The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JANUARY 18, 2007 VOL. 356 NO. 3

The Prognostic Role of a Gene Signature from Tumorigenic Breast-Cancer Cells

Rui Liu, Ph.D., Xinhao Wang, Ph.D., Grace Y. Chen, M.D., Ph.D., Piero Dalerba, M.D., Austin Gurney, Ph.D., Timothy Hoey, Ph.D., Gavin Sherlock, Ph.D., John Lewicki, Ph.D., Kerby Shedden, Ph.D., and Michael F. Clarke, M.D.

ABSTRACT

BACKGROUND

Breast cancers contain a minority population of cancer cells characterized by CD44 expression but low or undetectable levels of CD24 (CD44+CD24−/low) that have higher tumorigenic capacity than other subtypes of cancer cells.

METHODS

We compared the gene-expression profile of CD44+CD24−/low tumorigenic breast-cancer cells with that of normal breast epithelium. Differentially expressed genes were used to generate a 186-gene "invasiveness" gene signature (IGS), which was evaluated for its association with overall survival and metastasis-free survival in patients with breast cancer or other types of cancer.

RESULTS

There was a significant association between the IGS and both overall and metastasis-free survival ($P<0.001$, for both) in patients with breast cancer, which was independent of established clinical and pathological variables. When combined with the prognostic criteria of the National Institutes of Health, the IGS was used to stratify patients with high-risk early breast cancer into prognostic categories (good or poor); among patients with a good prognosis, the 10-year rate of metastasis-free survival was 81%, and among those with a poor prognosis, it was 57%. The IGS was also associated with the prognosis in medulloblastoma ($P=0.004$), lung cancer ($P=0.03$), and prostate cancer ($P=0.01$). The prognostic power of the IGS was increased when combined with the wound-response (WR) signature.

CONCLUSIONS

The IGS is strongly associated with metastasis-free survival and overall survival for four different types of tumors. This genetic signature of tumorigenic breast-cancer cells was even more strongly associated with clinical outcomes when combined with the WR signature in breast cancer.

From the Departments of Internal Medicine (R.L., G.Y.C., P.D., M.F.C.) and Statistics (K.S.), University of Michigan, Ann Arbor; Oncomed Pharmaceuticals, Mountain View, CA (X.W., A.G., T.H., J.L.); the Stanford Institute for Stem Cell Biology and Regenerative Medicine, Palo Alto, CA (P.D., M.F.C.); and the Department of Genetics, Stanford University School of Medicine, Stanford, CA (G.S.).

Drs. Liu and Wang contributed equally to this article, and Drs. Chen and Dalerba contributed equally to this article.

N Engl J Med 2007;356:217-26.

A growing body of evidence obtained from studies of different types of cancer strongly suggests that only a small subclass of cancer cells within a tumor are actually tumorigenic.[1-4] We have previously shown that in breast cancer, a small population of cancer cells characterized by CD44 expression but low or undetectable levels of CD24 (CD44+CD24−/low) have a high tumorigenic capacity when injected into immunodeficient mice.[1] The rest of the cancer cells, called nontumorigenic breast-cancer cells, have little or no such ability. Tumors in mice that originate from purified tumorigenic breast-cancer cells contain a mixture of both tumorigenic and nontumorigenic breast-cancer cells. Thus, the CD44+CD24−/low population shares with normal stem cells the capacity for self-renewal. The clinical implications of this finding are substantial, because tumorigenic breast-cancer cells may have a high potential to invade and metastasize.

We used gene-expression profiling of tumorigenic breast-cancer cells to develop prognostic tools to assess survival in patients with breast cancer. Gene-expression profiling has been used to predict clinical outcomes in breast cancer.[5-11] We identified 186 genes that are differentially expressed in tumorigenic breast-cancer cells and normal breast epithelium and generated a gene signature that differs substantially from previously reported gene signatures in breast cancer.[5,6,12] Since this signature indicated the likelihood of a tumor to metastasize — a process involving invasion — we called it the "invasiveness" gene signature (IGS). An important finding was that the IGS was associated with the risk of death and metastasis not only in breast cancer but also in lung cancer, prostate cancer, and medulloblastoma. This finding suggests that the IGS represents general biologic features shared by several different types of tumor.

Methods

The methods used to isolate tumorigenic breast-cancer cells and normal breast epithelium are described in detail in the Supplementary Appendix, available with the full text of this article at www.nejm.org. Those used for RNA amplification, microarray analysis, real-time polymerase chain reaction, normalization of microarray data, and generation of the IGS are also described in the Supplementary Appendix.

The flow cytometric and molecular biologic experiments were performed by an academic author, who analyzed the results. An employee of the sponsor analyzed the microarray data, identified the IGS, and performed the statistical analysis. Both the academic authors and the employees of the sponsor had access to and held the microarray data and contributed to the study design, the interpretation of the results, and the writing of the manuscript. Dr. Clarke designed the study and contributed to the analysis of the results and the writing of the manuscript; he made the decision to publish the manuscript and vouches for the completeness of the data and for the analysis.

