < 0.001						

TABLE V – AGE DISTRIBUTION AMONG DIFFERENT GROUPS ($\chi^2 = 33.554, P = 0.004$)

Group	Patient age in years				Total
	≤ 35	36-45	46-55	> 55	
Group 1	15 (4.5%)	58 (17.3%)	103 (30.7%)	160 (47.6%)	336
Group 2	6 (3.3%)	26 (14.4%)	67 (37.2%)	81 (45%)	180
Group 3	7 (3%)	43 (18.4%)	71 (30.3%)	113 (48.3%)	234
Group 4	0	1 (25%)	0	3 (75%)	4
Group 5	16 (8.7%)	49 (26.8%)	56 (30.6%)	63 (33.9%)	183
Group 6	4 (2.9%)	25 (18%)	34 (24.5%)	76 (54.7%)	139

have been noticed between the clusters identified in this series and 2 important prognostic parameters, grade and tumour size. In addition, there were significant differences in the distribution of the clusters by histologic tumour types. These findings indicate that different molecular pathways in breast carcinogenesis are associated with specific tumour morphologies, as has been suggested by previous genetic studies.^{40,41} It is worthwhile to note that the molecular grouping in the present series was determined by the expression of a group of proteins not by the expression of a single marker in isolation. This highlights the importance of studying groups of marker proteins in concert instead of focusing on a certain gene, protein or pathway. A significantly different age distribution was also noticed amongst groups where young patient age was more frequently seen in group 5 compared to other groups. In this series, group 5 exhibited the most aggressive behaviour compared to other groups that support previous observations that breast cancer in young women have lower hormone receptor level, higher proliferation and worse prognosis compared to those in older ages.⁴²⁻⁴⁴

Survival analyses in our study revealed significant differences in DFS and OS among clusters. Group 5 represents a distinct entity that was associated with a p53-positive and basal-positive phenotype. Both of these variables are known to be associated with aggressive behaviour and poor outcome.^{5,33,38,45} In addition, BRCA1 alteration was a common feature in that group. An association between decreased BRCA1 mRNA and increased metastatic potential in sporadic breast cancer has been reported,⁴⁶ which is consistent with the aggressive behaviour of this group in the present series. The same observation could explain the absence of metastases or cancer-related death in group 4, which was characterized by strong positive expression of BRCA1 in spite of the strong expression of other poor prognostic markers such as p53,

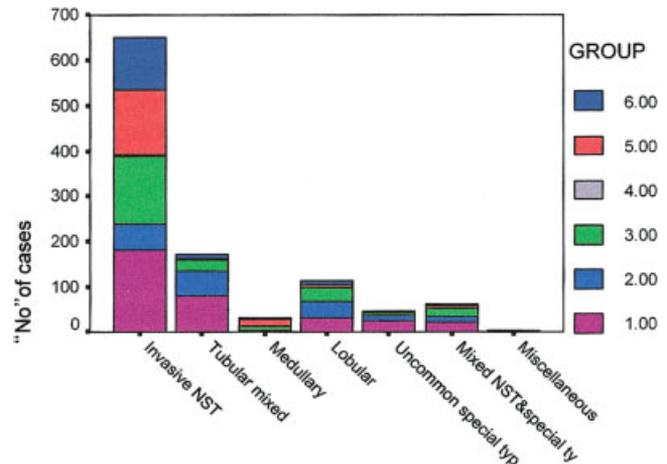


FIGURE 5 – Clustering group distribution among different histologic tumour types ($\chi^2 = 231.479, p < 0.001$).

c-erbB-2 and EGFR, together with the negative hormone receptor phenotype. However, the limited number of cases in that group precludes more definitive extrapolation.

