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

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 subdivide *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 sporadic cases, alternate modes of inheritance, and influence of low penetrance or modifying factors. We have previously shown that the genotypes of hereditary breast cancers are reflected in different gene expression profiles (7), and this has also been illustrated in the study by van't Veer *et al.* (8). These findings indicate that the separation of the heterogeneous group of *BRCAx* breast cancers into subgroups based on gene expression profiling may be possible, and may facilitate the future search for breast cancer predisposing genes in more homogeneous groups of families. Histopathological studies have revealed that *BRCAx* tumors are cytologically heterogeneous, but are generally of

lower grade, with less nuclear pleomorphism and lower mitotic activity as compared with *BRCA1* and *BRCA2* tumors and unselected controls (9). Moreover, unlike the preponderance 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-

Table 1. Characteristics of *BRCAx* breast cancer samples

Group*	Family	Sample	Age†	Histological type	Histological grade (score)‡	Ploidy§	SPF¶	ER‖	PgR‖
A	L5**	9275††	57	Ductal	2 (6)	D	4.6	+++	+++
	L16	7676††	60	Ductal	1 (5)	NA	NA	+++	+++
		9981	79	Metaplastic	3 (8)	ND	4.1	−	−
		10785	75	Ductal	3 (9)	ND	10	−	−
	L99	6874††	54	DCIS	x	D	5.5	+	++
		7650	45	Ductal	1 (5)	NA	NA	−	+
		12096	57	Lobular	2 (7)	ND	3.4	++	+++
B	L101	10463	50	Ductal	3 (8)	ND	20	−	−
		12237	78	Medullary	3 (9)	ND	23	−	−
	L111	12143††‡‡	43	Ductal	2 (7)	ND	11	++	++
		15564††	47	Ductal	2 (7)	ND	5.1	++	++
	L414	15401	57	Mucinous, ductal	3 (8)	ND	4.2	++	−
		15478††	58	DCIS, ductal	2 (7)	D	3.8	+++	++
	L502	8984	36	Ductal	3 (8)	D	4.9	++	−
		11954††	39	Metastasis	x	D	5.1	−	−
	L505	12314	36	Ductal	2 (7)	D	7	++	+++
		14316††	62	Ductal	3 (9)	ND	22	++	++

*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.
‖ER and PgR status, respectively: −, negative; +, weakly positive; ++, positive; +++, strongly positive.
**This family was previously found to be linked to 13q21 (3).
††Included in CGH analysis.
‡‡Not included in the expression analysis.

zation in TRIzol Reagent (Invitrogen) followed by RNeasy Maxi (Qiagen, Valencia, CA) according to the manufacturers' recommendations. The breast cancer cell line BT-474 (American Type Culture Collection, Manassas, VA) was harvested at subconfluency and used as a reference. Microarrays were hybridized 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, www.pnas.org. A common approach to finding a difference in gene expression patterns between known classes of samples is to show that there is an overabundance of genes that separate the classes, as compared with what would be expected by chance (7, 16, 18, 19). Subgroupings in classes of gene expression can thus be discovered by seeking partitions of samples with an overabundance of differentially expressed genes (20). We used the method by Ben-Dor *et al.* (20) modified as follows. For a given partition of samples into two classes (with n_A and n_B samples), a discriminative weight was calculated for each gene by using the signal-to-noise statistic (18). Genes were ranked according to this weight. To test the statistical significance of the weights, sample labels were randomly permuted and the weight for each gene was computed again. This random permutation of sample labels was performed 5,000 times to generate a weight distribution that could be expected for two classes with n_A and n_B samples under the null assumption of random gene expression. The weight values for the two actual classes were then assigned *P*s based on the weight distribution from the random permutations. Candidate partitions of the data were scored with the

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 (α) 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 α). We evaluated whether there was a significant overabundance of highly ranked genes in a given cytoband by using a null hyper-

geometric model (24) (see *Supporting Text*, which is published as supporting information on the PNAS web site). This method has recently been applied to identify regulatory elements (25, 26).

Chromosomal Mapping. To localize the chromosomal positions of the clones on our arrays, we aligned clone sequences with the draft human genome sequence (27), as assembled in the August 6, 2001, freeze of the Genome Browser at UCSC (ref. 28 and http://genome.ucsc.edu), using BLAT (ref. 29 and http://genome.ucsc.edu). We only aligned clones assigned to a Unigene cluster (build 142). If the cluster contained an mRNA sequence in the RefSeq dataset (ref. 30 and http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html), that sequence was used. Requiring at least 94% sequence identity, we aligned 7,216 of the 8,057 genes to the genome sequence.

