A single-gene biomarker identifies breast cancers associated with immature cell type and short duration of prior breastfeeding

W F Symmans1, D J Fiterman1, S K Anderson2, M Ayers5, R Rouzier3, V Dunmire1, J Stec5, V Valero3, N Sneige1, C Albarracin1, Y Wu1, J S Ross5,6, P Wagner3, R L Theriault3, B Arun3, H Kuerer4, K R Hess2, W Zhang1, G N Hortobagyi3 and L Pusztai3

Departments of 1Pathology, 2Biostatistics and Applied Mathematics, 3Breast Medical Oncology and 4Surgery, University of Texas M D Anderson Cancer Center, 1515 Holcombe Boulevard, Houston, TX 77030-4009, USA

5Millennium Pharmaceuticals, Inc., 75 Sidney Street, Cambridge, MA 02139, USA

6Department of Pathology, Albany Medical College, 47 New Scotland Avenue, Albany, NY 12208-3479, USA

Abstract

The pathogenesis of breast cancers that do not express estrogen receptors or Her-2/neu receptors (ER−/HER2− phenotype) is incompletely understood. We had observed markedly elevated gene expression of gamma-aminobutyric acid type A (GABAA) receptor subunit π (GABAπ, GABRP) in some breast cancers with ER−/HER2− phenotype. In this study, transcriptional profiles (TxPs) were obtained from 82 primary invasive breast cancers by oligonucleotide microarrays. Real-time reverse transcription–polymerase chain reaction (RT–PCR) was used to measure GABAπ gene expression in a separate cohort of 121 invasive breast cancers. GABAπ gene expression values from TxP and RT–PCR were standardized and compared with clinicopathologic characteristics in the 203 patients. GABAπ gene expression was increased in 16% of breast cancers (13/82 TxP, 20/121 RT–PCR), particularly in breast cancers with ER−/HER2− phenotype (60%), and breast cancers with basal-like genomic profile (60%). The profile of genes coexpressed with GABAπ in these tumors was consistent with an immature cell type. In multivariate linear regression analysis, the level of GABAπ gene expression was associated with ER−/HER2− phenotype (P < 0.0001), younger age at diagnosis (P = 0.0003), and shorter lifetime duration of breastfeeding (≤ 6 months) in all women (P = 0.017) and specifically in parous women (P = 0.013). GABAπ gene expression was also associated with combinations of high grade with ER−/HER2− phenotype (P = 0.002), and with Hispanic ethnicity (P = 0.036). GABAπ gene expression is increased in breast cancers of immature (undifferentiated) cell type and is significantly associated with shorter lifetime history of breastfeeding and with high-grade breast cancer in Hispanic women.

Introduction

Estrogen receptor (ER) and Her-2/neu receptor (HER2) are major determinants of growth and overall gene expression in breast cancer (Perou et al. 2000a, Pusztai et al. 2003). However, invasive breast cancers with negative ER and HER2 status (ER−/HER2−) are not well characterized with respect to the molecular determinants of cancer cell growth and selection of targeted molecular therapies.
Neelands & MacDonald 1999). Jiang et al. (2002) identified GABA<sub>π</sub> gene expression in breast cancers from subtracted cDNA libraries. Those authors detected GABA<sub>π</sub> transcripts in 10/27 primary breast cancers (37%) by semiquantitative reverse transcription–polymerase chain reaction (RT–PCR), but the tumor samples were not further characterized (Zehentner et al. 2002).

GABA<sub>π</sub> is not expressed in normal adult neuronal tissues (unlike other GABA<sub>A</sub> subunits), but it is expressed in endocrine and reproductive tissues (Hedblom & Kirkness 1997). Induction of GABA<sub>π</sub> gene expression has been described in the uterus at the time of implantation and also immediately before parturition (Majewska et al. 1989, Laurie et al. 1992, Fujii & Mellon 2001), possibly related to the levels of an active progesterone metabolite, allopregnanolone (Hedblom & Kirkness 1997, Fujii & Mellon 2001). Breast tissues undergo major physiologic changes during pregnancy and postpartum, so it is reasonable to propose that GABA<sub>π</sub> expression in breast tissues might be related to the endocrinology of reproduction or lactation. It follows, therefore, that elevated GABA<sub>π</sub> expression in a subset of breast cancers could be associated with specific patient characteristics, including reproductive and/or lactation history.

