Molecular classification of familial non-BRCA1/BRCA2 breast cancer

Ingrid Hedenfalk*†, Markus Ringnér*§, Amir Ben-Dor†, Zohar Yakhini†, Yidong Chen*, Gunilla Chebil*, Robert Ach†, Niklas Loman†, Håkan Olsson†, Paul Meltzer*, Åke Borg†, and Jeffrey Trent***††

* Cancer Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD 20892; †Department of Oncology, Lund University Hospital, SE-221 00 Lund, Sweden; ¶Agilent Laboratories, Palo Alto, CA 94303; ‡Department of Pathology, Helsingborg Hospital, SE-251 87 Helsingborg, Sweden; and **Translational Genomics Research Institute, Phoenix, AZ 85004

Communicated by Francis S. Collins, National Institutes of Health, Bethesda, MD, June 26, 2002 (received for review April 17, 2002)

In the decade since their discovery, the two major breast cancer susceptibility genes BRCA1 and BRCA2, have been shown conclusively to be involved in a significant fraction of families segregating breast and ovarian cancer. However, it has become equally clear that a large proportion of families segregating breast cancer alone are not caused by mutations in BRCA1 or BRCA2. Unfortunately, despite intensive effort, the identification of additional breast cancer predisposition genes has so far been unsuccessful, presumably because of genetic heterogeneity, low penetrance, or recessive/polygenic mechanisms. These non-BRCA1/2 breast cancer families (termed BRCAx families) comprise a histopathologically heterogeneous group, further supporting their origin from multiple genetic events. Accordingly, the identification of a method to successfully divide BRCAx families into recognizable groups could be of considerable value to further genetic analysis. We have previously shown that global gene expression analysis can identify unique and distinct expression profiles in breast tumors from BRCA1 and BRCA2 mutation carriers. Here we show that gene expression profiling can discover novel classes among BRCAx tumors, and differentiate them from BRCA1 and BRCA2 tumors. Moreover, microarray-based comparative genomic hybridization (CGH) to cDNA arrays revealed specific somatic genetic alterations within the BRCAx subgroups. These findings illustrate that, when gene expression-based classifications are used, BRCAx families can be grouped into homogeneous subsets, thereby potentially increasing the power of conventional genetic analysis.

Although germ-line mutations in BRCA1 and BRCA2 account for most familial breast-ovarian cancer cases, these mutations can only explain a small proportion of familial site-specific breast cancer susceptibility (1). Although a small proportion of non-BRCA1/2 hereditary breast cancers evolve in individuals with rare multicancer syndromes, very little is known about the genetic basis of non-BRCA1/BRCA2 breast cancer (BRCAx) families. Genetic linkage analysis of BRCAx families has been performed and identified several chromosomal regions potentially harboring a breast cancer susceptibility gene, including 8p12-p22 (2), 13q21 (3), and 2q31-q33 (P. Huusko, personal communication). However, these loci have either subsequently been excluded as major predisposing loci on a global perspective (4, 5), or remain to be confirmed, emphasizing genetic heterogeneity and population-specific effects within BRCAx kindreds (6). Other difficulties of linkage detection are a high rate of negative estrogen (ER) and progesterone (PgR) receptors within BRCA1 tumors (10), and positive hormone receptors in BRCA2 tumors. BRCAx tumors display variable levels of these receptors (11). Thus, it is likely that the different tumor phenotypes reflect the multiple genetic origins of BRCAx cancers.

Materials and Methods

Breast Cancer Specimens. Tumors were obtained from pathology departments within the southern Sweden health care region. The grossly dissected tumors were snap frozen in liquid nitrogen within 30–60 min of surgical excision and stored at −80°C until further processed. Touch imprints were made to confirm the presence of neoplastic cells in each tumor specimen. Patients diagnosed with breast cancer and with a family disease history compatible with a dominant mode of inheritance were referred to the Oncogenetic Clinic at the Department of Oncology, Lund University Hospital (Lund, Sweden) for genetic counseling or visited a research clinic for familial breast cancer. The patients provided blood samples, and mutation analysis for BRCA1 and BRCA2 was performed as described (12). Individuals with no mutations identified in these genes (non-BRCA1/2) are referred to as BRCAx. Eight families were included in the present study, and two or three tumors from each family were analyzed, with the exception of Lund 5, from which only one tumor was available for analysis. Tumors were from different individuals, except families Lund 111 (bilateral breast cancer, of which only one was included in the expression analysis) and Lund 502 (primary breast cancer and metastasis). The pathological review (Table 1) was performed by a single pathologist (G.C.). These studies were approved by the Research Ethical Committee of the Medical Faculty of Lund University (Lund, Sweden) and the Institutional Review Board of the National Human Genome Research Institute of the National Institutes of Health.