Patient Data

Patient information, including both clinical data and gene-expression data, was obtained from two independent sources: the Netherlands Cancer Institute database, which included data for 295 consecutive patients with early breast cancer[10,12] (available at http://www.ncbi.nlm.nih.gov/geo/) and the Erasmus Medical Center database, which included data for 286 patients with lymph-node–negative breast cancer[7,8] (available at http://microarray-pubs.stanford.edu/wound_NKI). (A description of the transformation and analysis of the patient data is in the Supplementary Appendix.)

Statistical Analysis

Average linkage clustering was performed with the open-source Cluster software (version 3.0),[13] and the results were visualized with the use of the open-source TreeView software (version 1.0.13).[14] A Pearson correlation coefficient for the correlation between the expression data for each patient and the average expression of the gene signature (the average for the six samples of tumorigenic breast-cancer cells used to generate the IGS) was calculated with the use of the expression level of each of the 186 genes (or a subgroup of the genes available in a specific database). Patients were grouped according to the correlation values, with 0 or the average correlation coefficient used as the threshold. For the breast-cancer databases, we chose the calculated Pearson correlation coefficient of 0 as the threshold to separate patients into two groups, one with gene-expression profiles that were related to the IGS (correlation coefficient, >0) and the other with profiles unrelated to the IGS (correlation coefficient, ≤0).

Because the number of patients with medul-

loblastoma, lung cancer, or prostate cancer for whom data were available was small, we divided the patients into two groups of approximately the same size, using the average correlation coefficient as the threshold. Overall survival and metastasis-free survival in the two groups were analyzed and compared by the Kaplan–Meier method. Differences in survival were tested for statistical significance by the log-rank test with the use of GraphPad Prism software, version 4.03 (GraphPad Software). Univariate and multivariate analyses of survival with the use of the Cox proportional-hazards method were performed with the use of the open-source R software, version 2.1.0 (www.r-project.org).

RESULTS

GENERATION OF THE IGS

We previously identified tumorigenic breast-cancer cells on the basis of their expression of the cell-surface proteins CD44 and CD24, which can be used to distinguish between tumorigenic and nontumorigenic breast-cancer cells.[1] In this study, we defined normal breast-epithelium cells as those that were positive for epithelial-specific antigen or CD10 (cell-surface proteins that mark luminal epithelial or myoepithelial mammary cells). On microarray analysis, we searched for genes differentially expressed between tumorigenic breast-cancer cells isolated from six different breast can-

Table 1. Classification of the 186 Genes in the Invasiveness Gene Signature (IGS).

Class	Genes
Apoptosis	*DPF2, CASP8, BCL2*
Calcium-ion binding	*SCGN, SWAP70, KIAA0276*
Cell cycle	*C10orf9, C10orf7, ALKBH, TOB2*
Cell-surface receptor	*XPR1, CD59, LRP2*
Chemotaxis	*PLP2, MAPK14, CXCL2*
Collagen catabolism	*MMP7*
Differentiation	*MGP, MLF1, FLNB*
Ion-channel activity	*SCNM1*
Membrane protein	*HSPC163, C5orf18, MGC4399, CDW92, TMC4, ZDHHC2, TICAM2, KDELR3*
Metabolism	*GNPDA1, THEM2, DBR1, FLJ90709, FLJ10774, C16orf33, GAPD, LDHA, MR-1, LARS, GTPBP1, PRSS16, WFDC2, AIM1, DHRS6, DHRS4, MGC15429, MGC45840, ECHDC2, GOLGIN-67, AFURS1, KIAA0436, CYP4V2, JTV1*
Methyltransferase	*ICMT, DNMT3A, HNMT, METTL7A, METTL2*
Morphology	*VIL2, TPD52, ARPC5*
Nucleotide binding	*NOL8, NSF, RAD23B, SRP54, HSPA2, PBP, THAP2, CIRBP, SNRPN, KIAA0052*
Phosphatase	*DUSP10*
Proliferation	*SSR1, ERBB4, EMP1, CHPT1, LRPAP1*
Protein binding	*FLJ11752, CSTF1, KLHL20, DNAJC13, APLP2, ARGBP2, DNAJB1, NEBL, SH3BGRL, NUDT5, GABARAPL1, MAPT, DCBLD1*
Protein kinase	*STK39, PAK2, CSNK2A1, PILRB, ERN1, SGKL, WEE1, MAST4, C11orf17*
Protein transport	*NUP37, CLTC, COPB2, SLC25A25*
Signal transduction	*ECOP, PDE8A, STAM, TUBB, SNX6, RAB23, PLAA, STC2, LTF*
Transcription factor	*ISGF3G, ATXN3, GTF3C3, GSK3B, KLF10, ELL2, ZBTB20, IRX3, ETS1, SERTAD1, MGC4251, MAFF, SFPQ, CITED4, CEBPD, EIF4E2*
Transferase	*HS2ST1, AGPS, PGK1, ATIC, ETNK1, ALG2*
Ubiquitination	*NCE2, MARCH8, CNOT4, RNF8, PSMA5, DPF2*
Function unknown	*AMMECR1, KIAA1287, LOC144233, LOC286505, PNAS-4, FLJ20530, THUMPD3, MGC45564, CAP350, ETAA16, HAN11, DNAPTP6, C7orf25, FLJ37953, FLJ10587, C7orf36, ELP4, NDEL1, NPD014, DKFZP564D172, FAM53C, IER5, LOC255783, KIAA0146, KIAA0792, LOC439994, LOC283481, CG018, LOC130576, NGFRAP1L1, KIAA1217, C4orf7, C21orf86, C9orf64, FLJ13456, KIAA1600, B7-H4, LOC80298, C7orf2, NUCKS, DKFZP566D1346, LOC388279, FLJ31795, C6orf107, FLJ12439, FLJ12806, FLJ39370*

