Gene Expression Profiling Shows Medullary Breast Cancer Is a Subgroup of Basal Breast Cancers

François Bertucci,1,2,5 Pascal Finetti,1 Nathalie Cervera,1 Emmanuelle Charafe-Jauffret,1,3,5 Emilie Mamessier,1 José Adélaïde,1 Stéphane Debono,1 Gilles Houvenaeghel,1,3,5 Dominique Maraninchi,1,3 Patrice Viens,1,3 Colette Charpin,1,7 Jocelyne Jacquemier,1,3 and Daniel Birnbaum1

1Institut de Cancérologie de Marseille, Département d’Oncologie Moléculaire, Institut Paoli-Calmettes et Unité Mixte de Recherche 599
Institut National de la Santé et de la Recherche Médicale, Biopathologie, and ‘Chirurgie,
Institut Paoli-Calmettes, 1Faculté de Médecine, Université de la Méditerranée; 2Ipogen S.A.; and 3Département d’Anatomopathologie, Hôpital Nord, Marseilles, France

Abstract
Medullary breast cancer (MBC) is a rare but enigmatic pathologic type of breast cancer. Despite features of aggressiveness, MBC is associated with a favorable prognosis. Morphologic diagnosis remains difficult in many cases. Very little is known about the molecular alterations involved in MBC. Notably, it is not clear whether MBC and ductal breast cancer (DBC) represent molecularly distinct entities and what genes/proteins might account for their differences. Using whole-genome oligonucleotide microarrays, we compared gene expression profiles of 22 MBCs and 44 grade III DBCs. We show that MBCs are less heterogeneous than DBCs. Whereas different molecular subtypes (luminal A, luminal B, basal, ERBB2-overexpressing, and normal-like) exist in DBCs, 95% MBCs display a basal profile, similar to that of basal DBCs. Supervised analysis identified gene expression signatures that discriminated MBCs from DBCs. Discriminator genes are associated with various cellular processes related to MBC features, in particular immune reaction and apoptosis. As compared with MBCs, basal DBCs overexpress genes involved in smooth muscle cell differentiation, suggesting that MBCs are a distinct subgroup of basal breast cancer with limited myoepithelial differentiation. Finally, MBCs overexpress a series of genes located on the 12p13 and 6p21 chromosomal regions known to contain pluripotency genes. Our results contribute to a better understanding of MBC and of mammary oncogenesis in general. (Cancer Res 2006; 66(9): 4636-44)

Introduction
Medullary breast cancer (MBC) accounts for <2% of breast cancers. Diagnosis of MBC is based on five pathologic criteria established by Ridolfi et al. (1) almost three decades ago. The reproducibility (2) and clinical relevance of such diagnosis have been questioned and other criteria have been proposed (3) but Ridolfi’s criteria remain the most appropriate.
Table 1. Patients and tumor characteristics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Total (N = 66)</th>
<th>DBCs (n = 44)</th>
<th>MBCs (n = 22)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Median (range)</td>
<td>54 (31-82)</td>
<td>60 (35-82)</td>
<td>49 (31-78)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Pathologic tumor size (64)</td>
<td></td>
<td></td>
<td></td>
<td>0.182</td>
</tr>
<tr>
<td>pT1</td>
<td>18 (28%)</td>
<td>11 (26%)</td>
<td>7 (33%)</td>
<td></td>
</tr>
<tr>
<td>pT2</td>
<td>35 (55%)</td>
<td>22 (51%)</td>
<td>13 (62%)</td>
<td></td>
</tr>
<tr>
<td>pT3</td>
<td>11 (17%)</td>
<td>10 (24%)</td>
<td>1 (5%)</td>
<td></td>
</tr>
<tr>
<td>Pathologic axillary lymph node (64)</td>
<td></td>
<td></td>
<td></td>
<td>0.316</td>
</tr>
<tr>
<td>Negative</td>
<td>37 (58%)</td>
<td>23 (54%)</td>
<td>14 (67%)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>27 (42%)</td>
<td>20 (46%)</td>
<td>7 (33%)</td>
<td></td>
</tr>
<tr>
<td>Density of lymphocyte infiltrate (62)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Low</td>
<td>12 (19%)</td>
<td>12 (30%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>Moderate</td>
<td>24 (39%)</td>
<td>20 (50%)</td>
<td>4 (18%)</td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>26 (42%)</td>
<td>8 (20%)</td>
<td>18 (82%)</td>
<td></td>
</tr>
<tr>
<td>Angio-invasion (56)</td>
<td></td>
<td></td>
<td></td>
<td>0.019</td>
</tr>
<tr>
<td>Negative</td>
<td>40 (71%)</td>
<td>25 (63%)</td>
<td>15 (94%)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>16 (29%)</td>
<td>15 (37%)</td>
<td>1 (6%)</td>
<td></td>
</tr>
<tr>
<td>ER status (66)</td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Negative</td>
<td>44 (67%)</td>
<td>22 (50%)</td>
<td>22 (100%)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>22 (33%)</td>
<td>22 (50%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>PR status (66)</td>
<td></td>
<td></td>
<td></td>
<td>0.003</td>
</tr>
<tr>
<td>Negative</td>
<td>48 (73%)</td>
<td>27 (61%)</td>
<td>21 (96%)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>18 (27%)</td>
<td>17 (39%)</td>
<td>1 (4%)</td>
<td></td>
</tr>
<tr>
<td>ERBB2 status (59)</td>
<td></td>
<td></td>
<td></td>
<td>0.032</td>
</tr>
<tr>
<td>Negative</td>
<td>48 (81%)</td>
<td>33 (75%)</td>
<td>15 (100%)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>11 (19%)</td>
<td>11 (25%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>EGFR status (31)</td>
<td></td>
<td></td>
<td></td>
<td>0.199</td>
</tr>
<tr>
<td>Negative</td>
<td>17 (55%)</td>
<td>16 (59%)</td>
<td>1 (25%)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>14 (45%)</td>
<td>11 (41%)</td>
<td>3 (75%)</td>
<td></td>
</tr>
<tr>
<td>P53 status (55)</td>
<td></td>
<td></td>
<td></td>
<td>0.589</td>
</tr>
<tr>
<td>Negative</td>
<td>27 (49%)</td>
<td>21 (51%)</td>
<td>6 (43%)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>28 (51%)</td>
<td>20 (49%)</td>
<td>8 (57%)</td>
<td></td>
</tr>
<tr>
<td>Ki67 status (57)</td>
<td></td>
<td></td>
<td></td>
<td>0.062</td>
</tr>
<tr>
<td>Negative</td>
<td>9 (16%)</td>
<td>9 (21%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>Positive</td>
<td>48 (84%)</td>
<td>34 (79%)</td>
<td>14 (100%)</td>
<td></td>
</tr>
</tbody>
</table>

