Repeated observation of breast tumor subtypes in independent gene expression data sets

Therese Sørlie*, Robert Tibshirani†, Joel Parker‡, Trevor Hastie§, J. S. Marron¶, Andrew Nobel¶, Shibing Deng‖, Hilde Johnsen**, Robert Pesich*, Stephanie Geisler††, Janos Demeter*, Charles M. Perou‡,‡‡, Per E. Lønning††, Patrick O. Brown§§, Anne-Lise Børresen-Dale**, and David Botstein*¶¶

Departments of *Genetics, †Health, Research and Policy, and Statistics, §Statistics and Health, and §§Biochemistry and Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA 94305; Departments of ‡Genetics and ‡‡Pathology and Laboratory Medicine, Lineberger Comprehensive Cancer Center, and Departments of ¶Statistics and ‖Biostatistics, University of North Carolina, Chapel Hill, NC 27599; ††Department of Medicine, Section of Oncology, Haukeland University Hospital, 5021 Bergen, Norway; and **Department of Genetics, Norwegian Radium Hospital, 0310 Oslo, Norway

Contributed by David Botstein, May 5, 2003

Characteristic patterns of gene expression measured by DNA microarrays have been used to classify tumors into clinically relevant subgroups. In this study, we have refined the previously defined subtypes of breast tumors that could be distinguished by their distinct patterns of gene expression. A total of 115 malignant breast tumors were analyzed by hierarchical clustering based on patterns of expression of 534 "intrinsic" genes and shown to subdivide into one basal-like, one *ERBB2*-overexpressing, two luminal-like, and one normal breast tissue-like subgroup. The genes used for classification were selected based on their similar expression levels between pairs of consecutive samples taken from the same tumor separated by 15 weeks of neoadjuvant treatment. Similar cluster analyses of two published, independent data sets representing different patient cohorts from different laboratories, uncovered some of the same breast cancer subtypes. In the one data set that included information on time to development of distant metastasis, subtypes were associated with significant differences in this clinical feature. By including a group of tumors from *BRCA1* carriers in the analysis, we found that this genotype predisposes to the basal tumor subtype. Our results strongly support the idea that many of these breast tumor subtypes represent biologically distinct disease entities.

Gene expression profiling has come into use as a way of defining, at the molecular level, the phenotypes of many kinds of tumors. In the past, we have studied genome-wide expression patterns in several cancers including lymphoma, breast, lung, liver, ovarian cancer and soft tissue sarcomas (1–5). One of the common features of these studies has been the emergence, through hierarchical clustering analysis, of tumor subtypes with distinct gene expression patterns for each of these cancers. The differences in gene expression patterns among these subtypes are likely to reflect basic differences in the cell biology of the tumors. On this basis, one might consider these molecular subtypes as separable diseases.

The molecular differences between the tumor subtypes are often accompanied by differences in clinical features, such as statistically robust differences in relapse-free and overall survival (1, 3, 6, 7). When an alternative approach (i.e., analysis supervised by outcome data) was used, many studies found handfuls of individual genes whose expression is associated with prognosis (8–13). These genes define potential prognostic molecular markers without respect to the biological diversity represented by the subtypes.

Previously we reported a molecular taxonomy of breast cancer based on variation in global gene expression patterns measured by cDNA microarrays (5, 6). In this report, we attempt to reconcile the definition of those subtypes, and the accompanying differences in disease outcome.

Materials and Methods

Tumor samples from two independent studies of response to chemotherapy of locally advanced breast cancer (14, 15) were analyzed on cDNA microarrays as described (6). Details of these and the additional tissue samples analyzed in this study can be found in Table 2, which is published as supporting information on the PNAS web site, www.pnas.org. Altogether, 122 tissue samples were included in the analysis, of which 77 carcinomas and 7 nonmalignant tissues were previously published. Data for all experiments are stored in the Stanford Microarray Database (http://genome-www.stanford.edu/MicroArray/) and can be accessed at (http://genome-www.stanford.edu/breast_cancer/). In addition to these samples, we reanalyzed published data from two independent studies: van't Veer *et al.* (13) and West *et al.* (16).

An "intrinsic" gene list was selected (534 genes represented by 552 clones; 500 of these correspond to a single unique UniGene cluster in SOURCE (http://source.stanford.edu) (17), consisting of those genes whose expression varied the least in successive samples from the same patient's tumor but which showed the most variation among tumors of different patients. For each data set, as many of these genes as possible were used for clustering (534 for the Norway/Stanford cohorts, 461 for the van't Veer *et al.* cohort, and 242 for the West *et al.* cohort).

Centroids (i.e., profiles consisting of the average expression for each of the 500 genes) were computed for each of the five classes found in the Norway/Stanford data. To be conservative, only those tumors that showed the highest correlation with each other within a subtype were used for this calculation (Table 3, which is published as supporting information on the PNAS web site). We then computed the correlation of each sample from the two additional published data sets to each of these five centroids.

