A SAGE (Serial Analysis of Gene Expression) View of Breast Tumor Progression

Dale A. Porter, Ian E. Krop, Selim Nasser, Dennis Sgroi, Carolyn M. Kaelin, Jeffrey R. Marks, Gregory Riggins, and Kornelia Polyak

Department of Adult Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115 [D. A. P., I. E. K., K. P.]; Harvard Medical School, Boston, Massachusetts 02115 [D. A. P., I. E. K., S. N., D. S., C. M. K., K. P.]; Department of Pathology, Massachusetts General Hospital, Charlestown, Massachusetts 02129 [S. N., D. S.]; Department of Surgery, Brigham and Women’s Hospital, Boston, Massachusetts 02115 [C. M. K.]; and Departments of Surgery [J. R. M.] and Pathology [G. R.], Duke University Medical Center, Durham, North Carolina 27710

Abstract

To identify molecular alterations involved in the initiation and progression of breast carcinomas, we analyzed the global gene expression profiles of normal mammary epithelial cells and in situ, invasive, and metastatic breast carcinomas using serial analysis of gene expression (SAGE). We identified sets of genes expressed only or most abundantly in a specific stage of breast tumorigenesis or in a certain subtype of tumors through the pair-wise comparison and by hierarchical clustering analysis of these eight SAGE libraries (two/stage). On the basis of these comparisons, we made the following observations: Normal mammary epithelial cells showed the most distinct and least variable gene expression profiles. Many of the genes highly expressed in normal mammary epithelium and lost in carcinomas encoded secreted proteins, cytokines, and chemokines, implicating abnormal paracrine and autocrine signaling in the initiation of breast tumorigenesis. Very few genes were universally up-regulated in all tumors regardless of their stage and histological grade, indicating a high degree of diversity at the molecular level that likely reflects the clinical heterogeneity characteristic of breast carcinomas. Tumors of different histology type and stage had very distinct gene expression patterns. No genes seemed to be specific for metastatic or for in situ carcinomas. We found that the most dramatic and consistent phenotypic change occurred at the normal-to-in situ carcinoma transition. This observation, combined with the fact that many of the genes involved encode secreted, cell-nonautonomous factors, implies that the normal epithelium-to-in situ carcinoma transition may be the most promising target for cancer prevention and treatment.

Introduction

Breast carcinoma is the second leading cause of cancer-related death in women of the Western world. In the United States alone, over 180,000 new cases are diagnosed annually (1). Despite improvements in cancer therapy, about one-quarter of the patients diagnosed with invasive breast carcinoma will eventually die from their disease. Thus, the identification of genes and biochemical pathways involved in breast oncogenesis are of utmost importance for the development of rational, molecularly based preventative and therapeutic approaches. Similar to other cancer types, breast tumorigenesis is a multistep process starting with benign then atypical hyperproliferation, progressing into in situ then invasive carcinomas, and culminating in metastatic disease (2). This tumor progression is driven by somatic genetic changes and is reflected in phenotypic changes such as altered gene expression profiles. Therefore, comprehensive analysis of gene expression profiles may identify key alterations that are important for the acquisition and maintenance of the cancerous phenotype and thus represent ideal targets for cancer prevention and treatment.

To identify such alterations we generated SAGE4 libraries from normal mammary epithelial cells and in situ, invasive, and metastatic carcinomas. SAGE analyzes 14-bp tags derived from a defined position of the cDNAs without a priori knowledge of the sequence of the genes expressed (3). The SAGE tag numbers directly reflect the abundance of the mRNAs; therefore, SAGE data are highly accurate and quantitative. Although several previous studies analyzed the gene expression profiles of breast carcinomas, no study used an unbiased comprehensive gene expression profiling approach and highly purified, uncultured, patient-derived tissues as starting material (4–6). Therefore, we believe that our study is a faithful description of the in vivo gene expression profiles of normal and cancerous mammary epithelial cells.