*Numbers in parentheses are numbers of evaluated cases among 66 tumors.

(positivity cutoff values of 20%). MBCs displayed characteristics similar to series in literature with 100% of samples SBR grade III, ER-negative, EBBB2-negative, Ki67-positive, and 57% P53-positive. Lymphocyte infiltrate was dense in 20 samples and moderate in two samples. DBCs and MBCs did not differ in the distribution of pathologic size or nodal status. After surgery, patients were treated using a multimodality approach according to standard guidelines.

A second series of 205 samples, which contained 73 SBR grade III breast cancer samples, including 6 MBCs and 67 DBCs, was profiled on Ipsogen microarrays for validation (Institut Paoli-Calmettes/Ipsogen data set; results are in the Supplementary Data).

Other samples. Eleven normal breast samples pooled in four RNA samples (NB0, NB1, NB2, and NB3, representing one sample from four women from Val d’Aurelle Hospital, and three commercial pools of, respectively, 1, 2, and 4 normal breast RNA; Clontech, Palo Alto, CA) were profiled. Eighteen cell lines represented various cell types: breast epithelium (BT-474, BT-483, HCC1500, HCC1954, HME-1, carcinosarcoma-derived Hs578T, MCF-7, MCF-10A, MDA-MB-134, MDA-MB-453, T47D, UACC-812, SUM-149, SUM-225, and HMEC-derived 184B5), fibroblasts (HFFB), B (Daudi) and T (Jurkatt) lymphocytes. All breast cell lines are derived from carcinomas except MCF-10A, HME-1, and 184B5. All cell lines, except SUM-149 and SUM-225, were obtained from American Type Culture Collection (Manassas, VA). They were grown as recommended by the supplier.

RNA extraction. Total RNA was extracted from frozen samples by using guanidium isothiocyanate and cesium chloride gradient, as previously described (11). Its integrity was controlled by Agilent analysis (Bioanalyzer, Palo Alto, CA).

Gene expression profiling with DNA microarrays. Gene expression analyses were done with Affymetrix U133 Plus 2.0 human oligonucleotide microarrays. Preparation of cRNA, hybridizations, washes, and detection were done as recommended by the supplier. For each sample, synthesis of the first-strand cDNA was done from 3 μg total RNA by T7-oligo(dT) priming, followed by second-strand cDNA synthesis. After purification, in vitro transcription associated with amplification generated cRNA-containing biotinylated pseudouridine. Biotinylated cRNA was purified, quantified and chemically fragmented (95°C for 35 minutes), then hybridized to microarrays in 200 μL hybridization buffer at 45°C for 16 hours. Automated washes and staining with streptavidin-phycoerythrin.

