Identification of Gene Expression Profiles That Predict the Aggressive Behavior of Breast Cancer Cells

Deborah A. Zajchowski, Marty F. Bartholdi, Yan Gong, Lynn Webster, Hsiao-Lai Liu, Alexander Munishkin, Catherine Beauheim, Susan Harvey, Stephen P. Ethier, and Paul H. Johnson

Departments of Cancer Research [D. A. Z., Y. G., L. W., H.-L. L., P. H. J.] and Genomics and Gene Therapy [M. F. B., C. B., S. H.], Berlex Biosciences, Richmond, California 94804; Clontech Laboratories, Palo Alto, California 94303 [A. M.]; and Department of Radiation Oncology, University of Michigan Medical Center, Ann Arbor, Michigan 48109 [S. P. E.]

ABSTRACT

With the goal of identifying genes that have an expression pattern that can facilitate the diagnosis of primary breast cancers (BCs) as well as the discovery of novel drug leads for BC treatment, we used CDNA hybridization arrays to analyze the gene expression profiles (GEPs) of nine weakly invasive and four highly invasive BC cell lines. Differences in gene expression between weakly and highly invasive BC cells were identified that enabled the definition of consensus GEPS for each invasive phenotype. To determine whether the consensus GEPS, comprising 24 genes, could be used to predict the aggressiveness of previously uncharacterized cells, gene expression levels and comparative invasive and migratory characteristics of nine additional human mammary epithelial cell strains/lines were determined. The results demonstrated that the GEP of a cell line is predictive of its invasive and migratory behavior, as manifest by the morphology of its colonies when cultured on a matrix of basement membrane constituents (i.e., Matrigel). We found that the expression of keratin 19 was consistently elevated in the less aggressive BC cell lines and that vimentin and fos-related antigen-1 (FRA-1) were consistently overexpressed in the more highly aggressive BC cells. Moreover, even without these three genes, the GEP of a cell line still accurately predicted the aggressiveness of the BC cell, indicating that the expression pattern of multiple genes may be used as BC prognosticators because single markers often fail to be predictive in clinical specimens.

INTRODUCTION

Advances in the technologies available to evaluate simultaneously the expression of thousands of genes (1–3) have fueled efforts to characterize the molecular basis of many diseases, including cancer. Our studies were aimed at the identification of a GEP4 that would be predictive of the metastatic ability of BC cells. We expected that such a profile could produce new diagnostic/prognostic markers for the clinical staging of BC as well as provide a screening strategy for discovering drugs that would be effective against tumor progression and metastasis.

A hallmark of metastatic tumor cells is the ability to invade and traverse basement membranes of the endothelium and secondary organ sites (4, 5). Because strong migratory and invasive abilities are also characteristic of cells of mesenchymal origin, it has been suggested that epithelial cell-derived tumors acquire a mesenchymal character during tumor progression by a process resembling the EMT that occurs during embryonic development (6, 7). Indeed, in cell culture systems modeling tumor progression in BC, the tumor cells with the greatest motile and invasive activity (e.g., chemoinvasion through the reconstituted basement membrane, Matrigel, in the modified Boyden chamber) are positive for the mesenchymal marker, vimentin (8). In addition, these highly aggressive cells have no ERα or E-cadherin expression, consistent with the hormone-independent growth and reduced intercellular junctional communication that is often associated with aggressive breast tumors. To further elucidate the nature and extent of the molecular changes associated with acquisition of a more aggressive cellular phenotype, we designed experiments to identify GEPS that would distinguish highly invasive breast tumor cells from their less aggressive counterparts. We hypothesized that although advanced stage BCs have acquired numerous mutations during their progression, the acquisition of strong invasive and motile ability would be correlated with the coordinate expression of a consensus set of genes in many tumors. The products of these genes would function as mediators of invasive or migratory activities themselves or may serve solely as markers of increased cellular aggressiveness.

To identify the GEPS, we measured gene expression levels in a series of normal and tumor-derived mammary epithelial cell lines that have been extensively characterized for their in vitro growth characteristics on Matrigel and invasive ability in modified Boyden chamber assays (8, 9) and their in vivo tumorigenic and metastatic capabilities (8, 10). On the basis of their relative activities in the chemoinvasion assay, MDA231, Hs578T, MDA435, and BT549 have been classified as highly invasive and other BC cell lines such as MCF7, ZR75-1, MDA453, and SKBR3 as weakly invasive (Table 1). Importantly, the highly invasive BC cell lines are more responsive to chemotactic stimuli than are the weakly invasive BC cell lines (8, 9) and also differ in their growth characteristics on Matrigel. The Matrigel outgrowth assay measures cellular proliferative, migratory, adhesive, and invasive ability in a gel comprised of basement membrane components (i.e., laminin, type IV collagen, and heparin sulfate proteoglycans; Ref. 11). The highly invasive BC cell lines form branching or stellate colonies when cultured on Matrigel, whereas the weakly invasive BC cell lines grow as colonies of clustered cells. In vivo, only highly invasive BC cell lines form metastases in a 60-day period after implantation into the mammary fat pad of nude, athymic mice (8, 10).

