Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications

Therese Sørlie[a,b,c], Charles M. Perou[a,d], Robert Tibshirani[e], Turid Aas[f], Stephanie Geisler[g], Hilde Johnsen[b], Trevor Hastie[e], Michael B. Eisen[h], Matt van de Rijn[i], Stefanie S. Jeffrey[j], Thor Thorsen[k], Hanne Quist[l], John C. Matese[c], Patrick O. Brown[m], David Botstein[c], Per Eystein Lønning[g], and Anne-Lise Børresen-Dale[b,n]

Departments of [b]Genetics and [l]Surgery, The Norwegian Radium Hospital, Montebello, N-0310 Oslo, Norway; [d]Department of Genetics and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, NC 27599; Departments of [e]Health Research and Policy and Statistics, [c]Genetics, [i]Pathology, [j]Surgery, and [m]Biochemistry and Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; Departments of [g]Medicine (Section of Oncology), [f]Surgery, and [k]Biochemical Endocrinology, Haukeland University Hospital, N-5021 Bergen, Norway; and [h]Life Sciences Division, Lawrence Orlando Berkeley National Laboratories, and Department of Molecular and Cellular Biology, University of California, Berkeley, CA 94720

Contributed by David Botstein, July 17, 2001

The purpose of this study was to classify breast carcinomas based on variations in gene expression patterns derived from cDNA microarrays and to correlate tumor characteristics to clinical outcome. A total of 85 cDNA microarray experiments representing 78 cancers, three fibroadenomas, and four normal breast tissues were analyzed by hierarchical clustering. As reported previously, the cancers could be classified into a basal epithelial-like group, an _ERBB2_-overexpressing group and a normal breast-like group based on variations in gene expression. A novel finding was that the previously characterized luminal epithelial/estrogen receptor-positive group could be divided into at least two subgroups, each with a distinctive expression profile. These subtypes proved to be reasonably robust by clustering using two different gene sets: first, a set of 456 cDNA clones previously selected to reflect intrinsic properties of the tumors and, second, a gene set that highly correlated with patient outcome. Survival analyses on a subcohort of patients with locally advanced breast cancer uniformly treated in a prospective study showed significantly different outcomes for the patients belonging to the various groups, including a poor prognosis for the basal-like subtype and a significant difference in outcome for the two estrogen receptor-positive groups.

The biology of breast cancer remains poorly understood. Although lymph node metastases (1), histologic grade (2), expression of steroid and growth factor receptors (3, 4), estrogen-inducible genes like cathepsin D (5), protooncogenes like _ERBB2_ (6), and mutations in the _TP53_ gene (7, 8) all have been correlated to prognosis, knowledge about individual prognostic factors provides limited information about the biology of the disease. Thus, because of their internal correlations in multivariate analysis, the prognostic value of many of these parameters fades away (9, 10).

The cellular and molecular heterogeneity of breast tumors and the large number of genes potentially involved in controlling cell growth, death, and differentiation emphasize the importance of studying multiple genetic alterations in concert. Systematic investigation of expression patterns of thousands of genes in tumors using cDNA microarrays, and their correlation to specific features of phenotypic variation, might provide the basis for an improved taxonomy of cancer (11–14).

Recently, we reported that variations in gene expression patterns in 40 grossly dissected human breast tumors analyzed by cDNA microarrays and hierarchical clustering provided a distinctive "molecular portrait" of each tumor, and that the tumors could be classified into subtypes based solely on differences in these patterns (14). The present work refines our previous classifications by analyzing a larger number of tumors and explores the clinical value of the subtypes by searching for correlations between gene expression patterns and clinically relevant parameters. We found that classification of tumors based on gene expression patterns can be used as a prognostic marker with respect to overall and relapse-free survival in a subset of patients that had received uniform therapy. One finding was the separation of estrogen receptor (ER)-positive tumors into at least two distinctive groups with characteristic gene expression profiles and different prognosis.