Materials and Methods

Normal and Cancerous Breast Tissue. All human tissue was collected following NIH guidelines and using protocols approved by the Institutional Review Boards. One of the normal mammary tissues (N1) was obtained from reduction mammoplasty tissue of a healthy patient 43 years of age undergoing plastic surgery at Brigham and Women’s Hospital. One of the DCIS samples (DCIS1) was obtained from a patient 67 years of age undergoing mastectomy for extensive, high-grade comedo DCIS and IDC in the same breast at Brigham and Women’s Hospital. The other normal and DCIS libraries (N2 and DCIS2) were obtained from a patient 35 years of age undergoing mastectomy of the left breast for intermediate-grade extensive DCIS and IDC (source of DCIS tissue) and undergoing prophylactic mastectomy of the right breast (source of normal tissue) at Massachusetts General Hospital. In both cases, cells for SAGE library generation were derived from extensive DCIS, and histology was confirmed by microscopic examination of H&E-stained sections adjacent to the parts used for SAGE. Both normal and the DCIS2 libraries were generated from cells purified using anti-BetEP4-coupled magnetic beads (Epithelial Enrich, Dynal). DCIS1 was generated from macroscopically dissected 20-µm frozen sections. The two cases of paired invasive and metastatic breast carcinomas were obtained from patients 53 years of age (IDC2 and MET2) and 57 years of age (IDC1 and MET1) who were operated on at Duke University Medical Center. Both cases were of high nuclear grade, and one (IDC1 and MET1) was estrogen and progesterone receptor-negative, whereas the other one (IDC2 and MET2) was estrogen and progesterone receptor-positive. SAGE libraries were generated from macroscopically dissected 20-µm frozen sections.

Generation and Analysis of SAGE Libraries. For immunomagnetic purification, minced breast tissue was digested in DMEM/F12 medium (Life Technologies, Inc., Rockville, MD) supplemented with 1% fetal bovine serum, 4 The abbreviations used are: SAGE, serial analysis of gene expression; DCIS, ductal carcinoma(s) in situ; IDC, invasive ductal carcinoma; MET, metastasis; HIN-1, high in normal-1; IBC-1, invasive breast carcinoma-1; C/EBP, CCAAT/enhancer binding protein; ER, estrogen receptor.
2 mg/ml collagenase I (C0130; Sigma Chemical Co.), and 2 mg/ml hyaluronidase (H3506; Sigma Chemical Co.) at 37°C for 2 h. Cells were collected by centrifugation; trypsinized; and resuspended in PBS, 1% BSA, and 2 mM EDTA and purified using the Epithelial Enrich kit (Dynal, Oslo, Norway) following the manufacturer’s recommendations. In the case of libraries derived from frozen sections, mRNA was prepared from OCT-embedded tissue using MACS kit (Miltenyi Biotec). SAGE libraries were generated following a modified micro-SAGE protocol, including a 1% SDS washing/heating step after each enzymatic reaction to ensure complete inactivation of the enzymes (7). As part of the Cancer Gene Anatomy Project SAGE consortium, SAGE libraries were arrayed at the Lawrence Livermore National Laboratories and sequenced at the Washington University Human Genome Center or at the National Institute Sequencing Center (NIH, Bethesda, MD). The data have been posted on the Cancer Gene Anatomy Project web-site as part of the SAGEmap database (8, 9). Approximately 50,000 SAGE tags were obtained from each library. The exact numbers were: N1 (49,351), N2 (38,371), D1 (42,556), D2 (29,768), IDC1 (39,996), IDC2 (67,703), MET1 (45,539), and MET2 (60,975). SAGE libraries were analyzed using the SAGE 2000 software (derivation of tags, library comparisons, and Monte Carlo analysis for statistical significance); the data were subsequently transferred into Microsoft Access (link to Unigene database) and Excel (to sort data and convert it into tab-delimited files). For comparisons, only SAGE tags that occurred at least twice/library/50,000 tags were included. To compensate for the unequal tag numbers in the eight different libraries, tag numbers were normalized to 50,000-tags/library before comparisons. Hierarchical clustering was applied to data using the Cluster program developed by Eisen et al. (10). Data were log-transformed and filtered for at least one observation abs Val ≥ and Maxval-Minval. Using these settings, 3,987 genes (of 16,808 total) were included in the analysis. Results in Fig. 1 were displayed with the TreeView program (10).

Cell Lines, RNA Preparation, and Northern Blot Analysis. Breast cancer cell lines were obtained from American Type Culture Collection or were generously provided by Drs. Steve Ethier (University of Michigan Medical Center, Ann Arbor, MI), Gail Tomlinson (University of Texas Southwestern Medical Center, Dallas, TX), and Arthur Pardee (Dana-Farber Cancer Institute, Boston, MA). Cells were grown in media recommended by the provider. However, 2 days before RNA collection, all cell lines were cultured in DMEM/F12 medium supplemented with 10% fetal bovine serum to minimize gene expression differences attributable to different culture conditions. RNA isolation, reverse transcription-PCR, and Northern blot analyses were performed as described (11).

