Biol.

A

Molecular portraits of human breast tumours

Charles M. Perou[*†], Therese Sørlie[†‡], Michael B. Eisen[*], Matt van de Rijn[§], Stefanie S. Jeffrey[‖], Christian A. Rees[*], Jonathan R. Pollack[¶], Douglas T. Ross[¶], Hilde Johnsen[‡], Lars A. Akslen[#], Øystein Fluge[☆], Alexander Pergamenschikov[*], Cheryl Williams[*], Shirley X. Zhu[§], Per E. Lønning[**], Anne-Lise Børresen-Dale[‡], Patrick O. Brown[¶††] & David Botstein[*]

*

Human breast tumours are diverse in their natural history and in their responsiveness to treatments[1]. Variation in transcriptional programs accounts for much of the biological diversity of human cells and tumours. In each cell, signal transduction and regulatory systems transduce information from the cell's identity to its environmental status, thereby controlling the level of expression of every gene in the genome. Here we have characterized variation in gene expression patterns in a set of 65 surgical specimens of human breast tumours from 42 different individuals, using complementary DNA microarrays representing 8,102 human genes. These patterns provided a distinctive molecular portrait of each tumour. Twenty of the tumours were sampled twice, before and after a 16-week course of doxorubicin chemotherapy, and two tumours were paired with a lymph node metastasis from the same patient. Gene expression patterns in two tumour samples from the same individual were almost always more similar to each other than either was to any other sample. Sets of co-expressed genes were identified for which variation in messenger RNA levels could be related to specific features of physiological variation. The tumours could be classified into subtypes distinguished by pervasive differences in their gene expression patterns.

We proposed that the phenotypic diversity of breast tumours might be accompanied by a corresponding diversity in gene expression patterns that we could capture using cDNA microarrays. Systematic investigation of gene expression patterns in human breast tumours might then provide the basis for an improved molecular taxonomy of breast cancers. We analysed gene expression patterns in grossly dissected normal or malignant human breast tissues from 42 individuals (36 infiltrating ductal carcinomas, 2 lobular carcinomas, 1 ductal carcinoma *in situ*, 1 fibroadenoma and 3 normal breast samples). Fluorescently labelled (Cy5) cDNA was prepared from mRNA from each experimental sample. We prepared cDNA, labelled using a second distinguishable fluorescent nucleotide (Cy3), from a pool of mRNAs isolated from 11 different cultured cell lines (see Supplementary Information Table 1); this common 'reference' sample provided an internal standard against which the gene expression of each experimental sample was compared[2,3].

Twenty of the forty breast tumours examined were sampled twice, as part of a larger study on locally advanced breast cancers (T_3/T_4 and/or N_2 tumours; see ref. 4). After an open surgical biopsy to obtain the 'before' sample, each of these patients was treated with doxorubicin for an average of 16 weeks (range 12–23), followed by resection of the remaining tumour. In addition, primary tumours