Materials and Methods

Patients and Tumor Specimens. A total of 78 breast carcinomas (71 ductal, five lobular, and two ductal carcinomas _in situ_ obtained from 77 different individuals; two independent tumors from one individual diagnosed at different times) and three fibroadenomas were analyzed in this study. These include 40 tumors that were previously analyzed and described (14). Four normal breast tissue samples from different individuals also were included, three of which were pooled normal breast samples from multiple individuals (CLONTECH). In summary, 85 tissue samples representing 84 individuals were analyzed. Tissue samples were snap-frozen in liquid N_2 and stored at $-170°C$ or $-80°C$. All tumor specimens analyzed contained more than 50% tumor cells. Fifty-one of the patients were part of a prospective study on locally advanced breast cancer (T_3/T_4 and/or N_2 tumors) treated with doxorubicin monotherapy before surgery followed by adjuvant tamoxifen in the case of positive ER and/or progesterone receptor (PgR) status (15). All but three patients were treated with tamoxifen. ER and PgR status was determined by using ligand-binding assays, and mutation analysis of the _TP53_ gene was performed as described (15). All common polymorphisms were recorded, but are considered wild type in this study. A detailed list of all samples and clinical data for the patients is included in Table 1, which is published as supporting information on the PNAS web site, www.pnas.org.

Microarray Analysis. Total RNA was isolated by phenol-chloroform extraction (Trizol, GIBCO/BRL), and mRNA was purified by either magnetic separation using Dynabeads (Dynal) or the Invitrogen FastTrack 2.0 Kit. All experiments and the production of microarrays were performed as described (14), with detailed protocols available at http://cmgm.

Abbreviations: ER, estrogen receptor; SAM, significance analysis of microarrays.

[a]T.S. and C.M.P. contributed equally to this work.

[n]To whom reprint requests should be addressed. E-mail: alb@labmed.uio.no.

Stanford.edu/pbrown/and http://genome-www.stanford.edu/molecularportraits/. Fluorescent images of hybridized microarrays were obtained by using a ScanArray 3000 (General Scanning, Watertown, MA) or a GenePix 4000 (Axon Instruments, Foster City, CA) scanner. The primary data tables and the image files are stored in the Stanford Microarray Database (http://genome-www4.stanford.edu/MicroArray/SMD/). Average-linkage hierarchical clustering was applied by using the CLUSTER program and the results were displayed by using TREEVIEW (software available at http://genome-www4.stanford.edu/MicroArray/SMD/restech.html).

The cDNA microarrays used in this study were from several different print runs that all contained the same core set of 8,102 genes. In total, the 85 microarray experiments were carried out by using six different batches of microarrays and three different batches of common reference, each independently produced. These variations in experimental materials produced microarray artifacts that were readily detected in our analysis. For example, batch CRB of the common reference was slightly deficient in the fibroblast-like cell line Hs578T, and hence, all samples that were analyzed by comparison to the CRB batch showed slightly elevated levels of most stromal cell genes on average (data not shown). Notwithstanding these known artifacts, mathematical searches for genes that correlated with outcome using the SAM (significance analysis of microarrays) algorithm (16) identified a set of genes, which were not influenced by these experimental artifacts. As described (14), the identification and selection of the intrinsic set of genes was based on a set of data collected with the same print run of microarrays and by using the same batch of common reference samples, and hence, it was not influenced by these experimental artifacts.

The 456 cDNA clones (427 unique genes) in the intrinsic gene list that formed the basis for the classification initially were selected from the 8,102 genes to include those with significantly greater variation in expression between different tumors than between paired samples from the same tumor (14). This subset of genes therefore should represent inherent properties of the tumors themselves rather than just differences between different samplings.

Statistical Analysis. We applied a recently described analytical method called SAM (16), to search for genes that correlated with patient survival. A total of 1,753 genes were used for this analysis, which represents all of the genes whose expression varied by at least 4-fold from the median red/green ratio in at least three or more of the samples included in the previously described sample set of 84 microarrays/40 breast tumor samples (14). Briefly, SAM computes a score for each gene that measures the strength of its correlation with survival. This score is the maximum-likelihood score statistic from Cox's proportional hazards model. When the score is negative, higher expression correlates with longer survival, whereas a positive score indicates that higher expression correlates with shorter survival. A threshold value was chosen to give a reasonably low false positive rate, as estimated by repeatedly permuting the survival times and counting the number of genes that were called significant at each threshold.

Cluster analysis and classifications were based on the total set of 78 malignant breast tumors. For the survival analyses, we used the subgroup of 49 patients (of the 51) with locally advanced tumors and no distant metastases (two of the 51 patients from this prospective study were retrospectively recorded to have a minor lung deposit and a liver metastasis, respectively) who were treated with neoadjuvant chemotherapy and adjuvant tamoxifen according to a prospective protocol (15). Kaplan–Meier plots were calculated by using WINSTAT EXCEL plug-in software from R. Fitch Software (http://www.winstat.com/). Median

follow-up time was 66 months. Deaths due to causes other than breast cancer were treated as censored observations.