Results

Generation and Analysis of SAGE Libraries. Genes and biochemical pathways underlying phenotypic changes during breast tumorigenesis can be identified in a number of different ways. We chose an unbiased, comprehensive gene expression profiling approach (SAGE) to create a transcriptome map of normal and cancerous mammary epithelial cells. Eight different SAGE libraries were generated and analyzed for this study: two independent cases of normal luminal epithelial cells, two different histological types of DCIS, two IDCs, and two lymph node METs. Approximately 50,000 SAGE tags were obtained from each library, leading to the analysis of over 50,000 unique transcripts. We first performed pair-wise comparisons and Monte Carlo analysis to identify SAGE tags, with statistically significantly different expression (P < 0.001) between normal-DCIS, DCIS-invasive, and invasive-metastatic libraries. Transcripts that were expressed in each pair-wise comparison were defined as normal, cancer, DCIS, invasive, or metastasis-specific genes (Table 1, A and B). In addition, we also identified genes with an expression pattern that correlated with histological grade or the presence of the ER...
Table 1  Genes highly abundant in normal and in cancerous mammary epithelial cells

<table>
<thead>
<tr>
<th>Unigene ID no.</th>
<th>Gene description</th>
<th>N1</th>
<th>N2</th>
<th>DCIS1</th>
<th>DCIS2</th>
<th>IDC1</th>
<th>IDC2</th>
<th>MET1</th>
<th>MET2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Genes highly abundant in normal mammary epithelial cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75765</td>
<td>Chemokine GROβ</td>
<td>67</td>
<td>138</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>789</td>
<td>Chemokine GROα</td>
<td>219</td>
<td>252</td>
<td>5</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>624</td>
<td>Chemokine CXCα (IL-8)</td>
<td>205</td>
<td>196</td>
<td>4</td>
<td>21</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>81328</td>
<td>IKα</td>
<td>75</td>
<td>84</td>
<td>3</td>
<td>21</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>17778</td>
<td>SOD2α</td>
<td>172</td>
<td>99</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>3337</td>
<td>Transmembrane 4 super family member 1</td>
<td>74</td>
<td>53</td>
<td>5</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>75969</td>
<td>Proline-rich protein</td>
<td>21</td>
<td>24</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2250</td>
<td>Leukemia inhibitory factor</td>
<td>35</td>
<td>75</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>23582</td>
<td>Membrane component, surface marker 1</td>
<td>61</td>
<td>66</td>
<td>10</td>
<td>20</td>
<td>5</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>241507</td>
<td>Ribosomal protein S6</td>
<td>143</td>
<td>108</td>
<td>51</td>
<td>40</td>
<td>38</td>
<td>16</td>
<td>28</td>
<td>16</td>
</tr>
<tr>
<td>75498</td>
<td>Chemokine MIP3α</td>
<td>24</td>
<td>16</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>25829</td>
<td>ras-related protein</td>
<td>25</td>
<td>16</td>
<td>3</td>
<td>0</td>
<td>5</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>NA</td>
<td>NO MATCH (CTTCTCTGTGA)</td>
<td>275</td>
<td>127</td>
<td>5</td>
<td>11</td>
<td>13</td>
<td>8</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>76722</td>
<td>CEP68</td>
<td>86</td>
<td>62</td>
<td>21</td>
<td>25</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>91539</td>
<td>ESTs</td>
<td>29</td>
<td>27</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>101382</td>
<td>TNF-α induced protein 2</td>
<td>36</td>
<td>18</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>NA</td>
<td>HIN-1</td>
<td>69</td>
<td>24</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>Genes highly abundant in cancerous mammary epithelial cells</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>82961</td>
<td>Trefoil factor 3 (intestinal)</td>
<td>19</td>
<td>3</td>
<td>285</td>
<td>477</td>
<td>206</td>
<td>97</td>
<td>159</td>
<td>136</td>
</tr>
<tr>
<td>83190</td>
<td>Fatty acid synthase</td>
<td>9</td>
<td>2</td>
<td>29</td>
<td>35</td>
<td>93</td>
<td>18</td>
<td>141</td>
<td>25</td>
</tr>
<tr>
<td>149923</td>
<td>X-box binding protein 1</td>
<td>44</td>
<td>32</td>
<td>82</td>
<td>109</td>
<td>136</td>
<td>118</td>
<td>334</td>
<td></td>
</tr>
<tr>
<td>265882</td>
<td>INF-α-inducible protein, IFI-6-16</td>
<td>2</td>
<td>0</td>
<td>9</td>
<td>359</td>
<td>72</td>
<td>95</td>
<td>7</td>
<td>293</td>
</tr>
<tr>
<td>170711</td>
<td>Glutamine synthase</td>
<td>6</td>
<td>3</td>
<td>41</td>
<td>30</td>
<td>2</td>
<td>6</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>74316</td>
<td>Desmoplakin</td>
<td>4</td>
<td>9</td>
<td>50</td>
<td>30</td>
<td>7</td>
<td>14</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>6335</td>
<td>PI4P5 kinase, type II</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>28</td>
<td>22</td>
<td>40</td>
<td>4</td>
</tr>
<tr>
<td>3343</td>
<td>3-phosphoglycerate dehydrogenase</td>
<td>6</td>
<td>3</td>
<td>11</td>
<td>8</td>
<td>51</td>
<td>59</td>
<td>26</td>
<td>106</td>
</tr>
<tr>
<td>268571</td>
<td>Apolipoprotein C-I</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>0</td>
<td>48</td>
<td>32</td>
<td>45</td>
<td>15</td>
</tr>
<tr>
<td>7744</td>
<td>NA HIN-1</td>
<td>49</td>
<td>8</td>
<td>3</td>
<td>33</td>
<td>42</td>
<td>27</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>169476</td>
<td>GADPH</td>
<td>27</td>
<td>24</td>
<td>16</td>
<td>21</td>
<td>87</td>
<td>73</td>
<td>127</td>
<td>92</td>
</tr>
<tr>
<td>84298</td>
<td>CD74 antigen</td>
<td>4</td>
<td>18</td>
<td>16</td>
<td>3</td>
<td>88</td>
<td>115</td>
<td>113</td>
<td>40</td>
</tr>
<tr>
<td>119206</td>
<td>KIF7</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>3</td>
<td>27</td>
<td>35</td>
<td>30</td>
<td>6</td>
</tr>
<tr>
<td>74823</td>
<td>NA HIN-1</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>36</td>
<td>15</td>
<td>23</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>2186</td>
<td>EEF1G</td>
<td>48</td>
<td>45</td>
<td>58</td>
<td>31</td>
<td>628</td>
<td>568</td>
<td>954</td>
<td>76</td>
</tr>
<tr>
<td>302063</td>
<td>Immunoglobulin μ</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>96</td>
<td>39</td>
<td>178</td>
<td>7</td>
</tr>
<tr>
<td>181125</td>
<td>Immunoglobulin A gene cluster</td>
<td>0</td>
<td>3</td>
<td>22</td>
<td>3</td>
<td>146</td>
<td>100</td>
<td>191</td>
<td>9</td>
</tr>
<tr>
<td>180142</td>
<td>Catheolin-like skin protein</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>26</td>
<td>0</td>
<td>26</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>NA</td>
<td>IBC-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>98</td>
<td>56</td>
<td>110</td>
<td>0</td>
</tr>
</tbody>
</table>