**Materials and methods**

**Biopsy samples**

A cohort of 203 newly diagnosed invasive breast cancers were evaluated, 82 by Affymetrix GeneChip oligonucleotide gene expression microarrays (TxP), and 121 by real-time RT–PCR (TaqMan). Fine-needle aspiration (FNA) samples were collected from 82 patients with newly diagnosed pretreatment invasive breast carcinomas and were prepared for genomic analysis (Pusztai et al. 2003, Symmans et al. 2003). This study was conducted at the University of Texas M D Anderson Cancer Center (MDACC) and was approved by the institutional review board of MDACC (IRB protocol LAB-99-402). All patients signed an informed consent for voluntary participation. FNA samples were obtained with 23 gauge needles (1–2 passes), mixed in 1.0 ml RNA<sub>later</sub> RNA stabilization reagent (Ambion, Inc., Austin, TX, USA), held at room temperature for 30 min and stored at −80°C. Cytologic samples were also collected from 121 surgical resection specimens containing newly diagnosed, invasive breast carcinoma (IRB protocol LAB-03-432), using 23 gauge needles (six passes) or cytologic scrapes with a no. 15 surgical blade, mixed into 1.0 ml RNA<sub>later</sub> and stored at −80°C.

All 203 samples were also annotated with the corresponding results of standard pathologic assays for ER with immunohistochemistry (monoclonal antibody clone 6F11, Novoceastra Labs, Newcastle, UK) and HER2 with immunohistochemistry (monoclonal antibody clone Ab8, Lab Vision, Freemont, CA, USA) and/or fluorescent in situ hybridization (FISH) with probes to detect c-erbB-2 and the centromere of chromosome 17 (Vysis, Downers Grove, IL, USA). (Harvey et al. 1999, Elledge et al. 2000, Ogura et al. 2003, Sapino et al. 2003). ER-positive status was defined as immunohistochemical (IHC) staining in ≥10% of cancer cell nuclei. HER2-positive status was defined as an IHC score of 3+ membranous staining and/or a FISH score of relative gene copy number of >2.0 (ratio of c-erb-B2/centromere 17). All tumors with 1+ or 2+ membranous staining for HER2 were tested for gene copy number by FISH.

**RNA extraction**

Total RNA was extracted from 82 thawed in vivo FNA samples with the RNeasy Kit (Qiagen). At least 1 μg total RNA with a 260/280 ratio of >1.8 by spectrophotometry (DU640 UV; Beckman Coulter, Fullerton, CA, USA) was required for transcriptional profiling. Total RNA was extracted from 121 ex vivo cytologic samples, using a modified protocol of the RNAqueous RNA isolation system (Ambion). Briefly, FNA samples (1.0 ml) were homogenized in 3.0 ml lysis-binding solution (cell pellets were homogenized in 1.0 ml). An equal volume of ethanol (64%) was then mixed into each sample, the mixture was passed through the filter cartridges provided, and the rest of the extraction protocol was followed (Ambion). RNA yield and quality were assessed with an Agilent 2100 Bioanalyzer (Agilent, Palo Alto, CA, USA). A ratio of 28s/18s RNA in the range 1.5–2.5 indicated high-quality RNA free of degradation. RNA samples were stored at −80°C.

**Transcriptional profiles (TxP)**

Transcriptional profiling was performed at Millennium Pharmaceuticals (Cambridge, MA, USA) on 82 in vivo FNA biopsies of breast cancer. The GeneChip system (Affymetrix, Santa Clara, CA, USA) was used for hybridization, labeling, and scanning of the U133A and B arrays. Microarray Suite 5.0 was used for data acquisition and preliminary analysis. Generation of cRNA followed a standard T7 amplification protocol. Total RNA was reverse-transcribed with SuperScript.
II (Gibco, Carlsbad, CA, USA) in the presence of T7-(dT)24 primer to generate first-strand cDNA. A second-strand cDNA synthesis was performed in the presence of DNA polymerase I, DNA ligase, and RNase H (Gibco). The resulting double-stranded cDNA was blunt-ended with T4 DNA polymerase. This double-stranded cDNA was then transcribed into cRNA in the presence of biotin-ribonucleotides, using the BioArray High Yield RNA transcript labeling kit (Enzo Life Sciences, Farmingdale, NY, USA). The amplified, biotin-labeled cRNA was purified with RNeasy columns (Qiagen), quantified and fragmented at 94°C for 35 min in the presence of fragmentation buffer (×1). Fragmented cRNA was hybridized to Affymetrix U133A chips overnight at 42°C. To control for hybridization efficiency, a standard probe cocktail supplied by Affymetrix was spiked into the hybridization mix. The software checked for the expression of all known housekeeping genes represented on the chip. For housekeeping genes on the chip, a ratio of the signal obtained for 3′ and 5′ probes is used as an indicator of the efficiency of cRNA preparation. Ratios of 1.0–3.0 indicate acceptable cRNA preparation. Affymetrix data were quantified and normalized with dCHIP V1.3 (http://biosun1.harvard.edu/complab/dchip/data) (Li & Hung Wong 2001).