Gene Expression Analysis and Class Discovery. cDNA microarrays were constructed as described (13), and contained 6,500 sequence verified cDNA clones obtained under a CRADA with ResGen (Huntsville, AL). Gene names are listed according to UniGene (available at www.ncbi.nlm.nih.gov/UniGene). The 6,500 clones represent ~4,700 unique known genes and 1,700 ESTs.

RNA was extracted from the frozen tumors after homogeni-
A scaling analysis was performed as described (7, 15, 16). Breasted and scanned, and image analysis was performed as recommended. The breast cancer cell line BT-474 (American (Qiagen, Valencia, CA) according to the manufacturers' rec-

ommendations. The breast cancer cell line BT-474 (American

zation in TRizol Reagent (Invitrogen) followed by RNase Maxi (Qiagen, Valencia, CA) according to the manufacturers’ recom-

mendations. The breast cancer cell line BT-474 (American Type Culture Collection, Manassas, VA) was harvested at subconfluency and used as a reference. Microarrays were hy-

bridized and scanned, and image analysis was performed as described (13, 14). Hierarchical clustering and multidimensional scaling analysis was performed as described (7, 15, 16).

We filtered clones by requiring that a clone should have an average spot quality (17) >0.5 as well as tumor and reference intensities >20 across all experiments. The number of clones that passed this filter was 4,795. Each slide was then normalized, such that the log intensity ratios were mean-centered. The expression data for these 4,795 well measured clones are available in Table 3, which is published as supporting information on the PNAS web site, data for these 4,795 well measured clones are available in Table 3.

Table 1. Characteristics of BRCAx breast cancer samples

<table>
<thead>
<tr>
<th>Group</th>
<th>Family</th>
<th>Sample</th>
<th>Age</th>
<th>Histological type</th>
<th>Histological grade (score)</th>
<th>Ploidy</th>
<th>SPF</th>
<th>ER</th>
<th>PgR</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>L5**</td>
<td>9275**</td>
<td>57</td>
<td>Ductal</td>
<td>2 (6)</td>
<td>D</td>
<td>4.6</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>L16</td>
<td>7676**</td>
<td>60</td>
<td>Ductal</td>
<td>1 (5)</td>
<td>NA</td>
<td>NA</td>
<td>+++</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>9981</td>
<td>79</td>
<td>Metaplastic</td>
<td>3 (8)</td>
<td>ND</td>
<td>4.1</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10785</td>
<td>75</td>
<td>Ductal</td>
<td>3 (9)</td>
<td>ND</td>
<td>10</td>
<td>--</td>
<td>--</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L99</td>
<td>6874**</td>
<td>54</td>
<td>DCIS</td>
<td>x</td>
<td>D</td>
<td>5.5</td>
<td>+</td>
<td>++</td>
</tr>
<tr>
<td>B</td>
<td>L101</td>
<td>10463</td>
<td>50</td>
<td>Ductal</td>
<td>3 (8)</td>
<td>ND</td>
<td>20</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lobular</td>
<td>2 (7)</td>
<td>ND</td>
<td>3.4</td>
<td>++</td>
<td>+++</td>
</tr>
<tr>
<td></td>
<td>L111</td>
<td>12237</td>
<td>78</td>
<td>Medullary</td>
<td>3 (9)</td>
<td>ND</td>
<td>23</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>12143††</td>
<td>43</td>
<td>Ductal</td>
<td>2 (7)</td>
<td>ND</td>
<td>11</td>
<td>++</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15564††</td>
<td>47</td>
<td>Ductal</td>
<td>2 (7)</td>
<td>ND</td>
<td>5.1</td>
<td>++</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L414</td>
<td>15401</td>
<td>57</td>
<td>Mucinous, ductal</td>
<td>3 (8)</td>
<td>ND</td>
<td>4.2</td>
<td>++</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>15478††</td>
<td>58</td>
<td>DCIS, ductal</td>
<td>2 (7)</td>
<td>D</td>
<td>3.8</td>
<td>++</td>
<td>++</td>
<td></td>
</tr>
<tr>
<td></td>
<td>L502</td>
<td>8984</td>
<td>36</td>
<td>Ductal</td>
<td>3 (8)</td>
<td>D</td>
<td>4.9</td>
<td>++</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Metastasis</td>
<td>x</td>
<td>D</td>
<td>5.1</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>L505</td>
<td>12314</td>
<td>36</td>
<td>Ductal</td>
<td>2 (7)</td>
<td>D</td>
<td>7</td>
<td>++</td>
<td>++</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lobular</td>
<td>3 (9)</td>
<td>ND</td>
<td>22</td>
<td>++</td>
<td>++</td>
</tr>
</tbody>
</table>