* http://www.atcc.org/.
were done as recommended. Double signal amplification was done by biotinylated antistreptavidin antibody with goat-IgG blocking antibody. Scanning was done with Affymetrix GeneArray scanner and quantification with Affymetrix G COS software. Validation study using cDNA-spotted arrays was done with Ipsogen Nylon microarrays10 containing ~8,000 genes/expressed sequence tags (EST) as previously described (12).

**Gene expression data analysis.** Data were analyzed by the Robust Multichip Average method in R using Biocductor and associated packages (13). Robust Multichip Average did background adjustment, quantile normalization, and summarization of 11 oligonucleotides per gene. Before analysis, a filtering process removed from the data set the genes with low and poorly measured expression as defined by an expression value inferior to 100 units in all 66 breast cancer tissue samples, retaining 27,243 genes/ESTs with expression values ranging from 3 to ~16,000 (mean, 277). Expression data are available in Supplementary Table S1.

Before unsupervised hierarchical clustering, a second filter, based on the intensity of SD, was applied to exclude genes showing low expression variation across the 66 samples. For genes with minimal expression value inferior to 100 (our threshold for background) in one sample, SD was calculated on values superior to background and a minimal value floored to 100 (because discrimination of expression variation in this low range cannot be done with confidence). Such filter allowed eliminating genes with low expression values and low expression variation, retaining 10,375 genes/ESTs. Data were then log-transformed and submitted to the Cluster program (14) using data median-centered on genes, Pearson correlation as similarity metric and centroid linkage clustering. Results were displayed using TreeView program (14).

To identify and rank genes discriminating two subgroups of samples, supervised analysis was applied to the 27,243 genes/ESTs. A discriminating score (DS) was calculated for each gene (15) as DS = (M₁ - M₂) / (S₁ + S₂), where M₁ and S₁, respectively, represent mean and SD of expression levels of the gene in subgroup 1, and M₂ and S₂ in subgroup 2. Confidence levels were estimated by 100 random permutations of samples as previously described (16). A “leave-one-out” (LOO) procedure (15) was applied to estimate the accuracy of prediction of the signatures and the validity of our supervised analysis. The lists of discriminator genes were interrogated by Onto-Express (17).

**Immunohistochemistry on breast cancer tissue microarrays.** Two tissue microarrays (TMA1 and TMA2) were prepared after careful selection of a representative tumor area for each sample by analysis of a H&E-stained section of a donor block (18). Core cylinders (diameter of 0.6 mm) were punched from this area and deposited into a paraffin block using an arraying device (Alphelys, Puisley, France). Five-micrometer sections of the resulting array block were made and transferred to glass slides before immunohistochemistry analysis. TMA1 contained 547 consecutive early breast cancers (19) treated at the Institut Paoli-Calmettes, including 17 grade III DBCs used for comparison with MBCs. TMA2 included 40 MBCs (25 from Institut Paoli-Calmettes and 15 from Hôpital Nord). In addition, TMA2 contained normal breast tissues (n = 10) and cell line pellets. Three mouse monoclonal antibodies were used for immunohistochemistry: anti-α-SMA (clone 1A4, 1/200 dilution, DakoCytomation, Glostrup, Denmark), antimoesin (clone 38/87, 1/400 dilution, Biomedia, Foster City, CA), and anti-GATA3 (clone 2C-268, 1/100 dilution, Santa Cruz Biotechnology, Santa Cruz, CA). Immunohistochemistry was done on 5-mm sections using DAKO LSAB 2 kit in a DAKO Autostainer (DakoCytomation, Glostrup, Denmark). The antibodies were incubated for 1 hour in citrate buffer. After staining, slides were evaluated by two pathologists. Moesin and α-SMA showed cytoplasmic localization and GATA3 nuclear staining. Intensity of α-SMA staining was scored from 0 (no staining) to 3 (strong and diffuse staining). For GATA3 and moesin, the results were estimated by the percentage of positive cells (P, from 0% to 100%) and the intensity of the staining (I, from 0 to 3), and expressed by the Quick score (Q = P × I, from 0 to 300; ref. 18).

**Statistical analysis.** Correlations between sample groups and histoclinical variables were calculated with the Fisher’s exact test or χ² test when appropriate. Follow-up was measured from the date of diagnosis to the date of last news for live patients. Metastasis-free survival was calculated from the date of diagnosis until date of first distant metastasis using the Kaplan-Meier method and compared between groups with the log-rank test. All tests were two-sided at the 5% level of significance. Analysis was done using the SPSS software (version 10.0.5).