* SOD2, superoxide dismutase 2; NA, Not Applicable; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; IGF-BP7, Insulin-like growth factor binding protein 7.

(Table 2, A and B). Data were also analyzed using a clustering algorithm to delineate patterns of gene expression among all eight libraries (Fig. 1). Clusters of coexpressed genes suggested that the two normal mammary epithelial libraries, despite being derived from two different patients, were the most similar to each other. Only 43 tags were expressed differentially between the two normal libraries. The four invasive and metastatic libraries appeared to be quite similar to each other. Interestingly the two DCIS libraries were very dissimilar (225 tags showed an at-least 2-fold statistically significant difference) and clustered according to their predicted clinical behavior. High-grade DCIS have high recurrence rates and worse overall prognoses when compared with intermediate or low-grade DCIS. In agreement with this, the high-grade comedo DCIS (DCIS-1) was more similar to invasive tumors, whereas the intermediate DCIS was more similar to normal mammary epithelium. Although the two normal and the DCIS-2 libraries were generated from cells purified using anti-Ber-Ep4-coated magnetic beads, whereas all other libraries were derived from macroscopically dissected (>95% pure) tumors, it is unlikely that the presence of contaminating cells (<5%; mostly fibroblasts) in the dissected samples would be the only reason for this intriguing clustering pattern. However, this possibility cannot be excluded.

Genes Characteristic for Normal Mammary Epithelium. A large fraction of normal mammary epithelium-specific genes (Table 1A) encode secreted proteins, chemokines (IL-8, GROα, GROβ, and MIP3α), and cytokines (LIF, IL-6, and HIN-1) that may play autocrine and/or paracrine roles in the regulation of normal mammary epithelial cell growth, differentiation, and morphogenesis. HIN-1, for example, is a novel growth-inhibitory cytokine hypermethylated in a large fraction of breast cancers that may play a role in epithelial cell proliferation and branching morphogenesis. Several other chemokines (SCYA2, SCYA7, SCYA5, SCYB6, and SYD1) and chemokine receptors (IL-4R, IL-6R, and IL-15R) were also more abundant (5–10-fold) in normal mammary epithelial cells, but because of their low tag numbers, the P of this difference was above P = 0.001. It is unlikely that these chemokines and cytokines were derived from contaminating leukocytes, because SAGE libraries were generated from immunomagnetic purified normal luminal mammary epithelial cells, and the purity of the cells was confirmed by reverse transcription-PCR analysis using known luminal (HIN-1 and EMA) and myoepithelial (calponin and CALLA/C/CD10) cell-specific transcripts (data not shown). Preliminary mRNA in situ and immunohistochemical experiments also confirmed that chemokines and cytokines are expressed in normal luminal mammary epithelial cells. Similarly, a recently identified novel chemokine, mammary enriched chemokine, or MEC, was also shown to be expressed in normal mammary epithelial cells and found to be down-regulated in breast carcinomas (12).

