

Tumor Gene Expression and Prognosis in Breast Cancer Patients with 10 or More Positive Lymph Nodes

Melody A. Cobleigh,¹ Bitu Tabesh,¹ Pincas Bitterman,¹ Joffre Baker,² Maureen Cronin,² Mei-Lan Liu,² Russell Borchik,¹ Juan-Miguel Mosquera,¹ Michael G. Walker,² and Steven Shak²

Abstract Purpose: This study, along with two others, was done to develop the 21-gene Recurrence Score assay (Oncotype DX) that was validated in a subsequent independent study and is used to aid decision making about chemotherapy in estrogen receptor (ER)–positive, node-negative breast cancer patients.

Experimental Design: Patients with ≥ 10 nodes diagnosed from 1979 to 1999 were identified. RNA was extracted from paraffin blocks, and expression of 203 candidate genes was quantified using reverse transcription-PCR (RT-PCR).

Results: Seventy-eight patients were studied. As of August 2002, 77% of patients had distant recurrence or breast cancer death. Univariate Cox analysis of clinical and immunohistochemistry variables indicated that HER2/immunohistochemistry, number of involved nodes, progesterone receptor (PR)/immunohistochemistry (% cells), and ER/immunohistochemistry (% cells) were significantly associated with distant recurrence-free survival (DRFS). Univariate Cox analysis identified 22 genes associated with DRFS. Higher expression correlated with shorter DRFS for the HER2 adaptor *GRB7* and the macrophage marker *CD68*. Higher expression correlated with longer DRFS for tumor protein p53-binding protein 2 (TP53BP2) and the ER axis genes *PR* and *Bcl2*. Multivariate methods, including stepwise variable selection and bootstrap resampling of the Cox proportional hazards regression model, identified several genes, including *TP53BP2* and *Bcl2*, as significant predictors of DRFS.

Conclusion: Tumor gene expression profiles of archival tissues, some more than 20 years old, provide significant information about risk of distant recurrence even among patients with 10 or more nodes.

Breast cancers exhibit widely varying behavior with regard to the likelihood of distant recurrence and response to therapy. Staging breast cancer by the tumor-node-metastasis system provides important information about the risk of distant recurrence and breast cancer death, but it is by no means precise, especially for node-negative patients (1–6). Many node-negative patients receive unnecessary chemotherapy because of the lack of robust prognostic and predictive tests.

Study of tumor molecular characteristics has enhanced our understanding of both the risk of breast cancer recurrence and the response to therapy. Assays to determine estrogen receptor (ER), progesterone receptor (PR), and HER2 receptor status are routinely done and used in breast cancer treatment planning

(3, 7–11). Hundreds of other putative markers have been identified by immunohistochemistry or biochemical assays in studies that focused on one gene or a few genes at a time. For example, several studies found that elevated levels of urokinase-type plasminogen activator and type 1 plasminogen activator inhibitor proteins predict breast cancer recurrence (12, 13). However, few of the many putative markers have been clinically validated.

Two recent complementary advances enable a new approach for the development and clinical validation of prognostic and predictive biomarkers. First, sensitive and quantitative expression assays for thousands of genes in fresh frozen tissue using DNA microarrays (14–19) can be used to identify candidate markers for validation. Although frozen tumor tissue is available for hypothesis-generating microarray studies, relatively few large banks of frozen tumor tissue exist with associated long-term recurrence and survival information. Second, reverse transcription-PCR (RT-PCR) can be used to quantify the expression of tens to hundreds of genes from a few thin sections of a formalin-fixed, paraffin-embedded tissue (20–22). Thus, RT-PCR can be used to validate candidate gene expression profiles in clinical studies, as well as to report results to be used in clinical practice.

To develop a sensitive, specific, reproducible, and practical test for breast cancer prognosis, we have created and optimized a multigene gene quantitative RT-PCR expression assay (22)

Authors' Affiliations: ¹Rush University Medical Center, Chicago, Illinois and ²Genomic Health, Inc., Redwood City, California

Received 4/4/05; revised 8/13/05; accepted 9/1/05.

Grant support: Genomic Health, Inc., Redwood City, CA.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Presented in abstract form at the 2003 Annual Meeting of American Society of Clinical Oncology.

Requests for reprints: Melody A. Cobleigh, Rush University Medical Center, Professional Building, 1725 West Harrison, Suite 821, Chicago, IL 60612. Phone: 312-942-2242; Fax: 312-563-3123; E-mail: melody.cobleigh@rush.edu.

© 2005 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-05-0735

that measures the expression of 187 cancer-related "candidate" genes and five reference genes, typically using no more than three 10- μ m formalin-fixed, paraffin-embedded sections. The represented genes were selected by surveying the research literature, including published DNA array studies, for genes implicated in cancer pathology or prognosis.

In two studies reported elsewhere, we have explored the relationship between expression of these candidate genes and recurrence in early breast cancer in independent, mostly node-negative, cohorts of breast cancer patients (23, 24). Here, we studied the correlation between tumor gene expression and distant recurrence-free survival (DRFS) in a group of patients with poor prognosis. Tabesh et al. have identified women who were operated on at Rush University Medical Center and who had ≥ 10 nodes (25). A subset of these women was found to have prolonged disease-free survival. We hypothesized that examination of this node-positive cohort would assist in identifying genes that are consistently related to the likelihood of recurrence across the range of breast cancer stage. This study contributed to culling of the original candidate genes and to the creation of a Recurrence Score algorithm that was subsequently validated in a large cooperative group trial involving tamoxifen-treated patients with node-negative breast cancer (26).

Patients and Methods

Patients. Eligible patients were women with invasive breast cancer and ≥ 10 positive nodes with no evidence of metastatic disease who had surgery at Rush University Medical Center from 1979 to 1999. Of 138 eligible patients, archival fixed paraffin blocks were available from the primary tumor in 86 patients. If multiple blocks were available, the block most representative of the diagnosis was selected. A threshold of tumor cellularity (at least 5% of the area invasive cancer) was used for eligibility. Most of the sections contained extensive tumor; the proportion of tumor was $>50\%$ in 73% of the cases and $<20\%$ in 5% of the cases. This study was done after approval by the local Human Investigations Committee.