**Results**

**Global gene expression profiles of DBCs and MBCs.** A total of 88 samples representing 66 SBR grade III breast cancer tissue samples (22 MBCs and 44 DBCs), 4 normal breast tissue samples, and 18 cell lines were profiled using whole-genome DNA microarrays. Hierarchical clustering was applied to the 10,375 genes/ESTs with significant variation in expression level across the cancer tissue samples (Fig. 1).

As reflected by the dendrogram, the cancer tissue samples displayed heterogeneous expression profiles (Fig. 1A-B), and were sorted into four large groups I, II, III, and IV. Classification in three major groups (group IV, which included only one DBC sample and four normal breast samples, was excluded from analysis) correlated with the pathologic type, the lymphocyte infiltrate and the immunohistochemistry status of samples (χ² test). Two of seven (29%) samples in group I, 20 of 38 (53%) in group II, and 0 of 20 (0%) in group III were MBCs (P < 0.001). Twenty of 22 MBCs clustered in group II with DBCs and two in group I. Lymphocyte infiltrate was dense in 58% of samples in group II, 50% in group I, and 10% in group III (P = 0.002). Fourteen percent of samples were ER-positive in group I, 7% in group II, and 90% in group III (P < 0.001). The same was found for PR status with 0%, 5%, and 80%, respectively (P < 0.001). ERBB2 was positive in 57% of samples in group I, 10% in group II, and 15% in group III (P = 0.01), whereas EGFR was positive in, respectively, 50%, 79%, and 8% (P = 0.001) of samples. Ki67/MIB1 was positive in 100% of samples in group I, 97% in group II, and 60% in group III (P = 0.001). Finally, P53 status was positive in 86%, 59%, and 28% of cases (P = 0.02). Classification of samples by using other unsupervised clustering tools gave similar results with >95% of concordance.

Several clusters of related genes were evidenced. Consistent with previous studies, some of them defined expression signatures corresponding to cell types, pathways, or chromosomal locations (see colored bars to the right of Fig. 1A and zooms in Fig. 1B; genes of these clusters are listed in Supplementary Table S2). As expected, the cluster with a prominent role in the classification of samples was the luminal/ER cluster (412 genes, including ESR1 and PGR, which code for ER and PR, respectively). Variation in expression of ESR1 and PGR mRNA correlated with ER and PR immunohistochemistry status of samples. The basal cluster (111 genes) included genes more specific of basal/myoepithelial mammary cells, such as cytokeratins (KRT7, KRT8, KRT14, KRT15, and KRT17), and EGFR. Expression of this cluster correlated with the EGFR immunohistochemistry status of tumors. The stromal (203 genes) and immune cluster (720 genes) reflected variation in specific cell types (stromal cells, B and T lymphocytes). The latter correlated with the density of lymphocyte infiltrate. Other clusters represented activities of specific signaling and/or regulatory pathways. The early response cluster (18 genes) was overexpressed in normal breast samples overall compared with tumors. The proliferation cluster (276 genes) included the two proliferation markers MKI67 and proliferating cell nuclear antigen, as well as many genes involved in cell cycle and mitosis; its expression correlated with the Ki67 immunohistochemistry status of samples.

The ribosomal/metabolism cluster (411 genes) included many genes encoding ribosomal proteins and cytochrome c oxidase subunits. Other clusters of coexpressed genes represented presumptive amplicons: 11q13 (61 genes), 20q13 (62 genes), 8p11 (23 genes), 8q12-24 (91 genes), 12p13 (43 genes), \textit{ERBB2}, and 17q amplicons (64 genes and 42 genes). Variation in expression of the \textit{ERBB2} cluster correlated with the \textit{ERBB2} immunohistochemistry status of samples. Some of the gene clusters were globally differentially expressed in the four sample groups. Group I showed the highest expression of \textit{ERBB2} and 17q amplicons, immune, and stromal clusters. Group II exhibited the highest expression of basal, proliferation, and 12p13 clusters. In this group, the relative high expression of stromal and immune clusters concerned mainly the DBCs and MBCs, respectively, in agreement with the higher lymphoid stroma in MBCs than in DBCs. In groups I and II, the luminal cluster and the 20q13, 11q13, 8p11, and 8q12-24 clusters were underexpressed overall compared with group III. Group III displayed relative low expression of the basal, immune, and proliferation clusters. Group IV strongly expressed overall the early response and the ribosomal/metabolism clusters, as well as the basal cluster. Many of these differential expressions were in agreement with the phenotypical characteristics of DBCs and MBCs. For example, the relative low expression of the immune and proliferation clusters in group III was in agreement with, respectively, the less abundant lymphoid stroma and the lower proliferation index of ER-positive tumors (20).