C-erbB-2 is recognized as an important molecular marker associated with poor overall survival and disease-free survival.⁴⁷ It has a therapeutic value in selecting cases more likely to be responsive to c-erbB-2-directed immunotherapy.⁴⁸ In our series, c-erbB-2 was a driving marker into 2 groups (groups 3 and 6). Group 3 was characterized mainly by c-erbB-2 expression, MUC1 mucin overexpression and reduced E-cadherin expression as determined by the mean values in the groups and by weighting analyses. The increased tumour invasive and metastatic potential associated with c-erbB-2 overexpression⁴⁹ could explain the unfavorable prognosis found. Overexpression of MUC1, together with the negative or reduced E-cadherin expression, could be additional features that reduce cell to cell adhesion, thus facilitating cell detachment and metastases.^{37,50} The reverse was identified in group 6, into which c-erbB-2 was also a driving marker; however, within this group tumours tended to exhibit negative to weak MUC1 expression and relatively stronger E-cadherin expression compared to group 3. These differences in molecular pattern of groups 3 and 6 may explain the relatively poorer DFS

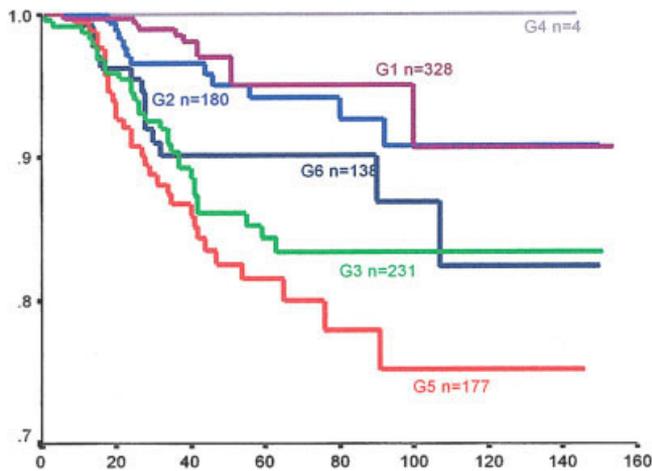


FIGURE 6 – Kaplan-Meier analysis of overall survival in different cluster groups.

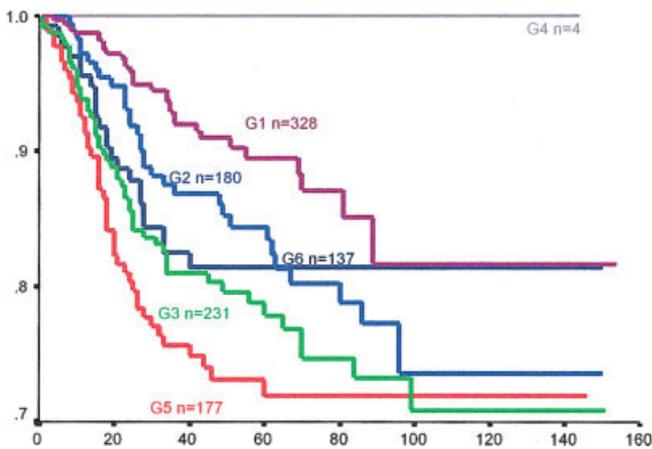


FIGURE 7 – Kaplan-Meier analysis of disease-free survival in different cluster groups.

and OS of those patients with tumours in group 3 compared to those in group 6.

Although groups 1 and 2 are broadly similar, we identified differences in patient outcome, especially obvious in DFS. A previous cDNA microarray series has also recorded that the luminal ER-positive group can be further divided into 2 groups with different outcomes (luminal A, luminal B).⁶ We have no clear explanation for these differences. However, we noticed overexpression of c-erbB-3 and c-erbB-4 in group 1 in contrast to group 2, where relatively weak expression was seen. The association of both c-erbB-3 and c-erbB-4 overexpression with the classical prognostic parameters and outcome has been poorly documented. However, some studies have reported an association with favorable outcome; longer overall survival has been reported in patients

TABLE VI – MULTIVARIATE COX SURVIVAL ANALYSES OF FACTORS RELATED TO OVERALL SURVIVAL

Variable	Relative risk (95% CI)	p-value
Grade		0.001
2 vs. 1	0.992 (0.351–2.807)	0.988
3 vs. 1	3.064 (1.188–7.907)	0.021
Lymph node stage		<0.001
2 vs. 1	1.268 (0.779–2.065)	0.339
3 vs. 1	4.369 (2.618–7.291)	<0.001
Tumour size	1.640 (1.173–2.293)	0.004
Group		0.028
2 vs. 1	2.461 (1.042–5.809)	0.040
3 vs. 1	3.134 (1.559–6.298)	0.001
4 vs. 1	0.000	0.966
5 vs. 1	3.157 (1.550–6.433)	0.002
6 vs. 1	2.039 (0.913–4.552)	0.082

