

## Commentary

# Molecular profiling of breast cancer: portraits but not physiognomy

James D Brenton\*, Samuel AJR Aparicio<sup>†</sup> and Carlos Caldas<sup>†</sup>

\*Department of Medical Oncology and Hematology, Ontario Cancer Institute/Princess Margaret Hospital, Toronto, Ontario, Canada

<sup>†</sup>Department of Oncology and Cambridge Institute for Medical Research/Wellcome Trust Centre for Molecular Mechanisms in Disease, University of Cambridge, Wellcome Trust/MRC Building, Cambridge, UK

**Correspondence:** JD Brenton, Department of Medical Oncology and Hematology, Ontario Cancer Institute/Princess Margaret Hospital, 610 University Avenue, Room 5-224, Toronto, Ont M5G 2M9, Canada. Tel: +1 416 946 2245; fax: +1 416 946 2082; e-mail: [jdb1003@uhnres.utoronto.ca](mailto:jdb1003@uhnres.utoronto.ca)

Received: 27 September 2000

Revisions requested: 17 October 2000

Revisions received: 31 October 2000

Accepted: 1 November 2000

Published: 18 December 2000

*Breast Cancer Res* 2001, **3**:77–80

© 2001 BioMed Central Ltd  
(Print ISSN 1465-5411; Online ISSN 1465-542X)

## Abstract

Breast cancers differ in response to treatment and may have a divergent clinical course despite having a similar histopathological appearance. New technology using DNA microarrays provides a systematic method to identify key markers for prognosis and treatment response by profiling thousands of genes expressed in a single cancer. Microarray profiling of 38 invasive breast cancers now confirms striking molecular differences between ductal carcinoma specimens and suggests a new classification for oestrogen-receptor negative breast cancer. Future approaches will need to include methods for high-throughput clinical validation and the ability to analyze microscopic samples.

**Keywords:** breast cancer, cDNA microarray, hierarchical clustering, molecular profiling

## Introduction

The decision to provide adjuvant treatment in breast cancer remains a clinical judgement based on classical pathology and surgical staging. The measurement of oestrogen receptor (ER) status is one of a very small number of molecular markers for solid tumours that have a large impact both on estimation of prognosis and the choice of therapy for the individual patient. Despite great efforts to identify better markers, the results in breast cancer have been unimpressive. This may be because many studies have focused on a single gene, protein or pathway. cDNA microarrays, in contrast, offer a systematic method to perform very extensive expression profiling within a single cancer specimen. The enabling technology has been the development of affordable robots that can spot thousands of gene probes onto glass microscope slides in high-density arrays. Fluorescently labelled tumour cDNA and a differently labelled control sample are hybridized together under a cover slip onto the array. The

ratio of the two intensities at each spot indicates the relative expression of that gene within the tumour. The technique is suitable for high-throughput analysis and this opens the way for molecular rather than phenotypic taxonomies, which will include new prognostic and treatment response features [1].

Perou *et al* use this technology to ask whether novel classifications in breast cancer can be identified from a series of 65 paired and non-paired breast specimens that included 36 ductal carcinomas and 2 lobular carcinomas [2]. Their results underscore the large variability in expression phenotype between individual cancers with the same histology and suggest a new subclassification of ER-negative breast cancer.

## Microarray profiles

The analysis of Perou *et al* used 84 arrays, each containing 8102 probes, producing over 680,000 data points. To

make sense of this large data set, the authors used an exploratory analysis tool called two-dimensional hierarchical clustering. In essence, this tool reorders genes and patient samples to reveal new groupings or clusters with similar patterns of gene expression (see Fig. 1) [3]. The result is displayed as a dendrogram in which the tumours with the greatest similarity are placed closest together on the tree. Tumours with greater differences are added to more distant nodes and leaves as the ordering process continues. A colour matrix summarizes the magnitude of expression for each gene across the tumours and allows the visual inspection of common expression patterns.

The initial clustering examined 1753 genes that showed at least a fourfold change in expression ratio in three or more of the samples. Despite the fact that most of the cases were invasive ductal carcinomas, there was great diversity of gene expression between cases and this involved many different gene groups. This disparity between tumours is perhaps not surprising given the complex genetic and epigenetic changes that occur in breast cancer secondary to loss of normal DNA repair and checkpoint control. However, there was very striking similarity between each of the 22 paired patient samples. Twenty of the pairs were collected before and after doxorubicin-based treatment in a clinical trial of primary medical therapy for advanced breast cancer, and two were synchronous breast primary and lymph node metastasis. The paired samples were consistently clustered together, indicating that, despite the effects of time and chemotherapy treatment, each cancer maintained an expression 'fingerprint'. This conservation of expression is not unique to breast cancer and has also been shown in lymphoma and melanoma samples [4,5]. However, as shown here, 'fingerprint' clusters may represent the most stable features of the cancer, such as histology or proliferative rate, that can remain constant over many years.