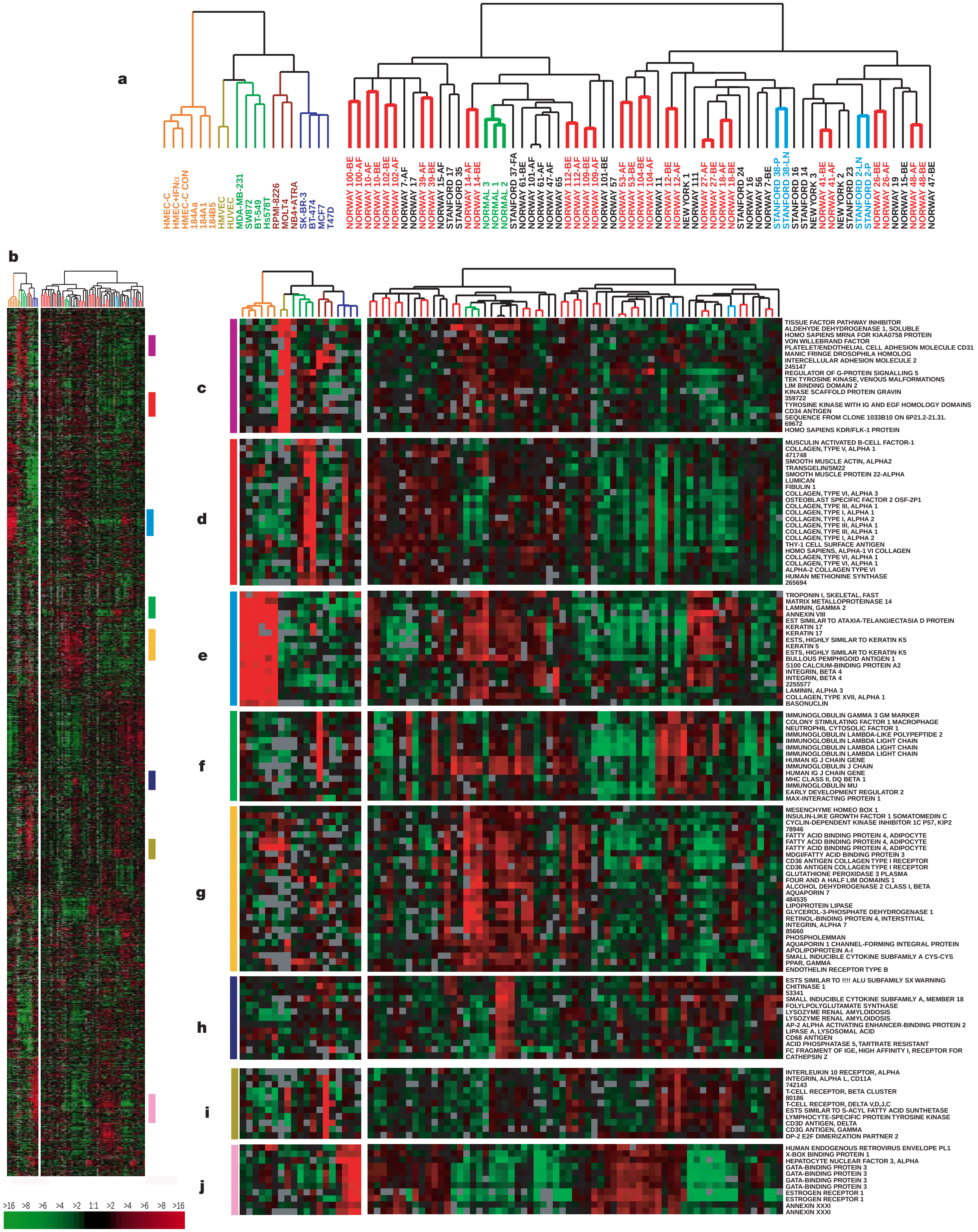

Figure 1 Variation in expression of 1,753 genes in 84 experimental samples. Data are presented in a matrix format: each row represents a single gene, and each column an experimental sample. In each sample, the ratio of the abundance of transcripts of each gene to the median abundance of the gene's transcript among all the cell lines (left panel), or to its median abundance across all tissue samples (right panel), is represented by the colour of the corresponding cell in the matrix. Green squares, transcript levels below the median; black squares, transcript levels equal to the median; red squares, transcript levels greater than the median; grey squares, technically inadequate or missing data. Colour saturation reflects the magnitude of the ratio relative to the median for each set of samples (see scale, bottom left; and Supplementary Information Fig. 4). **a**, Dendrogram representing similarities in the expression patterns between experimental samples. All 'before and after' chemotherapy pairs that were clustered on terminal branches are highlighted in red; the two primary tumour/lymph node metastasis pairs in light blue; the three clustered normal breast samples in light green. Branches representing the four breast luminal epithelial cell lines are shown in dark blue; breast basal epithelial cell lines in orange, the endothelial cell lines in dark yellow, the mesynchemal-like cell lines in dark green, and the lymphocyte-derived cell lines in brown. **b**, Scaled-down representation of the 1,753-gene cluster diagram; coloured bars to the right identify the locations of the inserts displayed in **c**–**j**. **c**, Endothelial cell gene expression cluster; **d**, stromal/fibroblast cluster; **e**, breast basal epithelial cluster; **f**, B-cell cluster; **g**, adipose-enriched/normal breast; **h**, macrophage; **i**, T-cell; **j**, breast luminal epithelial cell.

from two patients were also paired with a lymph node metastasis from the same patient. To help interpret the variation in expression patterns seen in the tumour samples, we also characterized 17 cultured cell lines (with one cell line cultured under three different conditions), which provided models for many of the cell types encountered in these tissue samples. In total, we analysed 84 cDNA microarray experiments (see Supplementary Information, Table 2; the primary data tables can be obtained at http://genome-www.stanford.edu/molecularportraits/).