TABLE VII – MULTIVARIATE COX REGRESSION ANALYSES OF FACTORS RELATED TO DISEASE-FREE SURVIVAL

Variable	Relative risk (95% CI)	p-value
Grade		0.004
2 vs. 1	0.876 (0.508–1.511)	0.635
3 vs. 1	1.666 (0.989–2.805)	0.055
Lymph node stage		<0.001
2 vs. 1	0.912 (0.638–1.305)	0.616
3 vs. 1	3.092 (2.067–4.624)	<0.001
Tumour size	1.209 (1.001–1.459)	0.049
Group		0.016
2 vs. 1	2.140 (1.298–3.530)	0.003
3 vs. 1	2.034 (1.291–3.204)	0.002
4 vs. 1	0.000	0.958
5 vs. 1	1.998 (1.231–3.243)	0.005
6 vs. 1	1.399 (0.804–2.434)	0.235

whose tumours overexpressed c-erbB-3 and c-erbB-4 mRNAs.⁵¹ More recently, another study has demonstrated c-erbB-4 overexpression to be associated with a favourable outcome.⁵² This is consistent with the better outcome we have noted in patients with cancers placed in group 1 compared to group 2.

The significant prognostic impact of the clustering on predicting outcome was independent in multivariate analysis from the 3 most important prognostic parameters in breast cancer: tumour grade, tumour size and lymph node stage. This further supports the relevance and importance of such a method of molecular classification and indicates that classification of breast cancer in a manner like this may be of value in the future in the evaluation and prediction of outcome in breast cancer.

In conclusion, the biologic and clinical significance of a protein expression-based classification that is described here is supported by its highly significant correlation with outcome in terms of overall survival and disease-free survival, its highly significant associations with the established prognostic parameters and by the broad similarities with previous taxonomies used for cDNA microarray interpretation in breast cancer. Our results show that TMA is an efficient and reliable tool for high-throughput screening, and our overall results support the feasibility of improving the classification of breast cancer based on a combination of morphologic and protein expression characteristics.

References

1. Elston CW, Ellis IO, Pinder SE. Pathological prognostic factors in breast cancer. *Crit Rev Oncol Hematol* 1999;31:209–23.
2. 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.
3. Beckmann MW, Niederacher D, Schnurch HG, Gusterson BA, Bender HG. Multistep carcinogenesis of breast cancer and tumour heterogeneity. *J Mol Med* 1997;75:429–39.
4. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747–52.

5. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci USA* 2001; 98:10869–74.
6. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, Deng S, Johnsen H, Pesich R, Geisler S, Demeter J, Perou CM, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci USA* 2003;100:8418–23.
7. Bertucci F, Houlgatte R, Benziene A, Granjeaud S, Adelaide J, Tagett R, Loriol B, Jacquemier J, Viens P, Jordan B, Birnbaum D, Nguyen C. Gene expression profiling of primary breast carcinomas using arrays of candidate genes. *Hum Mol Genet* 2000;9: 2981–91.
8. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP. Tissue microarrays for high-throughput molecular profiling of tumor specimens. *Nat Med* 1998;4:844–7.
9. Webb AR. Clustering, in statistical pattern recognition. Chichester, England: John Wiley & Sons, 2002.361–408.
10. Papik K, Moinar B, Schaefer R, Dombovari Z, Tulassay Z, Feher J. Application of neural network in medicine—a review. *Med Sci Monit* 1998;4:538–46.
11. Roadknight CM, Ball GR, Mills GE, Palmer-Brown D. Modeling complex environmental data. *IEEE Transact Neural Networks* 1997;8:852–62.
12. Ball G, Mian S, Holding F, Allibone RO, Lowe J, Ali S, Li G, McCauley S, Ellis IO, Creaser C, Rees RC. An integrated approach utilizing artificial neural networks and SELDI mass spectrometry for the classification of human tumours and rapid identification of potential biomarkers. *Bioinformatics* 2002;18: 395–404.
13. 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.
14. Ellis IO, Galea M, Broughton N, Locker A, Blamey RW, Elston CW. Pathological prognostic factors in breast cancer. II. Histological type. Relationship with survival in a large study with long-term follow-up. *Histopathology* 1992;20:479–89.
15. Pinder SE, Ellis IO, Galea M, O'Rourke S, Blamey RW, Elston CW. Pathological prognostic factors in breast cancer. III. Vascular invasion: relationship with recurrence and survival in a large study with long-term follow-up. *Histopathology* 1994;24:41–7.
16. Camp RL, Charette LA, Rimm DL. Validation of tissue microarray technology in breast carcinoma. *Lab Invest* 2000;80:1943–9.
17. Torhorst J, Bucher C, Kononen J, Haas P, Zuber M, Kochli OR, Mross F, Dieterich H, Moch H, Mihatsch M, Kallioniemi OP, Sauter G. Tissue microarrays for rapid linking of molecular changes to clinical endpoints. *Am J Pathol* 2001;159:2249–56.
18. Fergenbaum JH, Garcia-Closas M, Hewitt SM, Lissowska J, Sakoda LC, Sherman ME. Loss of antigenicity in stored sections of breast cancer tissue microarrays. *Cancer Epidemiol Biomarkers Prev* 2004; 13:667–72.
19. McCarty KS Jr., Miller LS, Cox EB, Konrath J, McCarty KS Sr. Estrogen receptor analyses. Correlation of biochemical and immunohistochemical methods using monoclonal antireceptor antibodies. *Arch Pathol Lab Med* 1985;109:716–21.
20. Rimm DL, Camp RL, Charette LA, Olsen DA, Provost E. Amplification of tissue by construction of tissue microarrays. *Exp Mol Pathol* 2001;70:255–64.
21. Callagy G, Cattaneo E, Daigo Y, Happerfield L, Bobrow LG, Pharoah PD, Caldas C. Molecular classification of breast carcinomas using tissue microarrays. *Diagn Mol Pathol* 2003;12:27–34.
22. Ginestier C, Charafe-Jauffret E, Bertucci F, Eisinger F, Geneix J, Bechlian D, Conte N, Adelaide J, Toiron Y, Nguyen C, Viens P, Muzicchianni MJ, et al. Distinct and complementary information provided by use of tissue and DNA microarrays in the study of breast tumor markers. *Am J Pathol* 2002;161:1223–33.
23. van't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530–6.
24. Fitzgibbons PL, Page DL, Weaver D, Thor AD, Allred DC, Clark GM, Ruby SG, O'Malley F, Simpson JF, Connolly JL, Hayes DF, Edge SB, et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med* 2000;124:966–78.
25. Burke HB, Goodman PH, Rosen DB, Henson DE, Weinstein JN, Harrell FE Jr., Marks JR, Winchester DP, Bostwick DG. Artificial neural networks improve the accuracy of cancer survival prediction. *Cancer* 1997;79:857–62.
26. Ball G, Palmer-Brown D, Sanders GE. Investigating microclimate influences on ozone injury in clover (*Trifolium subterraneum*) using artificial neural network. *New Phytol* 1996;132:271–80.
27. Olsson H. Tumour biology of a breast cancer at least partly reflects the biology of the tissue/epithelial cell of origin at the time of initiation—a hypothesis. *J Steroid Biochem Mol Biol* 2000;74: 345–50.
28. Korsching E, Packeisen J, Agelopoulos K, Eisenacher M, Voss R, Isola J, van Diest PJ, Brandt B, Boecker W, Buerger H. 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.