Sample preparation. Three 10- μ m sections were cut from each paraffin block and placed in a bar-coded microcentrifuge tube. One additional 5- μ m section was cut and stained with H&E. Tubes and slides were shipped to Genomic Health, Inc. (Redwood City, CA) at ambient temperature.

Gene expression analysis. Quantitative gene expression was determined by a multianalyte Taqman RT-PCR assay that was designed to accurately measure the small fragments of RNA present in archival tumor blocks (22).

We selected candidate genes by surveying the breast cancer literature for evidence of a significant role in cancer pathologic processes, including proliferation, invasion, sensitivity to apoptosis, metastasis, angiogenesis, immune surveillance, tumor suppression activity, oncogene activity, and differentiation status. Additionally, we included a number of genes identified in published DNA microarray studies of breast cancer, primarily from two groups of investigators. Sorlie et al. (27) used unsupervised cluster analysis to identify putative breast cancer cell subclasses. Van't Veer et al. (17) used supervised learning methods to identify genes whose expression correlated positively or negatively with breast cancer recurrence.

Paraffin was removed from specimens by xylene extraction. RNA was isolated using Epicentre Technologies, Inc. (Madison, WI) MasterPure RNA Purification kit. Total RNA content was measured using the RiboGreen RNA Quantitation Kit (Invitrogen, Carlsbad, CA). Residual genomic DNA contamination was assayed by Taqman PCR assay for β -actin DNA. Samples with contaminating DNA were resubjected to DNase I treatment and assayed again for DNA contamination.

Reverse transcription of the purified RNA was carried out using SuperScript II RT enzyme (Invitrogen) for first-strand cDNA synthesis.

Taqman reactions were carried out in 384-well plates according to manufacturer's instructions, using Applied Biosystems PRISM 7900HT Taqman instruments (Foster City, CA). A total of 192 genes (187 candidate genes and 5 reference genes) were tested initially. RNA extracts were subsequently retested for expression of 16 additional candidate genes. Expression of each gene was measured in duplicate and then normalized relative to the five reference genes (*β -actin*, *GAPDH*, *GUS*, *RPLPO*, and *TFRC*; ref. 22). Reference-normalized expression measurements typically range from 0 to 15, where a 1-unit change generally reflects a 2-fold change in RNA.

Clinical assessments and outcomes. Tumor size, nodal status, and systemic treatment (hormonal and/or chemotherapy) at initial diagnosis were entered prospectively into a database begun in 1979 and were reviewed for accuracy by examining case records. The primary end point of DRFS was determined from the dates of systemic disease recurrence, death, and/or last follow-up. Cause of death was classified as due to breast cancer, due to other cause, or of uncertain cause.

Other assessments. H&E-stained slides submitted to Genomic Health were reviewed by a pathologist for confirmation of the submitting diagnosis and for assessment of the proportion of tissue surface area composed of tumor. Immunohistochemistry for ER, PR, HER2, and Ki-67/MIB-1 protein was done by the avidin-biotin peroxidase technique, using reagents from DAKO (Carpinteria, CA). The threshold for categorizing immunohistochemistry as positive was 10% of cells. Tumor grade using the Bloom-Richardson criteria was independently assessed at Rush University Medical Center and at Genomic Health (28).

Statistical methods. Demographic and baseline characteristics were summarized by descriptive statistics. Cox proportional hazards models were applied in univariate, multivariate, and stepwise analysis of DRFS as a function of clinical variables and gene expression. For the primary analysis, death due to uncertain cause was treated as death due to breast cancer, and death due to causes other than breast cancer was censored at the time of death. Because the number of candidate genes relative to the number of patients was large, we did simulations in which we randomly shuffled the patient survival times versus gene expression to estimate the number of genes that would seem significant ($P < 0.01$ or $P < 0.05$) in the absence of a genuine association with survival (false-positive rate). Correlation analyses of gene expression used Pearson linear correlation. Cluster analysis used 1-Pearson R as the distance metric and single linkage hierarchical clustering. The Recurrence Score was calculated as described by Paik et al. (26) from the expression of 21 genes: 16 cancer-related genes (*Ki67*, *STK15*, *Survivin*, *CCNB1* or *cyclin B1*, *MYBL2*, *GRB7*, *HER2*, *ER*, *PGR*, *BCL2*, *SCUBE2*, *MMP11* or *stromelysin 3*, *CTSL2* or *cathepsin L2*,

GSTM1, *CD68*, and *BAG1*) and five reference genes (*ACTB* or β -actin, *GAPDH*, *RPLPO*, *GUS*, and *TFRC*).

Results

A total of 86 primary tumor samples were submitted to Genomic Health for tumor gene expression analysis. Total extracted RNA yield was sufficient for assay (>500 ng), and RT-PCR profiles were within the specifications of the assay for all patients. One patient was not clinically eligible (stage IV at diagnosis). Pathology review indicated that seven samples had little or no tumor tissue (<5% of the section) and were not evaluable. Sixty-eight of the 78 evaluable patients had infiltrating ductal carcinoma, nine had infiltrating lobular carcinoma, and one had both.

Characteristics and outcomes for the 78 evaluable patients are listed in Table 1. The mean tumor size was 4.4 ± 3.3 cm (mean \pm SD). Of note, 33% had T₁ tumors. The cohort had extensive nodal involvement. The median number of involved nodes was 15 (range, 10-40). Adjuvant tamoxifen was given in 54% of patients. During the 1980s, many premenopausal patients with ER-positive tumors did not receive tamoxifen. Adjuvant chemotherapy was administered in 80% of patients.