Results and Discussion

It has been anticipated for some time that *BRCAx* breast cancers constitute a heterogeneous group, most likely reflecting multiple subclasses. Kainu *et al.* (3) approached this heterogeneity by using chromosomal CGH and linkage analysis. In this study, we used the complementary strategy of global gene expression profiling followed by various mathematical methods to subclassify *BRCAx* families into genetically more homogeneous subgroups. Sixteen tumors from eight *BRCAx* families were analyzed by using cDNA microarrays (Table 1). Based on the gene expression profiles, our class discovery method (20) identified two distinct groups of tumors, comprising seven and nine samples, respectively. This classification was supported by 60 statistically significant ($P < 0.001$) genes (Fig. 1a), whereas only five genes were expected by chance (see *Materials and Methods* and Fig. 1b). Furthermore, all of the families in which multiple family members were examined remained intact. Given this data set, the probability that the families would remain intact in a random partitioning into two groups is only 1/256. Further studies are needed to pinpoint whether the partitioning reflects on a hereditary predisposition or is a manifestation of similarities in tumor progression within each group. Nevertheless, the family grouping supports an underlying hereditary cause for the partitioning. It is reasonable to expect the non-*BRCA1/BRCA2* tumors to comprise more than two distinct groups. However, given the small sample set, we did not attempt to further subdivide the samples into additional classes. Because we expect, for small sample sets, each partition to by random chance be supported by a large number of discriminatory genes, further partitioning of <10 samples will not result in robust classes using this class discovery approach.

The most sensitive sample in our partitioning was 11954, which in the second best partitioning was the only affected sample. In addition, it was the only sample sensitive to changing the level of significance to a less stringent selection of discriminatory genes. This finding may be explained by the fact that this sample was a brain metastasis from one of the breast cancers (8984). Nevertheless, the global gene expression profile of the brain metastasis was highly correlated with its primary tumor. Moreover, similar results were obtained when a rank-based gene discriminatory score was used, indicating that the partitioning is relatively insensitive to noise. Based on the 60 genes found to separate the *BRCAx* tumors into two groups, we generated a hierarchical dendrogram (31) illustrating the differences between the groups (Fig. 1a). Interestingly, many of the genes with increased expression in group A as compared with group B are ribosomal, possibly indicating different capacities for protein biosynthesis between these groups. A study of gene expression profiles in ovarian cancer found that a group of well differentiated tumors that clustered together with normal ovarian epithelial samples showed overexpression of a large number of ribosomal genes (32). Furthermore, up-regulation of ribosomal genes correlated with down-regulation of a cluster of proliferative