A hierarchical clustering method was used to group genes on the basis of similarity in the pattern with which their expression varied over all samples[5]. The same clustering method was used to group the experimental samples (cell lines and tissues separately) on the basis of similarity in their patterns of expression. We focus first on a set of 1,753 genes (about 22% of the 8,102 genes analysed), whose transcripts varied in abundance by at least fourfold from their median abundance in this sample set in at least three of the samples (Fig. 1; see Supplementary Information Fig. 4 for the complete cluster diagram).

Three striking features of the gene expression patterns of these tumours are evident in Fig. 1. First, the tumours show great variation in their patterns of gene expression. Second, this variation is multidimensional; that is, many different sets of genes show mainly independent patterns of variation. Third, these patterns have a pervasive order reflecting relationships among the genes, relationships among the tumours and connections between specific genes and specific tumours.

The hierarchical clustering algorithm organizes the experimental samples only on the basis of overall similarity in their gene expression patterns; these relationships are summarized in a dendrogram (Fig. 1a), in which the pattern and length of the branches reflects the relatedness of the samples[5]. Fifteen of the twenty before and after doxorubicin pairs (red dendrogram branches), and both primary tumour/lymph node metastasis pairs (light blue branches) were clustered together on terminal branches in the dendrogram; that is, despite an interval of 16 weeks, independent surgical procedures and cytotoxic chemotherapy, independent samples taken from the same tumour were in most cases recognizably more similar to each other than either was to any of the other samples. In three instances (Norway 47, 61 and 101), the 'after' chemotherapy specimens clustered in a branch of the dendrogram that also contained the three normal breast samples; we know from the clinical data that these tumours were 3 of the 20 tumours that were classified as doxorubicin 'responders' (data not shown). An analysis of the relationship between gene expression and correlations with clinical data will be reported elsewhere (T.S. *et al.*, manuscript in preparation).

The 'molecular portraits' revealed in the patterns of gene expression not only uncovered similarities and differences among the tumours, but in many cases pointed to a biological interpretation. Variation in growth rate, in the activity of specific signalling pathways, and in the cellular composition of the tumours were all reflected in the corresponding variation in the expression of specific subsets of genes. The largest distinct cluster of genes within the 1,753-gene cluster diagram was the 'proliferation cluster' (Supplementary Information Fig. 5), which is a group of genes whose levels of expression correlate with cellular proliferation rates[3,6]. Expression of this cluster of genes varied widely among the tumour samples, and was generally well correlated with the mitotic index. As one might expect, this cluster also included the genes encoding two widely used immunohistochemical markers of cell proliferation (Ki-67 and PCNA).

Several groups of co-expressed genes provided views of the activities of specific signalling and/or regulatory systems. A large cluster of genes regulated by the interferon pathway (including *STAT1*) showed substantial variation in expression among the tumours, as was previously observed in a smaller set of breast

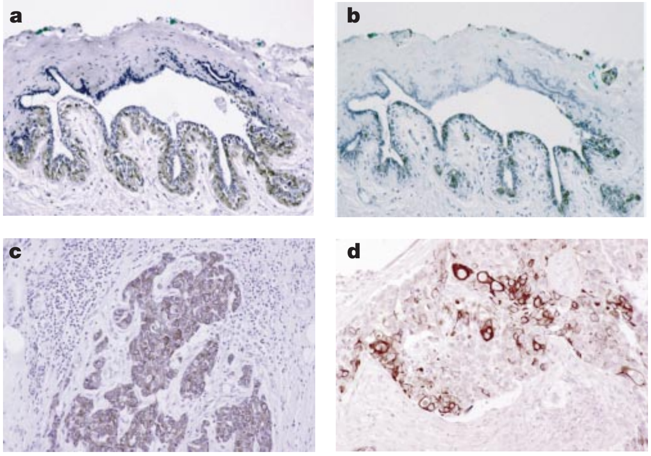

Figure 2 Breast tissue immunohistochemistry. **a**, Normal mammary duct using antibodies against the basal keratins 5/6. **b**, Normal mammary duct using antibodies against the luminal keratins 8/18 (adjacent tissues sections were used in **a** and **b**). **c**, Tumour Stanford 16 using antibodies against keratins 8/18. **d**, Tumour New York 3 using antibodies against keratins 5/6.