The GEPS of the four highly invasive BC cell lines and the nine weakly invasive BC cell lines were compared with the GEP of a “normal” reference mammary epithelial cell line, and consensus GEPS for each invasive phenotype were identified by the differences in gene expression that were characteristic of weakly and highly invasive BC cells. To determine whether these GEPS were predictive of BC aggressive behavior, the GEPS and comparative invasive and migratory characteristics of nine previously uncharacterized cell strains/lines were determined. The results demonstrated that the GEP of a cell line is diagnostic of the aggressive behavior of that cell line, as measured by combined invasive and motile activities and colony morphology on Matrigel.
GENE EXPRESSION PROFILES PREDICT TUMOR CELL AGGRESSIVENESS

Table 1 Characteristics of the breast cancer cell lines used to generate the consensus GEP

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Specimen origin</th>
<th>Tumorigenicity</th>
<th>Matrigel morph</th>
<th>Marker gene expression[^2]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weakly invasive</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T47D</td>
<td>Infiltrating ductal Ca; PE</td>
<td>+</td>
<td>+</td>
<td>ER-α</td>
</tr>
<tr>
<td>ZR75-1</td>
<td>Infiltrating ductal Ca; ascites</td>
<td>+</td>
<td>+</td>
<td>E-cad</td>
</tr>
<tr>
<td>MCF7</td>
<td>Breast adenocarcinoma; PE</td>
<td>+</td>
<td>+</td>
<td>Vim</td>
</tr>
<tr>
<td>MDA361</td>
<td>Breast adenocarcinoma; brain met</td>
<td>+</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>BT474</td>
<td>Invasive ductal Ca; PT</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>BT20</td>
<td>Breast adenocarcinoma; PT</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>MDA468</td>
<td>Metastatic breast adenocarcinoma; PE</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>SKBR3</td>
<td>Breast adenocarcinoma; PE</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
<tr>
<td>MDA453</td>
<td>Metastatic breast Ca; PE</td>
<td>+</td>
<td>+</td>
<td></td>
</tr>
</tbody>
</table>

[^1]: Specimen origin and pathological assessment information were obtained from the ATCC catalog. PT: primary tumor; PE: pleural effusion; Ca: carcinoma.

[^2]: Tumorigenicity data was reported in the ATCC catalog or in Ref. 44. +, palpable tumors produced as xenografts in athymic nude or SCID mice; −, nontumorigenic; met, metastatic cell lines as reported by Refs. 8 and 45.

[^3]: Description of the morphology of cells cultured in Matrigel and their activity in the Boyden chamber invasion assay were taken from Ref. 9.

[^4]: Expression of the mRNA and protein of marker genes ER-α, E-cadherin (E-cad), and vimentin (Vim) was taken from Ref. 9 and our unpublished data; ND, not determined; +, detectable; −, undetectable.

[^5]: Estrogen administration required for tumor formation.

MATERIALS AND METHODS

**Cells and Culture Conditions.** The 48RS HMEC strain and the 184B5 benzopyrene-immortalized HMEC line (12) were cultured in DFC1-1 medium (13). MCF10A (ATCC, Rockville, MD) cells were cultured as recommended by the ATCC. The HBL-100, T47D, ZR75-1, MCF7, BT483, MDA361, BT474, BT20, MDA68, SKBR3, MDA453, BT549, Hs578T, MDA231, and MDA4355 cell lines were obtained from the ATCC and routinely cultured in α-modified MEM supplemented with 1 mM HEPES, 2 mM glutamine, 0.1 mM glutathiamine, 0.1 mM sodium pyruvate, 50 μg/ml gentamicin, 1.0 μg/ml insulin (all of these from Life Technologies, Inc.), and 10% fetal bovine serum (Intergen, Purchase, NY; called α-MEM medium). These studies used derivatives of the MCF7 and MDA231 cells that were established from athymic nude mouse xenografts (14) instead of the parental MCF7 and MDA231 cells obtained from the ATCC. No significant differences were noted between the in vivo-selected cell lines and the parental cell lines for the in vitro behaviors assessed in this study. SUM44PE and SUM52PE cells were both derived from pleural effusion metastases of patients with advanced stage BC.

Both cell lines were routinely cultured in defined serum-free Ham’s F-12 medium supplemented with 1 mg/ml BSA, 1 μg/ml transferrin, 10 mM ethanalamine, 5 μg/ml insulin, and 1 μg/ml hydrocortisone, as described in detail previously (15, 16). SUM159PT cells were derived from a poorly differentiated primary BC. SUM 1315 Mo2 cells were derived from a second transplant generation mouse xenograft derived directly by transplantation of a primary tumor piece into immune-deficient mice. Both cell lines were cultured in serum-free Ham’s F-12 medium supplemented with either 5 μg/ml insulin and 1 μg/ml hydrocortisone (SUM159PT) or insulin and 10 ng/ml epidermal growth factor (SUM1315 Mo2). Refs. 17–19. SUM102PT cells were derived from a minimally invasive apocrine adenocarcinoma with extensive DCIS and was established from a BC lymph node metastasis by transformation with SV40 T antigen (21) and was cultured in the BRFF-BM2 medium as recommended.

To determine the steady state gene expression profiles of the normal and tumor-derived HMEC lines, cells were harvested at 70–80% confluence 4–5 days after plating and fed fresh medium 2 days before harvest.