Results and Discussion

Identification of Tumor Subtypes by Hierarchical Clustering. Using the intrinsic gene set of 456 cDNA clones, selected to optimally identify the intrinsic characteristics of breast tumors, the 78 carcinomas and seven nonmalignant breast samples were analyzed by hierarchical clustering (17) (Figs. 1A and 4, which is published as supporting information). As depicted in Fig. 1A, the tumors were separated into two main branches. The left branch contained three subgroups previously defined (14). These groups all were characterized by low to absent gene expression of the *ER* and several additional transcriptional factors expressed in the luminal/ER+ cluster. The basal-like subtype (Fig. 1A, red) was characterized by high expression of keratins 5 and 17, laminin, and fatty acid binding protein 7 (Fig. 1E), whereas the ERBB2+ subtype (Fig. 1A, pink) was characterized by high expression of several genes in the ERBB2 amplicon at 17q22.24 including *ERBB2* and *GRB7* (Fig. 1C). Tumor samples included in the normal breast-like group (Fig. 1A, green) showed the highest expression of many genes known to be expressed by adipose tissue and other nonepithelial cell types (Fig. 1F). These tumors also showed strong expression of basal epithelial genes and low expression of luminal epithelial genes.

Extension of the sample size allowed separation of the previously defined luminal/ER+ group into two or possibly three distinct subgroups (right-hand branch). The group of 32 tumors (termed luminal subtype A, Fig. 1A, dark blue) demonstrated the highest expression of the ER α gene, GATA binding protein 3, X-box binding protein 1, trefoil factor 3, hepatocyte nuclear factor 3 α, and estrogen-regulated *LIV*-1 (Fig. 1G). The second group of tumors positive for luminal-enriched genes could be broken into two smaller units, a small group of five tumors termed luminal subtype B (Fig. 1A, yellow) and the group of 10 tumors called luminal subtype C (Fig. 1A, light blue). Both of these groups showed low to moderate expression of the luminal-specific genes including the ER cluster. Luminal subtype C was further distinguished from luminal subtypes A and B by the high expression of a novel set of genes whose coordinated function is unknown (Fig. 1D), which is a feature they share with the basal-like and ERBB2+ subtypes.

Robustness of the Classification. To examine the robustness of the observed clustering patterns from the 78 heterogeneous carcinomas, a second hierarchical-clustering analysis was conducted by using the intrinsic gene set and the subset of 51 carcinomas from the single patient cohort (15), the three benign tumors, and the four normal breast samples (Fig. 5, which is published as supporting information). The resulting dendrogram produced with these 58 samples resembled closely (with somewhat less resolution, as anticipated) the dendrogram with all 85 samples (Fig. 2). The same major subtypes were seen, except that the position of the five luminal subtype B tumors changed from being grouped next to luminal subtype C to being grouped next to the ERBB2+ subtype. However, the luminal subtype B tumors do not overexpress *ERBB2* (Fig. 5). These results reflect the reality that higher-level branches of the dendrogram, which connect samples that have lower correlation coefficients (<0.2), are not always reflective of biologically meaningful relationships.