*Group as found by our gene expression-based class discovery analysis.
†Age at surgery of analyzed tumor sample.
‡Histological grade was based on the aggregate score for three variables (mitotic counts, nuclear pleomorphism, and tubular differentiation) and was as follows: grade 1 indicated a well-differentiated tumor (3–5 points); grade 2 indicated a moderately differentiated tumor (6 or 7 points); and grade 3 indicated a poorly differentiated tumor (8 or 9 points).
§D, diploid; ND, nondiploid.
¶Percent of cells in S-phase. NA, not available.
**This family was previously found to be linked to 13q21 (3).
††Included in CGH analysis.
‡‡Not included in the expression analysis.

number of statistically significant ($P < 0.001$) weights, i.e., the number of genes significantly different in expression between samples in the two classes. We used a simulated annealing (21) scheme, in which each step consists of changing the class of a randomly selected sample, to find the highest scoring partition of samples into two classes. This approach can easily be extended to accommodate more than two classes.

Copy Number Analysis Using cDNA Microarrays. cDNA microarrays used for CGH analysis were constructed at Agilent Technologies (Palo Alto, CA), and contained 11,367 cDNA clones (representing ~8,700 unique known genes and 2,000 ESTs) from the previously mentioned collection. DNA was extracted from breast tumors after the RNA extraction with TRizol according to the manufacturer’s recommendations. The DNA was further purified by several rounds of phenol-chloroform extractions before use. Random-prime labeling of DNA was performed according to the original protocol (22), using normal placental DNA as a reference. Scanning and image analysis was performed as for gene expression analysis.

We filtered clones by requiring that a clone should have an intensity larger than 100 fluorescence units in the reference channel across all experiments; 8,057 clones passed this filter. Each slide was then normalized, such that the log copy number ratios were mean-centered. By using the signal-to-noise statistic (18), genes were ranked based on differences in copy number between the BRCAx subgroups, and a weighted list of genes was generated. A permutation test was used to estimate the probability ($\alpha$) that a gene got a larger weight for a random labeling of the samples (23), as compared with the BRCAx subclasses.

The signal-to-noise statistic is designed to find genes consistently amplified within a group (as indicated by a small $\alpha$). We evaluated whether there was a significant overabundance of highly ranked genes in a given cytoband by using a null hyper-
genes correlated with down-regulation of a cluster of proliferative differentiated tumors that clustered together with normal ovarian gene expression profiles in ovarian cancer found that a group of well, many of the genes with increased expression in group A as illustrating the differences between the groups (Fig. 1 into two groups, we generated a hierarchical dendrogram (31) highly correlated with its primary tumor. Moreover, similar results less, the global gene expression profile of the brain metastasis was significance to a less stringent selection of discriminatory genes. About 10% of Caucasians have a highly inducible form of the enzyme that is associated with an increased risk of lung cancer in smokers (33), but which may also play a role in hormonal carcinogenesis. It would be of interest to investigate whether CYP1A1 overexpression and its susceptible genotype constitute a high risk for breast cancer development. In addition, many of the genes with low expression in group A are ESTs, and need to be further characterized.

To further exclude the possibility that our discriminatory genes were related to unidentified BRCA1 or BRCA2 mutations, we included a number of tumors from known BRCA1 and BRCA2 mutation carriers in our analysis. We performed multidimensional scaling analysis (15) and hierarchical clustering with the top genes that best separated the BRCAx tumors into two groups, and found that neither BRCA1 nor BRCA2 tumors were mixed with the BRCAx samples (Fig. 1c and d), supporting their underlying difference. Additionally, the proximity to the BRCA2 gene of the recently proposed novel breast cancer locus on 13q21 (3), has lead some investigators to believe that the families linked to this region are in fact BRCA2 families without identified mutations (5). One of the families included in our study was previously found to be linked to the 13q21 locus (Table 1); however, this sample was completely separated from the BRCA2 tumors. The generally lower grade of BRCAx breast tumors makes them histologically more similar to the moderate grade BRCA2 tumors than the high grade BRCA1 tumors. Histopathological reexamination of the BRCAx tumors revealed that, as expected, they were heterogeneous with respect to histological and clinical parameters (Table 1). However, of interest, in this study the percentage of high grade tumors, particularly with high mitotic counts, was found to be greater than has been previously described for this group of tumors (9). Numerous studies have shown that the most significant impact on gene expression profiles of sporadic breast cancers is from ER status and ER correlated genes (8, 34, 35). On the other hand, we have previously shown that BRCA1 and BRCA2 breast tumors can be correctly identified based on genes that are unable to separate ER-positive from ER-negative sporadic cases (7). Interestingly, the two groups identified here show variable hormone receptor levels. Therefore, the addition of sporadic samples in the class discovery would likely have a confounding effect on the analysis.