**cDNA Array Hybridization Studies.** The Atlas Human (7740-1) and Cancer (7742-1) arrays (Clontech Laboratories, Palo Alto, CA) containing 588 genes were used in these studies. Total RNA was isolated by the guanidinium-isothiocyanate-cesium chloride gradient procedure (22). The preparation and hybridization of radioactively labeled cDNA from total RNA (5 μg) were performed essentially as described in the Clontech Atlas Human cDNA array hybridization kit protocol. The only exceptions were the removal of unincorporated substrate and the exposure time were adjusted for the cDNA array hybridization kit protocol.

**Tumorigenicity data was reported in the ATCC catalog or in Ref. 44. +, palpable tumors produced as xenografts in athymic nude or SCID mice; −, nontumorigenic; met, metastatic cell lines as reported by Refs. 8 and 45.**
performed by measuring the signal intensity for specifically hybridized bands using a phosphorimager (Fujix; Fuji Medical Systems, Stanford, CT).

For the RT-PCR reactions, 1 μg of total RNA was reverse-transcribed by 200 units of Moloney murine leukemia virus RT (Amersham Pharmacia Biotech, Piscataway, NJ) for 1 h at 42°C in 50 mM Tris-HCl (pH 8.0), 50 mM KCl, 6 mM MgCl₂, and 5 mM DTT using 50 pmol of either oligo(dT)18 or random primers (Operon Technologies, Inc. Alameda, CA). Advantage PCR (Clontech) was used with a pair of gene-specific primers to amplify the cDNA for each gene (20 ng of cDNA/reaction) according to the protocol supplied with the kit. Each PCR cycle was 30 s at 94°C followed by 1 min at 68°C. PCR products were separated on 1.5% agarose gels at every cycle starting from cycle 18. The relative abundance of each mRNA in each cell line was obtained by determining the cycle number at which the PCR product was detectable after staining the gel with ethidium bromide.

Chemoinvasion and Migration Assays. The chemoinvasion assay was carried out using a modification of the method of Albini et al. (24). After trypsinization, cells (2 × 10⁵) were plated on Matrigel-coated 8-μm polyethylene terephthalate filter inserts in Boyden chambers (Biocoat Matrigel Invasion Chamber; Becton Dickinson, Bedford, MA). The bottom chamber contained 0.75 ml of NIH3T3-conditioned media, produced as described (24). BC cell lines obtained from the ATCC were trypsinized, centrifuged, and resuspended at 4 × 10⁵ cells/ml in RPMI medium containing 10% FBS.
remaining cell lines were resuspended in their regular growth medium. After 18–20 h, the cells remaining in the insert were removed with a cotton swab, and the cells on the bottom of the filter were fixed in Diff-quick (American Scientific Products, McGraw Park, IL) and treated with RNase A (at 50 μg/ml for 20 min at 37°C) before staining with propidium iodide (10 μg/ml in PBS) for 1 min at RT. The dried filters were removed and mounted on slides with Cytoseal 60 mounting media (Stephens Scientific, Kalamazoo, MI). Individual propidium iodide-stained nuclei on the filters were counted using a Laser Scanning Cytometer (Compucyte, Cambridge, MA). Fluorescence emission was measured using the 20× objective through a 625/28-band pass filter after exciting at 488 nm with an argon ion laser operating at 5 mW. Individual fluorescent cells were segmented by an intensity threshold, and a minimum exciting at 488 nm with an argon ion laser operating at 5 mW. Individual fluorescent cells were segmented by an intensity threshold, and a minimum area (40 μm²) was excluded to reduce background counts. Triplicate samples were counted in each experiment. Outlier values were eliminated from calculations of average invasive activity.

Migration activity was determined following the procedure described for the invasion assay except that the cells were plated on top of uncoated 8-μm pore polyethylene terephthalate filters in the Boyden chambers.

**Matrigel Outgrowth.** Determination of the morphology of cells grown on Matrigel was carried out as described previously (9). Briefly, cells (1.0 × 10³–1.6 × 10⁷/well of a 24-well plate) resuspended in 0.5 ml of culture medium were plated on top of a preset Matrigel coating consisting of 0.1 ml of Matrigel (Becton Dickinson) diluted to 10 mg/ml in α-MEM basal medium salts. Colony outgrowth was monitored over the course of the experiment and photographed at 7–10 days using a Zeiss Axiosvert 35 microscope equipped with a Polaroid DMC le digital camera.

**RESULTS**

To identify the genes that had expression levels that would distinguish weakly aggressive from highly aggressive BC, we used cDNA array hybridization analyses to determine the GEPs of the weakly invasive and highly invasive BC cell lines listed in Table 1. As the reference for these studies, gene expression was analyzed in MCF10A, a spontaneously immortalized “normal” HMEC line derived from a patient with fibrocystic disease (25). RNA from each of the cell lines was isolated and used to prepare a radiolabeled complex cDNA probe for hybridization to the Atlas Human and Cancer arrays, which contain spotted cDNA fragments corresponding to 940 genes (i.e., the number of nonredundant genes on both 588 gene arrays). Hybridization signals from 899 of the 940 genes were detected in the panel of cell lines, indicating that most of the genes present on these two arrays are expressed in “normal” or tumor-derived human mammary epithelial cells. In general, approximately 10% of the genes with detectable levels of mRNA were found to be differentially expressed in the BC cell lines relative to the reference MCF10A.