To further explore the relationships inferred by the dendrogram branching patterns, we computed an average expression profile for the samples contained within each of the five main subtypes from Fig. 1A (i.e., a core subtype profile) and the correlation of each sample to each of these five core expression profiles. The results are displayed in Fig. 6, which is published as supporting information, where the samples run from left to right following their hierarchical-cluster order as displayed in Fig. 1.
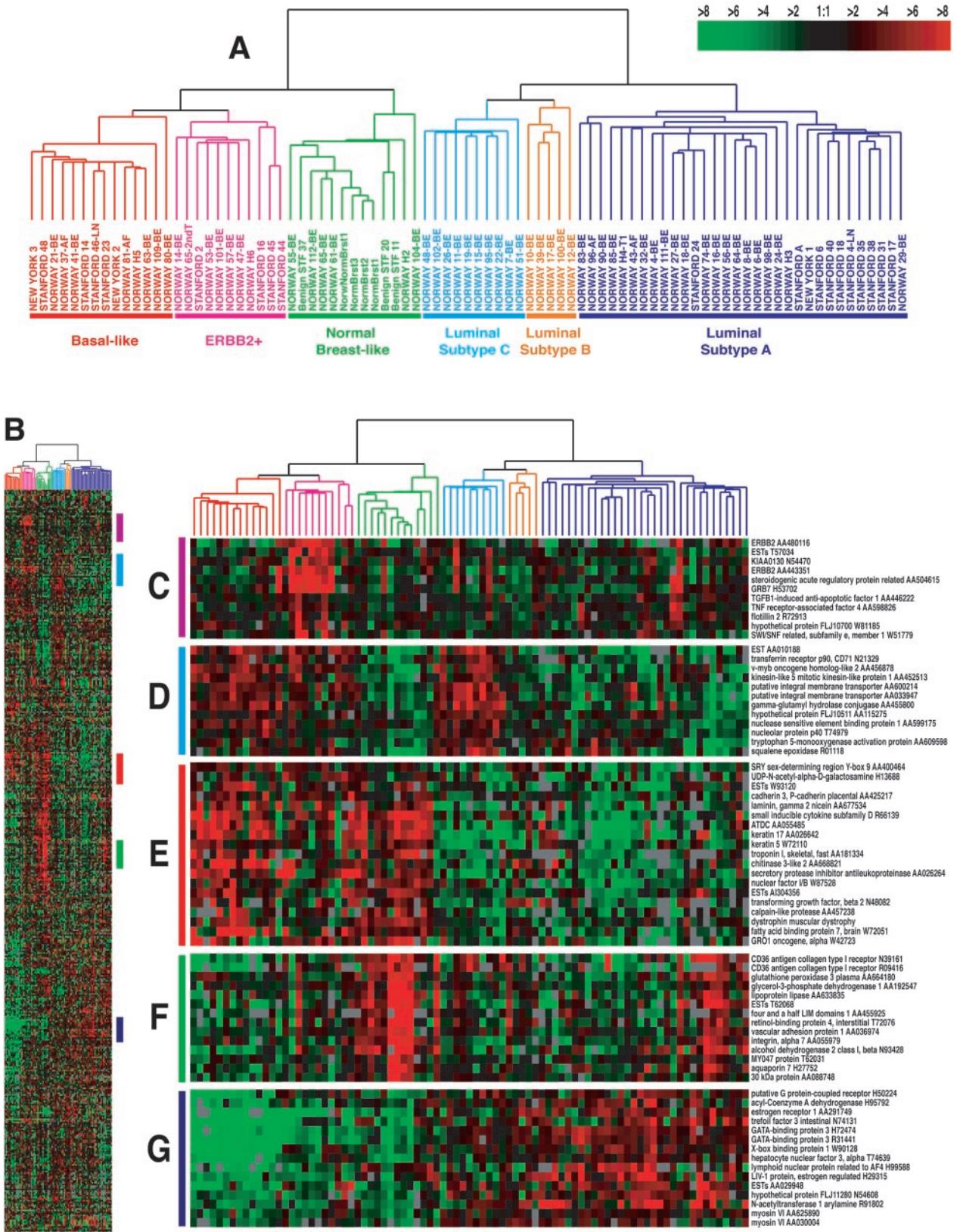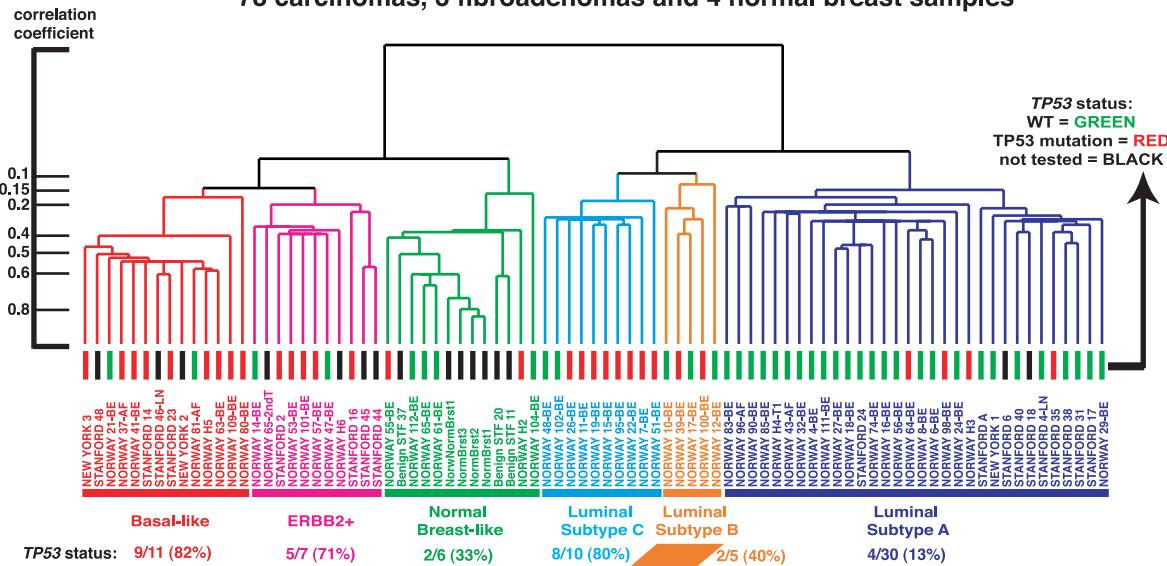
Fig. 1. Gene expression patterns of 85 experimental samples representing 78 carcinomas, three benign tumors, and four normal tissues, analyzed by hierarchical clustering using the 476 cDNA intrinsic clone set. (*A*) The tumor specimens were divided into five (or six) subtypes based on differences in gene expression. The cluster dendrogram showing the five (six) subtypes of tumors are colored as: luminal subtype A, dark blue; luminal subtype B, yellow; luminal subtype C, light blue; normal breast-like, green; basal-like, red; and ERBB2+, pink. (*B*) The full cluster diagram scaled down (the complete 456-clone cluster diagram is available as Fig. 4). The colored bars on the right represent the inserts presented in *C–G*. (*C*) *ERBB2* amplicon cluster. (*D*) Novel unknown cluster. (*E*) Basal epithelial cell-enriched cluster. (*F*) Normal breast-like cluster. (*G*) Luminal epithelial gene cluster containing *ER*.