Analyses of genomic aberrations have revealed several chromosomal regions that are common to most breast cancers, but also those that are unique to BRCA1, BRCA2, and BRCAx tumors (3, 36, 37). It should be noted that the regions of genomic aberrations characteristic for BRCA1 or BRCA2 tumors do not harbor the predisposing gene(s). This implies that the specific nature of the predisposing gene influences the subsequent somatic genetic tumor progression pathway, or alternatively, that the various hereditary tumors stem from cells of different lineage. We therefore determined the genomic content of eight individual tumors included in our study using microarray-based CGH analysis (Table 1). The limited number of samples available for CGH analysis precluded the possibility of discovering subclasses using only the CGH data. Instead, to further confirm the molecular differences between the subgroups, we identified
262 cDNA clones that displayed significant differences \((a < 0.02)\) in copy number ratio between the groups (see Table 4, which is published as supporting information on the PNAS website). These differences were not dominated by any individual sample, but were consistently found within the subgroups (see Table 4). The locations of these genes suggest the presence of common regions of alterations within the BRCAx subgroups (Table 2), confirming their molecular differences. In particular, a large number of clones were located on 8q24, and this region was significantly amplified \((P < 10^{-11})\) in group B as compared with group A (Fig. 2). Three of the tumors have previously been analyzed by using conventional CGH, and the
findings for chromosome 8 confirm our array-based results (data not shown). According to Kainu et al. (3), 8q23–24 amplification is seen in hereditary as well as unselected breast cancers. However, only approximately one-third of BRCAx cancers displayed 8q amplification in that study, whereas over half of BRCA1, BRCA2, and sporadic breast cancers showed gain of this chromosomal region (3, 36). This observation may be explained by the absence of this aberration in group A identified in our classification. Our finding that group A is a distinct and distant branch on the expression-based dendrogram (Fig. 1d) further supports this possibility. Moreover, it has been suggested that c-MYC is the target of this amplification, and 8q24 amplification was recently shown to be associated with poor survival in breast cancer (38). However, even though c-MYC had a significantly higher copy number ratio in this group ($\alpha = 0$), it did not differ in expression between the groups. Furthermore, c-MYC was on average amplified to a lesser extent (1.5-fold) than other clones (2-fold) located closer to the 8q telomere. Interestingly, a novel gene in this region (PRL-3) associated with metastasis of colorectal cancer has recently been identified as a candidate target for 8q amplification (39). Unfortunately, PRL-3 was not present on our arrays. It has recently been shown that although copy number has a pervasive effect on gene expression, only $\sim 10\%$ of the variation in gene expression in breast cancer can be attributed to copy number alterations (40, 41). Analysis of larger sample sets using microarrays that are denser in the regions of copy number change should identify candidate target genes for the observed amplifications. Taken together, these findings suggest that fine mapping with cDNA microarray-based CGH analysis can be used to delineate chromosomal aberrations at a high resolution, thereby enabling the identification of more specific amplicon boundaries. It remains to be seen whether group A completely lacks 8q amplification, or displays it more infrequently than other groups of breast cancer. However, the paucity of this chromosomal aberration further supports the less aggressive behavior of the tumors in this group.

In summary, we have used cDNA microarrays to identify two classes of familial BRCAx breast cancers that differ in their expression of a large number of genes. Tumors from individual patients/families remained clustered within the groups, and did not mix with tumors from BRCA1 and BRCA2 mutation carriers. Additionally, we showed that the identified subclasses were associated with differences in gene amplification patterns. The results clearly require additional experiments using a larger sample set from diverse populations to pinpoint and validate the optimal subclasses and their molecular characteristics. In particular, whether high expression of ribosomal genes is correlated with less aggressive familial breast cancer warrants further investigation. We expect that the groups identified may split into subgroups as additional samples are analyzed. Nevertheless, given the heterogeneity of BRCAx tumors, and the previous difficulty in subsetting this group in a meaningful way, our data suggest that using large-scale gene expression based class discovery, followed by conventional positional linkage/candidate gene analysis may be an effective approach to finally identify novel breast cancer predisposition genes.

We thank the South Sweden Breast Cancer Group for supplying tumors, B. Baldetorp for flow cytometric analyses, and M. Fernö for coordinating the pathological review. This study was supported by a grant from the Swedish Cancer Society. M.R. was supported by a postdoctoral fellowship from the Swedish Research Council.