Real-time quantitative RT–PCR

Real-time quantitative RT–PCR was performed with TaqMan 7700 technology (Applied Biosystems, Foster City, CA, USA). One-step RT and PCR was performed. A volume of 10 µL RNA (5 ng total) was mixed with 11.25 µL nuclease-free water (Invitrogen), 25 µL × 2 Master Mix without UNG (Applied Biosystems), 1.25 µL × 40 Multiscribe and RNase Inhibitor (Applied Biosystems), and 2.5 µL × 20 Assay-on-Demand (Applied Biosystems) predeveloped primer-probe sets for GABA\(\text{\textalpha}^{-}\)

Expression values of GABA\(\text{\textalpha}^{-}\) from the TxP and RT–PCR experiments underwent logarithmic transformation and standardization to a mean of zero and standard deviation of 1.0. Standardized GABA\(\text{\textalpha}^{-}\) expression values were ranked to demonstrate the threshold value that best separated higher from lower expression of GABA\(\text{\textalpha}^{-}\) in the TxP data and the RT–PCR data. These standardized expression values were used as a continuous variable for statistical comparisons with the clinical and pathologic variables in univariate analyses and in multivariate analyses using S-PLUS 6.1 for Windows software (Seattle, WA, USA). Linear regression models and ANOVA tests were used to determine the significance of the data, unless otherwise noted. Statistical significance was defined as \(P<0.05\).

Results

GABA\(\text{\textalpha}^{-}\) gene expression values were higher in ER−/HER2− breast cancers in the cohort of 82 breast cancer TxPs, using Affymetrix U133A microarrays (Fig. 1A), and also in an independent cohort of 121 breast cancers, using quantitative RT–PCR (Fig. 1B). The standardized gene expression values for GABA\(\text{\textalpha}^{-}\) from the TxP and RT–PCR data both demonstrated the greatest increase in GABA\(\text{\textalpha}^{-}\) expression at the 16.3 percentile, so a standardized GABA\(\text{\textalpha}^{-}\) gene expression value of >1.1 (in 16% of invasive breast cancers) was defined as high GABA\(\text{\textalpha}^{-}\) expression. The breast cancers with high GABA\(\text{\textalpha}^{-}\) expression (standardized value of >1.1) usually had ER−/HER2− phenotype (31/33, 94%) (Table 1), and were high grade (27/33, 81%) or intermediate grade (6/33, 19%). Overall, 26% (52/203) of invasive breast cancers had ER−/HER2− phenotype, but 60% of those tumors (31/52) also had high diagnosis (in years), race (white, black, Hispanic and all others), family history of breast or ovarian cancer in a first-degree relative (none, any), reproductive history (gravida (0, 1, 2, 3 or ≥4), parity (0, 1, 2, 3 or ≥4), abortion (0, 1 or ≥2) and age at first pregnancy (≤19, 20–24, 25–29 or ≥30 years)), and lactation history as lifetime duration of breastfeeding (none, ≤6 months, 6–12 months or ≥12 months) were recorded. Pathologic information included grade (low (1), intermediate (2) or high (3)) according to the modified Black’s nuclear grading system, primary tumor stage (T1–T4), regional nodal stage (N0–N3), ER status (positive if ≥10%), and HER2 status (positive if FISH ratio of >2.0 or IHC 3+ and FISH unknown).

Statistical methods

Expression values of GABA\(\text{\textalpha}^{-}\) from the TxP and RT–PCR experiments underwent logarithmic transformation and standardization to a mean of zero and standard deviation of 1.0. Standardized GABA\(\text{\textalpha}^{-}\) expression values were ranked to demonstrate the threshold value that best separated higher from lower expression of GABA\(\text{\textalpha}^{-}\) in the TxP data and the RT–PCR data. These standardized expression values were used as a continuous variable for statistical comparisons with the clinical and pathologic variables in univariate analyses and in multivariate analyses using S-PLUS 6.1 for Windows software (Seattle, WA, USA). Linear regression models and ANOVA tests were used to determine the significance of the data, unless otherwise noted. Statistical significance was defined as \(P<0.05\).
GABA\(\pi\) gene expression (Table 1). This was significantly more frequent than in other breast cancer phenotypes (ER\(+\)/HER2\(−\) (2/116, 2%), ER\(−\)/HER2\(+\) (0/20, 0%), or ER\(+\)/HER2\(+\) (0/15, 0%), \(P<0.0001\)) (Table 1). Table 1 also shows the Wilcoxon rank sum \(P\) values for testing GABA\(\pi\) gene expression in ER\(−\)/HER2\(−\) phenotype versus all others. The GABA\(\pi\) gene expression level was significantly higher for the ER\(−\)/HER2\(−\) phenotype overall and separately for TxP and RT–PCR patients.