**Identification of Consensus Gene Expression Profiles for Weakly Invasive versus Highly Invasive Cells.** These analyses identified 24 genes (Fig. 1; genes #1–24), the expression of which was differentially altered in the weakly invasive BC cell lines when compared with the highly invasive BC cell lines. The expression of K19 (#1), GATA-3 (#2), and insulin-like growth factor binding protein 1 (IGFBP-5) (#6) was elevated in most or all of the weakly invasive BC cell lines and not in the highly invasive cell lines. In contrast, the levels of plasminogen activator inhibitor-1 (PAI-1) (#16), c-jun (#22), and collagen VI α1 (#23) were increased in most of the highly invasive BC cell lines but not in the weakly invasive cell lines. The expression levels of some genes were altered in both weakly and highly invasive BC cells relative to the reference MCF10A, but either the magnitude or the direction of the change distinguished weakly from highly invasive BC cell lines; e.g., FRA-1 (#17) expression was reduced in the weakly invasive BC cells and increased in the highly invasive BC cells relative to the reference. Different preparations of cultured MCF10A cells showed only minor variation in expression levels for these genes (MCF10A column; Fig. 1). Thus, consensus patterns of gene expression were found that were characteristic of either the weakly invasive or highly invasive BC cell lines.

These analyses also identified 11 genes, the mRNA levels of which were changed in the majority of the tumor cell lines relative to the “normal” reference MCF10A, regardless of their invasive ability (Fig. 1; genes #25–35). Such genes might be useful as markers to distinguish normal and tumor cells but are not further discussed in this manuscript because they were not informative as discriminators between weakly and highly invasive BC cell lines.

The differential expression of 20 of these genes was evaluated and confirmed by Northern and RT-PCR analyses in a subset of the weakly invasive and highly invasive BC cell lines (Fig. 1 and Fig. 2). The expression differences measured for five other genes were supported by literature reports (listed in Fig. 1). In general, the data obtained by alternative methods correlated well with the results from the array hybridization studies (Fig. 1, confirmation column; and Fig. 2, A and B). The potential influence of biological variation on the reproducibility of the array hybridization data were evaluated for some genes by measuring gene expression by Northern analyses in RNA prepared from independent cell cultures (Fig. 2B). These data, combined with the reproducibility of the array results from different samples (data not shown), led us to conclude that these differentially expressed genes showed minimal variability in RNA prepared from different cultures.

The median gene expression ratios for the weakly and highly invasive BC cell lines relative to the reference MCF10A were calculated (Fig. 1, median columns) and used to generate consensus GEPs for each phenotype (Fig. 3A). Within these consensus profiles is a gene expression pattern that is positively associated with either weakly invasive BC (genes #1–8, “weakly aggressive BC discriminators”) or with highly invasive BC (genes #9–24, “highly aggressive BC discriminators”) cell lines.

**Gene Expression Profiles Predict the Aggressive Behavior of Previously Uncharacterized Normal and Tumor-derived Mammary Epithelial Cells.** To evaluate the predictive ability of the consensus GEPs for tumors of weakly or highly aggressive phenotypes, we performed cDNA array hybridizations with RNA from nine HMEC strains/lines, the invasive characteristics of which, relative to this panel of weakly invasive and highly invasive BC cell lines, were unknown. Included in this analysis were cells derived from reduction mammoplasty (normal, limited life span 48RS and benzopyrene-immortalized 184B5; Ref. 12) and human milk (HBL-100; Ref. 26). The other cell lines were early passage cultures derived from primary (i.e., SUM102PT; Ref. 20; and SUM159PT) and metastatic BC biopsies (i.e., SUM44PE, SUM52PE, SUM1315mo2, and BRF71T1). Expression levels for the 24 consensus genes in these cell lines relative to the reference MCF10A are shown in Fig. 3B. The GEPs of SUM44PE and SUM52PE resembled the weakly invasive BC consensus most closely, each having correlation coefficients with the weakly invasive BC consensus of 0.95. The GEPs of BRF71T1, HBL-100, SUM159PT, and SUM1315mo2 cells were most similar to the highly invasive BC consensus, with correlation coefficients ranging from 0.82 to 0.88 (Fig. 3B). Thus, each of these cell lines could be classified as weakly or highly aggressive based upon its correlation with the weakly or highly invasive BC GEP.

In contrast, the cells derived from reduction mammoplasty (i.e., 48RS and 184B5) had GEPs that were more similar to the MCF10A than to either the weakly or highly invasive BC consensus GEP (compare Fig. 3, A and B), as seen by the low number of genes with expression ratios that differed from MCF10A by more than 2-fold. One notable exception was the elevated level of osteonectin mRNA (#19) found in the 48RS, which was also observed in the very similar GEPs of two other cell strains derived from reduction mammoplasties
The correlation between 48RS and the highly invasive BC GEP decreased from 0.59 to 0.36 when osteonectin was excluded from the calculation. Interestingly, the DCIS-derived cell line SUM102PT had a GEP that was also more similar to the reference MCF10A than to either consensus, although we detected significant differences in the expression of other genes analyzed by the cDNA arrays (data not shown).