tumours[6]. Variation in expression of the oestrogen receptor-α gene (ER) correlated well with the direct clinical measurement of the ER protein levels in the tumours (Supplementary Information Table 3; concordance in 36/38 samples), and paralleled variation in the expression of a larger group of genes that included three other transcription factors (GATA-binding protein 3 (refs 7, 8), X-box binding protein 1 and hepatocyte nuclear factor 3α). *HER2/neu*, also known as *Erb-B2*, is overexpressed in 20–30% of all breast tumours, usually associated with DNA amplification of the *Erb-B2* locus[9,10]. Notably, most of the other genes contained within the

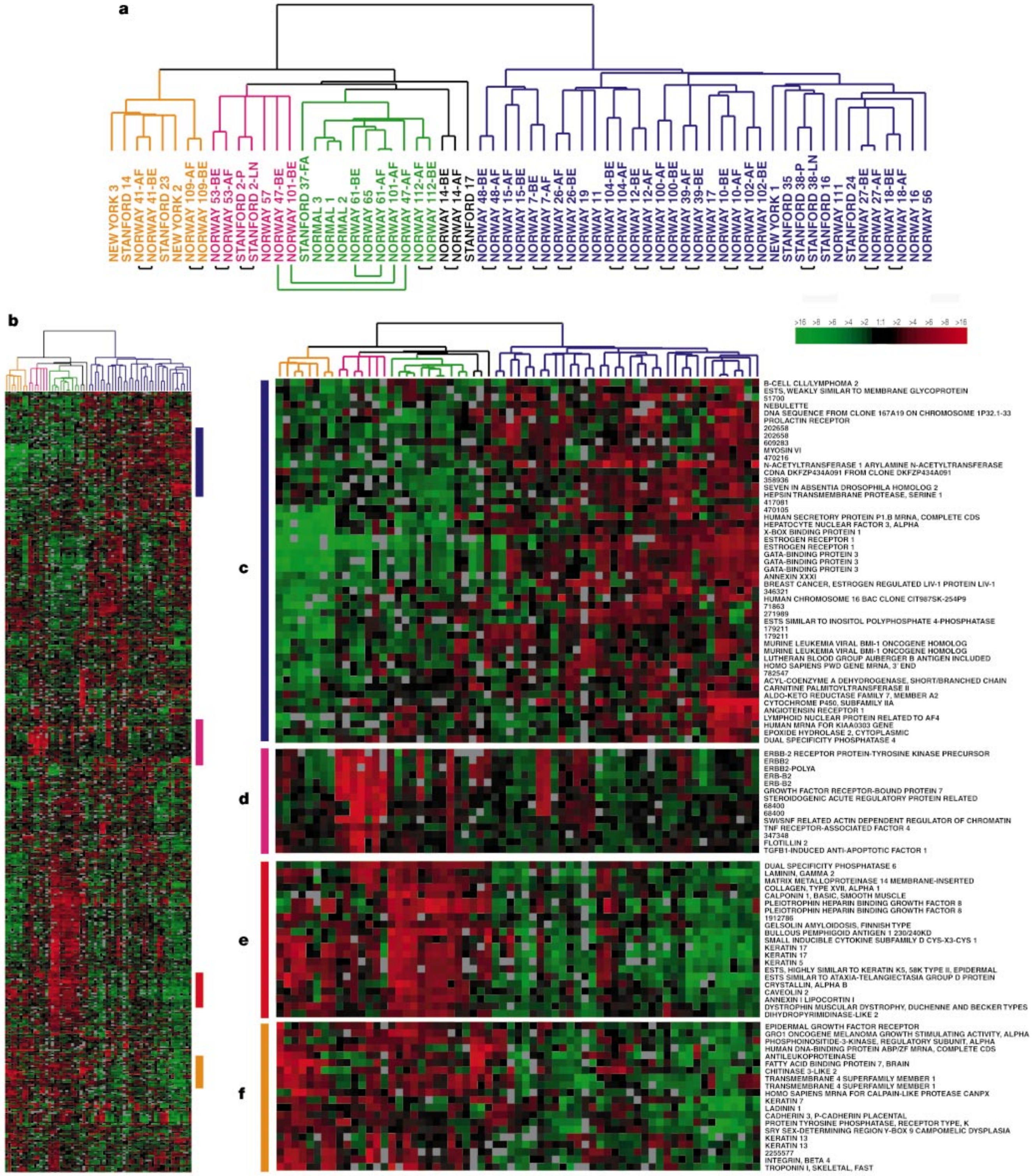

Figure 3 Cluster analysis using the 'intrinsic' gene subset. Two large branches were apparent in the dendrogram, and within these large branches were smaller branches for which common biological themes could be inferred. Branches are coloured accordingly: basal-like, orange; *Erb-B2*+, pink; normal-breast-like, light green; and luminal epithelial/ER+, dark blue. **a**, Experimental sample associated cluster dendrogram. Small black bars beneath the dendrogram identify the 17 pairs that were matched by this hierarchical clustering; larger green bars identify the positions of the three pairs that were not matched by the clustering. **b**, Scaled-down representation of the intrinsic cluster diagram (see Supplementary Information Fig. 6). **c**, Luminal epithelial/ER gene cluster. **d**, *Erb-B2* overexpression cluster. **e**, Basal epithelial cell associated cluster containing keratins 5 and 17. **f**, A second basal epithelial-cell-enriched gene cluster.

Erb-B2 cluster were located in this same region of chromosome 17, and were also amplified on the genomic DNA level (ref. 10; and J.R.P., unpublished data). Finally, a cluster of genes that included *c-Fos* and *JunB* co-varied in expression among the tumour specimens. We have found that this subset of genes is characteristically induced by prolonged handling of the samples after surgical resection (M.v.d.R. and C.M.P., unpublished data).