Class prediction was performed by using prediction analysis of microarrays (PAM), which is a variant of nearest-centroid classification with an automated gene selection step integrated into the algorithm (18). During cross-validation, a parameter Δ was iteratively increased, and a value that balanced prediction accuracy with a minimal set of genes was chosen for the final model. This value of Δ was used for training on the entire Norway/Stanford set and predicting the class of each sample in the van't Veer *et al.* and West *et al.* data sets (Table 4, which is published as supporting information on the PNAS web site). Details of all analyses are discussed in *Supporting Materials and Methods*, which is published as supporting information on the PNAS web site.

To evaluate the performance of the 231-gene metastasis predictor published by van't Veer *et al.* (13) on patients from the Norway cohort, we used 77 of the 231 genes (77 genes overlapped with the Norway/Stanford data set) and performed a 10-fold cross validation leave-one-out analysis on the data presented in van't Veer *et al.* This "training" resulted in 81% accuracy on their own data set, which is similar to the accuracy reported by the authors in a separate validation (13). We then applied this predictor to the Norway cohort (consisting of locally advanced breast tumors from two patient series) by using PAM as was described for the class-

Abbreviation: PAM, prediction analysis of microarrays.

¶¶To whom correspondence should be addressed. E-mail: botstein@genome.stanford.edu.

www.pnas.org/cgi/doi/10.1073/pnas.0932692100

Table 1. Distribution of tumors (%) from three different studies across five tumor subtypes

Data set	Basal	ERBB2	Luminal A	Luminal B	Normal breast-like	No subtype
Norway/Stanford	17	10	24	10	3	36
van't Veer *et al.*	21	6	25	18	5	25
West *et al.*	37	4	39	14	0	6

predictions above. Univariate Kaplan–Meier analysis was performed by using WINSTAT for EXCEL (R. Fitch Software, Staufen, Germany).

Results

Identification and Validation of Tumor Subtypes in Independent Data Sets. To test the generality of previously proposed subtypes of breast cancer described in Sørlie *et al.* (6), we analyzed three independent data sets: an extended Norway/Stanford cohort and two published data sets, those of van't Veer *et al.* (13) and West *et al.* (16). The genes used for clustering were obtained from an "intrinsic" gene list comprising 534 genes derived from 45 repeated samples in the extended Norway/Stanford cohort. For a comparable analysis of the two other data sets, we used as many of the same genes as possible. Expression centroids were calculated for the core members of each of the five subclasses in the Norway/Stanford data, and correlations with the centroid were calculated for each sample from the two other data sets. Core samples for each class are color-coded similarly in all dendrograms. Table 1 summarizes the distribution of the tumor samples from the different data sets across the subtypes. In an additional, supervised analysis, we trained a predictor on the Norway/Stanford data and used PAM to predict the tumor subtype of a given individual sample in the two other data sets.

Subclassification of Breast Tumors in the Norway/Stanford Data. Our previously published work contained data from 85 tissue samples, 84 of which were reanalyzed in this study. We added 38 more tumor samples from patients with locally advanced breast cancer, which totals 122 breast tissue samples analyzed in this study. The samples fell into five major subgroups, characterized by distinct variation in gene expression pattern, as was previously seen (5, 6) (for the complete cluster diagram, see Fig. 1 and Fig. 6, which is published as supporting information on the PNAS web site). Although the list of genes that was previously used and the gene list that forms the basis for this classification only have ≈200 genes in common, the subtypes are essentially the same and anchored by the same major gene clusters (Fig. 1 *C–G*). Similarly, with few exceptions, the samples that clustered together in the past still clustered together here (86% within the same groups).

The major distinction seen was between the tumors showing high expression of luminal epithelial specific genes including the *ESR1* (Fig. 1*G*) and all other tumors showing low or no expression of these genes. The basal subtype (red) was the most homogeneous cluster of tumors, as reflected by the relatively short branches linking the tumors in this cluster (node correlation >0.4) and the deep branch separating it from the other subtypes (Fig. 1*C*).

Breast Cancer Data from van't Veer *et al.* Gene expression data ($\log_{10}$ ratios) were available for 24,480 genes in a set of 117 tumors from young breast cancer patients (13). Hierarchical clustering was used, exactly as described above for the Norway/Stanford data, to display the expression patterns of 461 "intrinsic" genes in the 97 tumor samples that were obtained from patients diagnosed with sporadic cancer (Fig. 2). Individual dendrogram branches are colored according to the strongest correlation of the corresponding tumor with the subtype centroid as defined for the Norway/Stanford samples.

As in the Norway/Stanford data, the clearest discrimination was between tumors that expressed genes in the luminal A/ESR1 cluster at high levels (Fig. 2*C*) and the tumors that were negative for these genes and exhibited expression profiles characteristic of either the basal, the ERBB2+ or the luminal B subtypes (Fig. 2 *D–F*). It is noteworthy that all samples that showed the strongest correlation with the basal subtype (red branches) are all contained within the left branch of the dendrogram in a tight cluster. The luminal A/luminal B distinction, though less clear than the basal/luminal distinction, is also seen, with many of the luminal B tumors clustering near each other on the right branch of the dendrogram.