cers (see Table 1 in the Supplementary Appendix) and normal breast-epithelium cells derived from three reduction mammoplasties. In five samples of breast cancers with sufficient numbers of cells for analysis, we confirmed that the CD44+CD24−/low lineage cells, but not the other cancer cells, formed tumors in the mouse tumor assay (see Table 1 in the Supplementary Appendix). A set of 186 genes were selected on the basis of a significant difference, by a factor of 2, in the expression levels between tumorigenic breast-cancer cells and normal breast epithelium. A level of significance of 0.005 by Student's t-test was chosen for the comparison (for a description of the procedure, see the Methods section, Fig. 1, and Table 2, all in the Supplementary Appendix).

Table 1 presents the classification of the 186 genes in the IGS (for a detailed annotation of the IGS, see Gene Annotation in the Supplementary Appendix). The list includes the genes involved in the nuclear factor-κB pathway, the RAS–mitogen-activated protein kinase pathway, and epigenetic control of gene expression (see the Results section in the Supplementary Appendix). These pathways have been shown to play critical roles in tumorigenesis, cell differentiation, and development. We compared the IGS signature with previously reported gene signatures in breast cancer.[15] Of the 186 genes in the IGS, only 6 overlapped with those in the wound-response (WR) signature,[12] and none of the genes in the IGS were found in other signatures, suggesting that the IGS is a new signature.

THE IGS AND OUTCOME IN BREAST CANCER

A Pearson correlation coefficient for the correlation between the expression value of the genes in the sample obtained from each patient in the database and that of the genes in the IGS was calculated, and the correlation coefficient was tested for its association with clinical outcomes in a survival analysis with the use of a Cox proportional-hazards model. With the use of the Pearson correlation coefficient, patients were divided into two groups: one group with gene-expression profiles that were related positively to the IGS (correlation coefficient, >0) and the other group with gene-expression profiles that were related negatively to the IGS (correlation coefficient, ≤0). A positive correlation was associated with reduced metastasis-free survival and reduced overall survival; the univariate hazard ratio for metastasis or death was 1.3 (P<0.001) and 1.4 (P<0.001), respectively, for each increase of 0.1 in the correlation coefficient (Table 2). The estimated 10-year rates of overall survival and metastasis-free survival were 98% and 82%, respectively, among patients in the group with a correlation coefficient of 0 or less, and 62%

Table 2. Risk of Death or Metastasis among Patients with Breast Cancer (Univariate Analysis).*

Variable	Death		Metastasis	
	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value
IGS†	1.4 (1.0–1.4)	<0.001	1.3 (1.2–1.5)	<0.001
Tumor diameter‡	1.5 (1.2–1.8)	0.001	1.4 (1.2–1.7)	<0.001
Mastectomy vs. no mastectomy	1.2 (0.8–1.9)	0.40	1.4 (0.9–2.1)	0.10
Positive estrogen-receptor status vs. negative status	0.3 (0.2–0.5)	<0.001	0.5 (0.4–0.8)	0.005
Tumor grade				
Poor vs. good differentiation	4.7 (1.6–13.4)	0.005	2.4 (1.2–4.8)	0.01
Intermediate vs. good differentiation	10.2 (3.7–28.3)	<0.001	4.3 (2.2–8.2)	<0.001
Age§	0.6 (0.4–0.8)	0.004	0.5 (0.4–0.7)	<0.001
Positive lymph-node status vs. negative status	0.9 (0.6–1.4)	0.56	0.9 (0.6–1.3)	0.53
No adjuvant therapy vs. chemotherapy or hormonal therapy	1.2 (0.8–1.9)	0.40	1.3 (0.9–2.0)	0.19

* The analysis included data for 295 patients with breast cancer in the Netherlands Cancer Institute database.
† The correlation coefficient for the IGS was modeled as a continuous variable. The hazard ratio is for each increase of 0.1 in the correlation coefficient.
‡ Tumor diameter was modeled as a continuous variable. The hazard ratio is for each increase of 1 cm in diameter.
§ Age was modeled as a continuous variable. The hazard ratio is for each 10-year increase in age.

and 54%, respectively, in the group with a correlation coefficient greater than 0 (Fig. 1).