Table 1. Case characteristics and outcomes

Characteristic	No. patients/no. analyzed (%)
Mean age \pm SD (range), y	57 \pm 13 (33-86)
Tumor size (cm)	
0-2	26/78 (33%)
2-5	28/78 (36%)
>5	24/78 (31%)
No. lymph nodes at primary diagnosis	
0-9	0/78 (0%)
10-15	40/78 (51%)
15-20	18/78 (23%)
20-30	12/78 (15%)
>30	8/78 (10%)
Adjuvant tamoxifen	
Yes	42/78 (54%)
No	36/78 (46%)
Adjuvant chemotherapy	
Yes	62/78 (80%)
No	16/78 (20%)
Tumor grade (Rush)	
1	22/78 (29%)
2	28/78 (36%)
3	28/78 (36%)
Tumor grade (Genomic Health)	
1	11/78 (14%)
2	37/78 (47%)
3	30/78 (38%)
Vital status	
Distant recurrence, death due to breast cancer, or death due to unknown cause	55/78 (71%)
Alive without distant recurrence or death due to other cause	23/78 (29%)

Anthracycline-based regimens were used in 26 patients (five with the addition of a taxane); cyclophosphamide, methotrexate, and fluorouracil-based regimens in 29 patients; L-PAM/5-fluorouracil in 10 patients; and L-PAM alone in one case (four patients received multiple regimens). The use of diverse chemotherapeutic agents reflects the long follow-up of the cohort. The patients who did not receive chemotherapy were significantly older than the mean (only one was <60 years of age). Median follow-up was 15.1 years, and median time to event (distant recurrence or death due to breast cancer or uncertain cause) was 2.6 years.

Univariate analysis of clinical and pathologic variables and distant recurrence-free survival. The relative risk of distant recurrence conferred by the number of involved nodes, systemic treatment with hormonal therapy or chemotherapy, tumor size, and age at surgery is given in Table 2. Only the number of nodes involved was significantly associated ($P_{\text{unadjusted}} < 0.05$) with DRFS (4% increase in risk for each additional involved node). Tamoxifen treatment was marginally significant ($P = 0.073$), with a hazard ratio of 0.62, consistent with the known benefits of adjuvant hormonal therapy.

The relative risk of distant recurrence conferred by tumor grade, vascular invasion, and immunohistochemistry for ER, PR, HER2, and Ki-67/MIB-1 are given in Table 2. HER2 expression (assayed by immunohistochemistry; 2+ or 3+ in 17 of 78 patients) was significantly correlated with DRFS ($P < 0.001$) and had a hazard ratio of 3.66. Increased tumor grade was associated with a higher hazard ratio but was not statistically significant.

The Kaplan-Meier plots of the time to distant recurrence for nodal status, ER status by immunohistochemistry, and HER2 status by immunohistochemistry are shown in Fig. 1A-C.

Given the large number of multiple univariate analyses being made, we controlled for the expected proportion of falsely rejected hypotheses, or false discovery rate, using the procedure of Benjamini and Hochberg (29). This procedure adjusts the P s for the individual univariate tests for the number of erroneous rejections of the null hypothesis that could be expected given the total number of tests being done. With regard to the clinical and pathologic measurements, only HER2/immunohistochemistry remained significant ($P < 0.05$). The two pathologists agreed in the assessment of tumor grade in 57.6% of the patients.

Univariate analysis of gene expression and distant recurrence-free survival. The relative risk of DRFS for each of the candidate genes with $P_{\text{unadjusted}} < 0.1$ is given in Table 2. Four genes have a $P < 0.01$, and a total of 22 genes have a $P < 0.05$. Higher expression was associated with greater risk of distant recurrence for *GRB7*, *CD68*, *CA9*, and *CTSL*. Higher expression was associated with a lower risk of distant recurrence for, among other genes, *Bcl2*, *TP53BP*, *PR*, *PRAME*, *GSTM1*, *ER*, and *BAG1*.

The Kaplan-Meier plots of the time to distant recurrence for varying expression of the top two genes by RT-PCR, *Bcl2* and *GRB7*, are shown in Fig. 1D-E.

The procedure of Benjamini and Hochberg was again applied to control for false discovery rates (29). With regard to gene expression measurements of DRFS, four genes (*Bcl2*, *GRB7*, *TP53BP2*, and *PR*) remained significant.

Simulations based on randomly shuffling DRFS times, done to determine the number of genes that would be expected to

Table 2. Univariate Cox proportional hazard models of distant recurrence-free survival

	Hazard ratio (95% confidence interval)	Wald χ^2	df	P _{unadjusted}
Clinical variable				
Nodes involved	1.040* (1.006-1.075)	5.309	1	0.021
Adjuvant tamoxifen	0.615 (0.362-1.045)	3.225	1	0.073
Chemotherapy	0.803 (0.442-1.459)	0.519	1	0.471
Tumor size	1.024 (0.934-1.122)	0.253	1	0.615
Age > 50	1.153 (0.655-2.030)	0.245	1	0.621
Age at surgery	0.998 (0.977-1.020)	0.029	1	0.866
HER2 by IHC (positive/negative)	3.656 (1.899-7.039)	15.047	1	<0.001 [†]
ER (% cells)	0.993 [†] (0.986-1.000)	4.410	1	0.036
Moderate vs well, tumor grade (GHI)	2.005 (0.825-4.875)	2.356	1	0.125
Poor vs well, tumor grade (GHI)	2.389 (0.961-5.941)	3.512	1	0.061
PR (% cells)	0.988 [†] (0.979-0.998)	5.183	1	0.023
Moderate vs well, tumor grade (Rush)	0.964 (0.495-1.874)	0.012	1	0.913
Poor vs well, tumor grade (Rush)	1.617 (0.844-3.099)	2.099	1	0.147
ER (positive/negative)	0.594 (0.339-1.041)	3.311	1	0.069
Vascular invasion	1.361 (0.765-2.421)	1.098	1	0.295
Ki-67/MIB-1 (positive/negative)	1.327 (0.761-2.315)	0.992	1	0.319
PR (positive/negative)	0.817 (0.476-1.403)	0.538	1	0.463
Ki-67/MIB-1 (% cells)	1.002 (0.989-1.015)	0.119	1	0.730
Gene variable				
<i>Bcl2</i>	0.743 (0.611-0.903)	8.917	1	0.003 [†]
<i>GRB7</i>	1.261 (1.082-1.469)	8.850	1	0.003 [†]
<i>TP53BP2</i>	0.738 (0.604-0.902)	8.846	1	0.003 [†]
<i>PR</i>	0.857 (0.772-0.951)	8.426	1	0.004 [†]
<i>GSTM1</i>	0.795 (0.661-0.957)	5.905	1	0.015
<i>IGFBP2</i>	0.790 (0.652-0.957)	5.816	1	0.016
<i>PRAME</i>	0.743 (0.583-0.947)	5.763	1	0.016
<i>CD68</i>	1.507 (1.075-2.114)	5.660	1	0.017
<i>BAG1</i>	0.795 (0.656-0.964)	5.440	1	0.020
<i>ER</i>	0.908 (0.835-0.986)	5.221	1	0.022