Another gene expressed at high levels in normal mammary epithelium and lost in tumors is the mitochondrial superoxide dismutase, SOD-2. A decreased level of SOD-2 in cancer cells is well documented and it may contribute to the high level of reactive oxygen species and subsequent oxidative stress characteristic of breast cancers (13, 14). The expression levels of two transcriptional regulators, IKα and C/EBPβ, show a somewhat gradual decline from normal to in situ and then to invasive carcinomas. IKα sequesters NFκβ in the...
cytoplasm in an inactive form; therefore, loss of IKK expression may lead to activation of the NFκB pathway. C/EBPs are transcription factors that regulate cell growth and differentiation in a cell type-specific manner. C/EBPβ has been shown to be required for growth arrest in mammary epithelial cells (15); thus, the gradual decline of C/EBPβ mRNA levels during breast tumor progression may correlate with increased proliferation rates.

Genes Differently Expressed Among Various Tumors and Their Correlation with Histological Parameters. In contrast to the relatively high number of normal mammary epithelium-specific genes, only three genes have been clearly expressed in all breast carcinomas. These included trefoil factor 3, X-box binding protein 1, and fatty acid synthase (Table 1B). Both trefoil factor 3 and fatty acid synthase previously have been demonstrated to be up-regulated in breast carcinomas (16–18). The IFN-α-inducible protein, IFI-6 –16, is particularly interesting because of their restricted expression in invasive breast carcinomas. One corresponds to calmodulin-like skin gene, and the other, IFN-α-inducible protein, is likely attributable to increased protein synthesis and the possible presence of contaminating lymphocytes in the dissected samples.

