Immutable Functional Attributes of Histologic Grade Revealed by Context-Independent Gene Expression in Primary Breast Cancer Cells

Shanaz H. Dairkee,[1] Aejaz Sayeed,[1] Gloria Luciani,[1] Stacey Champion,[1] Zhenhang Meng,[1] Lakshmi R. Jakkula,[2] Heidi S. Feiler,[2] Joe W. Gray,[2] and Dan H. Moore[1]

[1]California Pacific Medical Center Research Institute, San Francisco, California and [2]Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California

Abstract

Inherent cancer phenotypes that are independent of fluctuating cross-talk with the surrounding tissue matrix are highly desirable candidates for targeting tumor cells. Our novel study design uses epithelial cell lines derived from low versus high histologic grade primary breast cancer to effectively diminish the breadth of transient variability generated within the tumor microenvironment of the host, revealing a "paracrine-independent expression of grade-associated" (PEGA) gene signature. PEGA members extended beyond "proliferation-driven" signatures commonly associated with aggressive, high-grade breast cancer. The calcium-binding protein *S100P* was prominent among PEGA genes overexpressed in high-grade tumors. A three-member fingerprint of *S100P*-correlated genes, consisting of *GPRC5A, FXYD3*, and *PYCARD*, conferred poor outcome in multiple breast cancer data sets, irrespective of estrogen receptor status but dependent on tumor size ($P < 0.01$). *S100P* silencing markedly diminished coregulated gene transcripts and reversed aggressive tumor behavior. Exposure to pathway-implicated agents, including the calmodulin inhibitor *N*-(6-aminohexyl)-5-chloro-1-naphthalenesulfonamide, phenothiazine, and chlorpromazine, resulted in rapid apoptotic cell death in high-grade tumor cells resistant to the chemotherapeutic drug cisplatin. This is the first comprehensive study describing molecular phenotypes intimately associated with histologic grade whose expression remains relatively fixed despite an unavoidably changing environment to which tumor cells are invariably exposed. [Cancer Res 2009;69(19):7826–34]

Introduction

It has long been known that the microscopic appearance and arrangement of tumor cells, defined as histologic grade, holds key information about malignant behavior and patient outcome. Criteria for the widely used Scarff-Bloom-Richardson system of grading and its modification by Elston and Ellis (1) subclassify primary breast cancer into low-grade or well-differentiated tumors (grade 1), intermediate-grade or moderately differentiated tumors (grade 2), and high-grade or poorly differentiated tumors (grade 3). High-grade tumors (HTG) result in early treatment failures, whereas low- and intermediate-grade tumors recur after longer time intervals, independently of clinical stage (2–4), with the exception of medullary carcinomas, which display HTG and have a relatively favorable prognosis (5). The 10-year survival is reduced from 76% for patients with grade l to 39% for those with grade 3 tumors (6). Based on the microscopic evidence of a striking proliferative differential, conventional systemic chemotherapy exploits the higher mitotic index of HTG. However, the effects of such drugs on nonmalignant tissues and organ systems are undesirable and often intolerable. Additionally, therapeutic resistance is a major obstacle in achieving complete response. Current research strives to address these deficiencies through the identification of molecular markers at diagnostic presentation, which predict long-term efficacy of systemic therapy in both adjuvant and neoadjuvant settings.

In this regard, gene expression profiles offer the scope and sensitivity to unveil aberrant changes underlying tumor aggressiveness and therapeutic vulnerability. A critical limitation in the clinical translation of such phenotypes, displayed as a single snapshot in time, includes the incorporation of transient, localized stromal influences within the afflicted tissue. Although differentially expressed signatures, such as the "grade gene index," have contributed toward improving grade-based subclassification (7–9), due to inordinate proliferative differences between the sample classes, functional insights required for tumor targeting remain limited. We have implemented an approach, which overrides the proliferative differential between tumor cells of varying histologic grade, thereby revealing additional functional phenotypes that contribute collectively to the biological basis of this key clinical parameter. Technical roadblocks that have previously persisted in the execution of this goal include maintenance of pristine, continuously proliferating epithelial cell populations from the full spectrum of primary breast cancer. Even when isolated from fresh tumor tissue, such malignant populations routinely display a finite *in vitro* life span in contrast to the common belief that cell immortalization is an inevitable consequence of tumorigenesis. On the other hand, rare spontaneously immortalized breast cancer cell lines have originated exclusively from metastatic or high-grade primary carcinoma (10–16). To provide a representation of the histologic and clinical heterogeneity of human breast disease, we have used experimental models of early and late clinical stage, and low-, intermediate-, and high-grade primary breast tumors (17–21), and more recently of nonmalignant high-risk breast tissue (22), developed through incremental improvements in selective cell isolation, short-term propagation, genotypic and phenotypic characterization, and immortalization. This study describes the molecular profiles of clinically based novel cellular models of low- and intermediate-grade primary breast cancer, as yet unavailable in the field. It uses these toward the

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Shanaz H. Dairkee, California Pacific Medical Center Research Institute, 475 Brannan Street, San Francisco, CA 94107. Phone: 415-600-1653; Fax: 415-600-1725; E-mail: dairkes@cpmcri.org or Dan H. Moore, California Pacific Medical Center Research Institute, 475 Brannan Street, San Francisco, CA 94107. Phone: 415-600-1567; Fax: 415-600-1725; E-mail: mooredx@cpmcri.org.

doi:10.1158/0008-5472.CAN-09-1564

identification of a constitutively expressed gene signature to facilitate an understanding of the functional biology of histologic grade for application in drug discovery efforts. A practical application of the tumor biology revealed through this innovative approach is exemplified by the *in vitro* recapitulation of differential *S100P* expression associated with tumor grade, its role in clinical stratification with novel coregulated gene fingerprints, and in the functional verification of predictive target response to pathway-guided therapy.

Materials and Methods

Tissue Culture, RNA Isolation, and Microarrays

Epithelial cultures were derived from clinical primary breast cancer specimens under Institutional Review Board–approved guidelines as previously described (18, 19). For cell immortalization, 16 independent cases were transduced with the *hTERT* gene as described (21). RNA isolated from subconfluent cultures of the immortalized cell lines was subjected to routine quality control measures and hybridized to the HG-U133A chip (Affymetrix) according to the manufacturer's protocols. Affymetrix software was used for image analysis and probe quantification to generate CEL files. Data collection and analysis are as described by Chin and colleagues (23). For quantitative real-time reverse transcription-PCR analysis, cDNA was synthesized and analyzed as before (21) by an Applied Biosystems 5700 Sequence Detection System. For primer sequences, see Supplementary Table S1.

S100P was silenced by small interfering RNA (siRNA) transfection of the sense strand sequence 5′-ACAAGGAUGCCGUGGAUAA-3′ (Dharmacon Research) using Lipofectamine 2000 (Invitrogen). siCONTROL nontargeting siRNA#1 served to evaluate off-target effects. *S100P* expression was normalized to that of the reference gene *ACTB* to control for differences in the amount and/or quality of total RNA isolated from different cell pellets.

Immunofluorescence and fluorescence-activated cell sorting analysis

Expression of S100P protein in tumor cell lines was measured by indirect immunofluorescence of a mouse monoclonal primary antibody to S100P (BD Biosciences) and detection with Alexa Fluor 488–conjugated secondary antibody (Invitrogen).