The aggressiveness of each of these cell lines was evaluated by measuring their activities in the modified Boyden chamber chemoinvasion and migration assays (24) and their growth characteristics when cultured on Matrigel. Some of the cell lines reported to be representative of weakly aggressive (i.e., MCF7, ZR75-1, SKBR3, and MDA468) and highly aggressive (i.e., MDA435S, Hs578T, and MDA231) phenotypes were evaluated for comparative purposes. In agreement with other reports (8), Hs578T, MDA435S, and MDA231 formed colonies with the stellate morphology characteristic of highly invasive cells when cultured in Matrigel and grew as branched (MDA231) or stellate (MDA435S and Hs578T) structures on this...
matrix (Fig. 4A and data not shown). In contrast, MCF7 and ZR75-1 formed fused colonies, and SKBR3 and MDA468 grew as grape-like clusters (referred to as “spherical”) on Matrigel (Fig. 4A and data not shown). It has been suggested (27) that these distinct types of aggregate cell colonies reflect the degree of cell-cell adhesive and junctional communication abilities and are dependent on E-cadherin function. The stellate or branched colony formation of the highly invasive BC cells most likely reflects an ability to invade as well as migrate on the Matrigel matrix.

The nine HMEC strains/lines fell into three distinct groups when analyzed for their growth characteristics on Matrigel. The 48RS, MCF10A, and 184B5 formed concave sphere-like “pseudo-acinar” structures, whereas SUM102PT, SUM44PE, and SUM52PE created fused and/or spherical colonies on this matrix, as are typical for the weakly invasive BC cell lines (Fig. 4B). The HBL-100, SUM1315 Mo2, SUM159PT, and BRF71T1 grew as stellate and branched colonies on Matrigel, as did the highly invasive BC cell lines (Fig. 4B).

Thus, with the exception of SUM102PT, the GEP of which predicted neither invasive phenotype, the weakly invasive or highly invasive BC prediction derived from the GEP of each of these cell lines correlated with the morphological appearance of their colonies in the Matrigel outgrowth assay.

The invasive activity of these cell lines was measured in the modified Boyden chamber chemoinvasion assay. In agreement with others (8, 9), we found that the MDA231 and MDA435S cell lines were much more active than the weakly invasive BC cell lines, MCF7, ZR75-1, and SKBR3 (Fig. 5A). The cell lines with GEPs that best correlated with the consensus highly invasive BC GEP were also more active than most of the reported weakly invasive cell lines, but significantly less active than the MDA231 cells. The SUM159PT cells were the most invasive from this group of cell lines (i.e., 50% of MDA231), with BRF71T1, SUM1315 Mo2, and HBL-100 being only 20–25% as active as the MDA231. However, these moderate invasive activities were difficult to distinguish from another reported weakly invasive BC cell line, MDA468 (8, 9). These data suggest that the highly invasive BC consensus GEP correlates more strongly with colony morphology in the Matrigel outgrowth assay than it does with invasive activity as measured by traversal of the Matrigel-coated filters.

Cell migration was evaluated in Boyden chambers in the absence of the Matrigel barrier. As expected, the MDA231, MDA435S, and Hs578T were considerably more motile than the weakly invasive MCF7, ZR75-1, and SKBR3. Unlike the results from the invasion assay, all four of the cell lines that were predicted to have highly aggressive behavior based upon their GEPs (i.e., HBL-100, SUM159PT, SUM1315 Mo2, and BRF71T1) were clearly more motile than the weakly invasive BC cell lines (Fig. 5B). These data suggest that the consensus highly invasive BC GEP is predictive of strong migratory activity, although the significant migratory behavior of the MCF10A is a noteworthy exception. When both migratory and chemoinvasive responses were considered, the most active cell lines were those with GEPs similar to those of highly invasive BC cells (Fig. 5C). Thus, the GEPs that we have identified are predictive for the structures assumed during colony formation on Matrigel and for the combined migratory and invasive abilities assessed in the modified Boyden chambers.

DISCUSSION

With the ultimate aim of defining a set of genes that had mRNA expression levels in BC cells that would be predictive of the aggressiveness of primary breast tumors, cDNA array hybridization analyses of 940 genes were performed using a panel of 13 BC cell lines with known invasive properties. We found that the expression pattern of 24 genes provided GEPs that were characteristic of either the weakly or the highly invasive BC cell lines. The predictive ability of these consensus GEPs was then tested by evaluating the gene expression and invasive properties of nine previously uncharacterized cell lines/strains derived from reduction mammoplasties, human milk, DCIS, and both primary and metastatic lesions from patients with invasive ductal breast carcinomas. These studies demonstrated that the in vitro aggressive behavior of cells cultured from human BC specimens could be successfully predicted by correlating the GEP of the “unknown” cell line with the consensus GEP derived for the weakly invasive or highly invasive BC phenotypes.