Table 2 Genes highly abundant in high-grade and in intermediate-grade/ER+ tumors

<table>
<thead>
<tr>
<th>Unigene ID no.</th>
<th>Gene description</th>
<th>N1</th>
<th>N2</th>
<th>DCIS1</th>
<th>DCIS2</th>
<th>IDC1</th>
<th>IDC2</th>
<th>MET1</th>
<th>MET2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Genes highly abundant in high-grade tumors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>172928</td>
<td>Collagen type I, ε1 (ACC(AAAACCC)</td>
<td>1</td>
<td>2</td>
<td>157</td>
<td>1</td>
<td>51</td>
<td>39</td>
<td>85</td>
<td>18</td>
</tr>
<tr>
<td>172928</td>
<td>Collagen type I, ε1 (TGGA(AATGACA)</td>
<td>1</td>
<td>1</td>
<td>106</td>
<td>0</td>
<td>102</td>
<td>50</td>
<td>140</td>
<td>48</td>
</tr>
<tr>
<td>111779</td>
<td>Osteonectin</td>
<td>2</td>
<td>0</td>
<td>65</td>
<td>1</td>
<td>62</td>
<td>53</td>
<td>91</td>
<td>18</td>
</tr>
<tr>
<td>2132</td>
<td>EGF receptor pathway substrate 8</td>
<td>4</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td>28</td>
<td>19</td>
<td>50</td>
<td>3</td>
</tr>
<tr>
<td>177486</td>
<td>Amyloid A4</td>
<td>9</td>
<td>11</td>
<td>43</td>
<td>0</td>
<td>30</td>
<td>16</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>293441</td>
<td>SNC73</td>
<td>3</td>
<td>13</td>
<td>80</td>
<td>0</td>
<td>215</td>
<td>111</td>
<td>197</td>
<td>28</td>
</tr>
<tr>
<td>5734</td>
<td>KIAA0679 protein</td>
<td>5</td>
<td>3</td>
<td>22</td>
<td>1</td>
<td>23</td>
<td>19</td>
<td>25</td>
<td>27</td>
</tr>
<tr>
<td>164170</td>
<td>Clone HeP06932</td>
<td>3</td>
<td>7</td>
<td>54</td>
<td>3</td>
<td>33</td>
<td>22</td>
<td>46</td>
<td>27</td>
</tr>
<tr>
<td>184108</td>
<td>Ribosomal protein S21</td>
<td>45</td>
<td>95</td>
<td>123</td>
<td>36</td>
<td>138</td>
<td>124</td>
<td>124</td>
<td>72</td>
</tr>
<tr>
<td>169793</td>
<td>Ribosomal protein L32</td>
<td>273</td>
<td>375</td>
<td>679</td>
<td>124</td>
<td>362</td>
<td>278</td>
<td>349</td>
<td>314</td>
</tr>
<tr>
<td>300067</td>
<td>Immunoglobulin κ</td>
<td>0</td>
<td>0</td>
<td>30</td>
<td>0</td>
<td>402</td>
<td>371</td>
<td>807</td>
<td>60</td>
</tr>
<tr>
<td>205233</td>
<td>NO MATCH (TGAACAGCAT)</td>
<td>1</td>
<td>0</td>
<td>56</td>
<td>1</td>
<td>32</td>
<td>22</td>
<td>30</td>
<td>12</td>
</tr>
<tr>
<td>B. Genes highly abundant in intermediate-grade/ER+ tumors</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>274404</td>
<td>Phosphoglycerate kinase activator</td>
<td>10</td>
<td>10</td>
<td>3</td>
<td>63</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>2006</td>
<td>Glutathione S-transferase M3</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>26</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td>75410</td>
<td>HSP70.5</td>
<td>20</td>
<td>16</td>
<td>8</td>
<td>94</td>
<td>48</td>
<td>15</td>
<td>29</td>
<td>34</td>
</tr>
<tr>
<td>76067</td>
<td>HSP27</td>
<td>19</td>
<td>9</td>
<td>4</td>
<td>132</td>
<td>11</td>
<td>11</td>
<td>0</td>
<td>7</td>
</tr>
<tr>
<td>79516</td>
<td>Neuronal tissue-enriched protein</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>16</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2340</td>
<td>Plakoglobin</td>
<td>30</td>
<td>24</td>
<td>24</td>
<td>99</td>
<td>25</td>
<td>19</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td>75243</td>
<td>Female sterile homeotic-related gene</td>
<td>5</td>
<td>10</td>
<td>10</td>
<td>35</td>
<td>2</td>
<td>6</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>77886</td>
<td>Lamin A/C</td>
<td>32</td>
<td>41</td>
<td>8</td>
<td>67</td>
<td>6</td>
<td>3</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>182265</td>
<td>Keratin 19</td>
<td>18</td>
<td>19</td>
<td>32</td>
<td>92</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>31439</td>
<td>Keratin 18</td>
<td>4</td>
<td>1</td>
<td>10</td>
<td>35</td>
<td>31</td>
<td>63</td>
<td>10</td>
<td>77</td>
</tr>
<tr>
<td>202833</td>
<td>Heme oxygenase</td>
<td>41</td>
<td>6</td>
<td>3</td>
<td>95</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>65114</td>
<td>Keratin 18</td>
<td>31</td>
<td>31</td>
<td>21</td>
<td>63</td>
<td>25</td>
<td>22</td>
<td>21</td>
<td>36</td>
</tr>
<tr>
<td>222341</td>
<td>EF18</td>
<td>0</td>
<td>13</td>
<td>21</td>
<td>52</td>
<td>1</td>
<td>17</td>
<td>0</td>
<td>9</td>
</tr>
<tr>
<td>74631</td>
<td>Basigin</td>
<td>8</td>
<td>5</td>
<td>8</td>
<td>21</td>
<td>0</td>
<td>6</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>79144</td>
<td>Clone MGC2479</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>20</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>72222</td>
<td>Clone PPI3795</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>18</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NA</td>
<td>NO MATCH (CAGACTTTTT)</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>30</td>
<td>1</td>
<td>5</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>NA</td>
<td>NO MATCH (TGTTAGCGA)</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>NA</td>
<td>NO MATCH (TGGAAATGAC)</td>
<td>1</td>
<td>1</td>
<td>106</td>
<td>0</td>
<td>102</td>
<td>50</td>
<td>140</td>
<td>48</td>
</tr>
<tr>
<td>NA</td>
<td>NO MATCH (ACCAAAAACC)</td>
<td>1</td>
<td>2</td>
<td>157</td>
<td>1</td>
<td>51</td>
<td>39</td>
<td>85</td>
<td>18</td>
</tr>
<tr>
<td>56892</td>
<td>CSORF4</td>
<td>11</td>
<td>19</td>
<td>2</td>
<td>77</td>
<td>3</td>
<td>10</td>
<td>1</td>
<td>13</td>
</tr>
<tr>
<td>NA</td>
<td>NO MATCH (CGCCGATATA)</td>
<td>4</td>
<td>6</td>
<td>2</td>
<td>20</td>
<td>1</td>
<td>0</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>NA</td>
<td>NO MATCH (CCGCTGCGAG)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>41</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>NA</td>
<td>NO MATCH (GTATTTTCTTC)</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NA</td>
<td>NO MATCH (GGGAAGCAGA)</td>
<td>43</td>
<td>57</td>
<td>19</td>
<td>87</td>
<td>7</td>
<td>33</td>
<td>6</td>
<td>9</td>
</tr>
<tr>
<td>NA</td>
<td>NO MATCH (GCCCTCTCAC)</td>
<td>9</td>
<td>13</td>
<td>1</td>
<td>33</td>
<td>7</td>
<td>3</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>17409</td>
<td>Cysteine-rich protein (intestinal)</td>
<td>18</td>
<td>2</td>
<td>11</td>
<td>36</td>
<td>3</td>
<td>27</td>
<td>1</td>
<td>33</td>
</tr>
<tr>
<td>278613</td>
<td>INF-γ-inducible protein 27</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>20</td>
<td>0</td>
<td>11</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>82932</td>
<td>Cyclin D1</td>
<td>4</td>
<td>1</td>
<td>10</td>
<td>35</td>
<td>31</td>
<td>63</td>
<td>10</td>
<td>77</td>
</tr>
</tbody>
</table>