A similar analysis was performed on the data from the Erasmus Medical Center for 109 of the 186 genes in the IGS; this database includes information only on metastasis-free survival. This analysis showed that the risk of metastasis was significantly higher among patients with an expression profile that correlated with the IGS (correlation coefficient, >0) than among those with

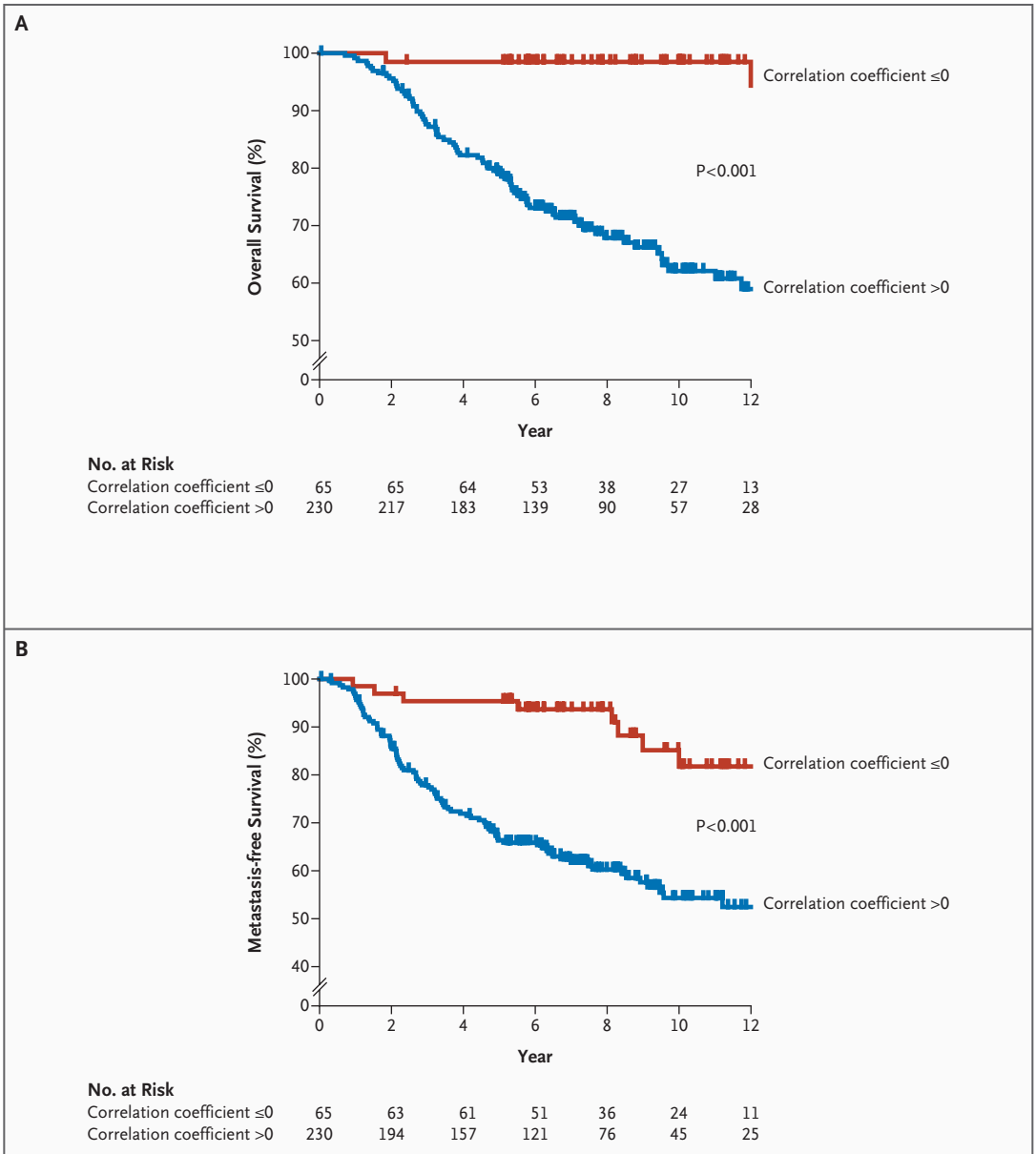


Figure 1. Association between the IGS and Survival in Patients with Breast Cancer.

A Pearson correlation coefficient was calculated for the correlation between the IGS and each of the 295 tumors included in the Netherlands Cancer Institute database on the basis of the expression values of the 186 genes included in the gene signature. Tumors were separated into two groups according to the correlation values, with 0 used as the threshold. Kaplan–Meier survival curves for the two groups were compared, with overall survival (Panel A) and metastasis-free survival (Panel B) as the clinical end points. Patients with tumors with a gene-expression pattern that was similar to the IGS (correlation coefficient, >0) had worse outcomes than those with tumors with a gene-expression pattern that was not similar to the IGS (correlation coefficient, ≤0).

an expression profile that did not correlate with the IGS (correlation coefficient, ≤0) and with relapse rates of 43% (96 of 224 patients) and 16% (10 of 62 patients), respectively (P=0.001 by the chi-square test).

THE IGS AND CLINICAL AND PATHOLOGICAL CRITERIA

To determine whether the association between the IGS and the clinical outcome in patients with breast cancer was independent of standard clinical and pathological criteria, the 295 patients with breast cancer identified in the Netherlands Cancer Institute database were stratified according to tumor size, lymph-node status, histologic grade, and estrogen-receptor status. A univariate Cox proportional-hazards model was used to evaluate the association of the IGS with the clinical outcome in each category (Table 3). The association between the IGS and the risk of death or metastasis was significant regardless of tumor size or lymph-node status (P<0.05). Furthermore, the IGS could be used to stratify tumors with intermediate differentiation into good and poor prognostic subcategories (hazard ratio for a poor prognosis, 1.6; 95% confidence interval [CI], 1.2 to 2.1; P<0.001), but was less useful for stratifying tumors with poor and good differentiation (P>0.05). The association with the outcome was significant for tumors that were positive for estrogen receptor but not for those that were negative for estrogen receptor. In a multivariate Cox proportional-hazards analysis (Table 4), the association between the IGS and death or metastasis was independent of tumor differentiation, patients' age, and estrogen-receptor status (P<0.05).