We were also able to demonstrate the association of GABA\(\pi\) gene expression with ER\(−\)/HER2\(−\) tumors by comparing the relative gene expression levels of GABA\(\pi\) (GABRP, probe set 205044), ER (ER-\(\alpha\) gene, ESR1, probe set 205225) and HER2 (c-erbB-2 gene, ERBB2, probe set 216836) in the 82 TxPs from Affymetrix U133A microarrays. The relative expression of GABA\(\pi\), ESR1 and ERBB2 genes in these 82 tumors appeared to be divergent, although some tumors have low expression of all three genes (Fig. 2).

For further investigation of whether high GABA\(\pi\) expression represents a particular lineage of differentiation, we classified the 82 transcriptional profiles into

<table>
<thead>
<tr>
<th></th>
<th>% Total</th>
<th>GABA(\pi) high (&gt;1.1)</th>
<th>Overall</th>
<th>% GABA(\pi) high</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TnP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER(−)/Her2(−)</td>
<td>26</td>
<td>13</td>
<td>21</td>
<td>62</td>
<td>(P&lt;0.001)</td>
</tr>
<tr>
<td>ER(−)/Her2(+)</td>
<td>17</td>
<td>0</td>
<td>14</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ER(+)/Her2(−)</td>
<td>45</td>
<td>0</td>
<td>37</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ER(+)/Her2(+)</td>
<td>12</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>13</td>
<td>82</td>
<td>16</td>
<td></td>
<td>(P&lt;0.001)</td>
</tr>
<tr>
<td>RT–PCR</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ER(−)/Her2(−)</td>
<td>26</td>
<td>18</td>
<td>31</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>ER(−)/Her2(+)</td>
<td>5</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ER(+)/Her2(−)</td>
<td>65</td>
<td>2</td>
<td>79</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>ER(+)/Her2(+)</td>
<td>4</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>20</td>
<td>121</td>
<td>17</td>
<td></td>
<td>(P&lt;0.001)</td>
</tr>
</tbody>
</table>

All samples (standardized TnP and RT–PCR)

<table>
<thead>
<tr>
<th></th>
<th>% Total</th>
<th>GABA(\pi) high (&gt;1.1)</th>
<th>Overall</th>
<th>% GABA(\pi) high</th>
<th>(P) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ER(−)/Her2(−)</td>
<td>26</td>
<td>31</td>
<td>52</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>ER(−)/Her2(+)</td>
<td>10</td>
<td>0</td>
<td>20</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>ER(+)/Her2(−)</td>
<td>57</td>
<td>2</td>
<td>116</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>ER(+)/Her2(+)</td>
<td>7</td>
<td>0</td>
<td>15</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Totals</td>
<td>33</td>
<td>203</td>
<td>16</td>
<td></td>
<td>(P&lt;0.001)</td>
</tr>
</tbody>
</table>
luminal, HER2+, normal-like and basal-like subsets, using the intrinsic gene set described by Perou et al. (2000b) and Sorlie et al. (2001, 2003). All 13 tumors with high GABA expression had basal-like genomic profile, and those constituted 60% (13/22) of that genomic group. The majority of basal-like tumors (18/22, 82%) had ER+/-HER2x phenotype, although three tumors had ER-/HER2+, and one tumor had ER+/HER2− phenotype. Conversely, the genomic classification of 22 ER-/HER2− breast cancers included 18 basal-like profiles (82%), three normal-like profiles and one HER2+ profile. Thirteen breast cancers had high GABA expression, and all of those tumors had ER-/HER2− phenotype and basal-like genomic profile. High GABA expression was identified in 72% (13/18) of basal-like tumors that were also ER-/HER2−. GABA gene expression alone accurately identified basal-like tumors with positive predictive value of 0.93, specificity of 0.98, negative predictive value of 0.87 and sensitivity of 0.59. The 82 transcriptional profiles were also used to identify genes that were strongly coexpressed with GABA. Fifteen genes had a Spearman rank correlation coefficient of ≥0.6 (Table 2).