***BRCA1* Mutations Are Associated with Basal-Type Tumors.** The van't Veer *et al.* study included tumors from 18 carriers of *BRCA1* mutations and two carriers of *BRCA2* mutations. We did not include this familial subset in the analysis above in order not to risk sample bias in estimating the frequency of different tumor subtypes (Table 1). When we included these 20 tumors along with the 97 samples, we saw little difference in the overall pattern, except for the striking result that all of the tumors from patients carrying *BRCA1* mutations fell within the basal subgroup (Fig. 3). This indicates that a mutation in the *BRCA1* gene predisposes for the basal tumor subtype, which is associated with lack of expression of the estrogen receptor and poor prognosis. As also reported previously, none of the *BRCA1* tumors showed evidence of *ERBB2* amplification (20). A distinct expression profile in *BRCA1* tumors was also noted by van't Veer *et al.* as well as by others (21). Two *BRCA1* tumors were reported by the authors to be estrogen receptor positive (sample 86 and sample 95); these two nevertheless clustered on the left branch in the dendrogram that contains the basal tumors, but were present on less correlated dendrogram branches. van't Veer *et al.* also analyzed tumor samples from two *BRCA2* carriers, both of which showed luminal A expression patterns.

Breast Cancer Data Set from West *et al.* Expression levels of a total of 7,129 genes were measured in 49 breast tumor samples by West *et al.* (16) using Affymetrix oligonucleotide arrays. Data were transformed to a compatible format by normalizing to the median experiment (see supporting information for details). Expression values for 242 "intrinsic" genes were used in a hierarchical cluster analysis exactly as was done for the Norway/Stanford and the van't Veer *et al.* data sets. Correlation coefficients to the five subtype centroids were calculated for each of the 49 tumors, and the branches of the dendrogram were colored according to the nearest centroid (Fig. 4).

Again, the main discrimination seen was between the tumors that highly expressed genes in the luminal A/ESR1 cluster and those that were clearly negative for these genes (Fig. 4 *C* and *D*). Data for only about half of the genes from the intrinsic gene list were found in this study, which may explain why a luminal A/luminal B distinction was not seen in this data set. Two tumors showed strong correlation to the ERBB2+ class (Fig. 4*E*), but they did not cluster together in this analysis.

Prediction of Tumor Subtypes By Using Norway/Stanford Data As a Training Set. Hierarchical clustering analysis is a powerful technique for class discovery; however, we wished also to apply a more supervised analysis that could make a prediction as to what subtype a single sample would belong to when considered only by itself. To accomplish this goal, we used the 115 Norway/Stanford tumor samples and the overlapping "intrinsic" genes for both data sets, respectively, as a training set to develop a breast tumor class predictor using PAM (see *Supporting Materials and Methods* for details). When we compared these calculations with the results of the hierarchical clustering described above, there was strong agreement, ranging from 79% to 89% for both the van't Veer *et al.* and