Table 3. IGS as a Prognostic Factor, According to Characteristics of Breast Cancer.*

Variable	Death		Metastasis	
	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value
Tumor size†				
T1	1.4 (1.1–1.7)	0.002	1.3 (1.1–1.6)	0.003
T2	1.3 (1.1–1.5)	0.004	1.2 (1.1–1.4)	0.008
Lymph-node involvement				
No	1.4 (1.2–1.7)	<0.001	1.4 (1.1–1.6)	<0.001
Yes	1.3 (1.1–1.6)	0.003	1.3 (1.1–1.5)	0.005
Differentiation				
Poor	1.0 (0.9–1.3)	0.72	1.1 (0.9–1.3)	0.62
Intermediate	1.6 (1.2–2.1)	<0.001	1.5 (1.2–1.8)	<0.001
Good	1.3 (0.7–2.5)	0.34	1.1 (0.8–1.6)	0.48
Estrogen-receptor status				
Negative	1.2 (0.9–1.5)	0.14	1.3 (1.0–1.7)	0.08
Positive	1.4 (1.1–1.6)	<0.001	1.3 (1.1–1.5)	<0.001

* The analysis included data for the 295 patients with breast cancer in the Netherlands Cancer Institute database, with the prognostic role of the IGS tested within each patient category. The correlation coefficient for the IGS was modeled as a continuous variable. Confidence intervals were calculated per 0.1 increase in the correlation coefficient.

† T1 denotes a tumor with a diameter less than or equal to 2.0 cm, and T2 a tumor with a diameter greater than 2.0 cm.

THE IGS AND BENEFIT OF ADJUVANT TREATMENT

The use of adjuvant chemotherapy in breast cancer is based on the prognostic criteria of the NIH (http://consensus.nih.gov) and the St. Gallen criteria.[16] We identified two groups of patients in the database of the Netherlands Cancer Institute who were at high risk for death or metastasis (284 identified according to the NIH prognostic criteria and 273 according to the St. Gallen criteria) and who would ordinarily be treated with chemotherapy. With the use of the Pearson correlation coefficient and the Kaplan–Meier method, the 284 patients identified on the basis of the NIH criteria were stratified according to the IGS. Of those patients, 21% (60 patients) had tumors with gene-expression profiles that correlated negatively with the IGS (correlation coefficient, ≤0); among these 60 patients, metastases developed in 13% (8 patients), and the 10-year rate of metastasis-free survival was 81% (see Fig. 2A and 2B in the Supplementary Appendix). By contrast, the gene-expression profile was strongly correlated with the IGS (correlation coefficient, >0) in 224 of the 284 patients; among these patients, metastatic disease developed in 41% (91 patients), and the 10-year rate of metastasis-free survival was 57% (P<0.001). When the same analysis was performed for the 273 high-risk patients identified on the basis of the St. Gallen criteria, the results were almost identical (P<0.001) (see Fig. 2C and 2D in the Supplementary Appendix).

Of the 295 patients with breast cancer in the Netherlands Cancer Institute database, 185 never received adjuvant chemotherapy. In this subgroup we evaluated the IGS independently of the confounding effects of adjuvant treatment. With the correlation threshold set at 0, the IGS signature divided the 185 patients into two groups: 42 patients with IGS-negative tumors (correlation coefficient, ≤0) and 143 patients with IGS-positive

Table 4. Risk of Death or Metastasis among Patients with Breast Cancer (Multivariate Analysis).*

Variable	Death		Metastasis	
	Hazard Ratio (95% CI)	P Value	Hazard Ratio (95% CI)	P Value
IGS†	1.2 (1.0–1.4)	0.03	1.2 (1.1–1.4)	0.004
Tumor diameter‡	1.2 (0.9 –1.5)	0.16	1.2 (1.0–1.6)	0.06
Mastectomy vs. no mastectomy	1.2 (0.8–1.9)	0.40	1.3 (0.9–2.0)	0.23
Positive estrogen-receptor status vs. negative status	0.5 (0.3–0.8)	0.006	0.9 (0.6–1.5)	0.70
Tumor grade				
Poor vs. good differentiation	4.8 (1.6–14.2)	0.005	2.3 (1.1–4.7)	0.03
Intermediate vs. good differentiation	3.9 (1.3–11.3)	0.01	2.1 (1.0–4.2)	0.04
Age§	0.6 (0.4–0.9)	0.02	0.5 (0.4–0.8)	0.0007
Positive lymph-node status vs. negative status	1.1 (0.5–2.2)	0.84	1.2 (0.7–2.3)	0.49
No adjuvant therapy vs. chemotherapy or hormonal therapy	1.1 (0.5–2.3)	0.80	1.5 (0.8–2.9)	0.20

* The analysis included the 295 patients with breast cancer in the Netherlands Cancer Institute database.
† The correlation coefficient for the IGS was modeled as a continuous variable. The hazard ratio is for each increase of 0.1 in the correlation coefficient.
‡ Tumor diameter was modeled as a continuous variable. The hazard ratio is for each increase of 1 cm in diameter.
§ Age was modeled as a continuous variable. The hazard ratio is for each 10-year increase in age.

tumors (correlation coefficient, >0). Kaplan–Meier survival curves showed that the relapse rates at 10 years differed significantly in the two groups: 12% (5 patients) and 43% (62 patients), respectively (P<0.001).