To measure growth rate after *S100P* knockdown, starting on day 4 after transfection with *S100P* siRNA or siCONTROL, triplicate sets of bromodeoxyuridine (BrdUrd)–incorporated cell suspensions were immunostained with anti-BrdUrd, acquired by FACScan (BD Biosciences), and analyzed using CellQuest Pro software. Ten thousand cells were acquired in each of three independent runs. Apoptotic cell fractions were analyzed 48 h after transfection using Annexin V-FITC and propidium iodide kit (MBL International) followed by FACScan analysis as described above.

Drug Response

Cells were plated at a density of 200,000 per well in a six-well plate and, 24 h later, treated with 200 μmol/L cisplatin, 10 μmol/L chlorpromazine, or *N*-(6-aminohexyl)-5-chloro-1-naphthalenesulfonamide (also known as W7; Sigma) for 24 h. Cells were then stained with Annexin V-FITC and propidium iodide and analyzed by FACScan as above.

Migration Assay

Cells were placed in inserts with hanging geometry (Becton Dickinson) in six-well plates and allowed 4 d for migration through the 8-μm pores in the polyester membrane bottom of the insert. Inserts were removed, and migrated cells in the plate bottom were propagated as colonies for an additional 4 d, fixed, stained with crystal violet, and counted. Migration potential was determined by the number of colonies at the end of the experiment divided by the number of cells initially seeded in the insert.

Coculture of primary tumor cell lines and fibroblasts was set up in a similar format with the exception that the insert pore size was reduced to 0.4 μm. Tumor cell RNA was isolated after 3 d of continuous coculture.

Statistical Analysis

Further analyses of annotated and filtered data were based on 6,053 probe sets that displayed ≥2-fold variation across low-grade versus high-grade primary breast cancer cell lines. Validation of the array data as described below follows published guidelines (24).

Paracrine-independent expression of grade-associated signature. We used the R program *samr* [significance analysis of microarrays (SAM); ref. 25] to select genes whose expression levels were significantly correlated with the pathologist-assigned histologic grade of the tumor tissue from which each of the cell lines was derived. The initial comparison included three grade 1 versus four grade 3 cell lines. Setting the *samr* parameter δ equal to 0.35 gave a false discovery rate of 1.19% for 105 induced and 0.79% for 536 repressed probe IDs. The expression pattern associated with high-grade cells was designated as the paracrine-independent expression of grade-associated (PEGA) gene signature. The validity of the association between cell culture–based PEGA genes and histologic grade of patient tumors was tested on data sets mentioned below. Using a hierarchical tree program in conjunction with the top 10 induced and 10 repressed PEGA probe IDs, false-positive and false-negative rates were calculated to show agreement between true grade and the PEGA classifier-based grade assignment in published

Table 1. Clinicopathologic characteristics of novel primary breast cancer cell lines

Sample ID	Histologic type	Age at diagnosis (y)	TNM stage	Histologic grade	Current *in vitro* passages
CCdl22	Invasive ductal carcinoma	NA	1	Low	171
CCdl67	Invasive ductal carcinoma	61	2	Low	55
CCdl68	Invasive ductal carcinoma	52	2	Low	100
CCdl61	Invasive lobular carcinoma	62	3	Intermediate	79
CCdl66	Invasive ductal carcinoma	47	2	Intermediate	57
CCdl78	Invasive ductal carcinoma	78	1	Intermediate	50
CCdl631	Invasive ductal/lobular carcinoma	49	3	Intermediate	105
CCdl1570	Invasive ductal/lobular carcinoma	66	2	Intermediate	70
CCdl1599	Invasive ductal carcinoma	64	1	Intermediate	90
CCdl329	Invasive lobular carcinoma	42	2B	Intermediate	37
CCdl54	Invasive ductal carcinoma	59	4	High	99
CCdl257	Invasive ductal carcinoma	41	2B	High	94
CCdl672	Invasive ductal carcinoma	60	2	High	80
CCdl675	Invasive ductal carcinoma	29	2B	High	79

Abbreviation: TNM, tumor-node-metastasis.