genes, further supporting a less aggressive phenotype in these ovarian cancers. In our data, the two groups were not significantly different from each other with respect to clinical and histopathological factors (Table 1). However, although not statistically significant, individuals in group A tended to have a later age at onset of disease and the tumors had a low percentage of cells in S-phase. This may indicate a less aggressive appearance of these tumors, in agreement with the findings for ovarian tumors reported by Welsh *et al.* (32). Moreover, groups A and B manifested significantly different expression of the *CYP1A1* gene, which encodes a phase I cytochrome P450 enzyme known to play a central role in the metabolism of a wide range of compounds, including steroids. 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 60 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. 1 c 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
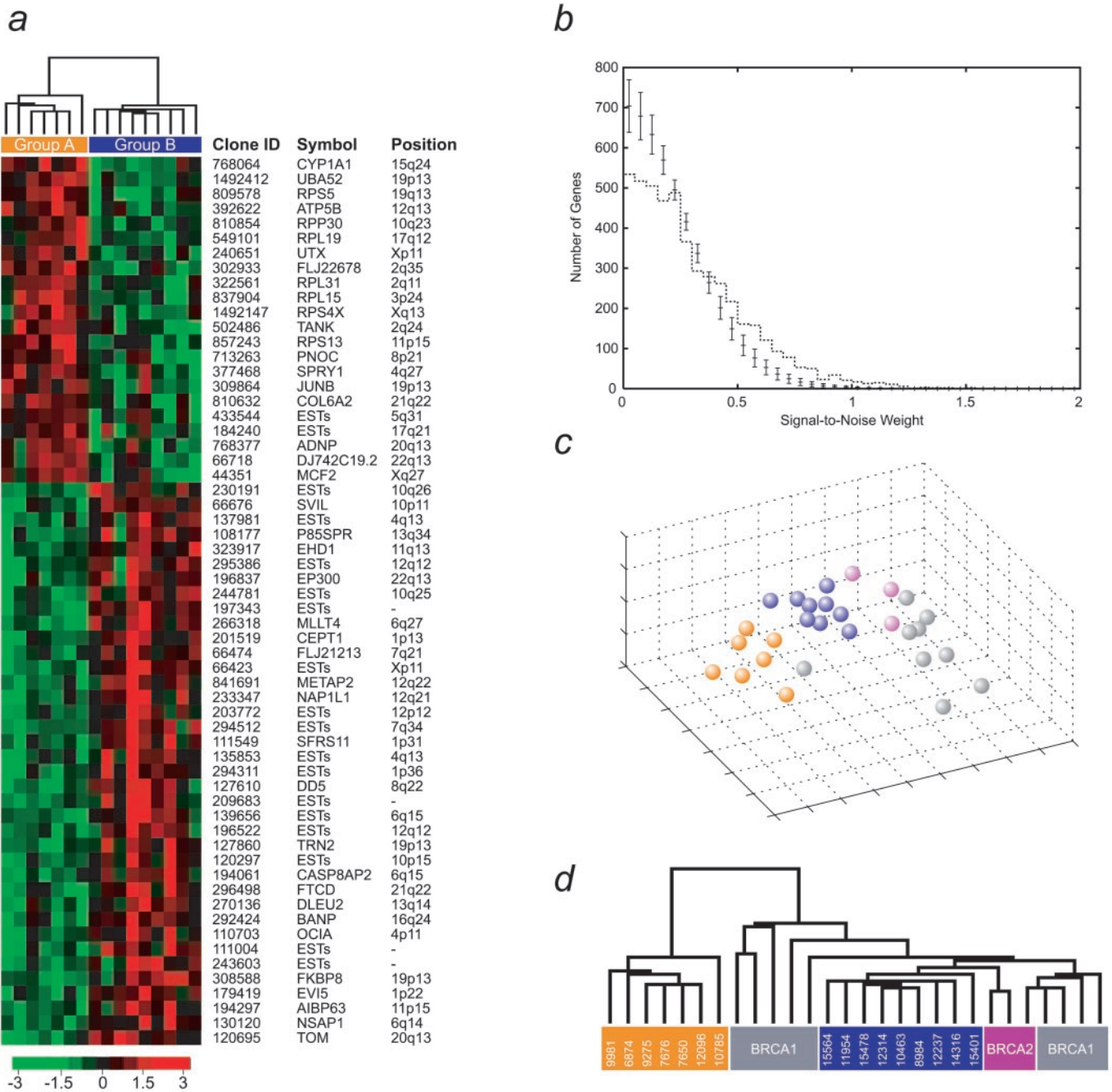
Clone ID	Symbol	Position
768064	CYP1A1	15q24
1492412	UBA52	19p13
809578	RPS5	19q13
392622	ATP5B	12q13
810854	RPP30	10q23
549101	RPL19	17q12
240651	UTX	Xp11
302933	FLJ22678	2q35
322561	RPL31	2q11
837904	RPL15	3p24
1492147	RPS4X	Xq13
502486	TANK	2q24
857243	RPS13	11p15
713263	PNOC	8p21
377468	SPRY1	4q27
309864	JUNB	19p13
810632	COL6A2	21q22
433544	ESTs	5q31
184240	ESTs	17q21
768377	ADNP	20q13
66718	DJ742C19.2	22q13
44351	MCF2	Xq27
230191	ESTs	10q26
66676	SVIL	10p11
137981	ESTs	4q13
108177	P85SPR	13q34
323917	EHD1	11q13
295386	ESTs	12q12
196837	EP300	22q13
244781	ESTs	10q25
197343	ESTs	
266318	MLLT4	6q27
201519	CEPT1	1p13
66474	FLJ21213	7q21
66423	ESTs	Xp11
841691	METAP2	12q22
233347	NAP1L1	12q21
203772	ESTs	12p12
294512	ESTs	7q34
111549	SFRS11	1p31
135853	ESTs	4q13
294311	ESTs	1p36
127610	DD5	8q22
209683	ESTs	-
139656	ESTs	6q15
196522	ESTs	12q12
127860	TRN2	19p13
120297	ESTs	10p15
194061	CASP8AP2	6q15
296498	FTCD	21q22
270136	DLEU2	13q14
292424	BANP	16q24
110703	OCIA	4p11
111004	ESTs	-
243603	ESTs	-
308588	FKBP8	19p13
179419	EVI5	1p22
194297	AIBP63	11p15
130120	NSAP1	6q14
120695	TOM	20q13