COMBINED USE OF THE IGS AND WR GENE SIGNATURE

The 512-gene WR signature is derived from transcriptional profiling of serum-stimulated fibroblasts. This signature correlates with overall survival and metastasis-free survival in patients with breast cancer.[8] The IGS and the WR signature are representations of different biologic phenomena and are based on nonoverlapping lists of genes. We compared the signatures and determined whether the result would be better with the two combined than with either signature alone. For this purpose, we used data on 262 patients in the Netherlands Cancer Institute database. When each signature was used alone, the significance of the association with the outcome was similar (see the Results section, Fig. 3, and Table 3a, all in the Supplementary Appendix). In a multivariate Cox proportional-hazards analysis the IGS and the WR signature performed independently, with a hazard ratio for metastasis of 1.3 (95% CI, 1.1 to 1.5; P=0.001) and 1.2 (95% CI, 1.1 to 1.4; P=0.003),

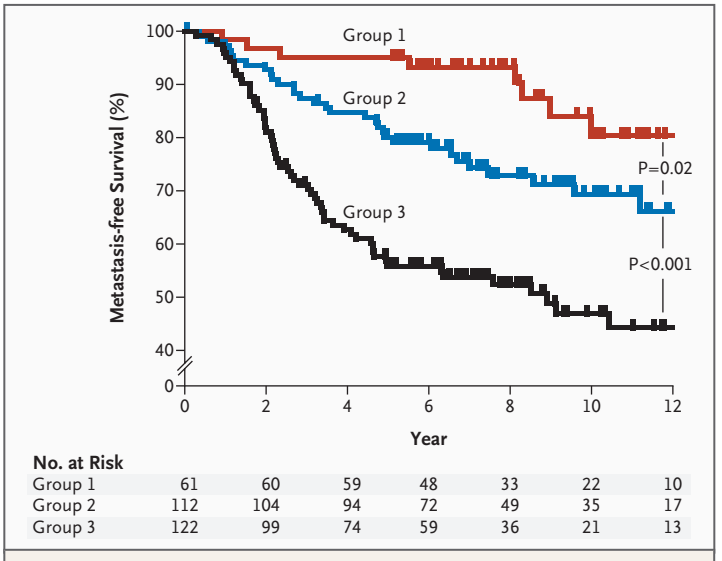


Figure 2. Risk of Metastasis According to Combined Use of the IGS and WR Signature.

Group 1 included low-risk patients with a quiescent WR signature (as defined by Chang et al.[8,12]; see the Methods section in the Supplementary Appendix) and a negative IGS (correlation coefficient, ≤0), denoting a good prognosis. Group 2 included intermediate-risk patients with either an activated WR signature or a positive IGS (correlation coefficient, >0), denoting a poor prognosis. Group 3 included high-risk patients with both an activated WR signature and a positive IGS (correlation coefficient, >0), denoting a poor prognosis. At 10 years, relapse-free survival in the low-risk group, the intermediate-risk group, and the high-risk group was 80%, 69%, and 47%, respectively.

respectively (see Table 3b in the Supplementary Appendix). This result suggested that the two signatures combined would perform better than either alone. Using the full data on the 295 patients from the Netherlands Cancer Institute database and both the IGS and the WR signature, we found that after 10 years of follow-up, metastatic disease had developed in 20%, 31%, and 53% of patients with tumors that were negative for both signatures, positive for one signature, or positive for both signatures, respectively (Fig. 2). The results were almost identical when the analysis was performed for the subgroup of patients at high risk for relapse, according to the prognostic criteria of both the NIH and St. Gallen (see Fig. 4 in the Supplementary Appendix).

Figure 3 (facing page). Association between the IGS and Survival among Patients with Lung Cancer, Medulloblastoma, or Prostate Cancer (Panel C).

A Pearson correlation coefficient was calculated for the correlation between the IGS and each tumor, on the basis of the expression values of the 186 genes included in the IGS or the subclass of genes available in each data set. Tumors were separated into two groups according to the correlation values, with the average used as the threshold. Kaplan–Meier survival curves for the two groups were compared, with overall survival (Panels A and B) and relapse-free survival (Panel C) as the clinical end points. Patients with tumors with a gene-expression pattern that was more similar to the IGS (a correlation coefficient that was higher than the average value) had worse outcomes than those with tumors with a gene-expression pattern that was less similar to the IGS (a correlation coefficient that was the same as or less than the average value).