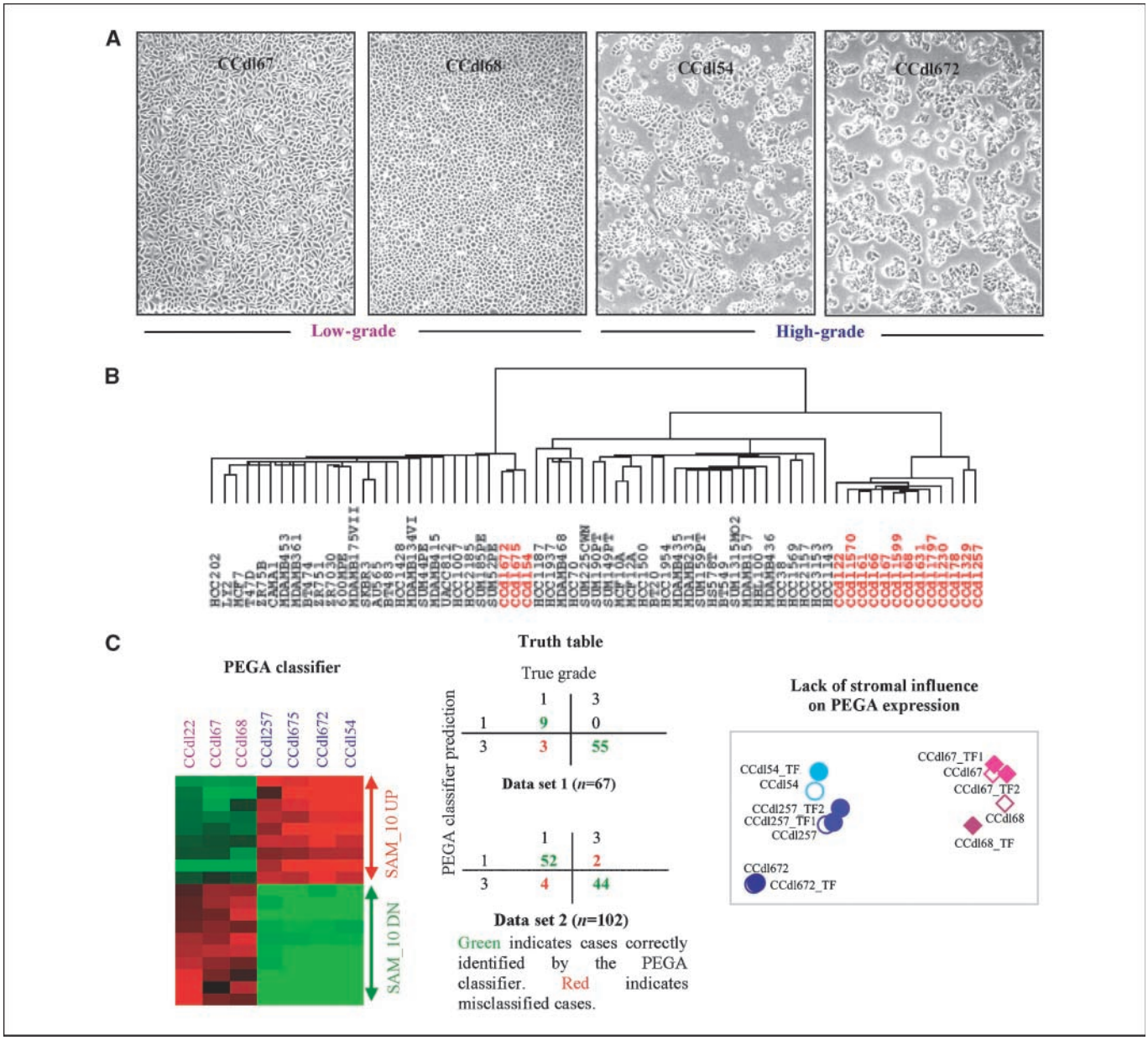

Figure 1. Distinctive profiles of proliferating primary breast tumor cells of varying histologic grade. *A,* microscopic phenotype of histologic grade in four independent breast epithelial cell lines derived from LTG (*left* and *middle left*) and HTG (*right* and *middle right*). *B,* hierarchical clustering of 67 independent breast epithelial cell lines based on global gene expression profiling. Sixteen novel primary tumor cell lines encompassing low, intermediate, and high histologic grade are displayed in red lettering (CCdl1797 and CCdl230 are ungraded). *C, left,* SAM-identified, PEGA-based top 10 up-regulated (*UP*) and down-regulated (*DN*) probe IDs distinguish proliferating tumor cells of low grade (*purple*; $n = 3$) versus high grade (*blue*; $n = 4$). *Middle,* cross-classification analysis of independent clinical breast cancer data sets using the PEGA classifier depicted by the adjoining heat map. *Right,* plot of the first two principal components of the top five up-regulated and down-regulated PEGA gene expression levels in HTG and LTG cell lines in the presence or absence of tumor fibroblast (*TF*) coculture. TF1 and TF2 are independent cases. *Open symbols,* control tumor cell lines; *filled symbols,* tumor plus fibroblast cocultures. *Circles,* HTG cell lines; *diamonds,* LTG cell lines.

cancer data sets. Kaplan-Meier survival curves for recurrence-free survival (RFS) were computed from follow-up data in these data sets.

S100P-COR. Genes specifically associated with S100P in primary tumor cell lines in the 95th percentile (most correlated) or in the 5th percentile (most anticorrelated) were identified. For tumor tissue data, using a criterion of 0.3 or greater correlation with S100P, genes common to two data sets were identified. Both gene lists were evaluated for enriched Gene Ontology terms. Kaplan-Meier analysis was used to compute RFS in these data sets. The statistical significance of the resultant hazard ratios was determined by the *P* value of the likelihood ratio test.

Patient array data sets. The following publicly available primary breast cancer microarray data sets matching our array platform were used for clinical validation of PEGA-derived genes. For clinical follow-up, data set 1[3] (E-TABM-158; ref. 23) was composed of 112 cases with 30 recurrences and median follow-up of 7.9 y, and data set 2[4] (GSE 6532; ref. 26) was composed of 317 cases with 106 recurrences and median follow-up of 8.9 y. For grade distribution, data set 1 was composed of 12 grade 1 versus 55 grade 3 tumors, and data set 2 was composed of 61 grade 1 versus 52 grade 3 tumors.

[3] http://www.ebi.ac.uk/arrayexpress
[4] http://www.ncbi.nlm.nih.gov/geo

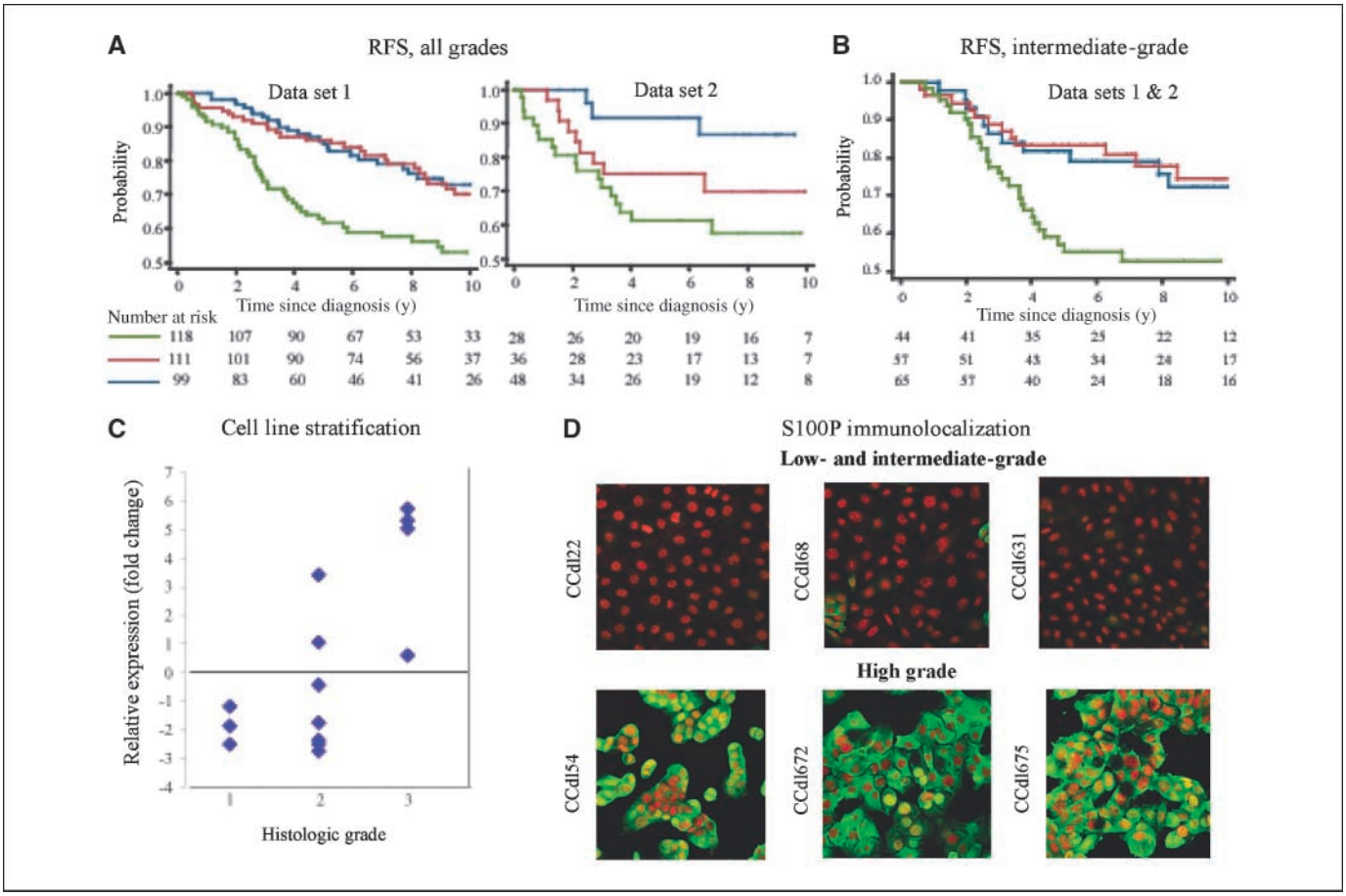

Figure 2. Identification of aggressive primary breast cancer by expression levels of the PEGA gene *S100P. A,* Kaplan-Meier analysis. *S100P* expression divided into tertiles: *blue,* low; *red,* intermediate; *green,* high. Log-normalized data values for each group: low risk, 3.0 to 7.07; intermediate risk, 7.08 to 9.46; high risk, 9.47 to 13.2. RFS for all grades, displayed independently for data sets 1 and 2. *B,* RFS for intermediate-grade tumors in data sets 1 and 2 combined. *C,* tumor cell lines of varying grade accurately stratified by *S100P* expression levels relative to *ACTB* and grade of original tumor tissue. *D,* immunofluorescence localization of anti-S100P (*green*) in LTG or ITG (*top*) versus HTG (*bottom*) cell lines. Nuclei counterstained with propidium iodide (*red*).