**Identification of molecular subtypes in MBC.** Tumor subtypes (luminal A, luminal B, basal, \textit{ERBB2}-overexpressing, and normal-like) have been recently identified using an intrinsic set of \( \sim 500 \) genes in DBC (8, 9), including inflammatory breast cancer (10). We looked whether these subtypes were present in our samples by
using the 476 genes common to the intrinsic 500-gene set and our 27,243 filtered genes/ESTs.

Hierarchical clustering of the available expression data for these 476 genes in the 122 samples from Sorlie et al. (9) discriminated the same five molecular subtypes (Supplementary Fig. S1), allowing the definition of typical expression profile of each subtype (thereafter designated centroid). The core samples of each subtype are color-coded in the dendrogram: They included 31 samples as luminal A, 10 luminal B, 14 ERBB2, 19 basal, and 10 normal-like (correlation superior to 0.33, 0.29, 0.39, 0.44 and 0.33 respectively), with 96% of concordance with the centroids defined by Sorlie et al.

The centroid expression for each subtype was computed as the median expression for each of the 476 genes in the corresponding core samples. We then measured the correlation of each of our 70 tissue samples with each centroid (Fig. 2): 16 DBCs were closer to the luminal A centroid; 2 DBCs were closer to the luminal B centroid; 37 samples, including 16 DBCs and 21 MBCs, were closer to the basal centroid; 6 DBCs were closer to the ERBB2 centroid; and 6 samples, including 1 DBC, 1 MBC, and all 4 normal breast samples, were closer to the normal-like centroid. Three DBCs displayed a correlation inferior to 0.15 with any centroid and were not attributed any subtype. Thus, 21 of 22 MBCs (95%) were basal.

To estimate the robustness of this result, we used alternative analyses and gene sets. First, global clustering based on 10,375 genes/ESTs (Fig. 2A) sorted 20 MBCs in group II, which also included all 16 basal DBCs, whereas the two other MBCs were in group I close to group II. Second, to test the hypothesis that the signatures of the different subtypes were equivalent in DBCs and MBCs, we applied supervised analysis based on the 27,244 genes/ESTs. The DBC

Figure 2. Molecular subtypes of MBCs and DBCs. A, top, dendrogram of tissue samples, pathologic type, and molecular subtype as depicted in Fig. 1B. Bottom, correlation of expression of the 70 tissue samples to five subtype centroids. Each subtype centroid represents the median expression profile of Sorlie et al.'s samples (9) from the corresponding subtype based on 476 genes common to the Stanford/Norway intrinsic 500-gene set and our 27,243 genes/ESTs. Correlation coefficients are plotted by colors indicating the centroid: dark blue for luminal A, light blue for luminal B, red for basal, pink for ERBB2+ and green for normal-like. The horizontal orange line represents the threshold for coefficient correlation (0.15). MBCs are assigned basal for 21 of 22 samples, and normal for one. B, top, hierarchical clustering based on expression of the 1,496 genes identified in the 3-subtype supervised analysis (luminal A, basal and ERBB2+) of 38 DBCs and applied to the 38 DBCs (left) and to the 22 MBCs (right). Expression levels are depicted according to the color scale used in Fig. 1. Branches of the dendrograms are color-coded according to the centroid-based subtype defined in (A). Vertical colored bars, location of the basal (red bar), luminal A (dark blue bar), and ERBB2 (pink bar) gene clusters. Bottom, under each data matrix, correlation coefficient of each sample to the median profile of the luminal DBCs (dark blue curve), of the basal DBCs (red curve), and of the ERBB2+ DBCs (pink curve).
samples were used as learning set to develop a molecular signature discriminating three centroid-based subtypes: luminal A (16 samples), basal (16 samples), and ERBB2-overexpressing (6 samples). Luminal B and normal breast-like subtypes were excluded from analysis because of the low number of samples. Using a DS and permutation tests, we identified 1,496 discriminator genes. The resulting classification of DBCs was in strong agreement with the centroid-based subtype (Fig. 2B, top, left): all luminal A samples were in the left group and all basal samples were in the right group. The ERBB2-overexpressing samples, located in the two groups, were more dispersed, representing a more heterogeneous subtype. Similar clustering applied to the 22 MBC samples (Fig. 2B, top, right) confirmed the relatively homogeneous basal profile of MBCs.

**Identification of a gene expression signature for MBC within basal breast cancers.** Supervised analysis searched for a gene expression signature (GES) that would discriminate between the 21 basal MBCs and 16 basal DBCs. The chosen significance threshold for DS ensured that the number of genes selected by chance, given 100 iterative random permutations, never exceeded the number of identified discriminator genes. A second analysis compared all MBCs and DBCs (see Supplementary Data for Results and Discussion of this "global" GES).