a NA, Not Applicable.

The expression of certain genes appeared to correlate with the histological grade of the tumor and/or the presence of the ER. All four invasive/metastatic and one of the DCIS tumors were high nuclear grade, and most of the genes highly expressed in these tumors encode extracellular matrix proteins, such as various collagens, osteonectin, and BIGH3, a TGFβ-induced gene. Although some of these genes could be expressed in contaminating fibroblasts, previous studies demonstrated that osteonectin is produced by most ER-negative breast carcinomas and may contribute to the aggressive phenotype and the presence of microcalcifications characteristic for these tumors. To demonstrate that these secreted factors can be produced by normal and/or cancerous epithelial cells themselves, we performed Northern blot analysis of normal organoids (purified breast ducts) and various breast cancer cell lines (Fig. 2). Correlating with previous studies, breast cancer cell lines do not seem to reproduce the in vivo expression pattern of these differentially expressed genes. However, some cells clearly express high levels of genes encoding extracellular matrix components and other secreted proteins, indicating that these proteins can be produced by the tumor cells themselves.

We were not able to define an “intermediate grade-specific” gene cluster, because we analyzed only one intermediate-grade tumor (DCIS2). However, the expression of several STAT transcription targets, such as heat shock proteins and IFN-α inducible protein 27, appeared to be significantly higher in DCIS2 (Table 2B). Despite the fact that we analyzed three ER-negative (DCIS1, IDC1, and MET1) and three ER-positive (DCIS2, IDC2, and MET2) tumors, the only ER target that clearly correlated with the ER status of the tumors was cyclin D1. This could be attributable to the fact that both IDC2 and MET2 were of high nuclear grade, whereas most ER-positive tumors are of low-intermediate grade. All other known ER target genes were either present in all (cathepsin D and bcl-2) or in a subset of libraries (LIV-1-DCIS2 only; Trefoil factor 1, both normal and both DCIS).

Discussion

Characterization of global gene expression profiles may help elucidate important biological processes in both normal and cancer cells. Although these studies are largely descriptive, they may reveal the molecular basis of already-known phenotypic observations, and they may suggest new hypotheses that could stimulate future experiments. In agreement with this, our SAGE analysis of gene expression patterns at various stages of mammary tumorigenesis led to several important observations that revealed or confirmed the molecular basis of known biochemical findings and also identified novel ones that are likely to open up new avenues of future studies. In addition, because SAGE does not require prior knowledge of genes, novel transcripts can be identified. In fact, several of the differentially expressed SAGE tags we identified currently have no ESTs or other database match (Table 2, A and B). This is likely attributable to the restricted expression pattern of these genes, which makes them excellent candidate tumor markers. Searching against the completed human genome sequence, these SAGE tags will likely lead to the identification of novel transcripts. These transcripts could then be included in DNA arrays for high throughput analysis of breast and other tumors to determine their usefulness for cancer diagnosis and/or prognostication.