MEDICAL SCIENCES
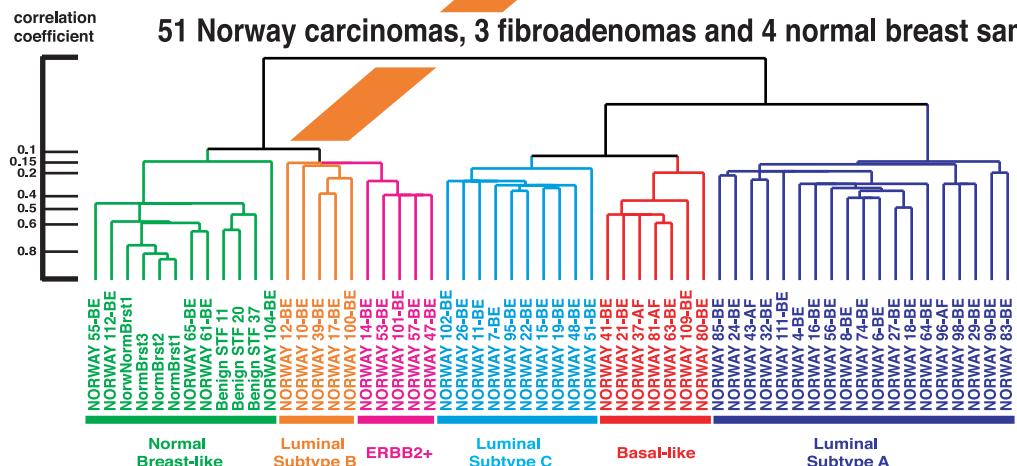
Fig. 2. Comparison of experimental sample-associated dendrograms from two different hierarchical clustering analyses. (*Upper*) Dendrogram is taken from Fig. 1 (85 samples) with the status of *TP53* indicated by the color of the terminal dendrogram line. Red lines indicate tumors with mutated *TP53* genes, green lines wild-type *TP53*, and black lines samples not tested. (*Lower*) The experimental dendrogram from the hierarchical-clustering analysis using the 51 Norway carcinomas, three benign tumors, and four normal breast tissues (58 samples). The subgroups are colored accordingly and show that the group of tumors highlighted in orange changed position compared with the dendrogram from Fig. 1*A*. Furthermore, the basal-like tumors shown in red are inserted in between luminal subtypes A and C. To the left are shown the correlation coefficients for the dendrogram branches.

As expected, the correlation was usually the highest with the expression profile for the subgroup containing that sample. This was the case for all except two samples present in the luminal subtype A cluster. The samples that were on the most distant dendrogram branches within any one of the five subgroups had lower correlations to the core profile. All but one of the samples in luminal subtype B had the lowest correlation to the combined subtype B + C core profile. This might explain why this set of samples changed location when comparing the 51-tumor clustering pattern to the 78-tumor clustering pattern and supports the identity of luminal subtype B tumors as an independent group. Hence, these data suggest that the groupings into the five subtypes are reasonably robust with most (>75%) of the tumor samples staying together in the same groups when using different sample sets for the analysis.