correlate with DRFS by chance, indicated that two genes would be expected to have a $P < 0.01$, and 10 genes have a $P < 0.05$.

Correlated gene expression. The assay revealed groups of coexpressed genes that were consistent with literature prediction and/or publicly available gene expression databases, indicating that the assay accurately measured the relative levels for the tested mRNAs. For example, expression of the following pairs of genes was highly correlated, as predicted by the work of Sorlie et al. (27): *cytokeratin 5* and *cytokeratin 17* ($r = 0.86$), *LPL* and *RBP4* ($r = 0.82$), *HER2* and *GRB7* ($r = 0.82$). For the *ER* gene, two independent probe-primer sets were tested and yielded very similar values for ER mRNA level ($r = 0.96$).

Unsupervised cluster analysis of the genes that correlated with DRFS ($P < 0.1$) revealed a cluster of genes known to coexpress with ER, including *Bcl2*, *ER*, *IGF1R*, *SCUBE2*, *PR*, and *IGFBP2*; a group of proliferation-related genes, including *SKT15* and *Chk1*; a group of HER2-related genes, including *HER2* and *GRB7*; and a group of macrophage-related genes, including *CD68* and *cathepsin L2*.

Immunohistochemistry and reverse transcription-PCR for estrogen receptor, progesterone receptor, HER2, and Ki-67/MIB-1. Immunohistochemistry and RT-PCR measurements of tumor

expression for ER, PR, HER2, and Ki-67/MIB-1 were compared (Fig. 2A-D). The concordance between RT-PCR (mRNA) and immunohistochemistry (protein) was high for both ER ($\kappa = 0.83$) and HER2 ($\kappa = 0.67$) and somewhat lower for PR ($\kappa = 0.40$). In contrast, the correlation between RT-PCR for Ki-67 and immunohistochemistry for Ki-67 was poor ($\kappa = 0.22$). Similarly, correlation was poor between RT-PCR for other proliferation-related genes, such as *PCNA* and *TOPO 2A*, and immunohistochemistry for Ki-67.

Multivariate analysis of clinical variables, pathology variables, gene expression, and distant recurrence-free survival. Stepwise variable selection of the Cox proportional hazards regression model was done to identify clinical variables (listed in Table 2) and gene expression measurements that were significant predictors of DRFS. The results, provided in Table 3, identified four genes that were significantly related to DRFS. None of the clinical variables were significantly related to DRFS after accounting for expression of these four genes.

A limitation of stepwise regression is the fact that the number of variables available for development of the model is much larger than the number of cases, resulting in overly optimistic estimates. Furthermore, due to the multiplicity of testing of the stepwise variable selection procedure, P s may become uninterruptible. Consequently, internal validation of the resulting

Table 2. Univariate Cox proportional hazard models of distant recurrence-free survival (Cont'd)

	Hazard ratio (95% confidence interval)	Wald χ^2	df	<i>P</i> _{unadjusted}
<i>MYBL2</i> [§]	1.310 (1.030-1.650)	4.956	1	0.026
<i>KLK10</i>	0.810 (0.671-0.979)	4.762	1	0.029
<i>KRT14</i>	0.868 (0.763-0.989)	4.541	1	0.033
<i>p27</i>	0.715 (0.522-0.978)	4.414	1	0.036
β -Catenin	0.838 (0.707-0.993)	4.162	1	0.041
<i>AKT2</i>	0.696 (0.490-0.987)	4.124	1	0.042
<i>SCUBE2</i>	0.909 (0.828-0.997)	4.110	1	0.043
<i>ER</i> (alternate)	0.918 (0.845-0.998)	4.041	1	0.044
<i>TBP</i>	0.701 (0.494-0.993)	4.008	1	0.045
<i>TOP2B</i>	0.600 (0.363-0.993)	3.955	1	0.047
<i>ZNF217</i>	0.725 (0.528-0.996)	3.933	1	0.047
<i>IGF1R</i>	0.854 (0.731-0.998)	3.920	1	0.048
<i>CA9</i>	1.446 (1.000-2.091)	3.831	1	0.050
<i>Bclx</i>	0.662 (0.435-1.007)	3.710	1	0.054
<i>CTSL</i>	1.460 (0.990-2.153)	3.653	1	0.056
<i>HER2</i>	1.153 (0.995-1.336)	3.600	1	0.058
<i>Chk1</i>	1.265 (0.982-1.628)	3.321	1	0.068
γ -Catenin	0.795 (0.620-1.019)	3.280	1	0.070
<i>CD9</i>	0.718 (0.499-1.033)	3.186	1	0.074
<i>Cyclin B1</i> [§]	1.170 (0.980-1.400)	3.106	1	0.078
<i>Ki-67</i> [§]	1.270 (0.970-1.660)	3.005	1	0.083
<i>Cathepsin L2</i> [§]	1.190 (0.980-1.440)	2.929	1	0.087
<i>FOXM1</i>	0.656 (0.402-1.070)	2.849	1	0.091
<i>BBC3</i>	0.823 (0.656-1.032)	2.840	1	0.092
<i>STK15</i>	1.236 (0.958-1.594)	2.648	1	0.104

Abbreviations: IHC, immunohistochemistry; GHI, Genomic Health, Inc.; df, degree of freedom.