Other investigators have used cDNA microarrays bearing thousands of gene sequences (28, 29) or differential cloning methods (30) to generate gene expression profiles for BC by using clinical biopsy samples as the source of RNA. In our study, we used cDNA hybridization filters containing 940 genes representing broad cellular functions, including oncogenes, tumor suppressor genes, and genes involved in cell cycle control, cell-cell interactions, apoptosis, and signal transduction pathways. The use of a well-characterized set of genes enabled us to ask whether specific signal transduction pathways were activated or repressed in highly aggressive BC cells as well as identify a GEP that could be useful in predicting the severity of the malignant phenotype in clinical specimens. We also chose to use cultured cells for our analyses because they are a more homogeneous cellular population than the actual tumor and would also be the tool for any future drug discovery screen.

From cDNA hybridization array data obtained from nine weakly invasive and four highly invasive BC cell lines, we identified 24 genes that were differentially expressed in either weakly invasive or highly invasive BC cell lines and formed the basis of consensus GEPs for weakly versus highly aggressive BC. To test the predictability of these consensus GEPs, we determined the GEPs of nine additional HMEC strains/lines. Each cell line was classified as weakly or highly aggressive by the correlation of their individual GEPs with the weakly or highly invasive BC consensus GEP. By this evaluation, the SUM44PE and SUM52PE cell lines were predicted to be poorly aggressive BC cells. Their growth on Matrigel as fused or fused/spherical colonies as well as their poor activities in the invasion and migration assays confirmed this prediction. The GEPs for HBL-100, SUM159PT, BRF71T1, and SUM1315 Mo2 strongly predicted a highly aggressive BC phenotype that was also confirmed by the Matrigel outgrowth assay, in which all of these cell lines formed branched or stellate colonies. The SUM159PT cell line was shown recently (18) to be highly invasive in vitro and to form aggressively growing tumors with metastatic capability in nude, athymic mice. The categorization of HBL-100 as a highly invasive BC cell line was unexpected because it was derived from cells in human milk (26). However, a number of reports (31) have shown the expression of SV40 T antigen in these immortal cells and their transformed behavior as measured by anchorage-independent growth in soft agar. Others (32) have provided evidence for the tumorigenic progression of HBL-100 sublines selected for by extensive passage in culture.

The “normal” cells derived from reduction mammoplasties had GEPs that were different from either the weakly or highly invasive BC consensus GEP. In agreement with the GEP prediction, the colonies formed by these cells did not resemble either the fused/spherical colonies of the weakly invasive BC cell lines or the stellate/branched forms of the highly invasive BC cell lines. Indeed, others (33) have shown that the structures formed in Matrigel by normal HMEC are hollow “pseudo-acinar-like,” growth-arrested colonies with polarized epithelium and an internal lumen. More extensive evaluation of additional cell strains derived from reduction mammoplasties as well as...
Fig. 3. Categorization of normal and tumor-derived HMECs by their GEPs. A, consensus GEPs for weakly invasive and highly invasive BC cell lines are graphically represented. The expression ratio values correspond to the median fold-change relative to the MCF10A reference observed for each gene from data in Fig. 1. “Discriminators of weakly invasive BC” (genes #1–8) have expression levels that are greater in weakly invasive BC cell lines than in highly invasive BC cell lines. “Discriminators of highly invasive BC” (genes #9–24) have expression levels that are greater in highly invasive BC cell lines than in weakly invasive BC cell lines.
their immortalized derivatives will be necessary to confirm an association between GEP and the formation of “pseudo-acinar-like” colonies. However, our studies suggest that the “normal” cell colony morphology in the Matrigel outgrowth assay, as well as the morphologies of the weakly and highly invasive BC cell colonies, are predicted by the GEP of the cell.

The cell lines with highly invasive BC GEPs were more active in the chemoinvasion assay than the weakly invasive MCF7 and ZR75-1 but not as active as the highly invasive MDA435S and MDA231 cell lines. However, the moderate invasive activities of these cell lines were difficult to distinguish from the reported weakly invasive BC cell line MDA468, suggesting that the highly invasive BC consensus GEP does not always correlate with the degree of invasive activity. Migration ability was more highly correlated with the GEP predictions because all of the cell lines with a highly invasive BC GEP were more motile than the other BC cell lines tested. When both invasive and migratory capabilities were considered, the highly invasive BC GEP identified the cell lines with high combined invasive and migratory activities. These observations and the strong correlation between the GEP prediction and Matrigel outgrowth suggest that the consensus GEPs that we have identified are predictive of tumor cell aggressiveness. They also suggest that the combination of an enhanced capability for migration with moderate or very high invasiveness through Matrigel-coated filters enables branched/stellate colony formation on Matrigel and enhanced metastatic potential.

The consensus weakly invasive and highly invasive BC GEPs in this study were generated using nine weakly invasive BC cell lines, four highly invasive BC cell lines, and a “normal” reference. After analysis of the additional six BC cell lines, two with weakly invasive BC GEPs and four with highly invasive BC GEPs, the 24 genes were re-evaluated to determine their value as indicators of weakly or highly aggressive BC phenotypes. To measure the predictive value of each gene for each phenotype, we counted the number of cell lines from each phenotype that exhibited an expression change similar to the median fold-change for all of the cell lines with the same phenotype (Table 2, #Obs column). All of the genes retained their association with either weakly or highly aggressive BC phenotypes, although some of the genes were more strongly correlated with either phenotype than other genes. It is noteworthy that from all of these common gene expression changes, only K19 (#1), FRA-1 (#17), and vimentin (#18) were consistently associated with all of the cell lines of either phenotype (Table 2). These data suggest that K19, FRA-1, or vimentin might be valuable markers for BC diagnosis. A number of studies (34, 35) have shown an association between luminal keratin expression (i.e., K8, 18, and 19) and favorable prognostic indicators, such as ER-positivity and lower proliferative index, but the prognostic value of K19 expression alone has not been reported. In a study of the expression pattern of AP-1 family members in BC extracts, FRA-1 protein levels were found to be negatively associated with ER status and differentiation but not with clinical stage or histological grading (36). Although many studies (for review, see Ref. 7) have documented the expression of vimentin in breast carcinomas, the relationship between vimentin expression and poor prognosis in BC has not been clearly established.