The high levels of SOD-2, IKKα, and C/EBPβ in normal mammary epithelial cells and the loss of expression in breast cancers correlates with previous studies (13, 14, 23). One of the novel findings reported here was that several of the genes highly expressed in normal mammary epithelial cells and down-regulated in tumors encode secreted proteins, chemokines, and cytokines. Chemokines are mediators of immune cell trafficking, and, although they are also involved in regulating cell movements during morphogenesis, their principal targets are bone marrow-derived cells (24, 25). Recent data indicate that in addition to fibroblasts and adipocytes, macrophages, eosinophils, and endothelial cells are also required for normal mammary gland development (26, 27). Although the function of leukocytes and endothelial cells in this process has not been fully elucidated, they may be necessary for the formation of terminal end buds and for branching morphogenesis. The high abundance of GROα, GROβ, MIP3α, IL-8, and, to a lesser degree, SCYA2, SCYA7, SCYB5, SCYB6, and SYD1 chemokines in normal mammary epithelial cells suggests that epithelial cells may actively recruit bone marrow-derived and endothelial cells. Alternatively, chemokines may play an as yet uncharacterized role in interepithelial cell communication, a hypothesis that deserves additional investigation.

There were very few genes that were universally up-regulated in all of the tumors examined, indicating a high degree of heterogeneity of breast carcinomas at the molecular level. Even based on the limited number of tumors analyzed, it is clear that tumors of different histological stage (invasive) or nuclear grade (high or intermediate) clearly show distinct gene expression patterns that are likely to reflect their clinical behavior. The same is true for ER-positive and -negative tumors. However, we found no genes that were up-regulated in both lymph node metastases that were not already expressed in invasive carcinomas. Similarly, there were very few genes that were up-regulated in both DCIS but not in invasive carcinomas. This could be attributable to the limited number of samples analyzed, or it may indicate that preinvasive and metastatic lesions may be even more heterogeneous than invasive ones. Alternatively, genes responsible for invasive and metastatic behavior may already be expressed in preinvasive and invasive lesions, respectively. Analysis of large numbers of in situ carcinomas without adjacent invasive lesions and analysis of many invasive carcinomas with and without lymph node metastasis would be required to answer this question.

Some of the cancer-specific genes we identified could be used for

![Fig. 2. Northern blot analysis of differentially expressed genes in breast cancer cell lines and in normal organoids. The following genes were analyzed: collagen 1α2, connective tissue growth factor (CTGF), psoriasin/S100A7, osteonectin, IFN-induced clone IFI-6–16; chemokines GROα, GROβ, interleukin-8 (IL-8); and MIP3α, leukemia inhibitor factor (LIF), trefoil factor 3, and IKKα.](image-url)
cancer diagnosis and/or molecular-based anticancer therapy. For example, breast and other cancer cells have very high levels of fatty acid synthase (Table 1B; Ref. 18), and its inhibition leads to apoptosis in in vitro cell cultures and in xenograft models (28). Similarly, two of the transcripts highly expressed in invasive breast cancers, calmodulin-like skin protein, and IBC-1, a novel gene, have restricted expression patterns that make them good candidate invasive breast cancer tumor markers.

There were several similarities and differences between our findings and previously published data on the molecular profiling of breast tumors (4–6). Correlating with our data, trefoil factor, cyto-keratin 18, X-box binding protein, cyclin D1, and transcriptional targets of the STAT/IFN pathway are highly up-regulated in a subset of breast cancers (4–6). In contrast, our “normal cluster” is very different from that of previous studies. There are several possible explanations of this difference: (a) in contrast with this study, no previous study used purified, uncultured, normal luminal mammary epithelial cells; (b) in previous cDNA array experiments the reference RNA was derived from a mix of cancer cell lines, potentially masking differences between normal and cancerous mammary epithelial cells; and (c) some of the genes we identified, especially the ones with no EST matches, may not be present on the arrays. Future studies analyzing the same samples on different platforms (SAGE versus arrays) are required to resolve the observed differences.

In summary, using comprehensive gene expression profiling, we identified several genes and pathways that have not previously been implicated in breast cancer and which may play important roles in the initiation and progression of breast carcinomas. Because we analyzed a limited number of specimens, additional experiments using high-throughput techniques, such as mRNA in situ hybridization or immunohistochemical analysis on tissue-microarrays, are required to determine how commonly these genes are differentially expressed. Additional analysis of these genes and the biochemical pathways in which they are involved will not only further our understanding of breast oncogenesis, but will also provide new and valuable targets for translational research.

We thank Drs. Daniel Haber and Jeff Sklar for their critical review of this manuscript.

References
A SAGE (Serial Analysis of Gene Expression) View of Breast Tumor Progression


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/61/15/5697

Cited articles
This article cites 27 articles, 12 of which you can access for free at:
http://cancerres.aacrjournals.org/content/61/15/5697.full#ref-list-1

Citing articles
This article has been cited by 41 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/61/15/5697.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.