We identified 534 genes (534-GES) as discriminator (theoretical number of produced false positives is 54) between basal MBCs and basal DBCs, with 269 genes overexpressed and 265 underexpressed in MBCs. They represented 426 different sequences, corresponding to 365 characterized genes and 61 ESTs (Supplementary Table S3). “Immune response” (GO:0006955; 17 genes, P < 0.001) was the Onto-Express biological process the most represented in MBCs versus basal DBCs. Other significant processes included "apoptosis" (GO:0006915; 10 genes, P < 0.001), “induction of apoptosis” (GO:0006917; 7 genes, P < 0.001), “proteolysis and peptidolysis” (GO:0006508; 14 genes, P < 0.001), “ubiquitin-dependent protein catabolism” (GO:0006511; 6 genes, P < 0.01), “cell proliferation” (GO:0008283, 5 genes, P < 0.02). Conversely, five Onto-Express biological profiles were strongly represented (P < 0.01) in basal DBCs versus MBCs: “muscle development” (GO:0007517; 11 genes), “cell adhesion” (GO:0007155; 14 genes), “smooth muscle contraction” (GO:0006939; 4 genes), “transmembrane receptor tyrosine kinase signaling pathway” (GO:0007517; 5 genes), and “actin cytoskeleton organization and biogenesis” (GO:0007155; 11 genes, P < 0.001). The analysis revealed overrepresentation of genes located at 12p13 among genes overexpressed in MBCs (P < 10-13, Fisher’s exact test). Seven of 24 unique 12p13 genes were on the p13.31 band. The 6p21.3 band was also strongly represented within genes up-regulated in MBCs (8 genes). The classification power of this GES is illustrated in Fig. 3A. With a threshold of 0 (orange line in Fig. 3A), the two classes defined by the signature (“predicted MBC class,” positive scores; “predicted basal DBC class,” negative scores) correlated with the pathologic type: 19 of 21 MBCs classified in the predicted MBC class, and all but one basal DBCs in the predicted basal DBC class (P < 0.001, Fisher’s exact test). Interestingly, a blinded histologic reevaluation of two misclassified samples (one of the two MBCs and one DBC) suggested the diagnosis of atypical MBC for the MBC (with an incomplete syncytial aspect inferior to 50% of tumor and a moderate lymphocyte infiltrate) and revealed an important syncytial aspect for the DBC. By LOO cross-validation, 70% of samples were correctly assigned by the predictors, and on average 86% of the genes of our signature were conserved.

**Figure 3.** Supervised classification of basal breast cancer samples based on the MBC/DBC molecular signature. A, classification of 37 samples using the 534 genes identified as discriminator between the 21 basal MBCs and the 16 basal DBCs. Top, expression data and pathologic type as in Fig. 1. Genes are ordered from top to bottom by their decreasing DS (the position of ACTG2/SMA is indicated). Tumor samples are numbered from 1 to 37 and are ordered from left to right according to the decreasing correlation coefficient of their expression profile with the median profile of the MBCs (bottom). Solid orange line, threshold 0 that separates the two classes of samples, predicted MBC class (left of the line) and predicted DBC class (right to the line). B, examples of immunohistochemical staining for α-SMA (stromal myofibroblastic cells: double arrows) are shown in DBC and MBC samples (magnification, ×100). α-SMA staining is positive in the basal DBC sample and low in the basal MBC sample.
As a validation study of this GES, we measured by immunohistochemistry on TMA the protein expression of ACTG2/α-SMA in 30 samples defined by gene profiling as basal DBCs (n = 15) and basal MBCs (n = 15; Fig. 3B). Consistent with RNA data, the α-SMA protein was overexpressed in basal DBCs compared with basal MBCs (P = 0.003, Fisher’s exact test).

Discussion

MBC is a fascinating but controversial entity. A better molecular characterization may not only provide diagnosis markers but also contribute to a better understanding of this ill-defined disease. Using DNA microarrays, we monitored the mRNA expression levels of ~47,000 transcripts and variants in 22 MBCs and 44 DBCs. To avoid the detection of expression differences related to grade, all profiled DBCs were grade III SBR. Similarly and with respect to ER status, half of them were ER-negative.