Human breast tumours are histologically complex tissues, containing a variety of cell types in addition to the carcinoma cells[11]. In analysing the gene expression patterns in solid human tumours, we used two lines of reasoning to infer the lineage of the cells that accounted for the apparently cell-type-specific expression of particular clustered groups of genes. First, such clusters included genes whose expression patterns have been previously characterized and that consistently pointed to a specific cell type. Second, these inferences were often corroborated by comparable expression of the same cluster in one or more of the cultured cell lines. Thus, eight independent clusters of genes appeared to reflect variation in specific cell types present within the tumours (Fig. 1c–j).

(1) Endothelial cells: a cluster of genes characteristically expressed by endothelial cells, including CD34, CD31 and von Willebrand factor were also strongly expressed in the two endothelial cell lines HUVEC and HMVEC (Fig. 1c). (2) Stromal cells: a previously characterized cluster of genes that included several isoforms of collagen showed significant variation in expression among samples (Fig. 1d)[3,6]. (3) Adipose-enriched/normal breast cells: a cluster of genes including fatty-acid binding protein 4 and PPARγ may represent the presence of adipose cells (Fig. 1g). (4) B lymphocytes: variation in expression of a cluster of genes that were highly expressed in the multiple myeloma-derived cell line RPMI-8226, including many immunoglobulin genes, appears to represent variable B-cell infiltration (Fig. 1f). (5) T lymphocytes: a cluster of genes including CD3δ and two subunits of the T-cell receptor were highly expressed in the T-cell leukaemia-derived cell line MOLT-4 and probably indicate T-cell infiltrates (Fig. 1i). (6) Macrophages: a cluster of genes that appeared to be markers of macrophage/monocytes included CD68, acid phosphatase 5, chitinase and lysozyme (Fig. 1h).

Two distinct types of epithelial cell are found in the human mammary gland: basal (and/or myoepithelial) cells and luminal epithelial cells[11,12]. These two cell types are conveniently distinguished immunohistochemically; basal epithelial cells can be stained with antibodies to keratin 5/6 (Fig. 2a), whereas luminal epithelial cells stain with antibodies against keratins 8/18 (Fig. 2b). Many genes were expressed by one of these two cell lineages, but not by the other (Fig. 1e and j). The gene expression cluster characteristic of basal epithelial cells included keratin 5, keratin 17, integrin-β4 and laminin (Fig. 1e)[11]. The gene expression cluster characteristic of the luminal cells was anchored by the previously noted cluster of transcription factors that included ER (Fig. 1j).