Perou *et al* then reanalyzed the data to better discriminate between the tumours, taking advantage of the similarity information from the paired tumours. They defined a 'within-between' score to identify genes that showed sig-

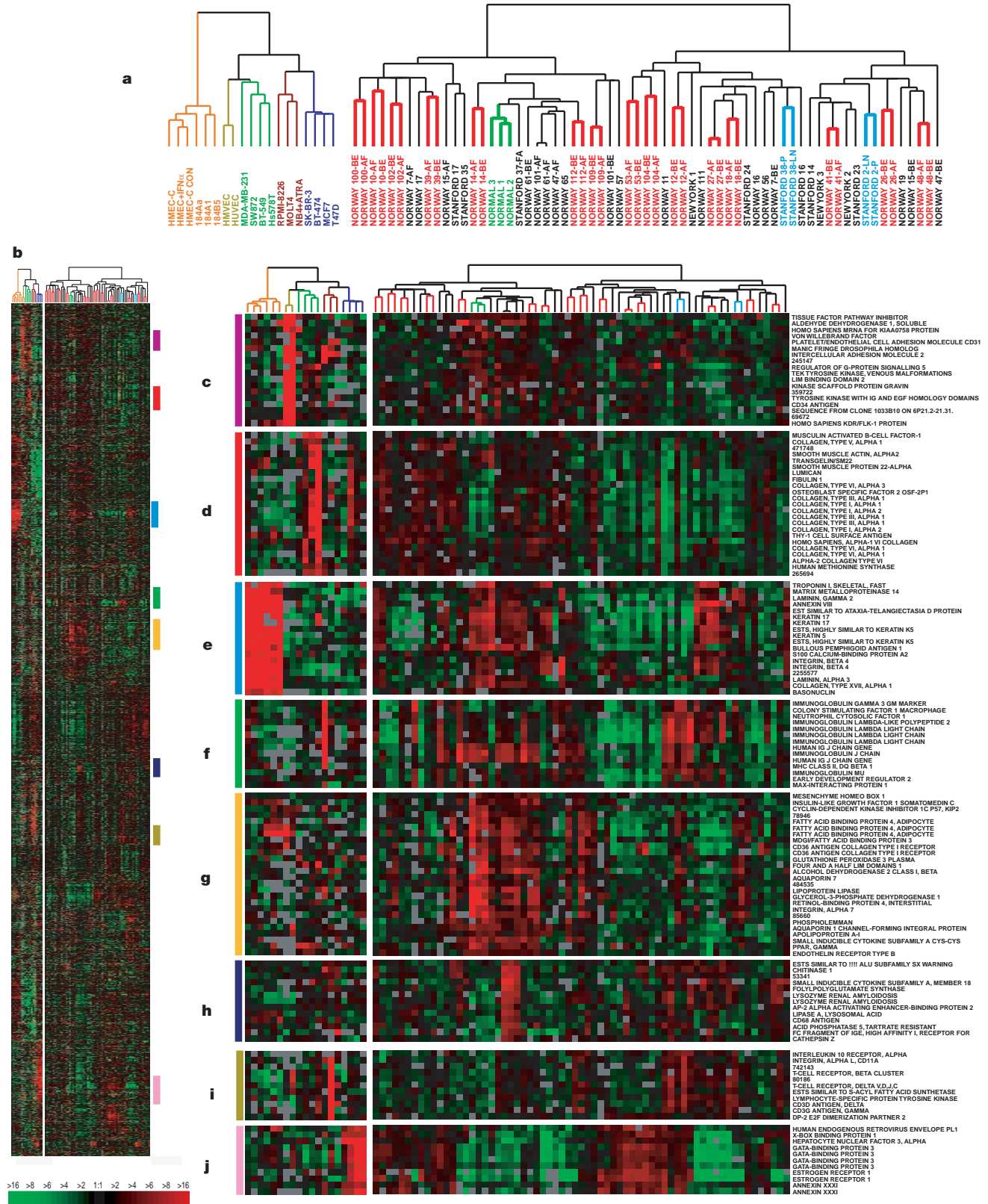
nificantly greater variation in expression between separate cases as compared to within each pair samples. These scores were then used to recluster the complete set. The reclustering with this 'intrinsic' subset of 496 genes suggested four new breast cancer classifications: ER+ with luminal epithelial cell expression; ER- with basal epithelial expression; *Erb-B2*+ and ER-/ER low; and normal breast with basal epithelial and adipose cell expression. Most breast cancers arise from luminal ductal epithelium and only 3–15% of cases are thought to arise from basal (myoepithelial) epithelium. The differing origins can be distinguished by examining specific cytokeratins, and the microarray findings were confirmed by performing immunohistochemistry on paraffin material from the parent tumours. This suggestion that ER- cancers may encompass two separate groups is very intriguing, especially as there is some independent evidence that basal epithelial cancers have a worse outcome [6]. Correlation with clinical and outcome data for this series, however, is not yet available, and it is possible that this clustering may not be found in a larger series. Perou *et al* did, however, identify the same classification by reclustering with a separate set of genes only expressed by epithelial cells (this new set had a 25% overlap with the intrinsic group). It is also likely, given the large size of the data set, that other important patterns will be discovered. Indeed, two-dimensional hierarchical clustering is only one of several mathematical methods that can be used to cluster the data set and has the disadvantage that its agglomerative method can prevent the identification of complex gene relationships.