The distribution of clinical and pathologic characteristics that were evaluated in the 203 women was as follows: median age at diagnosis, race (70% white, 11% black, 12% Hispanic and 7% other; n=203), family history of breast or ovarian cancer in a first-degree relative (83% none and 17% any; n=202), reproductive history (n=199: gravida (11% G0, 8% G1, 30% G2, 25% G3 and 28% ≥G4), parity (15% P0, 10% P1, 36% P2, 23% P3 and 17% ≥P4), abortion (68% AB0, 20% AB1 and 12% AB2), age at first pregnancy (12% G0, 22%, ≤19, 31% 20–24, 23% 25–29 and 11% ≥30; n=180), lactation history as lifetime duration of breastfeeding (52% none, 18% ≤6 months, 14% 6–12 months and 15% ≥12 months;

Table 2 Genes with expression most strongly correlated with GABA expression (Spearman rank correlation coefficient of ≥0.6) in the 82 TxPs are annotated by GenBank accession and gene symbol

<table>
<thead>
<tr>
<th>Correlation coefficient</th>
<th>Accession no.</th>
<th>Gene name</th>
<th>Gene symbol</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.771</td>
<td>NM_003012.2</td>
<td>Secreted frizzled-related protein 1</td>
<td>SFRP1</td>
</tr>
<tr>
<td>0.725</td>
<td>NM_002423.2</td>
<td>Matrix metalloproteinase 7</td>
<td>MMP7</td>
</tr>
<tr>
<td>0.698</td>
<td>NM_002888.1</td>
<td>Retinoic acid receptor responder 1</td>
<td>RARRES1</td>
</tr>
<tr>
<td>0.696</td>
<td>NM_000424.1</td>
<td>Keratin 5</td>
<td>KRT5</td>
</tr>
<tr>
<td>0.695</td>
<td>AF016004.1</td>
<td>Glycoprotein M6B</td>
<td>GPM6B</td>
</tr>
<tr>
<td>0.689</td>
<td>BC003610.1</td>
<td>Milk fat globule-EGF factor 8 protein</td>
<td>MFGE8</td>
</tr>
<tr>
<td>0.684</td>
<td>AI367319</td>
<td>SRY (sex-determining region Y)-box 10</td>
<td>SOX10</td>
</tr>
<tr>
<td>0.659</td>
<td>NM_006533.1</td>
<td>Melanoma inhibitory activity</td>
<td>MIA</td>
</tr>
<tr>
<td>0.637</td>
<td>NM_000222.1</td>
<td>v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog</td>
<td>KIT</td>
</tr>
<tr>
<td>0.628</td>
<td>AI3631452</td>
<td>Keratin 6B</td>
<td>KRT6B</td>
</tr>
<tr>
<td>0.626</td>
<td>NM_003986.1</td>
<td>Butyrobetaine (gamma), 2-oxoglutarate dioxygenase (gamma-butyrobetaine hydroxylase)</td>
<td>BBOX1</td>
</tr>
<tr>
<td>0.625</td>
<td>NM_000422</td>
<td>Keratin 17</td>
<td>KRT17</td>
</tr>
<tr>
<td>0.62</td>
<td>NM_002852.1</td>
<td>Pentaxin-related gene</td>
<td>PTX3</td>
</tr>
<tr>
<td>0.616</td>
<td>NM_017578</td>
<td>Ropporin, rhophilin-associated protein 1</td>
<td>ROP1N</td>
</tr>
<tr>
<td>0.612</td>
<td>NM_002639.1</td>
<td>Serine proteinase inhibitor, clade B (ovalbumin), member 5</td>
<td>SERPINB5</td>
</tr>
<tr>
<td>0.609</td>
<td>BC02690.1</td>
<td>Keratin 14</td>
<td>KRT14</td>
</tr>
<tr>
<td>0.6</td>
<td>NM_007231</td>
<td>Solute carrier family 6 (neurotransmitter transporter), member 14</td>
<td>SLC6A14</td>
</tr>
</tbody>
</table>
Linear regression analyses were used to compare the standardized GABA\(\pi\) gene expression values with each of the clinical and pathologic characteristics in univariate ANOVA analyses. GABA\(\pi\) gene expression level was positively associated with ER\(−/\)HER2\(−\) phenotype (\(P<0.0001\)), younger age at diagnosis (\(P=0.0021\)), and lifetime duration of breastfeeding of \(\leq 6\) months in all women (\(P=0.036\)), and specifically in parous women (\(P=0.012\)). The mean age of 33 patients whose breast cancer had high GABA\(\pi\) expression was 52.8 years, compared with 55.7 years in the other 170 patients. Higher grade had borderline positive association with GABA\(\pi\) gene expression level (\(P=0.063\)). There was no association with tumor stage, nodal stage, grade, age at diagnosis, race, number of pregnancies, births or abortions, age at first gestation, or family history of first-degree relative with breast or ovarian cancer.