Our data show that several other genes were also overexpressed in all of the eight cell lines with highly invasive BC GEPs relative to the cell lines with weakly invasive BC GEPs [i.e., caveolin-1 (#10), GST P (#12), MLH1 (#13), and c-jun (#22)], whereas others were found elevated in seven highly invasive BC cell lines (Table 2). In each of these cases, at least one of the weakly invasive BC cell lines expressed comparable levels of mRNA with the highly invasive BC cells such that none of these genes would qualify as a single marker for predicting the aggressive BC phenotype. A similar conclusion can be reached for the weakly aggressive BC discriminators that were found elevated in at least 10 of the 11 weakly invasive BC cell lines [i.e., GATA-3 (#2), K18 (#3), RARα1 (#4), IGFBP-5 (#6), and PIG7 (#8)]. Yet, together with the other indicators in the context of the entire weakly invasive versus highly invasive BC consensus GEPs, these genes are powerful predictors. To determine the extent of the contribution of K19, vimentin, and FRA-1 to the predictive power of the weakly invasive and highly invasive BC consensus GEPs, the data for these genes were eliminated, and the correlation coefficients were recalculated. The results showed the same statistically significant assignment of cell lines to the weakly aggressive or highly aggressive BC groups.
GENE EXPRESSION PROFILES PREDICT TUMOR CELL AGGRESSIVENESS

(data not shown). Thus, the predictive ability of the weakly invasive and highly invasive BC GEPs is not solely dependent upon these three markers.

These results emphasize the potential utility of gene expression patterns in diagnosis. One advantage in using multiple markers is to identify phenotypically similar breast tumors that express a significant fraction, but not all, of the markers. A number of investigators (37–39) have used gene expression profiling to characterize clinical cancer specimens with the goal of identifying gene expression patterns for diagnostic use. Martin et al. (37) measured gene expression in 18 breast tumor specimens to identify four gene clusters containing 35 genes that grouped the tumors according to ER status, clinical stage, or tumor size. However, the set of 35 genes was not tested on additional tumor samples to determine the predictive ability of that gene set. Our studies have demonstrated the predictive ability of the 24 weakly aggressive versus highly aggressive BC discriminators in in vitro analyses. Future efforts will determine whether the expression pattern of the consensus set of genes in clinical specimens is correlated with clinical stage and disease outcome.

It is possible that the distinct GEPs associated with the weakly and highly aggressive BC reflect different cell origins for these phenotypically distinct tumors. In such a case, the weakly aggressive cells that express K19 would be derivatives of the luminal cell lineage, whereas the highly aggressive BC cells expressing vimentin would be derived from a basal cell lineage. But it is also possible that these GEPs are reflective of coordinate regulation of gene expression that results from an EMT. Support for the c-jun homodimeric AP-1 transcription factor as a key component of the molecular switch in the EMT comes from the studies of Smith et al. (40). They demonstrated the involvement of c-jun in the development of the mesenchymal phenotype by showing that overexpression of c-jun in the weakly invasive BC cell line MCF7 resulted in increased invasiveness, migration, and hormone-independent tumor formation. Concomitant with those phenotypic changes were increased levels of FRA-1 and vimentin, which are consistent with the reported involvement of FRA-1 and vimentin in the transcriptional control of these genes.

The role of vimentin as a mediator of enhanced migration has been suggested by observations that treatment of MDA231 cells with antisense oligonucleotides against vimentin caused reduced in vitro migratory activity (41). However, the evidence that vimentin plays a critical role in mediating the EMT is not definitive, as two reports (41, 42) describing effects of exogenous vimentin expression in MCF7