TP53 Status in the Tumor Subtypes. The coding region of the *TP53* gene (exons 2–11) was screened for mutations in all but 12 tumor samples (DNA or RNA were not available from these cases) (15). Thirty of the 69 tumors analyzed were found to harbor mutations in the *TP53* gene. The distribution of mutations is illustrated in Fig. 2 (*Upper*) and showed a significant difference in the frequency of *TP53*-mutated tumors among the subclasses ($P < 0.001$, two-sided). Luminal subtype A contained only 13% mutated tumors, whereas the ERBB2+ and basal-like subclasses had 5/7 (71%) and 9/11 (82%), respectively. As the *TP53* gene was not included in the intrinsic gene set, the distribution of *TP53* mutations among the different tumor groups nevertheless points to a significant role for this gene in determining the gene expression patterns in the various tumor subtypes. Previous studies have shown that mutations in the *TP53* gene predict poor
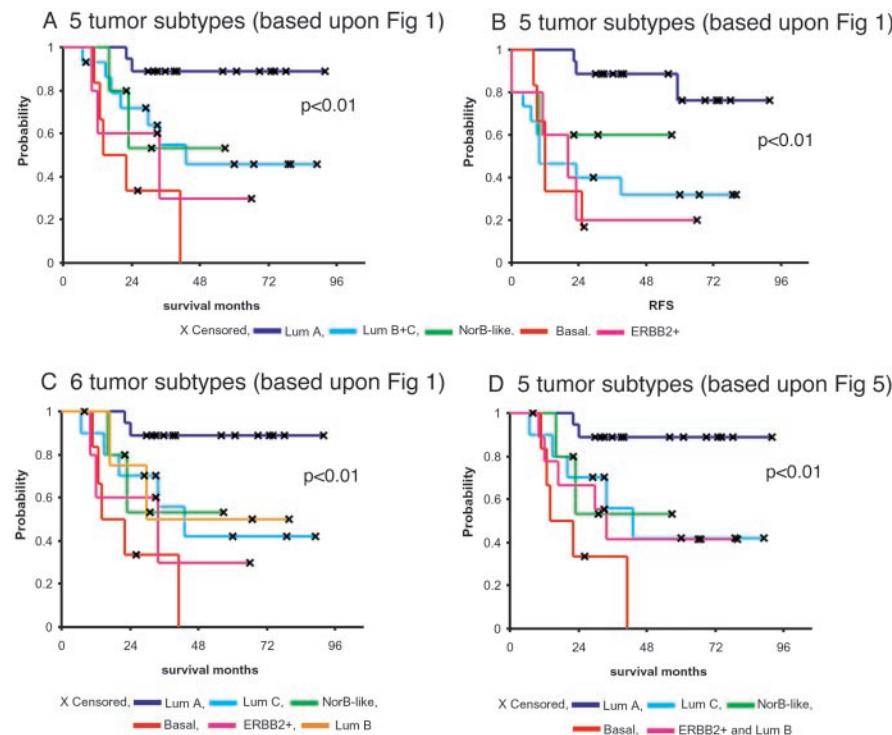
Fig. 3. Overall and relapse-free survival analysis of the 49 breast cancer patients, uniformly treated in a prospective study, based on different gene expression classification. (*A*) Overall survival and (*B*) relapse-free survival for the five expression-based tumor subtypes based on the classification presented in Fig. 1 (luminals B and C were considered one group). (*C*) Overall survival estimated for the six-subtype classification with the three different luminal subtypes presented in Fig. 1. (*D*) Overall survival based on the five-subtype classification presented in Figs. 2 *Lower* and 5.

prognosis and are associated with poor response to systemic therapy (7, 8, 18, 19). Our findings of *TP53* mutations in tumors simultaneously expressing genes in the ERBB2 amplicon at high levels supports previous observations of an interdependent role for *TP53* and *ERBB2* (15, 20).