*Hazard ratio for an increment of one node.

†Significant *P* value adjusting for false discovery rate using an experiment wise error rate of 0.16 and the method of Benjamini and Hochberg (29).

‡Hazard ratio for an increment of 1% of cells.

§Results upon retesting of extracted RNA for genes not included in the original 192-gene set.

multivariate model was done to confirm results were not spurious and to assess the performance of the resulting models with respect to potential over fitting.

Specifically, bootstrap resampling, based on 500 resamplings with replacement of the original data set, was done to assess the arbitrariness of the stepwise variable selection procedure. Results of this analysis identified 30 variables that occurred in at least 10% of the 500 stepwise regressions, of which 26 of these variables were gene expression measurements. Ten of these 26 genes (*BAG1*, *Bcl2*, *CD68*, *CTSL2*, *GRB7*, *GSTM1*, *MYBL2*, *PR*, *Survivin*, and *STK15*) are included in the calculation of Recurrence Score as described by Paik et al. (26). The top clinical variables identified were HER2/immunohistochemistry, adjuvant tamoxifen, number of nodes involved, and tumor size.

Analysis of the 21 genes in the recurrence score assay. Fourteen of the 16 cancer-related genes used in the Recurrence Score algorithm correlated with breast cancer recurrence, nine of them at *P* < 0.05 and 14 at *P* < 0.10. Five of the 16 genes belong to a group of proliferation genes (Pearson correlation coefficients range, 0.35-0.60); four of the 16 genes used in the algorithm belong to an ER group; the tightly linked *HER2* and *GRB7* genes and two invasion genes encoding extracellular matrix-degrading proteases are also represented.

Of the 78 evaluable patients in this study, 11 patients (14%) had a Recurrence Score of <18 and had a rate of distant recurrence at 10 years of 29% (95% confidence interval, 0-53%); 19 patients (24%) had a Recurrence Score between 18 and 31 and had a rate of distant recurrence at 10 years of 72% (95% confidence interval, 38-88%); and 48 patients (62%) had a Recurrence Score of \geq 31 and had a rate of distant recurrence at 10 years of 80% (95% confidence interval, 63-89%).

Discussion

The results of this study indicated that gene expression in formalin-fixed, paraffin-embedded tumor tissue obtained at the time of surgery done as long as 24 years ago was associated with the likelihood of distant recurrence in women with breast cancer and \geq 10 nodes. Clinical and pathologic variables, such as age, tumor size, number of involved nodes, and systemic treatment, had only modest correlation with the likelihood of recurrence. HER2 protein expression by immunohistochemistry was significantly correlated with DRFS. Tumor gene expression quantified by RT-PCR for 22 genes was significantly correlated with DRFS. The top genes were *Bcl2* and *TP53BP2*, for which higher expression was associated with longer DRFS, and *GRB7*,