Fig. 1. Gene expression-based class discovery of *BRCAx* breast cancers. (*a*) Based on 16 *BRCAx* tumors, the most significant separation into two classes (see *Materials and Methods*) resulted in classes with seven (group A, yellow) and nine (group B, blue) samples, respectively. Group A consists of families L5, L16, and L99, and group B of families L101, L111, L414, L502, and L505 (see Table 1). Sixty statistically significant genes ($P < 0.001$), which were found to separate the groups, are listed. Expression levels for each gene are normalized across the samples such that the mean is 0 and the variance is 1. Expression levels greater than the mean are pseudocolored red, and those below are pseudocolored green. The scale indicates SDs above or below the mean. (*b*) The number of genes separating *BRCAx* cancers into two subgroups (dotted line) is plotted as a function of the signal-to-noise weight. The bars (1 SD) show the number of genes expected by chance. There is a clear overabundance of genes separating the *BRCAx* subgroups. (*c* and *d*) Based on the 60 genes that best separated *BRCAx* tumors into two groups, multidimensional scaling analysis and hierarchical clustering of the 16 samples together with *BRCA1* (gray) and *BRCA2* (purple) tumors is shown. The *BRCAx* subgroups were separated from one another as well as from the *BRCA1* and *BRCA2* tumors, reflecting the difference between *BRCAx* and *BRCA1/2* tumors.

262 cDNA clones that displayed significant differences ($\alpha < 0.02$) in copy number ratio between the groups (see Table 4, which is published as supporting information on the PNAS web site). 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

Table 2. Regions of genomic alterations differing between *BRCAx* subgroups

Position*	P value†
8q24	1.9×10^{-12}
21q22	4.7×10^{-08}
17q21	3.8×10^{-07}

*Chromosome position as aligned to the draft of the human genome sequence (see *Materials and Methods*).

†P value as determined by using a method based on the hypergeometric function (see *Supporting Text*).

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. 1*d*) 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 ≈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 show that the identified subclasses
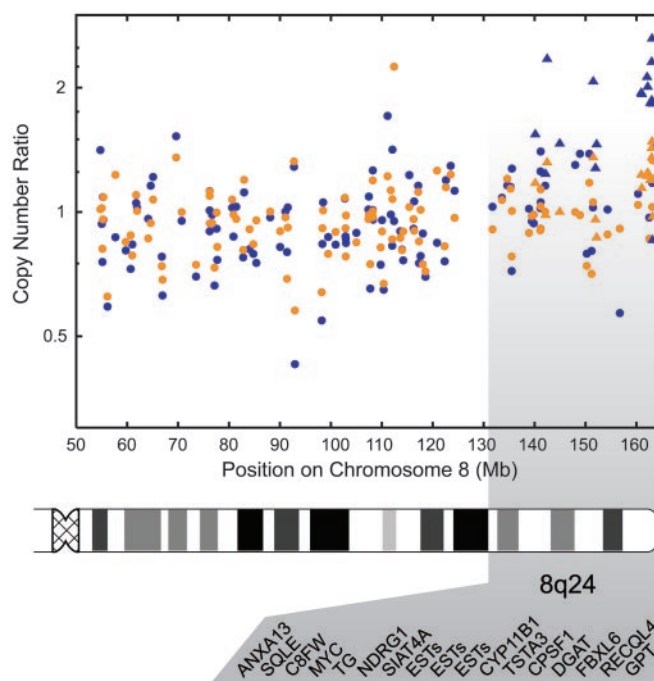

Fig. 2. cDNA microarray-based CGH analysis of chromosome 8q in *BRCAx* subgroups. Average copy number ratios for group A (yellow) and group B (blue) are shown for clones on chromosome 8q. The position of the clones on chromosome 8 is shown in Mbs from the 8p telomere, as given by alignment to the draft human genome sequence (see *Materials and Methods*). Seventeen clones on 8q24 that were found to be different ($\alpha < 0.05$) in copy number ratio between groups A and B are listed in positional order below and are highlighted (triangle). All but one (*RECQL4*) had higher ratios in group B than in group A.