One goal of this study was to develop a system for classifying tumours on the basis of their gene expression patterns. The subset of genes shown in Fig. 1 was not necessarily optimal for this purpose, as the choice of genes whose expression levels provided the basis for the ordering of the tumour samples determined which phenotypic relationships among the tumours were reflected in the clustering patterns. We therefore selected an alternative subset of genes to use as the basis for a new clustering analysis.

The rationale behind this alternative gene subset was that specific features of a gene expression pattern that are to be used to classify tumours should be similar in any sample taken from the same tumour, and they should vary among different tumours. The 22 paired samples provided a unique opportunity for a deliberate and systematic search for such genes. From the genes whose expression was well measured in the 65 tissue samples, we selected a subset of 496 genes (termed the 'intrinsic' gene subset) that consisted of genes with significantly greater variation in expression between different tumours than between paired samples from the same tumour (see Supplementary Information). When variation in expression of this set of genes was used to order the tissue samples (Fig. 3; and Supplementary Information Fig. 6), 17 of the 20 'before and after' doxorubicin pairs were grouped together as were both of the tumour/lymph node metastasis pairs. Qualitatively similar sample clustering patterns were obtained when a second gene subset that focused on genes expressed by epithelial cell types, and which had only 25% overlap with the intrinsic gene subset, was used (data not shown).

The division of the tissue samples into two subgroups was a striking feature of the intrinsic gene subset cluster analysis (Fig. 3a). As a test of the robustness of this division, we applied the 'weighted voting' method[13]. This algorithm recapitulated the sorting of the tissue samples between these two subgroups for all but 1 of the 65 samples (data not shown). It is important to note, however, that there is extensive residual variation in expression patterns within each of these two broad subgroups. Indeed, many of the finer subdivisions probably have important biological properties (see below).

The two dendrogram branches in Fig. 3 largely separate the tumour samples into those that were clinically described as ER positive (blue) and those that were ER negative (other colours). The tumours in the ER+ group were characterized by the relatively high expression of many genes expressed by breast luminal cells (Fig. 3c). This connection was further corroborated using immunohistochemical analysis and antibodies against the luminal cell keratins 8/18 (Fig. 2c). With one exception, none of the tumours in this group expressed *Erb-B2* at high levels (Fig. 3d).

Many of the genes characteristic of breast basal epithelial cells were also highly expressed in a group of six clustered tumours (Fig. 3e). To corroborate the 'basal-like' characteristics of these tumours, we carried out immunohistochemistry using antibodies against the breast basal cell keratins 5/6 and 17. All six of these tumours showed staining for either keratins 5/6 or 17 or both (Fig. 2d). Notably, these six tumours also failed to express ER and most of the other genes that were usually co-expressed with it (Fig. 3c). Breast tumours that stain positive for basal keratins have been described[14–16], and basal keratins may account for 3–15% of all breast tumours[15,17–19]; in this study, the incidence was 15% (6/40).

As mentioned above, overexpression of the *Erb-B2* oncogene was associated with the high expression of a specific subset of genes. We identified a cluster of tumours that was partially characterized by the high level of expression of this subset of genes (Fig. 3d). These tumours also showed low levels of expression of ER[20,21] and of almost all of the other genes associated with ER expression—a trait they share with the basal-like tumours.

Several tumour samples and the single fibroadenoma tested (Fig. 3, light green), were clustered with a group of samples that also contained the three normal breast specimens (Fig. 3a). The 'normal breast' gene expression pattern is typified by the high expression of genes characteristic of basal epithelial cells and adipose cells, and the low expression of genes characteristic of luminal epithelial cells.

The number of clearly different molecular phenotypes observed among the breast tumours suggests that we are far from having a complete picture of the diversity of breast tumours. When hundreds (instead of tens) of breast tumours have been characterized, a more defined tumour classification is likely, and statistically significant relationships with clinical parameters should be uncovered. We were, however, able to identify four groups of samples that might be related to different molecular features of mammary epithelial biology (that is, ER+/luminal-like, basal-like, *Erb-B2*+ and normal breast). An important implication of this study is that the clinical designation of 'oestrogen receptor negative' breast carcinoma encompasses at least two biologically distinct subtypes of tumours (basal-like and *ErB-B2* positive), which may need to be treated as distinct diseases.