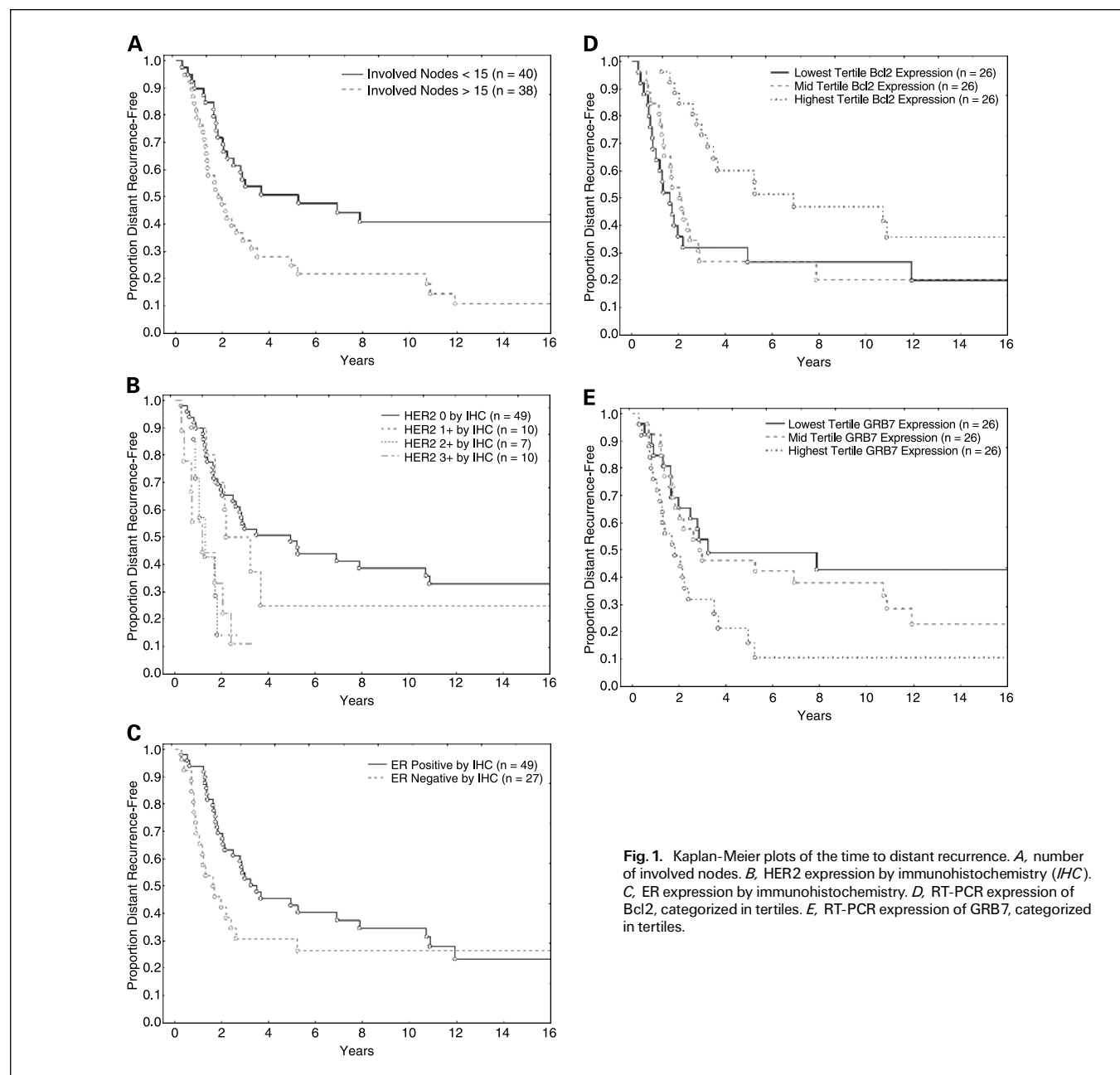


Fig. 1. Kaplan-Meier plots of the time to distant recurrence. *A*, number of involved nodes. *B*, HER2 expression by immunohistochemistry (IHC). *C*, ER expression by immunohistochemistry. *D*, RT-PCR expression of *Bcl2*, categorized in tertiles. *E*, RT-PCR expression of *GRB7*, categorized in tertiles.

for which higher expression was associated with shorter DRFS. Multivariate analysis of the clinical variables, pathology variables, and gene expression by RT-PCR for *Bcl2* and *TP53BP2* indicated that gene expression was independently associated with distant recurrence.

In the univariate analysis of gene expression by RT-PCR and DRFS, 22 of the 203 candidate genes were correlated with DRFS with a $P < 0.05$ (not adjusted for multiple comparisons; Table 2). Because of the relatively small sample size and multiple comparisons, it is likely that a number of these “significant” correlations are false positives, and that some important genes might not be detected. However, our strategy was to perform several independent studies with a similar panel of candidate genes to identify genes that consistently correlate with recurrence over a diverse range of patients and then to take this

shorter list into a large study, where the gene list and algorithm for predicting risk are prospectively defined. Of note, 14 of the 34 genes associated with DRFS as shown in Table 2 were also significantly associated with distant recurrence in two independent studies of a lower risk group of breast cancer patients (23, 24) and were included in the Recurrence Score algorithm used in the subsequent large clinical validation studies (26, 30, 31).

Of the genes in Table 2 that were correlated with DRFS in the univariate analysis ($P_{\text{unadjusted}} < 0.10$), many came from four multigene expression clusters that were defined by two criteria: (a) grouped by published gene expression analysis (15, 17, 27, 32) and/or by published evidence of related biological functions, and (b) coexpression in our study with correlation coefficients of >0.5 . Without exception, genes within a cluster that influenced risk did so in the same direction. Five of the

genes (*ER*, *PR*, *Bcl2*, *SCUBE2*, and *IGF1R*) came from an "ER cluster," where higher expression was associated with longer DRFS. Five of the genes (*CHK1*, *STK15*, *MYBL2*, *cyclin B1*, and *Ki-67*) came from a "proliferation cluster," where higher expression was associated with shorter DRFS. Two of the genes (*GRB7* and *HER2*, which lie within 70 kb on chromosome 17 and are frequently coamplified in breast cancer) had a correlation coefficient of coexpression of 0.82 and form a "HER2 cluster." As expected, higher expression of *GRB7* and *HER2* were associated with shorter DRFS. Finally, three of the genes (*CD68*, *CTSL*, and *CTSL2*) came from a "macrophage cluster," where higher expression was associated with shorter DRFS. This is consistent with previous reports that cancer cells stimulate macrophages to produce various growth factors, angiogenesis factors, and matrix-degrading enzymes that in turn promote tumor growth and survival (33, 34).

Of note, several tested genes that have been previously linked to breast cancer recurrence did not correlate significantly with recurrence in the present study. For example, large clinical trials have correlated urokinase-type plasminogen activator and type 1 plasminogen activator inhibitor with shortened survival (12, 13). Those studies used protein assays (by ELISA) rather than mRNA, a distinction that might explain why we did not find significant correlation with these genes. Van't Veer et al. (17), who also measured mRNA levels, also failed to detect a

significant relationship between urokinase-type plasminogen activator, type 1 plasminogen activator inhibitor, and recurrence. Seventy genes were identified in that microarray study on fresh frozen tissue. Two of the 34 genes identified in this study (Table 2; *SCUBE2* and *BBC3*) overlap with the previously reported 70-gene panel.

The accuracy and specificity of this RT-PCR assay of formalin-fixed, paraffin embedded tumor tissue was supported by comparison of the results of RT-PCR assay of RNA and immunohistochemistry assay of protein for ER, PR, and HER2. Overall, the concordance between the RT-PCR and immunohistochemistry for ER, PR, and HER2 determinations was high. In contrast, the concordance between the RT-PCR measurements and immunohistochemistry assay for Ki-67 was poor. It has been reported previously that it is difficult to standardize and control the immunohistochemistry assay for Ki-67 (35). The results of this study suggest that caution should be exercised when using the immunohistochemistry assay for Ki-67 for clinical decision making.