Results

An autonomously regulated profile represented by PEGA genes. Fourteen independent primary breast tumor cell cultures of known histologic grade (and 2 additional cell lines from ungraded primary breast tumors) were isolated and immortalized with *hTERT* transduction in a single step. These are designated as the CCdl series (Table 1). Morphologically, epithelial cultures originating from HTG were distinct from those of low tumor grade (LTG) even after the emergence of homogeneous, immortalized cell populations (Fig. 1*A*). The new tumor cell lines were rigorously authenticated by direct comparison with the patient's constitutive genotype obtained from nonmalignant tissue DNA (data not shown).

Global gene expression data of the CCdl lines were combined with that of 51 previously established breast epithelial cell lines and examined by unsupervised hierarchical clustering. Three of four HTG CCdl cell lines were well integrated with other known HTG-derived tumor cell lines of a putative luminal molecular subtype. The remaining 13 cell lines, representing 3 LTG, 7 intermediate tumor grade (ITG), 1 HTG, and 2 ungraded cases, clustered apart in a distinct subgroup, suggesting the presence of attributes that distinguish them from the malignant and nonmalignant breast epithelial cell lines represented in the two major clusters (Fig. 1*B*). SAM of HTG (n = 4) and LTG (n = 3) CCdl tumor cell line data identified 641 differentially expressed (105 induced and 536 repressed in HTG) probe IDs, designated as the PEGA signature.

Toward minimizing the effect of background noise, a SAM-defined classifier based on the top 10 induced and 10 repressed (in high-grade cultures) PEGA probe IDs (Fig. 1*C, left*) was used to confirm association with pathologist-designated histologic grade in clinical breast cancer samples. This PEGA classifier achieved 95% correct classifications for grade 1 (LTG) and grade 3 (HTG) in two independent breast cancer data sets (Fig. 1*C, middle*). In comparison, an arbitrary classifier based on marginal relative frequencies of the two grades provided only 42% correct classifications. Remarkably, grade-associated expression profiles of both HTG and LTG cell cultures were maintained despite the induction of cell immortalization.

To ascertain the absence of a significant effect induced by stromal fibroblasts on PEGA profiles, gene expression was measured in both HTG and LTG cell lines cocultured with one or more independent samples of tumor-derived fibroblasts. Principal component analysis based on the top 10 PEGA genes (5 up-regulated and 5 down-regulated), namely, *ASS1, ALDH2, S100P, CLU, FGF13, FYN, FLRT2, AKT3, F2R*, and *SNCA*, was used to evaluate changes in gene expression levels of cocultured tumor cells. Although HTG and LTG cell lines clustered in different parts of a two-dimensional space, however, each of the five controls was tightly grouped with its test sample counterpart exposed to paracrine signaling from stromal fibroblasts (Fig. 1*C, right*).

Overexpression of a PEGA gene, *S100P*, stratifies intermediate-grade tumors *in vivo* and *in vitro*. The top 10 overexpressed PEGA genes were further tested in conjunction with two independent clinical data sets for predicting patient outcome. Using backward stepwise selection, among the grade predictors retained for joint significance at $P < 0.05$, only *S100P*, a calcium ion binding and calcium-dependent protein binding protein, attained significance in predicting RFS at the single gene level. *S100P* expression tertiles defined for the combined data from two independent data sets were used to separately evaluate RFS in each data set. Significant associations were observed for both data sets [data set 1: hazard ratio, 1.8; 95% confidence interval (95% CI), 1.2–1.74; $P = 0.001$; data set 2: hazard ratio, 1.3; 95% CI, 1.1–1.6; $P = 0.002$; Fig. 2*A*].

We then asked whether expression of *S100P* could serve to dichotomize ITG. In tumor tissue, *S100P* expression was significantly associated with RFS at the cutpoint used above (data set 2: hazard ratio, 2.09; 95% CI, 1.2–3.4; $P = 0.004$; Fig. 2*B*). Similarly *in vitro*, *S100P* expression in seven ITG-derived cell lines ranged from HTG-like ($n = 2$) to LTG-like ($n = 5$) transcript levels (Fig. 2*C*). In agreement with the expression profiling data, indirect immunofluorescence of *S100P* protein in HTG cell lines displayed strong nuclear/cytoplasmic staining, whereas in LTG and some ITG cell lines, immunostaining was undetectable to minimal (Fig. 2*D*).

A role for *S100P*-correlated genes in the manifestation of tumor grade. In the search for a broader scope of biological processes, which link HTG and poor clinical outcome to *S100P*, we identified genes whose expression levels were significantly correlated to *S100P* in clinical tumor tissue (S100P-CORtt). A criterion of 0.3 or greater correlation with *S100P* identified 17 probe IDs (14 genes) common between two independent breast cancer data sets (Supplementary Table S2). A higher correlation threshold produced too few genes, whereas lower resulted in too many. Known functions for the 14-gene set of S100P-CORtt genes included calcium ion binding, transport, and regulation (*PLA2G10, CAPN9, CALML5, GPRC5A*, and *ST6GALNAC2*); glycosylation (*PMM2*); cell adhesion (*CEACAM6, MUC5B, CAPN9*, and *PLA2G10*); oxidoreductase activity (*KMO*); apoptosis (*PYCARD* and *NTHL1*); cell differentiation and transforming growth factor β (TGFβ) signaling (*SLC2A10, CEACAM6, CAPN9, FXYD3, CAPN9, CALML5*, and *CTSD*); and cell motility and invasion (*CTSD* and *CEACAM6*).