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.

1. Ford, D., Easton, D. F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D. T., Weber, B., Lenoir, G., Chang-Claude, J., *et al.* (1998) *Am. J. Hum. Genet.* **62,** 676–689.
2. Seitz, S., Rohde, K., Bender, E., Nothnagel, A., Kolble, K., Schlag, P. M. & Scherneck, S. (1997) *Oncogene* **14,** 741–743.
3. Kainu, T., Juo, S. H., Desper, R., Schaffer, A. A., Gillanders, E., Rozenblum, E., Freas-Lutz, D., Weaver, D., Stephan, D., Bailey-Wilson, J., *et al.* (2000) *Proc. Natl. Acad. Sci. USA* **97,** 9603–9608.
4. Rahman, N., Teare, M. D., Seal, S., Renard, H., Mangion, J., Cour, C., Thompson, D., Shugart, Y., Eccles, D., Devilee, P., *et al.* (2000) *Oncogene* **19,** 4170–4173.
5. Thompson, D., Szabo, C. I., Mangion, J., Oldenburg, R. A., Odefrey, F., Seal, S., Barfoot, R., Kroeze-Jansema, K., Teare, D., Rahman, N., *et al.* (2002) *Proc. Natl. Acad. Sci. USA* **99,** 827–831.
6. Nathanson, K. L. & Weber, B. L. (2001) *Hum. Mol. Genet.* **10,** 715–720.
7. Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O. P., *et al.* (2001) *N. Engl. J. Med.* **344,** 539–548.
8. van't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., *et al.* (2002) *Nature* **415,** 530–536.
9. Lakhani, S. R., Gusterson, B. A., Jacquemier, J., Sloane, J. P., Anderson, T. J.,

van de Vijver, M. J., Venter, D., Freeman, A., Antoniou, A., McGuffog, L., *et al.* (2000) *Clin. Cancer Res.* **6,** 782–789.

10. Johannsson, O. T., Idvall, I., Anderson, C., Borg, Å., Barkardottir, R. B., Egilsson, V. & Olsson, H. (1997) *Eur. J. Cancer* **33,** 362–371.

11. Loman, N., Johannsson, O., Bendahl, P.-O., Borg, Å., Fernö, M. & Olsson, H. (1998) *Cancer* **83,** 310–319.

12. Hakansson, S., Johannsson, O., Johansson, U., Sellberg, G., Loman, N., Gerdes, A. M., Holmberg, E., Dahl, N., Pandis, N., Kristoffersson, U., *et al.* (1997) *Am. J. Hum. Genet.* **60,** 1068–1078.

13. DeRisi, J., Penland, L., Brown, P. O., Bittner, M. L., Meltzer, P. S., Ray, M., Chen, Y., Su, Y. A. & Trent, J. M. (1996) *Nat. Genet.* **14,** 457–460.

14. Chen, Y., Dougherty, E. R. & Bittner, M. L. (1997) *J. Biomed. Optics* **2,** 364–374.

15. Khan, J., Simon, R., Bittner, M., Chen, Y., Leighton, S. B., Pohida, T., Smith, P. D., Jiang, Y., Gooden, G. C., Trent, J. M., *et al.* (1998) *Cancer Res.* **58,** 5009–5013.

16. Bittner, M., Meltzer, P., Chen, Y., Jiang, Y., Seftor, E., Hendrix, M., Radmacher, M., Simon, R., Yakhini, Z., Ben-Dor, A., *et al.* (2000) *Nature* **406,** 536–540.

17. Chen, Y., Kamat, V., Dougherty, E. R., Bittner, M. L., Meltzer, P. S. & Trent, J. M. (2002) *Bioinformatics* **18,** 1207–1215.

18. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., *et al.* (1999) *Science* **286,** 531–537.