A striking conclusion from these data concerns the stability, homogeneity and uniqueness of the 'molecular portraits' provided by the quantitative analysis of gene expression patterns. We infer that these portraits faithfully represent the 'tumour' itself, and not merely the particular tumour 'sample', because we could recognize the distinctive expression pattern of a tumour in independent samples. The finding that a metastasis and primary tumour were as similar in their overall pattern of gene expression as were repeated samplings of the same primary tumour, suggests that the molecular program of a primary tumour may generally be retained in its metastases. Finally, we have explicitly discussed only a tiny fraction of the genes whose expression patterns varied among these tumours. Attention to the thousands of individual genes that define the molecular portraits of each tumour, and learning to interpret their patterns of variation, will undoubtedly lead to a deeper and more complete understanding of breast cancers. □

Methods

Most of the techniques used in this work have been described elsewhere[2,3,22,23], and detailed protocols are available at ⟨http://cmgm.Stanford.EDU/pbrown/⟩. The methods and protocols are also included in the Supplementary Information, and the primary data tables can be obtained at ⟨http://genome-www.stanford.edu/molecularportraits/⟩.

Received 7 February; accepted 25 May 2000.

1. Tavassoli, F. A. & Schnitt, S. J. *Pathology of the Breast* (Elsevier, New York, 1992).
2. Eisen, M. B. & Brown, P. O. DNA arrays for analysis of gene expression. *Methods Enzymol.* **303,** 179–205 (1999).
3. Ross, D. T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nature Genet.* **24,** 227–235 (2000).
4. Aas, T. *et al.* Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. *Nature Med.* **2,** 811–814 (1996).
5. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95,** 14863–14868 (1998).
6. Perou, C. M. *et al.* Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl Acad. Sci. USA* **96,** 9212–9217 (1999).
7. Yang, G. P., Ross, D. T., Kuang, W. W., Brown, P. O. & Weigel, R. J. Combining SSH and cDNA microarrays for rapid identification of differentially expressed genes. *Nucleic Acids Res.* **27,** 1517–1523 (1999).
8. Hoch, R. V., Thompson, D. A., Baker, R. J. & Weigel, R. J. GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int. J. Cancer* **84,** 122–128 (1999).
9. Pauletti, G., Godolphin, W., Press, M. F. & Slamon, D. J. Detection and quantitation of HER-2/neu gene amplification in human breast cancer archival material using fluorescence in situ hybridization. *Oncogene* **13,** 63–72 (1996).
10. Pollack, J. R. *et al.* Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nature Genet.* **23,** 41–46 (1999).
11. Ronnov-Jessen, L., Petersen, O. W. & Bissell, M. J. Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. *Physiol. Rev.* **76,** 69–125 (1996).
12. Taylor-Papadimitriou, J. *et al.* Keratin expression in human mammary epithelial cells cultured from normal and malignant tissue: relation to in vivo phenotypes and influence of medium. *J. Cell Sci.* **94,** 403–413 (1989).
13. Golub, T. R. *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286,** 531–537 (1999).
14. Dairkee, S. H., Mayall, B. H., Smith, H. S. & Hackett, A. J. Monoclonal marker that predicts early recurrence of breast cancer. *Lancet* **1,** 514 (1987).
15. Dairkee, S. H., Puett, L. & Hackett, A. J. Expression of basal and luminal epithelium-specific keratins in normal, benign, and malignant breast tissue. *J. Natl Cancer Inst.* **80,** 691–695 (1988).
16. Malzahn, K., Mitze, M., Thoenes, M. & Moll, R. Biological and prognostic significance of stratified epithelial cytokeratins in infiltrating ductal breast carcinomas. *Virchows Arch.* **433,** 119–129 (1998).
17. Guelstein, V. I. *et al.* Monoclonal antibody mapping of keratins 8 and 17 and of vimentin in normal human mammary gland, benign tumors, dysplasias and breast cancer. *Int. J. Cancer* **42,** 147–153 (1988).
18. Gusterson, B. A. *et al.* Distribution of myoepithelial cells and basement membrane proteins in the normal breast and in benign and malignant breast diseases. *Cancer Res.* **42,** 4763–4770 (1982).
19. Nagle, R. B. *et al.* Characterization of breast carcinomas by two monoclonal antibodies distinguishing myoepithelial from luminal epithelial cells. *J. Histochem. Cytochem.* **34,** 869–881 (1986).
20. Berns, E. M. *et al.* Prevalence of amplification of the oncogenes c-myc, HER2/neu, and int-2 in one thousand human breast tumors: correlation with steroid receptors. *Eur. J. Cancer* **28,** 697–700 (1992).
21. Heintz, N. H., Leslie, K. O., Rogers, L. A. & Howard, P. L. Amplification of the c-erb B-2 oncogene and prognosis of breast adenocarcinoma. *Arch. Pathol. Lab. Med.* **114,** 160–163 (1990).
22. DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278,** 680–686 (1997).
23. Alizadeh, A. A. *et al.* Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403,** 503–511 (2000).