In CCdl cell lines, expression of genes at significantly higher correlation (≥0.9 or −0.9) to S100P, designated as S100P-CORcc, represented association with biological processes similar to those observed for the 14-member S100P-CORtt gene set, for example, *FGFR1, MYO1B, GNA15*, and *SFN* (calcium flux); *JAG1, S100A2*, and *EHD2* (calcium ion binding); *BSPRY* (ion transport); *TMEM8, CD44, AZGP1, FLRT3*, and *COL4A6* (cell adhesion); *ALDH5A1* (oxidoreductase activity); *SFN* and

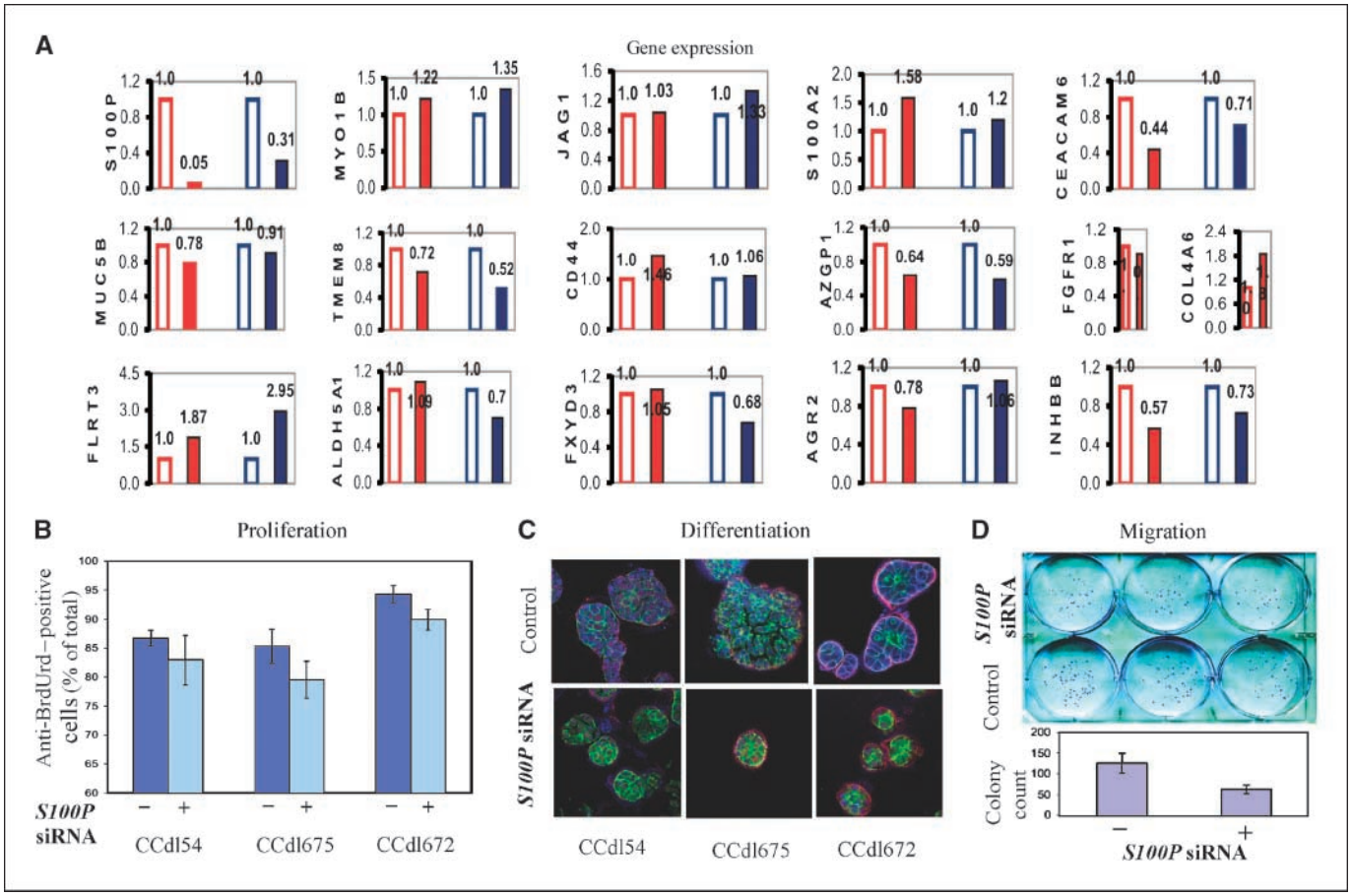

Figure 3. Pleiotropic effects of *S100P* silencing. *A,* expression levels of S100P-COR genes in HTG cell lines altered by *S100P* knockdown. *Red columns,* CCdl54; *blue columns,* CCdl675; *open columns,* control siRNA; *solid columns, S100P* siRNA. *B,* reduction in the number of anti–BrdUrd-stained proliferating cells in three HTG cell lines transfected with *S100P* siRNA. *Columns,* mean of triplicate assays; *bars,* SD. *C, top,* apolar morphology of HTG colonies propagated in Matrigel; *bottom,* normalized by *S100P* silencing to resemble polarized acinar structures. *Blue,* immunolocalization of anti-S100P; *green,* actin localization by phalloidin; *red,* immunolocalization of anti–integrin α6. *D,* reduction in migration potential induced by *S100P* siRNA in HTG cell lines. *Columns,* mean of triplicate assays; *bars,* SE.