The data generated in this study contributed to the development of the 21-gene assay and Recurrence Score algorithm that was subsequently tested in a prospective study on the tamoxifen-treated patients in NSABP Study B-14 who were node negative and had tumors that were ER positive (26). The sample size in this study of node-positive patients was not

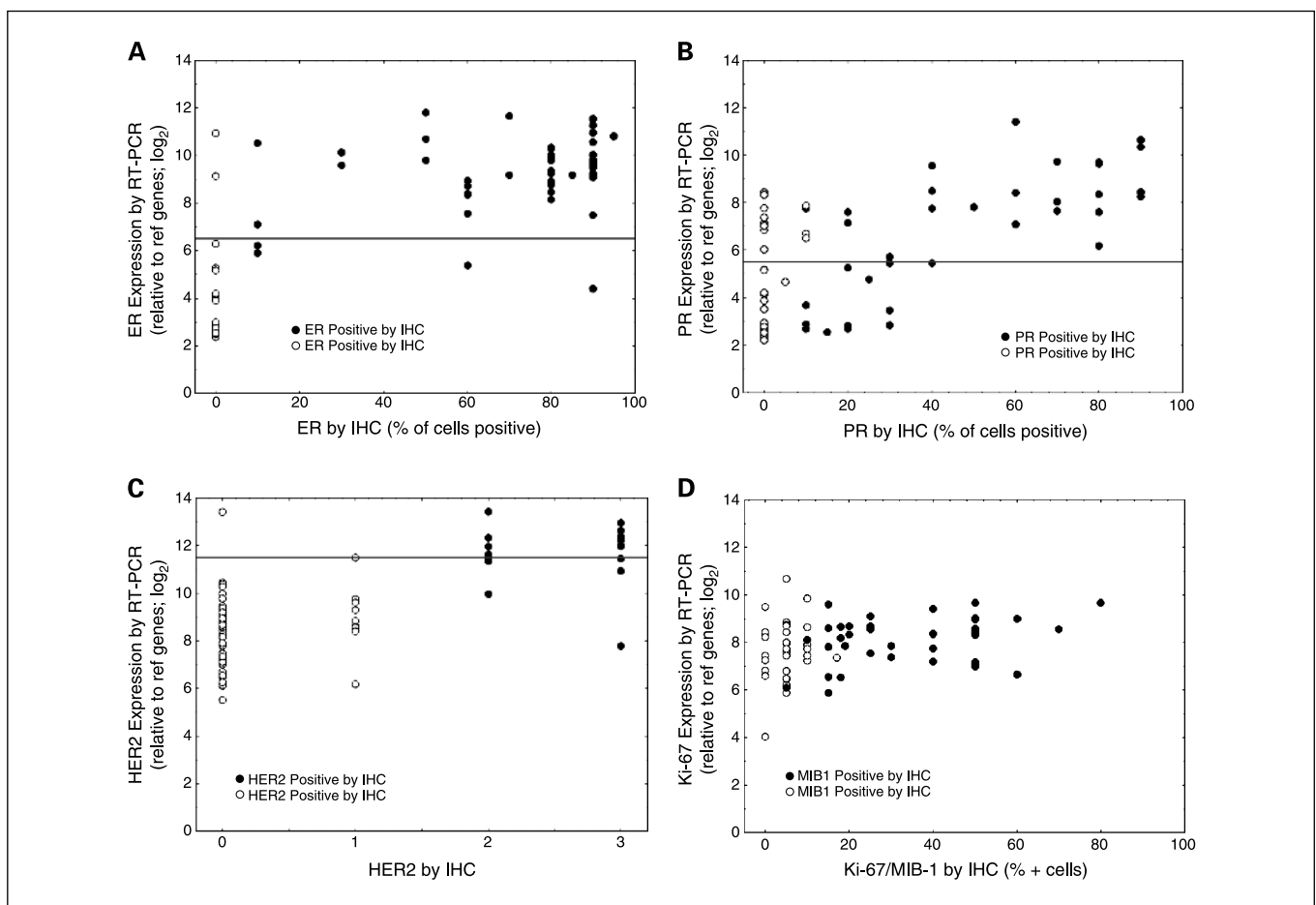


Fig. 2. Gene expression by RT-PCR and immunohistochemistry (IHC) for ER (A), PR (B), HER2 (C), and Ki-67/MIB-1 (D). RT-PCR expression cutoff points for "positive" and "negative" (ER = 6.5, PR = 5.5, and HER2 = 11.5) were derived from studies of other cohorts compared with immunohistochemistry.

Table 3. Stepwise Cox proportional hazards model for clinical variables and gene expression measurement by RT-PCR

Gene variable	Step entered	Hazard ratio* (95% confidence limit)	Wald χ^2	df	P
<i>TP53BP2</i>	1	0.510 (0.387-0.672)	22.881	1	<0.001
<i>Bcl2</i>	2	0.478 (0.355-0.645)	23.308	1	<0.001
<i>EPHX1</i>	3	1.317 (1.139-1.523)	13.775	1	<0.001
<i>PDGFRb</i>	4	1.950 (1.245-3.055)	8.516	1	0.004

NOTE: To better control false discovery rates, stepwise variable selection was done in PROC PH of SAS 9.1 using a significance level of 0.05 for entering an explanatory variable and a significance level of 0.01 for removal of an explanatory variable from the model.

Abbreviation: df, degree of freedom.

*Hazard ratio for every 1 unit change, where each unit represents an ~ 2-fold change in expression.

sufficient to address the relative performance of the Recurrence Score and standard measures, such as patient age, tumor size, and tumor grade. The larger independent NSABP B-14 study was also designed to determine whether the performance of the Recurrence Score exceeded standard measures.

It is important to emphasize that because this data set was used to develop the 21-gene assay, the performance of the Recurrence Score in this population can not be taken as validation. However, as might be expected, the proportion of patients with Recurrence Score of <18 (low-risk group) was much smaller in this cohort of node-positive patients than in cohorts of node-negative patients. In addition, for any Recurrence Score category, the risk of recurrence for patients with ≥ 10 nodes was higher than for patients with negative nodes. New larger studies are needed to more accurately define the relationship between Recurrence

Score and likelihood of recurrence in node-positive patients, especially for patients with one to three pathologically positive nodes or patients with nodes positive by immunohistochemistry alone.

In summary, this study found that tumor gene expression was correlated with the likelihood of distant recurrence in patients who have invasive breast cancer and ≥ 10 nodes and contributed to the development of the 21-gene Recurrence Score assay.