Multivariate ANOVA analysis of the standardized GABA\(\pi\) gene expression values included the following variables as main effects: phenotype (ER\(−/\)HER2\(−\), all others), tumor stage, nodal stage, grade, age at diagnosis, race, number of pregnancies, births, abortions, age at first gestation, family history of first-degree relative with breast or ovarian cancer, and duration of breast feeding. In this multivariate model (R-square = 0.52), positive association was observed between GABA\(\pi\) expression level and ER\(−/\)HER2\(−\) phenotype (\(P<0.0001\)), younger age at diagnosis (\(P=0.0003\)) and lifetime breastfeeding of \(\leq 6\) months (\(P=0.0173\)), including patients who were nulliparous, never breastfed, or reported a lifetime history of breastfeeding of \(\leq 6\) months. Higher grade had a negative association with GABA\(\pi\) expression (\(P=0.0058\)), but there was a significant interaction with ER\(−/\)HER2\(−\) phenotype (described below). Considering only the subset of parous women (\(P>0\)), we still found that GABA\(\pi\) expression was associated with lifetime breastfeeding of \(\leq 6\) months in this multivariate main effects model (\(P=0.013\)).

The finding of high GABA\(\pi\) gene expression in approximately 60% of ER\(−/\)HER2\(−\) invasive breast cancers from 82 Txs (Affymetrix U133A microarrays) was independently confirmed in an independent cohort of 121 samples by RT–PCR (Fig. 1 and Table 1). Expression of GABA\(\pi\), ESR1 (ER) and ERBB2 (HER2) genes in the Txs appears to be mutually exclusive (Fig. 2), suggesting that elevated GABA\(\pi\) expression represents a divergent lineage of differentiation from ER and HER2, or that it represents a relatively undifferentiated cell type. There was an association between GABA\(\pi\) gene expression and the least differentiated basal-like genomic profile (Perou et al. 2000a, Sorlie et al. 2001, 2003), and the genes that were coexpressed with GABA\(\pi\) also indicate an immature cell type.

Five genes that were strongly coexpressed with GABA\(\pi\) are known to distinguish the basal-like subset of breast cancers: KRT5, KRT6, KRT17, KIT and MFGE8 (Table 2). KRT5, KRT6 and KRT14 encode cytokeratins 5, 6 and 14, which are expressed in immature breast epithelial cells, although there is controversy as to whether expression of these cytokeratins is restricted to stem cells or includes other immature
progenitor cells (Bocker et al. 2002, Abd El-Rehim et al. 2004, Clarke et al. 2004). KIT is a recognized stem cell factor receptor (Baghestanian et al. 2002). MFGE8 (milk fat globulin-EGF factor 8 protein) (lactadherin) is highly expressed during lactation and plays a role in the attachment of basal cells to the basement membrane through αvβ3 integrins (Taylor et al. 1997). Other genes coexpressed with GABAP (Table 2) are not included in the basal-like profile, but are expressed in immature epithelial and/or neuroepithelial cells. SFRP1 enhances cellular survival and is a secreted inhibitor of Wnt-induced signaling (Han & Amar 2004). SFRP1 is expressed in specific epithelial cells during embryogenesis, including the lens and salivary gland (Leimeister et al. 1998, Chen et al. 2004).

Although SFRP1 expression was lost in approximately 80% of breast cancers (described as ER-positive), expression was seen in a minority of breast cancers that were high-grade, ER-negative and described as having medullary histologic appearance (Ugolini et al. 2001). GPM6B (glycoprotein M6B) is expressed early in embryonic development of the central nervous system (Narayanan et al. 1998). SOX-10 is a transcription factor that is associated with neural precursor cells in the peripheral nervous system, including the enteric nervous system (Suarez-Rodriguez & Belkind-Gerson 2004). MMP7 is a matrix metalloproteinase enzyme that contributes to embryogenesis and to tumor invasion by aiding in the degradation of the extracellular matrix (Hulboy et al. 2004, Kitoh et al. 2004). Coexpression of these genes with GABAP illustrates the immature cell type of this subset of breast cancers.