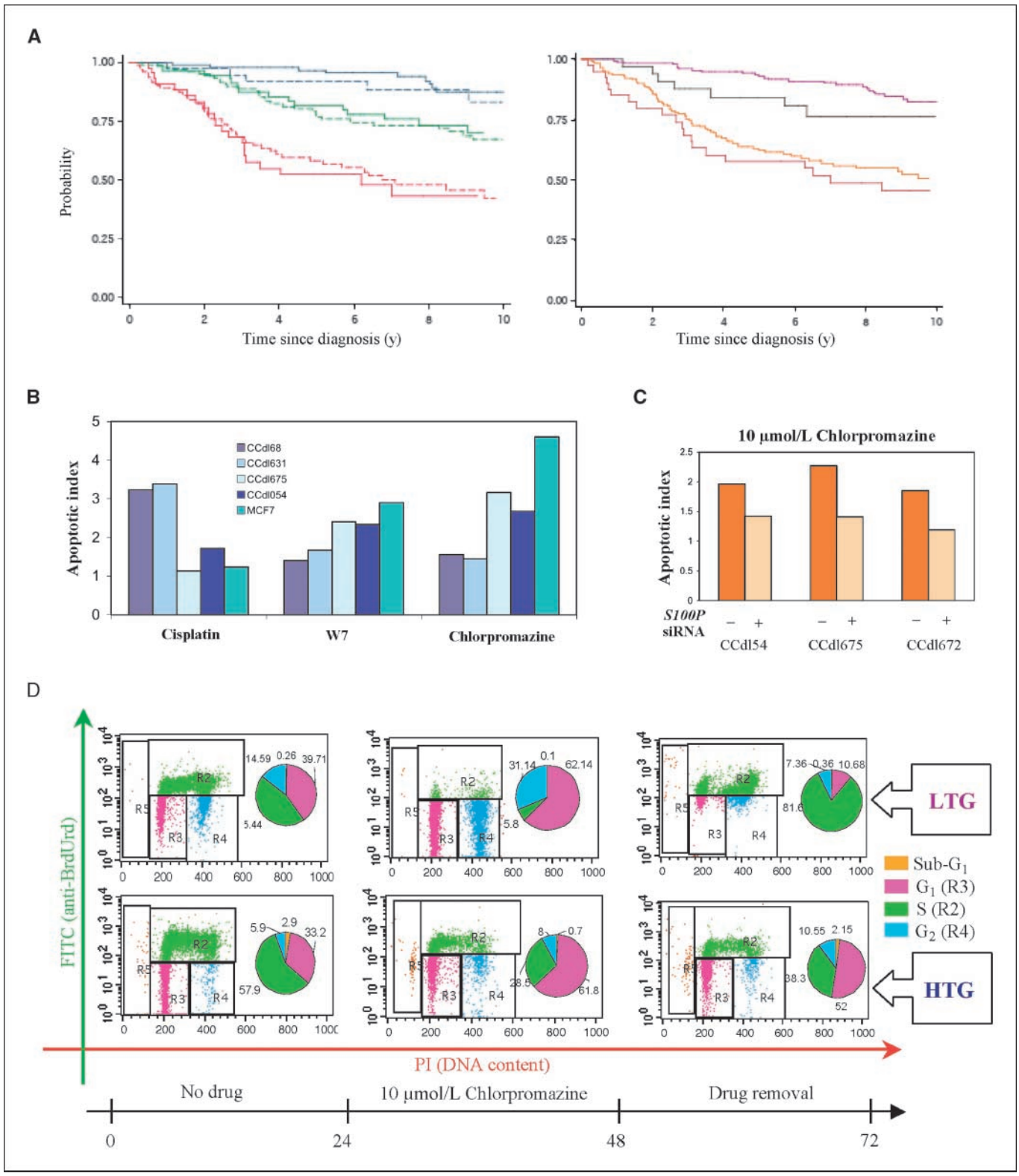

Figure 4. Stratification of primary breast cancer by tumor size plus S100P-COR three-gene fingerprint and chemotherapeutic response measurements. *A, left,* comparison of training versus test data sets. *Solid lines,* training set; *dashed lines,* test set. *Blue,* low-risk tertile; *green,* intermediate-risk tertile; *red,* high-risk tertile. *Right,* comparison of ER-negative versus ER-positive cases in the combined data sets based on a median value of the classifier. *Purple,* low-risk/ER-positive; *black,* low-risk/ER-negative; *yellow,* high-risk/ER-positive; *maroon,* high-risk/ER-negative. *B,* apoptosis induction measured as Annexin V–positive cells in tumor cell lines of varying grade exposed to 200 μmol/L cisplatin, 10 μmol/L chlorpromazine, or W7. Values represent statistically significant differences between unexposed controls and drug-treated cells ($P < 0.05$). *C, S100P* siRNA–induced loss of chlorpromazine sensitivity shown as declining number of Annexin V–positive apoptotic cells in three independent HTG cell lines. *D,* cell cycle analysis of LTG (CCdl68) and HTG (CCdl675) cell lines in the presence and absence of chlorpromazine. Representative fluorescence-activated cell sorting analysis of cells without drug treatment (*left*), with 24-h drug exposure (*middle*), or 24-h postrecovery from drug exposure (*right*). The numbers for pie chart segments indicate the percentage of cells within each cell cycle phase (sub-G_1, G_1, S, and G_2). *PI,* propidium iodide.

Table 2. RFS estimates for training and test data sets based on tumor size plus three-gene classifier

Data set	Risk group	5-y survival probability (%)	(95% CI)	10-y survival probability (%)	(95% CI)
Training	Low	97	(0.90–0.99)	88	(0.77–0.93)
	Medium	77	(0.67–0.84)	67	(0.56–0.76)
	High	58	(0.47–0.68)	42	(0.30–0.54)
Test	Low	92	(0.77–0.97)	83	(0.63–0.93)
	Medium	82	(0.69–0.90)	70	(0.55–0.81)
	High	52	(0.36–0.66)	43	(0.26–0.59)

TP73L (apoptosis); *GNAI1* and *RRAS* (G protein signaling); *INHBB, JAG1, AGR2*, and *INHBA* (cell differentiation and TGFβ signaling); and *MMP14* and *RRAS2* (cell migration).

Toward a functional understanding of the role of the PEGA gene *S100P* and genes correlated to it in tumor cells of varying histologic grade, targeted transcript reduction was induced using *S100P* siRNA. In these cells, a significant reduction in S100P protein levels was confirmed by indirect immunofluorescence of anti-S100P (data not shown). Together with a reduction in *S100P* transcript levels, siRNA-transfected HTG cell lines displayed a rapid reversal of several S100P-CORcc and S100P-CORtt genes, providing a functional association for the correlative data (Fig. 3*A*). Dramatic cellular effects induced by diminished levels of S100P-CORcc and S100P-CORtt transcripts included (*a*) an appreciable loss of the BrdUrd-incorporating proliferating cell fraction (Fig. 3*B*); (*b*) marked growth reduction in the three-dimensional substrate, Matrigel, resulting in significantly smaller colony size, and rearrangement of cell nuclei from a random, apolar configuration to a relatively organized central alignment (Fig. 3*C*); and (*c*) a significant reduction in migration potential (Fig. 3*D*).

Targeting high-grade primary breast tumors through a PEGA gene pathway. To test the effect of the functional attributes of tumor aggressiveness conferred by S100P-COR genes at a clinical level, we applied the 14-gene S100P-CORtt classifier toward outcome stratification, independently of *S100P*. First, classic clinical parameters, such as patient age at diagnosis, clinical stage, histologic grade, tumor size, lymph node status, and estrogen receptor (ER) status, were tested for RFS prediction by combining two independent data sets. These provided information for all six parameters in 440 tumors with 139 recurrent cases. Among the classic variables, only tumor size and ER status were found to be significantly related to outcome in this diverse group of patients. For further analysis, we divided these data into training (n = 292) and test (n = 148) sets using a uniform random number generator.

In the next step, the 14-gene classifier was added to ER status and tumor size for backward stepwise selection of significant predictors of RFS in the training set. Tumor size and 3 of 14 genes (*FXYD3, GPRC5A*, and *PYCARD*) maintained statistical significance (P < 0.01; Supplementary Table S3). The size plus three-gene predictor was combined linearly, using coefficients from the Cox proportional hazards model applied to the training set, to generate a risk score for each patient. The 33rd and 67th percentiles of the risk scores in the training set subdivided both training and test sets into patient groups at high, medium, and low risk of recurrence. Survival in the three risk subgroups differed significantly in each data set (P < 0.001 by log-rank test in each set; Fig. 4*A, left*). Moreover, the log-rank test was significant in each ER category (P = 0.0274 for ER-negative tumors and P < 0.001 for ER-positive tumors; Fig. 4*A, right*). Table 2 summarizes the results and shows that RFS in the test set, using the same classification scheme as the training set, gave closely similar results, thereby providing robust data validation.

CIs for 5- and 10-year survival in each subgroup of the training set include the survival estimates of the test subgroup and vice versa (Table 2). A log-rank test for survival differences within each subgroup revealed no differences between survival in the training and test group within each stratum (P = 0.98 for differences in the lowest risk, P = 0.54 for differences in the middle risk, and P = 0.84 for differences in the highest risk stratum). Our model was further validated using the PSEP test, defined as the difference between the predicted probability of survival in the best and worst groups at a selected time (27). In our study, at 5 years, survival was 0.97 for the low-risk group and 0.58 for the high-risk group, with a PSEP of 0.39 (0.97–0.58) in the training set. In the test set, PSEP was 0.40 (0.92–0.52). Thus, PSEP for training and test set survival was almost identical. Another validation test consisted of comparing Harrell's C statistic in the training and test sets. The statistic is equal to the number of times a pair of patients, selected at random, has their survival time correctly predicted by the model. In our training data, the model correctly predicted the survival order 70% of the time, whereas in the test set the correct order was predicted 68% of the time. Both of these percentages are considered to be quite high for breast cancer prediction (28).