Global expression profile of MBC and molecular basal subtype. Global clustering revealed that MBC is less heterogeneous than DBC. All MBCs clustered in two large neighbor groups, one of which included 91% of cases. As expected, MBCs displayed more similarities with ER-negative DBCs than with ER-positive DBCs. As compared with ER-positive DBCs, MBCs overexpressed the basal, immune, proliferation, and 12p13 gene clusters; and underexpressed the luminal, 11q13, 20q13, and 8p11 clusters. These results show that MBCs are related to the basal subtype much more frequently than DBCs. This was confirmed when we investigated, using the Stanford/Norwegian intrinsic gene set, whether the five molecular subtypes previously described in DBCs (8, 9) were also present in MBCs; 95% of MBCs had a basal profile. The use of alternative gene sets and methods showed the robustness and reliability of this taxonomy. Notably, a three-class supervised analysis identified a gene signature that classified MBC as basal. The same was true when we applied a gene signature defined on luminal versus basal breast cell lines (21). Altogether, these results showed that MBCs are related to molecular features of basal mammary epithelial cell lineage, and that basal DBCs and MBCs have similar global gene expression profiles. A similar degree of basal/myoepithelial differentiation has recently been shown at the protein level (22, 23) and agrees with the high proportion of BRCA1-mutated cases within MBCs (5) and the frequent basal phenotype of BRCA1-mutated tumors (9, 24).

Identification of a GES for MBC within basal breast cancers. To reveal differences not related to the molecular subtype, we focused a supervised analysis on basal samples and identified 534 genes differentially expressed between basal MBCs and basal DBCs.

Genes overexpressed in basal MBCs. Among 269 overexpressed genes, immune response was the most represented Onto-lexmap category. Tumor-infiltrating lymphocytes (TIL) are mainly composed of immature cell or undergo a block of differentiation at an early stage. Several up-regulated genes are involved in antigen processing and presentation: ART51, HCP5, BFX5, HLA-DOB; and HLA-F, HLA-C, HLA-DOA, HLA-DQB1, and HLA-DRB4 (fold change from 1.4 to 2.5), allowing increased interaction with TIL and cell targeting from cytotoxic T cells. ART51 encodes endoplasmic reticulum and serves as MHC class I epitopes (28). Several up-regulated genes code for proteins involved in the degradation of intracellular proteins followed by antigen loading on MHC class I molecule: proteasome proteins PSMB8 (29), PSMB10 and PSMA5, ATP-binding cassette transporters TAP1 and TAP2 (fold change = 1.7), TAP-binding protein TAPBP (fold change = 1.5; ref. 30), and TAP binding protein-like TAPBPL.

Genes underexpressed in basal MBCs. We identified 265 genes as underexpressed in basal MBCs. Several underexpressed genes are involved in the architecture and remodeling of cytoskeleton. They encode actins (ACTG2, ACTA2), α-actinin (ACTN1), myosin light chain (MYL9), β-tropomyosin (TMPO), and several regulators or associated proteins. Examples of myosin regulators include myosin light chain kinase (MYLK), myosin phosphatase-RHO interacting protein (M-RIP), and caldesmon (CALD1). Actin-associated proteins include filamins (FLNA, FLNC), pleckstrin homology domain containing, family C member 1 (PLEKHC1), smoothelin (SMTN), colocalized with α-smooth muscle actin (α-SMA/ACTG2) on stress fibers. Other regulators of cytoskeleton are MTTS1, diseased-associated activator of morphogenesis 1 (DAAM1), calponin 2 (CNN2), α-parvin (PARVA), skeletal muscle LIM protein 1 (FHL1), ADAM12, and transgelin (TAGLN).

Many of the genes overexpressed in basal DBCs code for smooth muscle-specific proteins (ACTG2/α-SMA, ACTA2, TPM2, MYL9, M-RIP, CALD1, CNN2, SMTN, KCNMB1, TAGLN, ACTN1, APEG1, and BOC). Several explanations, not mutually exclusive, may be proposed. First, basal cancers have a degree of smooth muscle differentiation and include myoepithelial cancers. Within the basal subtype, MBCs may represent a particular subgroup that has lost or never acquired this differentiation (underexpression of myoepithelial markers, such as ACTG2, CNN2, and, to a lesser extent, MYH11, CNN1, and TP73L/P63). MBCs could arise from a more mature cell or undergo a block of differentiation at an early
stage. Second, the stroma is richer in myofibroblasts in basal DBCs than in MBCs. Consistent with this hypothesis, we found a significantly stronger α-SMA staining in the basal DBC samples. According to this hypothesis, MBCs and basal DBCs differ mainly by the type of stroma, made of myofibroblasts in basal DBCs and of immune cells in MBCs. Possible sources of myofibroblasts in basal DBCs might be recruitment of myofibroblasts or epithelial-to-mesenchyme transition. Interestingly, some of the above-cited genes, such as TAGLN, ACTG2, FHL2 (31) and TPM2, ACTN1, CNN2 (32) are up-regulated by transforming growth factor β (TGF-β). Other genes associated to or induced by TGF-β were also overexpressed in DBCs: FSTL1 (33), BGN (34), and, to a lesser extent, CTGF, TGFBRS, and TGFBR2 (fold change from 0.5 to 0.6). TGF-β promotes metastasis at late stages of cancer (35). It stimulates the formation of reactive stroma, induces myofibroblastic differentiation with dense cytoskeletal fibers, extracellular matrix remodeling, and angiogenesis.