19. Ben-Dor, A., Bruhn, L., Friedman, N., Nachman, I., Schummer, M. & Yakhini, Z. (2000) *J. Comput. Biol.* **7,** 559–583.

20. Ben-Dor, A., Friedman, N. & Yakhini, Z. (2001) in *Fifth Annual International Conference on Computational Molecular Biology (RECOMB)*, eds. Lengauer, T., Sankoff, D., Istrail, S., Pevzner, P. & Waterman, M., (ACM Press, Montreal), pp. 31–38.

21. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. (1983) *Science* **220,** 671–680.

22. Pollack, J. R., Perou, C. M., Alizadeh, A. A., Eisen, M. B., Pergamenschikov, A., Williams, C. F., Jeffrey, S. S., Botstein, D. & Brown, P. O. (1999) *Nat. Genet.* **23,** 41–46.

23. Allander, S. V., Nupponen, N. N., Ringner, M., Hostetter, G., Maher, G. W., Goldberger, N., Chen, Y., Carpten, J., Elkahloun, A. G. & Meltzer, P. S. (2001) *Cancer Res.* **61,** 8624–8628.

24. DeGroot, M. (1986) *Probability and Statistics* (Addison–Wesley, New York).

25. Hughes, J. D., Estep, P. W., Tavazoie, S. & Church, G. M. (2000) *J. Mol. Biol.* **296,** 1205–1214.

26. Barash, Y., Bejerano, G. & Friedman, N. (2001) in *First International Workshop on Algorithms in Bioinformatics*, eds. Gascuel, O. & Moret, B. M. E. (Springer, Berlin), pp. 278–293.

27. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature* **409,** 860–921.

28. Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M. & Haussler, D. (2002) *Genome Res.* **12,** 996–1006.

29. Kent, W. J. (2002) *Genome Res.* **12,** 656–664.

30. Pruitt, K. D. & Maglott, D. R. (2001) *Nucleic Acids Res.* **29,** 137–140.

31. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14863–14868.

32. Welsh, J. B., Zarrinkar, P. P., Sapinoso, L. M., Kern, S. G., Behling, C. A., Monk, B. J., Lockhart, D. J., Burger, R. A. & Hampton, G. M. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 1176–1181.

33. McLemore, T. L., Adelberg, S., Liu, M. C., McMahon, N. A., Yu, S. J., Hubbard, W. C., Czerwinski, M., Wood, T. G., Storeng, R., Lubet, R. A., *et al.* (1990) *J. Natl. Cancer Inst.* **82,** 1333–1339.

34. Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., *et al.* (2000) *Nature* **406,** 747–752.

35. Gruvberger, S., Ringnér, M., Chen, Y., Panavally, S., Saal, L. H., Borg, Å., Fernö, M., Peterson, C. & Meltzer, P. S. (2001) *Cancer Res.* **61,** 5979–5984.

36. Tirkkonen, M., Johannsson, O., Agnarsson, B. A., Olsson, H., Ingvarsson, S., Karhu, R., Tanner, M., Isola, J., Barkardottir, R. B., Borg, Å., *et al.* (1997) *Cancer Res.* **57,** 1222–1227.

37. Monni, O., Bärlund, M., Mousses, S., Kononen, J., Sauter, G., Heiskanen, M., Paavola, P., Avela, K., Chen, Y., Bittner, M. L., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5711–5716.

38. Jain, A. N., Chin, K., Borresen-Dale, A. L., Erikstein, B. K., Eynstein Lonning, P., Kaaresen, R. & Gray, J. W. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 7952–7957.

39. Saha, S., Bardelli, A., Buckhaults, P., Velculescu, V. E., Rago, C., Croix, B. S., Romans, K. E., Choti, M. A., Lengauer, C., Kinzler, K. W., *et al.* (2001) *Science* **294,** 1343–1346.

40. Hyman, E., Kauraniemi, P., Hautaniemi, S., Wolf, M., Mousses, S., Rozenblum, E., Ringnér, M., Sauter, G., Monni, O., Elkahloun, A., *et al.* (2002) *Cancer Res.* **62,** 6240–6245.

41. Pollack, J. R., Sorlie, T., Perou, C. M., Rees, C. A., Jeffrey, S. S., Lonning, P. E., Tibshirani, R., Botstein, D., Borresen-Dale, A. L. & Brown, P. O. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 12963–12968.

GENETICS