29. Megha T, Ferrari F, Benvenuto A, Bellan C, Lalinga AV, Lazzi S, Bartolommei S, Cevenini G, Leoncini L, Tosi P. p53 mutation in breast cancer. Correlation with cell kinetics and cell of origin. *J Clin Pathol* 2002;55:461–6.
30. Wynford-Thomas D. Proliferative lifespan checkpoints: cell-type specificity and influence on tumour biology. *Eur J Cancer* 1997;33: 716–26.
31. Takei H, Iino Y, Horiguchi J, Kanoh T, Takao Y, Oyama T, Morishita Y. Immunohistochemical analysis of cytokeratin #8 as a prognostic factor in invasive breast carcinoma. *Anticancer Res* 1995;15: 1101–5.
32. 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.
33. Abd El-Rehim DM, Pinder SE, Paish CE, Bell J, Blamey R, Robertson JF, Nicholson RI, Ellis IO. Expression of luminal and basal cytokeratins in human breast carcinoma. *J Pathol* 2004;203: 661–71.
34. Goffin JR, Chappuis PO, Begin LR, Wong N, Brunet JS, Hamel N, Paradis AJ, Boyd J, Foulkes WD. Impact of germline BRCA1 mutations and overexpression of p53 on prognosis and response to treatment following breast carcinoma: 10-year follow up data. *Cancer* 2003;97:527–36.
35. Kondo K, Kohno N, Yokoyama A, Hiwada K. Decreased MUC1 expression induces E-cadherin-mediated cell adhesion of breast cancer cell lines. *Cancer Res* 1998;58:2014–9.
36. Hilkens J, Vos HL, Wesseling J, Boer M, Storm J, van der Valk S, Calafat J, Patriarca C. Is episialin/MUC1 involved in breast cancer progression? *Cancer Lett* 1995;90:27–33.
37. Wesseling J, van der Valk SW, Hilkens J. A mechanism for inhibition of E-cadherin-mediated cell-cell adhesion by the membrane-associated mucin episialin/MUC1. *Mol Biol Cell* 1996;7:565–77.
38. Yamashita H, Nishio M, Toyama T, Sugiura H, Zhang Z, Kobayashi S, Iwase H. Coexistence of HER2 over-expression and p53 protein accumulation is a strong prognostic molecular marker in breast cancer. *Breast Cancer Res* 2004;6:R24–30.
39. Bull SB, Ozcelik H, Pinnaduwa D, Blackstein ME, Sutherland DA, Pritchard KI, Tzontcheva AT, Sidlofsky S, Hanna WM, Qizilbash AH, Tweeddale ME, Fine S, et al. The combination of p53 mutation and neu/erbB-2 amplification is associated with poor survival in node-negative breast cancer. *J Clin Oncol* 2004; 22:86–96.
40. Buerger H, Mommers EC, Littmann R, Simon R, Diallo R, Poremba C, Dockhorn-Dworniczak B, van Diest PJ, Boecker W. 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.
41. Buerger H, Otterbach F, Simon R, Schafer KL, Poremba C, Diallo R, Brinkschmidt C, Dockhorn-Dworniczak B, Boecker W. Different genetic pathways in the evolution of invasive breast cancer are associated with distinct morphological subtypes. *J Pathol* 1999; 189:521–6.
42. Adami HO, Malmer B, Holmberg L, Persson I, Stone B. The relation between survival and age at diagnosis in breast cancer. *N Engl J Med* 1986;315:559–63.
43. Ferno M, Borg A, Johansson U, Norgren A, Olsson H, Ryden S, Sellberg G. Estrogen and progesterone receptor analyses in more than 4,000 human breast cancer samples. A study with special reference to age at diagnosis and stability of analyses. Southern Swedish Breast Cancer Study Group. *Acta Oncol* 1990;29:129–35.
44. Host H, Lund E. Age as a prognostic factor in breast cancer. *Cancer* 1986;57:217–21.
45. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, Torhorst J, Sauter G, Zuber M, Kochli OR, Mross F, Dieterich H, 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.
46. Seery LT, Knowlden JM, Gee JM, Robertson JF, Kenny FS, Ellis IO, Nicholson RI. BRCA1 expression levels predict distant metastasis of sporadic breast cancers. *Int J Cancer* 1999;84:258–62.

47. Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* 1987;235:177–82.
48. Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, Fleming T, Eiermann W, Wolter J, Pegram M, Baselga J, Norton L. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* 2001;344:783–92.
49. Tan M, Yao J, Yu D. Overexpression of the c-erbB-2 gene enhanced intrinsic metastasis potential in human breast cancer cells without increasing their transformation abilities. *Cancer Res* 1997;57:1199–205.
50. Heimann R, Lan F, McBride R, Hellman S. Separating favorable from unfavorable prognostic markers in breast cancer: the role of E-cadherin. *Cancer Res* 2000;60:298–304.
51. Pawlowski V, Revillion F, Hebbbar M, Hornez L, Peyrat JP. Prognostic value of the type I growth factor receptors in a large series of human primary breast cancers quantified with a real-time reverse transcription-polymerase chain reaction assay. *Clin Cancer Res* 2000;6:4217–25.
52. Witton CJ, Reeves JR, Going JJ, Cooke TG, Bartlett JM. Expression of the HER1-4 family of receptor tyrosine kinases in breast cancer. *J Pathol* 2003;200:290–7.