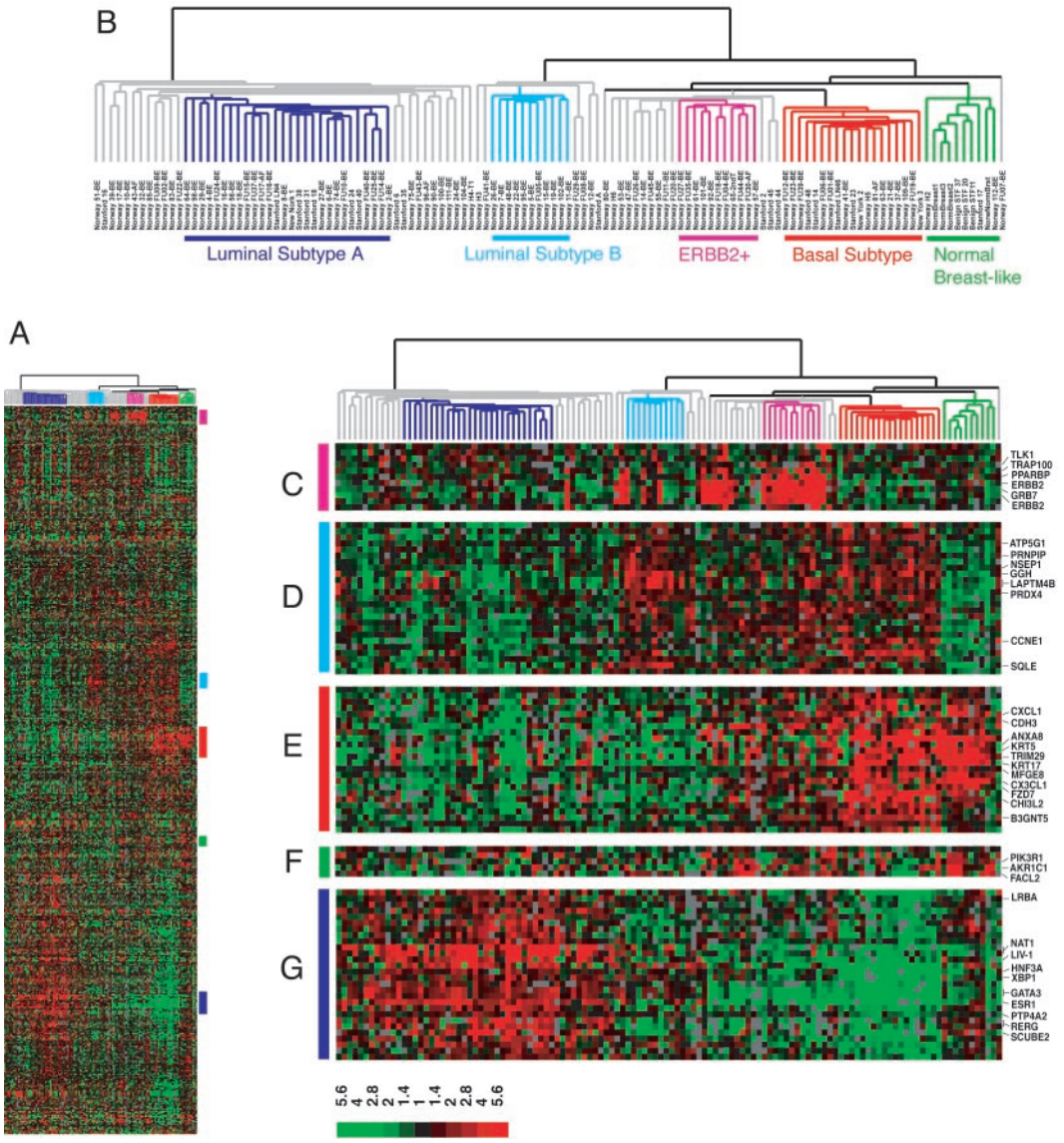

Fig. 1. Hierarchical clustering of 115 tumor tissues and 7 nonmalignant tissues using the ''intrinsic'' gene set. (*A*) A scaled-down representation of the entire cluster of 534 genes and 122 tissue samples based on similarities in gene expression. (*B*) Experimental dendrogram showing the clustering of the tumors into five subgroups. Branches corresponding to tumors with low correlation to any subtype are shown in gray. (*C*) Gene cluster showing the *ERBB2* oncogene and other coexpressed genes. (*D*) Gene cluster associated with luminal subtype B. (*E*) Gene cluster associated with the basal subtype. (*F*) A gene cluster relevant for the normal breast-like group. (*G*) Cluster of genes including the estrogen receptor (*ESR1*) highly expressed in luminal subtype A tumors. Scale bar represents fold change for any given gene relative to the median level of expression across all samples. (See also Fig. 6.)

West *et al.* data sets (Table 4). We note that prediction accuracies reported above are somewhat optimistic, as some of the genes used as predictors were used to define the test set groups in the first place.

Tumor Subtypes Are Associated with Significant Difference in Clinical Outcome. In our previous work, the expression-based tumor subtypes were associated with a significant difference in overall survival as well as disease-free survival for the patients suffering from locally advanced breast cancer and belonging to the same treatment protocol (6). To investigate whether these subtypes were also associated with a significant difference in outcome in other patient cohorts, we performed a univariate Kaplan–Meier analysis with time to development of distant metastasis as a variable in the data set comprising the 97 sporadic tumors taken from van't Veer *et al.* As shown in Fig. 5, the probability of remaining disease-free was significantly different between the subtypes; patients with luminal A type tumors lived considerably longer before they developed metastatic disease, whereas the basal and ERBB2+ groups showed much shorter disease-free time intervals. Although the methodological differences prevent a definitive interpretation, it is notable that the order of severity of clinical outcome associated with the several subtypes is similar in the two dissimilar cohorts. We could not carry out a similar analysis in the West *et al.* data because the necessary follow-up data were not provided.

Discussion

Breast Tumor Subtypes Represent Distinct Biological Entities. Gene expression studies have made it clear that there is considerable diversity among breast tumors, both biologically and clinically (5, 6,