Identification of Tumor Subtypes using SAM Supervised by Patient Survival. To search for additional sets of genes useful for tumor classification, we performed SAM (16), using patient survival as the supervising variable on the data set comprising the 76 carcinomas from which clinical data were available (i.e., excluding patient H6 and the second tumor in patient 65). Starting with their expression values from the set of 1,753 genes (14), this approach resulted in a list of 264 cDNA clones, using a significance threshold expected to produce fewer than 30 false positives. This SAM264 clone set was used to perform a hierarchical-clustering analysis on all samples, and the resulting diagram showed that almost all of the 264 cDNA clones that were selected in this analysis fell into three main gene expression clusters, the luminal/ER+ cluster, the basal epithelial cluster that contained keratins 5 and 17, and the previously described proliferation cluster (Figs. 7 and 8, which are published as supporting information). The branching patterns in the resulting dendrogram organized the tumors into four main groups. The largest group (Fig. 7, dark blue labels) consisted of tumors with the luminal/ ER+ characteristics and corresponded almost exactly to the luminal subtype A from Fig. 1. The genes comprising the ERBB2 amplicon from the intrinsic gene list were not included in the SAM clone set, which resulted in a merging of the ERBB2+ subtype with the basal-like tumors into a larger group (Fig. 7, red and pink sample names); notably, all but one of the basal-like tumors clustered together on a distal branch within this larger group. The luminal subtype C and the normal breast-like group were seen, whereas the luminal subtype B samples were spilt

between subtypes A and C. In conclusion, 71 of 78 carcinomas were organized into the same main subtypes when using the list of 264 survival-correlated cDNA clones as compared with using the intrinsic set of 456 clones (with only 81 genes overlap).

Correlations to Clinical Outcome. To investigate whether the five different groups identified by hierarchical clustering may represent clinically distinct subgroups of patients, univariate survival analyses comparing the subtypes with respect to overall survival and relapse-free survival were performed (Fig. 3). For all of the following analyses, only 49 of the patients from the prospective study with locally advanced disease and with no distant metastases were used (see *Statistical Analysis* section). Including the two patients with minor metastases did not influence the outcome of the survival analysis. The Kaplan–Meier curves based on the subclasses from Fig. 1 showed a highly significant difference in overall survival between the subtypes (Fig. 3*A*, $P < 0.01$), with the basal-like and ERBB2+ subtypes associated with the shortest survival times. Similar results were obtained with respect to relapse-free survival (Fig. 3*B*). These two tumor subtypes were characterized by distinct variations in gene expression that were different from the luminal subtype tumors. Overexpression of the ERBB2 oncoprotein is a well-known prognostic factor associated with poor survival in breast cancer, which also was found for the ERBB2+ group defined in this study. The basal-like subtype may represent a different clinical entity that is associated with shorter survival times and a high frequency of *TP53* mutations. Interestingly, the two deaths among the T_1/T_2 tumors (new york 2, new york 3) withdrawn from the data set for the purpose of the survival analysis, occurred in this subgroup of tumors; both harbored mutations in the *TP53* gene.

We observed a difference in outcome for tumors classified as luminal A versus luminal B + C. Whereas the ER protein value

(determined by ligand binding) differed between the two groups (mean 111 and 60, respectively), not all luminal A tumors showed high values (9 > 100 fmol/mg; 5, 30–100 fmol/mg; 4, 10–30 fmol/mg; 1 < 10 fmol/mg). It also should be noted that the ER+ protein category cases based on ligand binding were highly heterogeneous with respect to their gene expression profiles (18/19 were in luminal A, 5/5 in luminal B, 9/10 in luminal C, 2/7 in basal-like, 4/5 in ERBB2+, and 3/5 in normal breast-like tumors). The luminal subtype B + C tumors might represent a clinically distinct group with a different and worse disease course, in particular with respect to relapse (Fig. 3 *A* and *B*). Luminal subtype C was associated with the worst outcome of the three presumed subtypes when a six-subtype classification formed the basis for the survival analysis (Fig. 3*C*). The potential clinical significance of this molecular subtype is highlighted by the similarities in expression of some of the genes that are characteristic of the ER-negative tumors in the basal-like and ERBB2+ subtypes (Fig. 1*D*), which suggests that the high level of expression of this set of genes is associated with poor disease outcomes. The difference in outcome between the different subgroups was confirmed in a subanalysis based on the five subgroups identified when using the intrinsic gene list and only the 51 Norway carcinomas (Figs. 2 and 5), as seen in Fig. 3*D*.