Several genes encoding receptors involved in cell invasiveness were underexpressed in MBCs, such as DDR2, EPHA3, endothelin receptor EDNRA, and its ligand EDN2. Similarly, genes involved in cell adhesion were down-regulated. They code for ITGB5 (36), SPOCK, a member of a Ca²⁺-binding proteoglycan family, CX3CR1, δ-catenin (CTNND2), α-2-laminin (LAMA2), thrombospondin 4 (THBS4), podocadherin 18 (PCDH18), and a semaphorin (SEMA5A).

**MBC and 12p13 chromosomal region.** We found a highly significant overrepresentation of genes located in 12p13 among the genes overexpressed in MBCs. The 12p13 cluster was overexpressed in group II tumors, which exhibited the highest expression of the basal cluster, suggesting some relationship between basal subtype and overexpression of 12p13 genes. The 12p chromosomal location is a hotspot for structural chromosomal changes associated with germ cell tumors (37). Genes from 12p13 were even more overrepresented in MBCs versus basal DBCs. We could think of several explanations. First, one or several oncogenes are present in a potential MBC-specific 12p13 amplification. Second, several genes involved in stem cell biology (NANOG, GDF3, STELLA, DPPA3, CD9, EDR1; ref. 38), which all map to 12p13.31, are involved in MBC. These particular genes were not included in our analysis or the GES, but some 12p13 genes of the GES might be additional pluripotency genes. Interestingly, a recent study (39) revealed overrepresentation of genes from the 12p12.2-12p13.33 region among genes overexpressed in embryonic carcinoma cell lines. A gene of the GES that might be both an oncogene and expressed in progenitor cells is ETV6 (40). Genes from the 6p21.3 band were also overrepresented in the MBC signature. This band contains the OCFT4/POU5F1 transcription factor gene, which is a master pluripotency gene and whose expression correlates with that of above-cited 12p13 genes (38).

In conclusion, our study shows that MBC belongs to the basal molecular subtype. Basal subtype can be subdivided in at least two subgroups: MBCs and basal DBCs. Our expression profiles not only reinforce the different hypotheses put forward to explain the biological basis for the more favorable prognosis of MBCs (effective host immune response, enhanced tumor cell apoptosis, elevated levels of metastasis-inhibiting factors and low levels of metastasis-promoting factors), but also provide new insights in the underlying molecular mechanisms. TGF-β infiltration, and antigenic presentation suggest the existence of an antigen-dependent reaction (41), which could be directed against endogenous or exogenous molecules, perhaps viral proteins. Conversely, our data suggest some degree of epithelial-to-mesenchymal transition in basal DBCs with more developed cell migration system. These data agree with the classic prognostic difference between MBC and DBC, which we confirmed in our series. With a median follow-up of 41 months after diagnosis (range, 4-132) and a similar treatment, 27% of 16 basal DBC patients displayed metastatic relapse versus only 5% of 21 MBC patients (P = 0.05, Fisher’s exact test). Finally, our study points to 12p13 as an important region in basal breast oncogenesis and provides lists of molecules that could be used as markers or targets in the management of breast cancer.

**Acknowledgments**

Received 1/4/2006; revised 3/2/2006; accepted 3/3/2006.

**Grant support:** Institut Pauli-Calmettes, Institut National de la Santé et de la Recherche Médicale, Université de la Méditerranée, and grants from the Ministries of Health and Research (Cancéropôle) and Ligue Nationale contre le Cancer (label DB).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank L. Xerri, F. Birg, and C. Mawas for encouragement and C. Theillet (EMI229 INSERM, University of Montpellier I, Montpellier, France) for the gift of normal breast tissue and S.P. Ethier (Department of Biology, Wayne State University, Detroit, MI) for the gift of SUM cell lines.

**References**


Gene Expression Profiling Shows Medullary Breast Cancer Is a Subgroup of Basal Breast Cancers

François Bertucci, Pascal Finetti, Nathalie Cervera, et al.


Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/66/9/4636

Supplementary Material
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2006/05/12/66.9.4636.DC1

Cited articles
This article cites 41 articles, 20 of which you can access for free at:
http://cancerres.aacrjournals.org/content/66/9/4636.full.html#ref-list-1

Citing articles
This article has been cited by 29 HighWire-hosted articles. Access the articles at:
/content/66/9/4636.full.html#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.