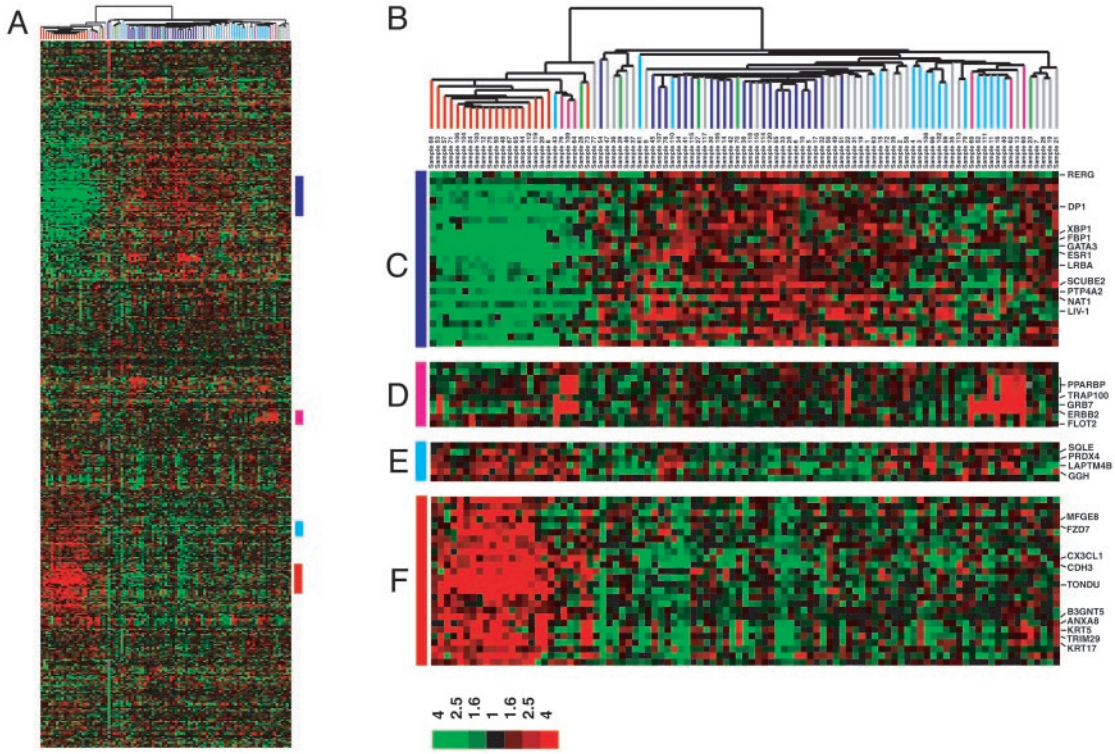

Fig. 2. Hierarchical clustering of gene expression data from van't Veer *et al.* (*A*) The full cluster of 461 genes across 97 sporadic tumors. (*B*) Experimental dendrogram displaying similarities between the tumors. Branches are color-coded according to the subtype to which the corresponding tumor sample shows the highest correlation. Tumors with low correlation (<0.1) to a specific subtype are indicated by gray branches. (*C*) Gene cluster associated with the luminal subtype A. (*D*) Gene cluster containing the *ERBB2* oncogene and coclustered genes. (*E*) Group of genes that tend to be highly expressed in luminal subtype B tumors. (*F*) Gene cluster characteristic of basal tumors. Scale bar represents fold change for any given gene relative to the median level of expression across all samples. (See also Fig. 7, which is published as supporting information on the PNAS web site.)

13, 16, 19, 21, 22). This is not a new idea, as epidemiological studies previously had inferred the existence of two or more subpopulations of breast cancer (23). A straightforward interpretation of the recurrent appearance of several different patterns of gene expression among tumors of similar anatomical origin is to regard each as representing a different biological entity. One possible basis for the consistent differences in these patterns between tumor subtypes might be that they originate from different cell types. Our findings provide some support for this interpretation, as we found breast tumor subtypes with patterns of gene expression similar to those of luminal epithelial cells (the cells that line the duct and give rise to the majority of breast cancers) and patterns of at least one other subtype (termed basal) that resembles the pattern found in basal epithelial cells of the normal mammary gland (characterized by expression of cytokeratins 5/6 and 17).

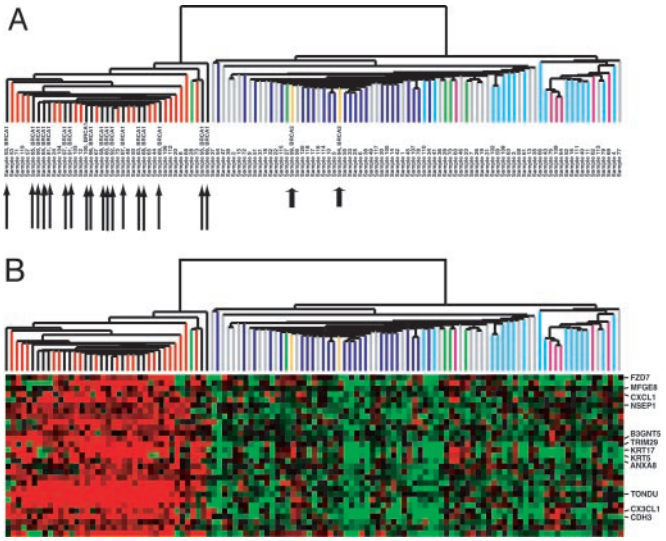

Fig. 3. BRCA1 tumors associated with a basal tumor profile. (*A*) Dendrogram showing all tumors from van't Veer *et al.*, including 18 tumors from *BRCA1* mutation carriers (black branches) and two tumors from *BRCA2* mutation carriers (yellow branches). *BRCA1* tumors are indicated with longer arrows; *BRCA2* tumors are indicated with shorter arrows. (*B*) Cluster of genes characteristic of basal tumors and highly expressed in tumors from *BRCA*1-carriers. (See also Fig. 8, which is published as supporting information on the PNAS web site.)