Acknowledgments

We thank Rick Baehner (University of California at San Francisco) for his contributions, Dr. David Roseman for computerizing the breast cancer database, Dr. Craig Silverton for funding the database at Rush University Medical Center, and Drew Watson, Debjani Dutta, Daniel Klaus, Mylan Pho, Anhthu Nguyen, Jennie Jeong, Stephanie Butler, Joel Robertson, Ken Stineman, and Claire Alexander (Genomic Health) for their efforts.

References

1. Singletary SE, Allred C, Ashley P, et al. Revision of the American Joint Committee on Cancer staging system for breast cancer. *J Clin Oncol* 2002;20:3628-36.
2. Yarbrow JW, Page DL, Fielding LP, Partridge EE, Murphy GP. American Joint Committee on Cancer prognostic factors consensus conference. *Cancer* 1999;86:2436-46.
3. Fitzgibbons PL, Page DL, Weaver D, et al. Prognostic factors in breast cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med* 2000;124:966-78.
4. Greene FL, Page DL, Fleming ID, et al. *AJCC cancer staging manual*, 6/e. New York: Springer-Verlag; 2002.
5. Hayes DF. Do we need prognostic factors in nodal-negative breast cancer? *Arbiter. Eur J Cancer* 2000;36:302-6.
6. Esteva FJ, Sahin AA, Cristofanilli M, Arun B, Hortobagyi GN. Molecular prognostic factors for breast cancer metastasis and survival. *Semin Radiat Oncol* 2002;12:319-28.
7. Clinical practice guidelines for the use of tumor markers in breast and colorectal cancer. Adopted on May 17, 1996 by the American Society of Clinical Oncology. *J Clin Oncol* 1996;14:2843-77.
8. Bast RC, Jr., Ravdin P, Hayes DF, et al. 2000 update of recommendations for the use of tumor markers in breast and colorectal cancer: clinical practice guidelines of the American Society of Clinical Oncology. *J Clin Oncol* 2001;19:1865-78.
9. Eifel P, Axelson JA, Costa J, et al. National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1-3, 2000. *J Natl Cancer Inst* 2001;93:979-89.
10. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Senn HJ. Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer. Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. *J Clin Oncol* 2001;19:3817-27.
11. Press MF, Slamon DJ, Flom KJ, Park J, Zhou JY, Bernstein L. Evaluation of HER-2/*neu* gene amplification and overexpression: comparison of frequently used assay methods in a molecularly characterized cohort of breast cancer specimens. *J Clin Oncol* 2002;20:3095-105.
12. Janicke F, Prechtl A, Thomssen C, et al. Randomized adjuvant chemotherapy trial in high-risk, lymph node-negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1. *J Natl Cancer Inst* 2001;93:913-20.
13. Look MP, van Putten WL, Duffy MJ, et al. Pooled analysis of prognostic impact of urokinase-type plasminogen activator and its inhibitor PAI-1 in 8377 breast cancer patients. *J Natl Cancer Inst* 2002;94:116-28.
14. Davis RE, Staudt LM. Molecular diagnosis of lymphoid malignancies by gene expression profiling. *Curr Opin Hematol* 2002;9:333-8.
15. Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000;406:747-52.
16. Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531-7.
17. van't Veer LJ, Dai H, van de Vijver MJ, He YD, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 2002;415:530-6.
18. 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.
19. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995;270:467-70.
20. Godfrey TE, Kim SH, Chavira M, et al. Quantitative mRNA expression analysis from formalin-fixed, paraffin-embedded tissues using 5' nuclease quantitative reverse transcription-polymerase chain reaction. *J Mol Diagn* 2000;2:84-91.
21. Specht K, Richter T, Muller U, Walch A, Werner M, Hofler H. Quantitative gene expression analysis in microdissected archival formalin-fixed and paraffin-embedded tumor tissue. *Am J Pathol* 2001;158:419-29.
22. Cronin M, Pho M, Dutta D, et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. *Am J Pathol* 2004;164:35-42.
23. Esteban J, Baker J, Cronin M, et al. Tumor gene expression and prognosis in breast cancer: multi-gene RT-PCR assay of paraffin-embedded tissue. *Proc Am Soc Clin Oncol* 2003;22:850.
24. Paik S, Shak S, Tang G, et al. Multi-gene RT-PCR assay for predicting recurrence in node negative breast cancer patients-NSABP studies B-20 and B-14. *Breast Cancer Res Treat* 2003;82:A16.
25. Tabesh B, Usha L, Oleske D, et al. Clinical characterization of a tissue repository of high-risk breast cancer patients. *Proc Am Soc Clin Oncol* 21. 2002.
26. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 2004;351:2817-26.
27. Sorlie T, Perou CM, Tibshirani R, et al. Gene

- expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* 2001;98:10869–74.
28. 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.
29. Benjamini Y, Hockberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc* 1995;57:289–300.
30. Esteva FJ, Sahin AA, Coombes K, et al. Multi-gene RT-PCR assay for predicting recurrence in node negative breast cancer patients: MD Anderson Clinical Validation Study. *Breast Cancer Res Treat* 2003;82:A16.
31. Habel LA, Quesenberry CP, Jacobs MK, et al. A large case-control study of gene expression and breast cancer death in the Northern California Kaiser Permanente population. *Breast Cancer Res Treat* 2004;88 [Supplement 1]:S118.
32. Gruberger S, Ringner M, Chen Y, et al. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* 2001;61:5979–84.
33. Bingle L, Brown NJ, Lewis CE. The role of tumour-associated macrophages in tumour progression: implications for new anticancer therapies. *J Pathol* 2002; 196:254–65.
34. Huang S, Van Arsdall M, Tedjarati S, et al. Contributions of stromal metalloproteinase-9 to angiogenesis and growth of human ovarian carcinoma in mice. *J Natl Cancer Inst* 2002;94:1134–42.
35. Mengel M, von Wasielewski R, Wiese B, Rudiger T, Muller-Hermelink HK, Kreipe H. Inter-laboratory and inter-observer reproducibility of immunohistochemical assessment of the Ki-67 labelling index in a large multi-centre trial. *J Pathol* 2002;198: 292–9.