The genomewide expression patterns of tumors are a representation of the biology of the tumors; the diversity in patterns reflects biological diversity. Thus, relating gene expression patterns to clinical outcomes is a key issue in understanding this diversity. Although many parameters have been explored in relation to breast cancer biology and outcome, the finding that patients with tumors expressing the *ER* have a relatively favorable prognosis, despite the fact that estradiol is a highly potent mitogen in receptor-positive cells, underlines the problems of correlating different parameters and extrapolating knowledge about the biological function of a single factor from its prognostic value. The ability to classify tumors into distinct entities by identifying recurrent gene expression patterns of hundreds or thousands of genes would further enable identification of combinations of marker genes that otherwise would be unrecognized by standard methods and help to get a deeper understanding of the function of gene interplay. In this article we have provided evidence for a relationship between five expression-based subclasses of breast tumors and patient outcome. Of particular interest is the finding that ER+ tumors may be subclassified into distinct subgroups with different outcomes. Furthermore, these studies set the stage for a larger and more elaborate study in which many additional breast tumors need to be examined and combined with detailed clinical information, which then will provide a means for identifying expression motifs that represent important clinical phenotypes, like resistance and sensitivity to specific therapies, invasiveness, or metastatic potential.

1. Fisher, E. R., Costantino, J., Fisher, B. & Redmond, C. (1993) *Cancer* **71,** 2141–2150.
2. Elston, C. W. & Ellis, I. O. (1991) *Histopathology* **19,** 403–410.
3. Torregrosa, D., Bolufer, P., Lluch, A., Lopez, J. A., Barragan, E., Ruiz, A., Guillem, V., Munarriz, B. & Garcia Conde, J. (1997) *Clin. Chim. Acta* **262,** 99–119.
4. Vollenweider-Zerargui, L., Barrelet, L., Wong, Y., Lemarchand-Beraud, T. & Gomez, F. (1986) *Cancer* **57,** 1171–1180.
5. Foekens, J. A., Look, M. P., Bolt-de Vries, J., Meijer-van Gelder, M. E., van Putten, W. L. & Klijn, J. G. (1999) *Br. J. Cancer* **79,** 300–307.
6. Slamon, D. J., Godolphin, W., Jones, L. A., Holt, J. A., Wong, S. G., Keith, D. E., Levin, W. J., Stuart, S. G., Udove, J., Ullrich, A., *et al.* (1989) *Science* **244,** 707–712.
7. Borresen, A. L., Andersen, T. I., Eyfjord, J. E., Cornelis, R. S., Thorlacius, S., Borg, A., Johansson, U., Theillet, C., Scherneck, S. & Hartman, S. (1995) *Genes Chromosomes Cancer* **14,** 71–75.
8. Bergh, J., Norberg, T., Sjogren, S., Lindgren, A. & Holmberg, L. (1995) *Nat. Med.* **1,** 1029–1034.
9. Battaglia, F., Scambia, G., Rossi, S., Panici, P. B., Bellantone, R., Polizzi, G., Querzoli, P., Negrini, R., Iacobelli, S. & Crucitti, F. (1988) *Eur. J. Cancer Clin. Oncol.* **24,** 1685–1690.
10. Howat, J. M., Barnes, D. M., Harris, M. & Swindell, R. (1983) *Br. J. Cancer* **47,** 629–640.
11. Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., *et al.* (2000) *Nature (London)* **403,** 503–511.
12. 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.
13. 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.
14. 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 (London)* **406,** 747–752.
15. Geisler, S., Lonning, P. E., Aas, T., Johnsen, H., Fluge, O., Haugan, D. F., Lillehaug, J. R., Akslen, L. A. & Børresen-Dale, A.-L. (2001) *Cancer Res.* **61,** 2505–2512.
16. Tusher, V. G., Tibshirani, R. & Chu, G. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 5116–5121. (First Published April 17, 2001; 10.1073/pnas.091062498)
17. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14863–14868.
18. Aas, T., Borresen, A. L., Geisler, S., Smith-Sorensen, B., Johnsen, H., Varhaug, J. E., Akslen, L. A. & Lonning, P. E. (1996) *Nat. Med.* **2,** 811–814.
19. Berns, E. M., Foekens, J. A., Vossen, R., Look, M. P., Devilee, P., Henzen-Logmans, S. C., van Staveren, I. L., van Putten, W. L., Inganas, M., Meijer-van Gelder, M. E., *et al.* (2000) *Cancer Res.* **60,** 2155–2162.
20. Nakopoulou, L. L., Alexiadou, A., Theodoropoulos, G. E., Lazaris, A. C., Tzonou, A. & Keramopoulos, A. (1996) *J. Pathol.* **179,** 31–38.