If indeed luminal and basal tumor subtypes are distinct biological entities, then the cognate expression patterns should be detectable in other genome-scale studies of breast cancer. As shown above, we found strong evidence for the universality of a distinction between basal-like and luminal-like subtypes in two additional, independent data sets comprising different patient populations whose gene expression profiles had been determined by using different microarray technology platforms. We found considerable evidence, in one of the studies, for the distinction between the luminal A and B subtypes. The fact that we could make these distinctions for the basal and luminal subtypes (less so for the luminal B subtype *vis a vis* luminal A) means that the substantial differences in the characteristics of the patients (e.g., age and tumor stage) are less important determinants of tumor expression phenotypes than intrinsic biology.

The statistical nature of the definition, the differences in the expression technologies, and, more importantly, the limited number of intrinsic genes held in common in particular for the West *et al.* data set, probably suffices to account for the failure to find coherent clusters for every subtype in each of the cohorts examined.

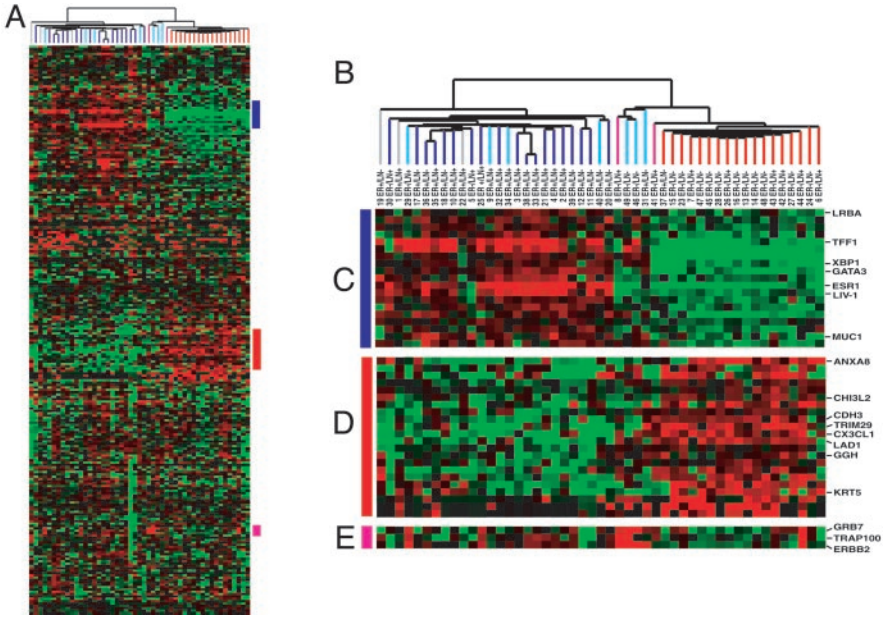

Fig. 4. Hierarchical clustering of gene expression data from West *et al.* (*A*) Scaled-down representation of the full cluster of 242 intrinsic genes across 49 breast tumors. (*B*) Dendrogram displaying the relative organization of the tumor samples. Branches are colored according to which subtype the corresponding tumor showed the strongest correlation with. Gray branches indicate tumors with low correlation (<0.1) to any specific subtype. (*C*) Luminal epithelial/estrogen receptor gene cluster. (*D*) Basal gene cluster. (*E*) ERBB2+ gene cluster. (See also Fig. 9, which is published as supporting information on the PNAS web site.)

Another expectation from the concept that the tumor subtypes represent different biological entities is that genetic predispositions to breast cancer might give rise preferentially to certain subtypes. This expectation is amply fulfilled by our finding in the data of van't Veer *et al.*, which shows that the women carrying *BRCA1*-mutated alleles all had tumors with the basal-like gene expression pattern.