Table 2. Frequency of consensus gene expression changes in all of the cell lines

<table>
<thead>
<tr>
<th>ID#</th>
<th>Gene</th>
<th>Weakly Invasive BC</th>
<th>Highly Invasive BC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>K19</td>
<td>21.2</td>
<td>1.0</td>
</tr>
<tr>
<td>2</td>
<td>GATA-3</td>
<td>9.8</td>
<td>1.2</td>
</tr>
<tr>
<td>3</td>
<td>K18</td>
<td>4.2</td>
<td>2.3</td>
</tr>
<tr>
<td>4</td>
<td>RARα1</td>
<td>4.3</td>
<td>1.4</td>
</tr>
<tr>
<td>5</td>
<td>RhoB</td>
<td>8.0</td>
<td>1.6</td>
</tr>
<tr>
<td>6</td>
<td>IGFBP5</td>
<td>26.3</td>
<td>1.8</td>
</tr>
<tr>
<td>7</td>
<td>Plakoglobin</td>
<td>1.1</td>
<td>7.6</td>
</tr>
<tr>
<td>8</td>
<td>PKG7</td>
<td>1.2</td>
<td>3.3</td>
</tr>
<tr>
<td>9</td>
<td>Integrin α-3</td>
<td>5.2</td>
<td>1.6</td>
</tr>
<tr>
<td>10</td>
<td>Caveolin-1</td>
<td>11.9</td>
<td>1.6</td>
</tr>
<tr>
<td>11</td>
<td>TIMP-3</td>
<td>8.7</td>
<td>1.5</td>
</tr>
<tr>
<td>12</td>
<td>GST P</td>
<td>351.1</td>
<td>3.6</td>
</tr>
<tr>
<td>13</td>
<td>MLH1</td>
<td>171.0</td>
<td>3.5</td>
</tr>
<tr>
<td>14</td>
<td>BIGH3</td>
<td>15.0</td>
<td>1.6</td>
</tr>
<tr>
<td>15</td>
<td>MT1-MMP</td>
<td>4.8</td>
<td>1.5</td>
</tr>
<tr>
<td>16</td>
<td>PAI-1</td>
<td>4.2</td>
<td>2.8</td>
</tr>
<tr>
<td>17</td>
<td>FRA-1</td>
<td>3.3</td>
<td>3.3</td>
</tr>
<tr>
<td>18</td>
<td>Vimentin</td>
<td>5.2</td>
<td>7.9</td>
</tr>
<tr>
<td>19</td>
<td>Osteonectin</td>
<td>2.2</td>
<td>64.6</td>
</tr>
<tr>
<td>20</td>
<td>TSP1</td>
<td>1.0</td>
<td>4.5</td>
</tr>
<tr>
<td>21</td>
<td>TIMP-2</td>
<td>1.2</td>
<td>2.5</td>
</tr>
<tr>
<td>22</td>
<td>c-jun</td>
<td>1.1</td>
<td>3.9</td>
</tr>
<tr>
<td>23</td>
<td>Collagen VI α-1</td>
<td>1.4</td>
<td>7.9</td>
</tr>
<tr>
<td>24</td>
<td>Collagen L α-2</td>
<td>1.3</td>
<td>13.1</td>
</tr>
</tbody>
</table>

**Median values for fold-change in expression levels of the indicated genes relative to the reference MCF10A were recalculated using all of the weakly aggressive and highly aggressive BC cell lines. The number of cell lines in which values for differential gene expression matched the median in magnitude or direction is indicated in the "#Obs" column. The arrows indicate the direction of the change in gene expression relative to MCF10A, with each additional arrow representing a 10-fold difference in gene expression.**
cells differed in their conclusions, although concurring that vimentin expression was not sufficient to induce the metastatic phenotype.

A role for FRA-1 in the transformation of epithelioid tumor cells to a more invasive, mesenchymal cell type was demonstrated by overexpression of FRA-1 (43), although similar studies in BC cells have not been reported. Our data, demonstrating a tight association of FRA-1 and vimentin expression with only the highly invasive BC cell lines as well as c-jun elevation in all of the highly invasive BC cells, support the hypothesis that AP-1 plays a role in tumor progression. Additional studies are needed to determine which of the other genes comprising the highly invasive BC consensus GEP are also modulated by AP-1 and which genes are reporters of other key signaling pathways.

Our studies have provided evidence that GEPs can be used in the prediction of in vitro characteristics associated with a highly aggressive BC phenotype. Future studies will determine the utility of these consensus GEPs in BC diagnosis and in drug discovery screens for inhibitors of the invasive phenotype.

ACKNOWLEDGMENTS

We thank Dr. M. Stämpfer for the 184B5 and 48S HMEC, Dr. A. Chenchik and Karim Hyder for their advice and assistance with the cDNA microarrays, and Drs. R. Humm and J. MacRobbie for cell culture assistance. We also thank Drs. S. Sriniwasan and R. Silva for their roles in developing software for using the gene expression database and Drs. H. Dinter, R. Feldman, I. Kuhn, C. Lin, M. Mamounas, and G. Parry for their helpful suggestions on the manuscript.

REFERENCES

11. Kleinman, H. K., McGarvey, M. L., Hassell, J. R., Star, V. L., Cannon, F. B., Laurie, A. C., Chenchik and Karim Hyder for their advice and assistance with the cDNA microarrays, and Drs. R. Humm and J. MacRobbie for cell culture assistance. We also thank Drs. S. Sriniwasan and R. Silva for their roles in developing software for using the gene expression database and Drs. H. Dinter, R. Feldman, I. Kuhn, C. Lin, M. Mamounas, and G. Parry for their helpful suggestions on the manuscript.


Identification of Gene Expression Profiles That Predict the Aggressive Behavior of Breast Cancer Cells

Deborah A. Zajchowski, Marty F. Bartholdi, Yan Gong, et al.

Cancer Res 2001;61:5168-5178.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/61/13/5168

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2001/07/02/61.13.5168.DC2

Cited articles
This article cites 52 articles, 25 of which you can access for free at:
http://cancerres.aacrjournals.org/content/61/13/5168.full#ref-list-1

Citing articles
This article has been cited by 50 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/61/13/5168.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.