THE IGS IN OTHER TYPES OF TUMORS

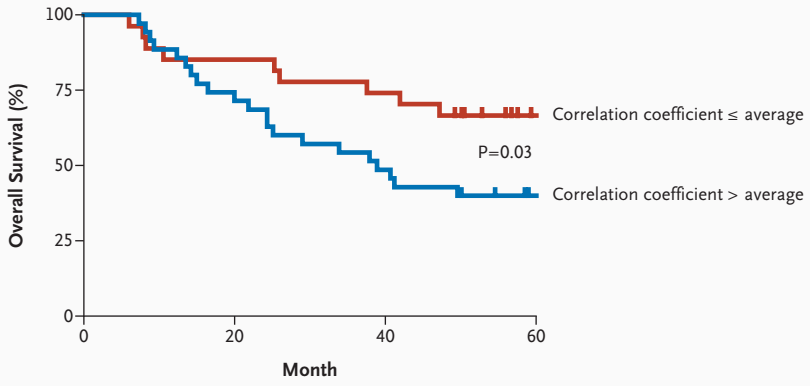
To evaluate the contribution to the prognostic capacity of the IGS of transcriptional features that are specific to tumorigenic breast-cancer cells, as compared with nontumorigenic breast-cancer cells, we compared the transcriptional profiles of autologous pairs of tumorigenic and nontumorigenic breast-cancer cell populations that had been directly isolated from the three primary samples (see the Methods section and Table 1 in the Supplementary Appendix). On the basis of the 186 genes in the IGS, gene-expression profiles derived from tumorigenic breast-cancer cells correlated significantly with overall survival and metastasis-free survival at 10 years ($P<0.001$ for both comparisons), as compared with expression profiles from nontumorigenic breast-cancer cells ($P=0.3$ for overall survival, and $P=0.06$ for metastasis-free survival) (see Fig. 5 in the Supplementary Appendix). To validate this finding, we compared the gene-expression profiles of both tumorigenic and paired primary nontumorigenic breast-cancer cells with those of normal breast epithelium and evaluated the genes differentially expressed (tumorigenic breast-cancer cells vs. normal breast epithelium, and nontumorigenic breast-cancer cells vs. normal breast epithelium). The gene-expression profiles generated from tumorigenic breast-cancer cells were more closely associated with the outcome than were the gene-expression profiles generated from nontumorigenic breast-cancer cells ($P<0.05$) (see Fig. 6 in the Supplementary Appendix).

On the basis of these observations, we investigated whether the IGS could be applied to other types of cancer by testing the association of the IGS with overall survival and relapse-free survival in patients with lung cancer, prostate cancer, or medulloblastoma. With the use of the Pearson correlation analysis, data on 62 patients with lung cancer, 60 with medulloblastoma, and 21 with prostate cancer were analyzed, and patients were categorized according to tumors that were IGS-negative (a correlation coefficient that was less than or equal to the average value) and tumors that were IGS-positive (a correlation coefficient that was greater than the average value). With all three types of cancer, a gene-expression profile that was closely correlated with the IGS signature was associated with rates of overall or relapse-free survival that were less than 50% at 5 years (Fig. 3A). Overall survival or relapse-free survival was higher among patients with a tumor that was negatively correlated with the IGS (>60% among those with lung cancer and >80% among those with medulloblastoma or prostate cancer; $P=0.03$, $P=0.004$, and $P=0.001$, respectively) (Fig. 3B and 3C).

DISCUSSION

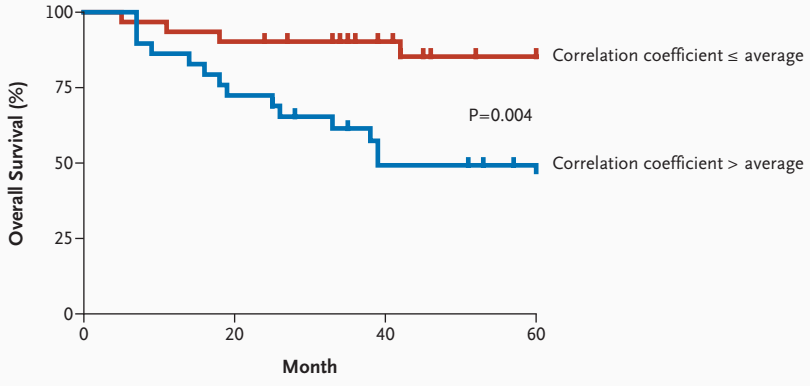
We compared the gene-expression profiles of CD44+CD24−/low tumorigenic breast-cancer cells and normal breast epithelium, using gene-expression microarrays. We identified a unique list of 186 differentially expressed genes and defined a gene signature that we named IGS. The association between the IGS and both overall survival and tumor recurrence was significant not only in patients

A Lung Cancer



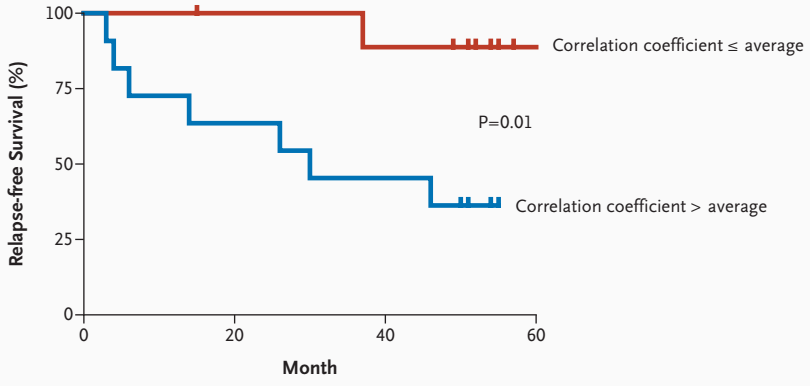
No. at Risk				
Correlation coefficient ≤ average	27	24	21	10
Correlation coefficient > average	35	25	17	10