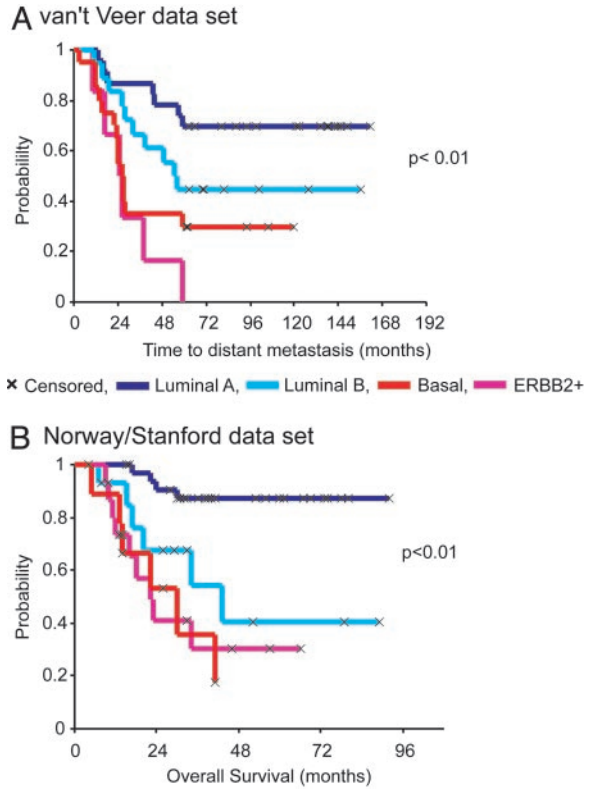

Fig. 5. Kaplan–Meier analysis of disease outcome in two patient cohorts. (*A*) Time to development of distant metastasis in the 97 sporadic cases from van't Veer *et al*. Patients were stratified according to the subtypes as shown in Fig. 2*B*. (*B*) Overall survival for 72 patients with locally advanced breast cancer in the Norway cohort. The normal-like tumor subgroups were omitted from both data sets in this analysis.

Tumor Subtypes and Clinical Outcome. Consistent with the results previously found in our data (6), we also found differences in clinical outcome associated with the different tumor subtypes in the data set produced by van't Veer *et al*. The outcomes, as measured here in time to development of distant metastasis, were strikingly similar to what we found previously: worst for basal (and ERBB2+), best for luminal A, and intermediate for luminal B subtypes. Recently, two reports corroborating the poor outcome of the basal subtype solely based on immunohistochemistry with antibodies against keratins 5 and 17 and Skp2, strongly supports our results (24, 25). The finding that our gene cluster profile was of similar prognostic importance in the van't Veer *et al.* cohort as among our patients is remarkable, taking into account differences regarding disease stage (locally advanced versus stage I primaries) and patient age, but in particular, the fact that the Norwegian patients had presurgical chemotherapy and all patients expressing *ESR1* received adjuvant endocrine treatment, whereas the patients from van't Veer *et al.* in general did not receive any systemic adjuvant treatment.

The observation that *BRCA1* mutations are strongly associated with a basal tumor phenotype indicates a particularly poor prognosis for these patients. *BRCA1*-associated breast cancers are usually highly proliferative and *TP53*-mutated, and usually lack expression of *ESR1* and *ERBB2* (20, 26). Status of *BRCA1* in familial cancers has failed to be an independent prognostic factor in several studies (reviewed in ref. 27), and is complicated by confounding factors such as frequent screening and early diagnosis.

Molecular Marker Identification. In a mixture of biologically distinct subtypes, it may well be that individual markers derived by supervised analysis will under-perform what is possible if tumor subtypes were separated before searching, in a supervised fashion, for prognostic indicators. Indeed, when we tested the prognostic impact of the 231 markers published by van't Veer *et al.* on the Norwegian cohort, we found that they performed less well (47%) in predicting recurrences within 5 years (see *Materials and Methods*). This may in part be due to differences in the patient cohorts and treatments as discussed above.

Both van't Veer *et al.* and West *et al.* showed the ability of gene

expression profiles to classify tumors into clinically relevant groups and to predict outcome by using supervised statistical analyses. Both reports, however, showed only how the gene expression signatures discriminated tumors based on previously known molecular and clinical parameters, such as *ESR1* status, lymph node status, and time to the development of distant metastasis. We have taken a less supervised approach and showed that there are several subtypes of tumors, which may be the result of alterations in different and independent regulatory pathways. The basal subtype was repeatedly recognized as a distinct group in each of three independent data sets, and should be considered as a separate disease with respect to treatment and follow up. The other subtypes are less clear, and require refinement of their molecular definition before they can be reliably defined and diagnosed.

To conclude, classification of breast cancer based on gene expression profiling captures the molecular complexity of tumors. It is for this reason that we believe that the patterns that distinguish subtypes appear to provide a more refined stratification of the patients compared with individual tumor markers. These results imply that the status of the transcriptional programs in the tumor cells and the underlying genetic alterations are major determinants of the tumorigenic potential and ultimately the disease outcome for the patient.

We thank Mike Fero and Stanford Functional Genomics Facility for the production of microarrays, and Michael Whitfield for help with analysis of the data produced by using Affymetrix arrays. This research was supported by National Cancer Institute Grants CA77097 and CA85129 (to D.B. and P.O.B.) and CA58223-09A1 and CA097769-01 (to C.M.P.). T.S. is a postdoctoral fellow of the Norwegian Cancer Society. P.O.B. is an Associate Investigator of the Howard Hughes Medical Institute.