There is a recognized overlap between inherited BRCA-1-associated breast cancer and tumors with the basal-like genomic profile. A recent study of Ashkenazi Jewish women from Montreal employed immunohistochemical (IHC) expression of CK 5/6 (KRT5 and KRT6) to define basal-like phenotype in 38% of 247 breast cancers (Foulkes et al. 2004). CK 5/6 expression was identified in 74% of their breast cancers in women who carried germ-line mutation of BRCA-1 gene (11% of the population studied), compared with 34% of breast cancers in women who did not carry BRCA-1 gene mutation (Foulkes et al. 2004). The same authors previously reported that 88% of ER−/HER2− breast cancers in BRCA-1 mutation carriers are CK 5/6-positive (basal-like), compared with 45% of ER−/HER2− breast cancers in noncarriers (Foulkes et al. 2003). Although higher GABAP expression was associated with younger age at diagnosis in our linear regression analyses, we did not identify any association with family history of breast or ovarian cancer in a first-degree relative, and the mean age of 52.8 years in those with high GABAP expression is older than might be expected for familial breast cancer. The BRCA-1 mutation carrier status of our patients is not known, but we observed no overlap between the genes that were coexpressed with GABAP in our study (Table 2) and the genes that were reported to characterize BRCA-1 associated breast cancer (Hedenfalk et al. 2001). The undifferentiated tumors with GABAP expression are diagnosed at a slightly younger age, but are probably not inherited.

Variable abrogation of cellular differentiation from oncogenesis produces breast cancers with different phenotypic and genomic characteristics. Furthermore, the molecular potency of a specific oncogenic insult, combined with the stage in cellular differentiation when breast epithelial cells are most susceptible to that specific oncogenic insult, has a profound effect on the level of differentiation of the breast cancer that ensues. That is why there is overlap among high-grade breast cancer, the ER−/HER2− phenotype, the basal-like genomic profile, breast cancer in BRCA-1 mutation carriers, and aberrations of important molecules such as p53, cyclin E, p27 and myc. These undifferentiated forms of breast cancer arise from early abrogation of cellular differentiation due to potent oncogenic insults. In this context, it is possible that elevated GABAP gene expression is simply a biomarker for cellular immaturity, and not overexpression of a putative oncogene. However, the markedly elevated expression levels of GABAP in undifferentiated breast cancers offers some hope that this molecule may have functional relevance that could one day be exploited for diagnostic, preventive and/or therapeutic uses. If GABAP is a biomarker of progenitor breast cells and can be manipulated to induce differentiation, it could present a useful chemoprevention target for populations at risk of undifferentiated breast cancer.

Numerous studies have reported that breastfeeding decreases the risk of breast cancer, and that the lifetime duration of breastfeeding is inversely associated with risk (Yoo et al. 1992, Enger et al. 1997, Newcomb 1997, Lipworth et al. 2000, Purwanto et al. 2000, Collaborative Group 2002). A recent meta-analysis estimated a decrease in lifetime risk of breast cancer of 7% for every birth and 4% for every 12 months of breastfeeding (Collaborative Group 2002). However, there are only limited data linking lactation history with the phenotype of subsequent breast cancer. A history of previous breastfeeding was significantly associated with ER-positive breast cancer in one study of 148 patients, but was not associated with ER status in two other studies (Hildreth et al. 1983, Montgomery et al. 1985, McTiernan et al. 1986). Breastfeeding for
≥12 months was associated with decreased risk of developing HER2-positive breast cancer (gene amplification), but not HER2-negative breast cancer, in one study of 577 women (Huang et al. 2000).

The biologic reasons for the protective effect of prolonged breastfeeding are unknown, but proposed mechanisms include decreased blood and/or breast fluid estrogen levels, delayed re-establishment of ovulation, excretion of carcinogens in milk, higher pH in milk and extended terminal differentiation of breast epithelium (Lipworth et al. 2000). During pregnancy, there is proliferative expansion of progenitor cells that express cytokeratins 5, 6 and 14 within breast lobules (Bocker et al. 2002). These genes (KRT5, KRT6 and KRT14) are coexpressed with GABA\(\pi\) (Table 2). Breastfeeding leads to progressive loss of this expanded progenitor cell population through terminal differentiation of epithelial cells in the lobules (Smith et al. 1990, Russo & Russo 1994, Bocker et al. 2002). Therefore, it follows that full-term pregnancy followed by a short duration of lactation may lead to retention of progenitor cells within the breast lobules in some women (Russo & Russo 1994). Malignant transformation of retained progenitor cells would halt cellular differentiation at an early stage and lead to undifferentiated breast cancer. Our results tend to support this hypothesis, but larger and more detailed biologic and molecular epidemiologic studies are still required.