Finally, toward the goal of defining candidate drugs, which efectively address molecular heterogeneity in breast cancer for targeting pertinent pathways, tumor cell response was compared between a conventional chemotherapeutic drug, cisplatin, and compounds whose mode of action is calcium dependent, such as phenothiazine, chlorpromazine, and the calmodulin inhibitor W7. Almost no apoptosis induction was observed in HTG primary breast cancer cell lines or in the MCF7 cell line in response to 200 μmol/L cisplatin. However, these cell lines displayed striking sensitivity to the calcium-binding drugs (Fig. 4*B*). Reversal of drug response in *S100P* siRNA–transfected HTG cell lines showed a direct role for *S100P* and coregulated genes in this regard (Fig. 4*C*). In these experiments, the sensitivity of HTG cell lines to both chlorpromazine and W7 was persistent over time (Fig. 4*D*). On removal of drugs after a 24-h exposure period, HTG cells continued to be eliminated through apoptotic cell death, whereas G_1-arrested LTG cells rapidly reentered the cell cycle.

Discussion

Here, our approach of evaluating global expression profiles of tumor cell cultures of extreme histologic grades yet proliferating at relatively comparable rates has significantly amplified the detectability of unsuspected grade-associated targets. Based on the widely accepted fact that routine tissue culture conditions are inadequate in representing the tumor microenvironment with regard to oxygen,

and glucose concentrations, and a variety of paracrine interactions, our experimental strategy uniquely afforded the level of resolution required to uncover a functional basis for the diagnostic differential in histologic differentiation patterns long used clinically for breast cancer subclassification and patient management. Such a context-independent profile of tumor aggressiveness as depicted by PEGA could assist in sifting apart cellular programs of cancer cell autonomy from the broad interactive dimension of an infiltrating malignant population amidst a wide spectrum of surrounding host cells. Clinically, this could enable rapid improvements in disease stratification and tumor targeting.

As shown, top-tier differentially expressed PEGA genes are a robust discriminant of histologic grade in tumor-derived cell cultures as well as in clinical tumor tissue. Among these, *CEACAM6, MUC5B*, and *S100P* are known indicators of patient outcome in many tumor types, including breast cancer (29–31). However, little is known about their functional role or mechanism of action during tumor progression. In previous microarray analyses of archived tumors with respect to grade or outcome, differential expression of PEGA genes was masked and not found to be as distinctive as proliferation-related genes. In fact, as shown in a recent meta-analysis (32), proliferation was the single common driving force for nine independent prognostic signatures (7, 33–40). In contrast, PEGA genes defined here are predictive without prior selection based on patient outcome and reflect the basic infrastructure of cell signaling pathways underlying the prognostic performance of proliferation genes. Uniquely, in the derivation of this context-independent signature, issues related to stromal and epithelial heterogeneity within archived tumor tissue, observed to be a limitation in the verification and clinical application of candidate prognostic cancer markers in general, are easily circumvented.

In view of the significance of the calcium ion in the cell differentiation process (41), the "backbone" of histologic grading schemes, we have highlighted the role of the PEGA member *S100P* and coregulated genes in the maintenance of the dedifferentiated, HTG phenotype both *in vivo* and *in vitro*. S100 proteins display cell type–specific expression patterns and regulate cytoplasmic Ca^{2+} levels (42). Consistent with this role, we have shown that transcript levels of *S100P* and coregulated genes associated with HTG in primary breast carcinoma parallel increased proliferation, decreased apoptosis, and greater migration potential in high-grade, tumor-derived cell cultures. Conversely, experimental silencing of *S100P* initiates the appearance of epithelial polarity, a phenotype characteristic of well-differentiated human breast epithelial cells (43).

Defining comprehensive phenotypes of histologic grade, which portray molecular and functional traits, is a pivotal objective toward unambiguous, quantitative measures of breast cancer stratification. Based on gene expression cutpoint, a grade gene index classification suggests that intermediate-grade tumors essentially represent high or low histologic grade and reflect the outcome predicted for the two extreme grades (8). Such an approach significantly enhances the usefulness of pathologic evaluation of malignant tumors. Further mechanistic insights about temporal events within the afflicted tissue that lead to a specific tumor grade gene index at clinical presentation could be invaluable. In this regard, induction of high-grade phenotypes in nonmalignant breast epithelial cells of high-risk individuals exposed to natural and synthetic estrogens suggests that specific grade-determining events occur early in tumorigenesis (22). This is consistent with the observation of S100P overexpression at early stages of breast cancer development (44).

It could be speculated that early in tumorigenesis, selective expansion of cells harboring PEGA phenotypes, such as altered calcium regulation, cell adhesion, oxidative stress management, apoptosis, cell differentiation, and cell migration, facilitates the manifestation of tumor dedifferentiation, recorded as microscopically determined histologic grade. Notably in our study, a three-gene fingerprint composed of *GPRC5A, PYCARD*, and *FXYD3* expression replaces ER status as a significant predictor of outcome and enhances the predictive power of tumor size, another long measured clinical parameter. Although it can be appreciated that underlying these highly predictive genes are key cellular processes related to calcium regulation, apoptosis, and differentiation, respectively, a biological basis for the contribution of tumor size to this algorithm, in particular, is presently unclear. In generating a reliable risk score for breast cancer patients, this three-gene fingerprint potentially serves as a molecular adjunct to two clinicopathologic variables: tumor size and histologic grade. High-grade primary breast tumors are generally treated with radiation, chemotherapy, and hormonal therapy (if ER positive). Response to these agents is highly variable. Consequently, a sizeable proportion of patients with HTG succumb to their disease. Prior knowledge of such cases could assist in more aggressive and targeted clinical management. By using cell lines spanning a wide spectrum of risk phenotypes identified by the PEGA profile, our data further show exquisite sensitivity of high-risk tumor cells to calcium-regulating drugs, whereas low-risk counterparts display resistance by rapidly reversing to a proliferative state on drug removal. These data provide robust functional evidence toward a therapeutic approach worthy of further consideration.

Toward fruitful drug discovery efforts, our experimental strategy advocates continued tumor cell characterization through the implementation of new model systems. Although routinely used cancer cell lines provide an unlimited supply of high-grade and metastatic cells at late stages of cancer progression, an important application of early stage and well-differentiated to moderately differentiated tumor cell models is to provide functional validation for presumed phenotypes of tumor aggressiveness. Using these model systems, a biologically derived gene signature of tumor cell immortalization, ImmSig, was previously identified, which portrays the importance of oxidative stress–reducing genes in maintaining continuously proliferating tumor populations (21), recently confirmed by others (45). In the search for new effective therapeutic targets against aggressive breast tumor cells, it is important to acknowledge that the successful application of autonomously regulated gene products, such as *ERBB2* and *EGFR*, represents the validity of our described approach. Moreover, because expression patterns of such genes are more accurately replicated in preclinical experimental models, the potential for clinical reagent development is likely to be realized within a shorter time frame. The PEGA signature sets forth a new paradigm and provides a clear rationale for using the full range of grade-associated molecular phenotypes and model systems for refining approaches in tumor targeting to ensure data reliability and clinical benefit.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

Received 4/28/09; revised 7/16/09; accepted 8/4/09; published OnlineFirst 9/29/09.