1. 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* **403,** 503–511.
2. Chen, X., Cheung, S. T., So, S., Fan, S. T., Barry, C., Higgins, J., Lai, K. M., Ji, J., Dudoit, S., Ng, I. O., *et al.* (2002) *Mol. Biol. Cell* **13,** 1929–1939.
3. Garber, M. E., Troyanskaya, O. G., Schluens, K., Petersen, S., Thaesler, Z., Pacyna-Gengelbach, M., van de Rijn, M., Rosen, G. D., Perou, C. M., Whyte, R. I., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 13784–13789.
4. Nielsen, T. O., West, R. B., Linn, S. C., Alter, O., Knowling, M. A., O'Connell, J. X., Zhu, S., Fero, M., Sherlock, G., Pollack, J. R., *et al.* (2002) *Lancet* **359,** 1301–1307.
5. Perou, C. M., Sørlie, 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.
6. Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., *et al.* (2001) *Proc. Natl. Acad. Sci. USA* **98,** 10869–10874.
7. Wigle, D. A., Jurisica, I., Radulovich, N., Pintilie, M., Rossant, J., Liu, N., Lu, C., Woodgett, J., Seiden, I., Johnston, M., *et al.* (2002) *Cancer Res.* **62,** 3005–3008.
8. Beer, D. G., Kardia, S. L., Huang, C. C., Giordano, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Gharib, T. G., Thomas, D. G., *et al.* (2002) *Nat. Med.* **8,** 816–824.
9. Jenssen, T. K., Kuo, W. P., Stokke, T. & Hovig, E. (2002) *Hum. Genet.* **111,** 411–420.
10. Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., Angelo, M., McLaughlin, M. E., Kim, J. Y., Goumnerova, L. C., Black, P. M., Lau, C., *et al.* (2002) *Nature* **415,** 436–442.
11. Rosenwald, A., Wright, G., Chan, W. C., Connors, J. M., Campo, E., Fisher, R. I., Gascoyne, R. D., Muller-Hermelink, H. K., Smeland, E. B., Giltnane, J. M., *et al.* (2002) *N. Engl. J. Med.* **346,** 1937–1947.
12. Shipp, M. A., Ross, K. N., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G. S., *et al.* (2002) *Nat. Med.* **8,** 68–74.
13. 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, K. K., Marton, M. J., Witteveen, A. T., *et al.* (2002) *Nature* **415,** 530–536.
14. Geisler, S., Lonning, P. E., Aas, T., Johnsen, H., Fluge, O., Haugen, D. F., Lillehaug, J. R., Akslen, L. A. & Borresen-Dale, A. L. (2001) *Cancer Res.* **61,** 2505–2512.
15. Geisler, S., Børresen-Dale, A.-L., Johnsen, H., Aas, T., Geisler, J., Akslen, L. A., Anker, G. & Lønning, P. E. (2003) *Clin. Cancer Res*., in press.
16. West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J. A., Jr., Marks, J. R. & Nevins, J. R. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 11462–11467.
17. Diehn, M., Sherlock, G., Binkley, G., Jin, H., Matese, J. C., Hernandez-Boussard, T., Rees, C. A., Cherry, J. M., Botstein, D., Brown, P. O., *et al.* (2003) *Nucleic Acids Res.* **31,** 219–223.
18. Tibshirani, R., Hastie, T., Narasimhan, B. & Chu, G. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 6567–6572.
19. van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., *et al.* (2002) *N. Engl. J. Med.* **347,** 1999–2009.
20. Grushko, T. A., Blackwood, M. A., Schumm, P. L., Hagos, F. G., Adeyanju, M. O., Feldman, M. D., Sanders, M. O., Weber, B. L. & Olopade, O. I. (2002) *Cancer Res.* **62,** 1481–1488.
21. 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.
22. Gruvberger, S., Ringner, M., Chen, Y., Panavally, S., Saal, L. H., Borg, A., Ferno, M., Peterson, C. & Meltzer, P. S. (2001) *Cancer Res.* **61,** 5979–5984.
23. Fox, M. S. (1979) *J. Am. Med. Assoc.* **241,** 489–494.
24. Signoretti, S., Di Marcotullio, L., Richardson, A., Ramaswamy, S., Isaac, B., Rue, M., Monti, F., Loda, M. & Pagano, M. (2002) *J. Clin. Invest.* **110,** 633–641.
25. van de Rijn, M., Perou, C. M., Tibshirani, R., Haas, P., Kallioniemi, O., Kononen, J., Torhorst, J., Sauter, G., Zuber, M., Kochli, O. R., *et al.* (2002) *Am. J. Pathol.* **161,** 1991–1996.
26. Armes, J. E., Trute, L., White, D., Southey, M. C., Hammet, F., Tesoriero, A., Hutchins, A. M., Dite, G. S., McCredie, M. R., Giles, G. G., *et al.* (1999) *Cancer Res.* **59,** 2011–2017.
27. Kennedy, R. D., Quinn, J. E., Johnston, P. G. & Harkin, D. P. (2002) *Lancet* **360,** 1007–1014.