B Medulloblastoma



No. at Risk				
Correlation coefficient ≤ average	31	28	19	13
Correlation coefficient > average	29	22	14	10

C Prostate Cancer



No. at Risk				
Correlation coefficient ≤ average	10	9	9	2
Correlation coefficient > average	11	7	5	0

with breast cancer but also in those with lung cancer, prostate cancer, or medulloblastoma. The contribution of the tumorigenic breast-cancer–cell component of the IGS was essential to its association with the clinical outcome. These results suggest the clinical relevance of the CD44+CD24–/low tumorigenic subclass of breast-cancer cells.

There are several possible explanations for the association between the IGS and clinical outcomes in four different types of tumors. First, the signature may detect the genetic fingerprints of invasion pathways that are activated in aggressive tumors. Second, since the signature was derived from cells that behave like stem cells in mouse xenograft assays, it may detect an increased number of "cancer stem cells" in tumors.[1-4] Tumors with a large number of cancer stem cells may be more likely to metastasize than tumors with a small number of cancer stem cells. The observation that the association between the IGS and the clinical outcome can be improved when the IGS is combined with the 512-gene WR signature is consistent with a model in which self-renewing cancer stem cells represent the seeds of cancer and the tumor microenvironment the soil that promotes the growth of those seeds.[17] Third, the IGS may detect transcriptional profiles associated with mutations that arrest cells in an immature state of differentiation and function as markers of more aggressive tumors.

In summary, we found that with the use of the IGS, patients with early breast cancer who are at high risk for metastasis or death can be stratified into two groups with substantially different 10-year relapse rates (13% vs. 41%), and more than 90% of patients in whom metastatic breast cancer develops can be identified. There was an association between the IGS and the clinical outcome in patients who had breast tumors with intermediate-grade differentiation, a group whose prognosis is difficult to assess. Further validation and refinement of the IGS by characterization of the gene subclass that acts synergistically with the WR signature or other gene signatures will help to establish and exploit the full clinical value of the IGS.

Supported by grants from the National Cancer Institute (T32 CA009357 to Dr. Chen, and CA104987 and CA126524 to Dr. Clarke), the Breast Cancer Research Foundation, the Morton Foundation, and the Virginia and D.K. Ludwig Foundation (to Dr. Clarke), and the Fondazione Italiana per la Ricerca sul Cancro and the California Institute for Regenerative Medicine (to Dr. Dalerba).

Drs. Wang, Gurney, Hoey, and Lewicki report being employees of Oncomed Pharmaceuticals, a biotechnology company that has applied for patents related to the gene signature described in this study; and Dr. Clarke, being a paid member of the advisory board of Oncomed Pharmaceuticals and owning stock in the company. No other potential conflict of interest relevant to this article was reported.

We thank T. Iwashita for discussions of the RNA amplification protocol; D. Adams, M. White, and A.-M. Des Lauriers for help with cell sorting; J. Washburn and J. MacDonald for help with the microarray analysis; and D. Renner-Chuey for help with collection of the tumor samples.

REFERENCES

1. Al-Hajj M, Wicha MS, Benito-Hernandez A, Morrison SJ, Clarke MF. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A 2003;100:3983-8. [Erratum, Proc Natl Acad Sci U S A 2003;100:6890.]
2. Lapidot T, Sirard C, Vormoor J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature 1994;367:645-8.
3. Reya T, Morrison SJ, Clarke MF, Weissman IL. Stem cells, cancer, and cancer stem cells. Nature 2001;414:105-11.
4. Singh SK, Hawkins C, Clarke ID, et al. Identification of human brain tumour initiating cells. Nature 2004;432:396-401.
5. 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.
6. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. Nature 2002;415:530-6.
7. Wang Y, Klijn JG, Zhang Y, et al. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet 2005;365:671-9.
8. Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A 2005;102:3738-43.
9. 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.
10. 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.
11. Glinsky GV, Berezovska O, Glinskii AB. Microarray analysis identifies a death-from-cancer signature predicting therapy failure in patients with multiple types of cancer. J Clin Invest 2005;115:1503-21.
12. Chang HY, Sneddon JB, Alizadeh AA, et al. Gene expression signature of fibroblast serum response predicts human cancer progression: similarities between tumors and wounds. PLoS Biol 2004;2:E7.
13. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics 2004;20:1453-4.
14. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95:14863-8.
15. Fan C, Oh DS, Wessels L, et al. Concordance among gene-expression-based predictors for breast cancer. N Engl J Med 2006;355:560-9.
16. Goldhirsch A, Glick JH, Gelber RD, Coates AS, Thurlimann B, Senn HJ. Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005. Ann Oncol 2005;16:1569-83.
17. Fidler IJ. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. Nat Rev Cancer 2003;3:453-8.