Hispanic ethnicity and the combined interaction of Hispanic ethnicity with high tumor grade were positively associated with GABA\(\pi\) gene expression. The incidence of breast cancer has historically been lower in Hispanic women in the USA, but it has risen in recent years more quickly in this ethnic group than in others (Hunter 2000). Other studies have also shown that Hispanic women present with breast cancer of higher stage of disease (Daly et al. 1985, Natarajan et al. 1985, Mandelblatt et al. 1991, Bentley et al. 1998, Hedeen & White 2001), and at a younger age than non-Hispanic white women (Elledge et al. 1994, Zaloznik 1997). Access to health care and socioeconomic status probably do not explain these observed differences (Zaloznik 1997, Bentley et al. 1998). The survival of Hispanic women with breast cancer is worse than for non-Hispanic white women, but better than for black women, when controlled for age, stage, histology and treatment (Boyer-Chammard et al. 1999). It has also been reported that breastfeeding is relatively protective against subsequent breast cancer in African-American women more than in Caucasian women, and in Caucasian women more than in Hispanic women (Mayberry & Stoddard-Wright 1992, Gilliland et al. 1998). However, prolonged breastfeeding was relatively protective against breast cancer in premenopausal Hispanic women (Gilliland et al. 1998). Undifferentiated breast cancer was independently associated with the combination of Hispanic ethnicity and high-grade invasive breast cancer in our multivariate analysis. This suggests that the lifetime duration of lactation, age at diagnosis or phenotype does not completely explain the association of GABA\(\pi\) expression with Hispanic ethnicity in our study.

There is no significant difference in the frequency of ER-negative status in Hispanic women compared with other groups (Elledge et al. 1994, Hedeen & White 2001). The frequency of HER2-positive and/or p53-positive invasive breast cancer is also similar in Hispanic and non-Hispanic white women (Hunter 2000). We note that Hispanic patients in our patient population include women who have always lived in Latin America, those who emigrated to the USA and those who were born in the USA. There may be differences in the epidemiology and pathology of breast cancers in these different groups of Hispanic patients (Hedeen & White 2001). We believe the observed association between GABA\(\pi\) gene expression and high-grade breast cancers in Hispanic women requires further study, but this does shed a different light on the molecular pathology of breast cancer in Hispanic women.

It is likely that elevated GABA\(\pi\) gene expression represents a relatively undifferentiated subset of ER−/HER2− breast cancers with high nuclear grade, presentation at a slightly younger age and association with shorter duration of lifetime breastfeeding. Shorter lifetime duration of breastfeeding may predispose to future development of undifferentiated and estrogen-independent breast cancer, possibly due to retention of progenitor cells and their failure to undergo terminal differentiation. Further study is needed to ascertain the biologic role of GABA\(\pi\) in the breast and to explore the potential therapeutic or preventive strategies for breast cancer.

**Funding**

This work was supported by institutional research funds (W F S), a research grant from Millennium Pharmaceuticals, Inc., Cambridge, MA (L P), and the Tobacco Settlement Fund as appropriated to the University of Texas M D Anderson Cancer Center by the Texas Legislature and by generous donations from the Kadoorie Foundation (Core Genomics Facility). The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.
References


Baghestanian M, Jordan LH, Kiener HP, Bevec D, Agis H, Fritsch G, Muller MR, Banki HC, Schrenthanner GH, Lechner K et al. 2002 Activation of human mast cells through stem cell factor receptor (KIT) is associated with expression of bcl-2. International Archives in Allergy and Immunology 129 228–236.


Han X & Amar S 2004 Secreted frizzled-related protein 1 (SFRP1) protects fibroblasts from ceramide-induced apoptosis. Journal of Biological Chemistry 279 2832–2840.

Harvey JM, Clark GM, Osborne CK & Allred DC 1999 Estrogen receptor status by immunohistochemistry is superior to the ligand-binding assay for predicting response to adjuvant endocrine therapy in breast cancer. Journal of Clinical Oncology 17 1474–1481.


Pusztai L, Ayres M, Stec J, Clark E, Hess K, Stivers D, Damokosh A, Sneige N, Buchholz TA, Esteva FJ *et al.* 2003 Gene expression profiles obtained from single passage fine needle aspirations (FNA) of breast cancer reliably identify prognostic/predictive markers such as estrogen (ER) and HER-2 receptor status and reveal large scale molecular differences between ER-negative and ER-positive tumors. *Clinical Cancer Research* **9** 2406–2415.