Supplementary information is available on *Nature*'s World-Wide Web site (http://www.nature.com) or as paper copy from the London editorial office of *Nature*.

Acknowledgements

We thank W. Gerald and L. Norton for the three New York tumour specimens; M. Stampfer and P. Yaswen for the 184 sample mRNAs; and members of the P. O. Brown, D. Botstein and A.-L. Børresen-Dale labs for discussions. We are grateful to the NCI and the Howard Hughes Medical Institute who provided support for this research. C.M.P. is a SmithKline Beecham Pharmaceuticals Fellow of the Life Sciences Research Foundation. T.S. is a research fellow of the Norwegian Cancer Society. M.B.E. is an Alfred P. Sloan Foundation Postdoctoral Fellow in Computational Molecular Biology. D.T.R. is a Walter and Idun Berry Fellow. P.O.B. is an Associate Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to D.B.
(e-mail: botstein@genome.stanford.edu) or P.O.B.
(e-mail: pbrown@cmgm.stanford.edu).

Potential for biomolecular imaging with femtosecond X-ray pulses

Richard Neutze*, Remco Wouts*, David van der Spoel*, Edgar Weckert†‡ & Janos Hajdu*

* *Department of Biochemistry, Biomedical Centre, Box 576, Uppsala University, S-75123 Uppsala, Sweden*

† *Institut für Kristallographie, Universität Karlsruhe, Kaiserstrasse 12, D-76128, Germany*

Sample damage by X-rays and other radiation limits the resolution of structural studies on non-repetitive and non-reproducible structures such as individual biomolecules or cells[1]. Cooling can slow sample deterioration, but cannot eliminate damage-induced sample movement during the time needed for conventional measurements[1,2]. Analyses of the dynamics of damage formation[3–5] suggest that the conventional damage barrier (about 200 X-ray photons per Å^2 with X-rays of 12 keV energy or 1 Å wavelength[2]) may be extended at very high dose rates and very short exposure times. Here we have used computer simulations to investigate the structural information that can be recovered from the scattering of intense femtosecond X-ray pulses by single protein molecules and small assemblies. Estimations of radiation damage as a function of photon energy, pulse length, integrated pulse intensity and sample size show that experiments using very high X-ray dose rates and ultrashort exposures may provide useful structural information before radiation damage destroys the sample. We predict that such ultrashort, high-intensity X-ray pulses from free-electron lasers[6,7] that are currently under development, in combination with container-free sample handling methods based on spraying techniques, will provide a new approach to structural determinations with X-rays.

Radiation damage is caused by X-ray photons depositing energy directly into the sample. At 1 Å wavelength, the photoelectric cross-section of carbon is about 10 times higher than its elastic-scattering cross-section, making the photoelectric effect the primary source of damage. The photoelectric effect is a resonance phenomenon in which a photon is absorbed and an electron ejected[8], usually from a low-lying orbital of the atom (about 95% of the photoelectric events remove K-shell electrons from carbon, nitrogen, oxygen and sulphur), producing a hollow ion with an unstable electronic configuration. Relaxation is achieved through an electron from a higher shell falling into the vacant orbital. In heavy elements this usually gives rise to X-ray fluorescence, whereas in light elements the falling electron is more likely to give up its energy to another electron, which is then ejected in the Auger effect. Auger emission is predominant in carbon, nitrogen, oxygen and sulphur ($> 95\%$)[9]; thus, most photoelectric events ultimately remove two electrons from these elements. These two electrons have different energies ($\sim$12 keV for photoelectrons and $\sim$0.25 keV for Auger electrons),