Grant support: Research in the S.H. Dairkee laboratory was supported by NIH CA109325. Work performed in the J.W. Gray laboratory was funded by NIH CA58207, and by the U.S. Department of Energy under Contract No. DEAC02-05CH11231.

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.

References

1. Elston CW, Ellis IO. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 1991;19:403–10.
2. Black MM, Barclay TH, Hankey BF. Prognosis in breast cancer utilizing histologic characteristics of the primary tumor. Cancer 1975;36:2048–55.
3. Rosen PP, Saigo PE, Braun DW, Jr., Weathers E, DePalo A. Predictors of recurrence in stage I (T1N0M0) breast carcinoma. Ann Surg 1981;193:15–25.
4. Hopton DS, Thorogood J, Clayden AD, MacKinnon D. Histological grading of breast cancer; significance of grade on recurrence and mortality. Eur J Surg Oncol 1989;15:25–31.
5. Ridolfi RL, Rosen PP, Port A, Kinne D, Mike V. Medullary carcinoma of the breast: a clinicopathologic study with 10 year follow-up. Cancer 1977;40:1365–85.
6. Pereira H, Pinder SE, Sibbering DM, et al. Pathological prognostic factors in breast cancer. IV. Should you be a typer or a grader? A comparative study of two histological prognostic features in operable breast carcinoma. Histopathology 1995;27:219–26.
7. Sotiriou C, Wirapati P, Loi S, et al. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst 2006;98:262–72.
8. Ivshina AV, George J, Senko O, et al. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. Cancer Res 2006;66:10292–301.
9. Yu K, Ganesan K, Miller LD, Tan P. A modular analysis of breast cancer reveals a novel low-grade molecular signature in estrogen receptor-positive tumors. Clin Cancer Res 2006;12:3288–96.
10. Lasfargues EY, Ozzello L. Cultivation of human breast carcinomas. J Natl Cancer Inst 1958;21:1131–47.
11. Soule HD, Vazguez J, Long A, Albert S, Brennan M. A human cell line from a pleural effusion derived from a breast carcinoma. J Natl Cancer Inst 1973;51:1409–16.
12. Cailleau R, Young R, Olivé M, Reeves WJ, Jr. Breast tumor cell lines from pleural effusions. J Natl Cancer Inst 1974;53:661–74.
13. Fogh J, Fogh JM, Orfeo T. One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice. J Natl Cancer Inst 1977;59:221–26.
14. Gazdar AF, Kurvari V, Virmani A, et al. Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. Int J Cancer 1998;78:766–74.
15. Ethier SP. Identifying and validating causal genetic alterations in human breast cancer. Breast Cancer Res Treat 2003;78:285–7.
16. Smith HS, Wolman, SR, Dairkee SH, et al. Immortalization in culture occurs at a late stage in progression of breast cancer. J Natl Cancer Inst 1987;78:611–5.
17. Dairkee SH, Deng G, Stampfer MR, Waldman FM, Smith HS. Selective cell culture of primary breast carcinoma. Cancer Res 1995;55:2516–19.
18. Dairkee SH, Paulo EC, Traquina P, Moore DH, Ljung B-M, Smith HS. Partial enzymatic degradation of stroma allows enrichment and expansion of primary breast tumor cells. Cancer Res 1997;57:1590–6.
19. Li Z, Bustos V, Miner J, et al. Propagation of genetically altered tumor cells derived from fine needle aspirates of primary breast carcinoma. Cancer Res 1998; 58:5271–4.
20. Dairkee SH, Ji Y, Ben Y, Moore DH, Meng Z, Jeffrey SS. A molecular 'fingerprint' of primary breast cancer cultures; patterns resembling tumor tissue. BMC Genomics 2004;5:47.
21. Dairkee SH, Nicolau M, Sayeed A, et al. Oxidative stress pathways are highlighted in an immortalization signature in breast cancer. Oncogene 2007;26:6269–79.
22. Dairkee SH, Seok J, Champion S, et al. Bisphenol A induces a profile of tumor aggressiveness in high-risk cells of breast cancer patients. Cancer Res 2008;68:2076–80.
23. Chin K, DeVries S, Fridlyand J, et al. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 2006;10:529–41.
24. Dupuy A, Simon RM. Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 2007;99:147–57.
25. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116–21.
26. Loi S, Haibe-Kains B, Desmedt C, et al. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. J Clin Oncol 2007;25:1239–46.
27. Altman DG, Royston P. What do we mean by validating a prognostic model? Stat Med 2000;19:453–73.
28. Chow E, Abdolell M, Panzarella T, et al. Predictive model for survival in patients with advanced cancer. J Clin Oncol 2008;26:5863–9.
29. Maraqa L, Cummings M, Peter MB, et al. Carcinoembryonic antigen cell adhesion molecule 6 predicts breast cancer recurrence following adjuvant tamoxifen. Clin Cancer Res 2008;14:405–11.
30. Yu CJ, Yang PC, Shun CT, Lee YC, Kuo SH, Luh KT. Overexpression of MUC5 genes is associated with early post-operative metastasis in non-small-cell lung cancer. Int J Cancer 1996;69:457–65s.
31. Wang G, Platt-Higgins A, Carroll J, et al. Induction of metastasis by S100P in a rat mammary model and its association with poor survival of breast cancer patients. Cancer Res 2006;66:1199–207.
32. Wirapati P, Sotiriou C, Kunkel S, et al. Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res 2008;10:R65.
33. 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.
34. 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.
35. Miller LD, Smeds J, George J, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A 2005;102:13550–5.
36. Whitfield ML, George LK, Grant GD, Perou CM. Common markers of proliferation. Nat Rev Cancer 2006;6:99–106.
37. 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.
38. Naderi A, Teschendorff AE, Barbosa-Morais NL, et al. A gene-expression signature to predict survival in breast cancer across independent data sets. Oncogene 2007;26: 1507–16.
39. Teschendorff AE, Naderi A, Barbosa-Morais NL, et al. A consensus prognostic gene expression classifier for ER positive breast cancer. Genome Biol 2006;7:R101.
40. 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.
41. Bikle DD, Oda Y, Xie Z. Calcium and 1,25(OH)2D: interacting drivers of epidermal differentiation. J Steroid Biochem Mol Biol 2004;89–90:355–60.
42. Becker T, Gerke V, Kube E, Weber K. S100P, a novel Ca(2+)-binding protein from human placenta: cDNA cloning, recombinant protein expression and Ca(2+) binding properties. Eur J Biochem 1992;207:541–7.
43. Petersen OW, Rønnov-Jessen L, Howlett AR, Bissell MJ. Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells. Proc Natl Acad Sci U S A 1992;89:9064–8.
44. Guerreiro Da Silva ID, Hu YF, Russo IH, et al. S100P calcium-binding protein overexpression is associated with immortalization of human breast epithelial cells *in vitro* and early stages of breast cancer development *in vivo*. Int J Oncol 2000;16:231–40.
45. Rai P, Onder TT, Young JJ, et al. Continuous elimination of oxidized nucleotides is necessary to prevent rapid onset of cellular senescence. Proc Natl Acad Sci U S A 2009;106:169–74.