## Conclusion

The validation of gene cluster information presents a key challenge for the best use of microarray data. There is no method at the present time for calculating the significance of a particular cluster prediction. The use of microarray technology is also expensive, both in amount of tumour material required and reagent costs. It will not be feasible for many investigators to extend this approach to very large series to improve correlation with clinical features or predictive value. An alternative approach may be to intensively investigate highly selected sample sets (chosen for

Variation in expression of 1753 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). (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 mesenchymal-like cell lines in dark green, and the lymphocyte-derived cell lines in brown. (b) Scaled-down representation of the 1753-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. The Figure and legend are reproduced here, with permission, from [2]. The "supplementary information Figure 4" referred to above is also part of [2], and can be accessed there.

Figure 1



commentary

review

reports

primary research

different outcome or response to treatment) in order to generate candidate clusters. A minimal set of genes can then be chosen [7] for validation on tissue microarrays [8]. These arrays are pathology slides that contain up to 600 minute paraffin sections, each from a different patient. High-throughput testing can be performed for each candidate marker in a manner analogous to expression arrays, using either immunohistochemistry or fluorescent *in situ* hybridization. These analyses should ideally be carried out on patient material garnered from prospective, randomized clinical trials so that statistical analysis is based on high-quality, independent data.

A disadvantage of profiling macroscopic specimens, as performed in this article, is that gene expression is averaged across all cells in the sample. The authors of the present study highlight the fact that their methods were sensitive enough to detect expression signatures from admixed non-cancer cells. Patterns unique to infiltrating lymphocytes and vascular endothelium, for example, were present as discrete clusters and could be correlated with signatures from lymphocyte and vascular cell lines (Fig. 1). It is inevitable despite this sensitivity that heterogeneity within the cancer itself will be diluted, which is important because there is considerable variation in chromosomal dosage between breast cancer cells when examined by fluorescent *in situ* hybridization. Striking changes in copy number can be seen within a single histological section and between primary cancers and their metastases (Gray *et al*, unpublished data). These dosage changes will give rise to clonal differences in gene expression. Whether distinct subclones within the cancer determine the prognosis will have to be addressed by analyzing microscopic biopsies from laser capture microdissection or fine needle sampling. Robust amplification methods that can preserve RNA expression ratios from such small samples are evolving [9]. It has recently been shown that comparative genome hybridization for chromosomal dosage may be performed at much higher resolution by hybridizing labelled genomic DNA to microarrays (array CGH) [10–12], and this is feasible from very small amounts of genomic DNA. Given the inherent variability of expression arrays [13], combining both techniques for the analysis of microscopic biopsies may offer a more rigorous framework for preliminary analysis of expression data [14].

Should this new classification of ER-negative cancers become part of the clinical assessment of breast cancer? Not yet, despite the interesting associations found in this work, because there is no independent evidence that new prognostic markers have been identified. We should not, however, have to wait too long for independent corroboration because the number of microarray profiles in the literature will rapidly increase along with the use of tissue microarrays. More fine brushwork should then allow the cancer portraits to reveal their underlying character.

## References

1. Lockhart DJ, Winzeler EA: **Genomics, gene expression and DNA arrays.** *Nature* 2000, **405**:827–836.
2. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO, Botstein D: **Molecular portraits of human breast tumours.** *Nature* 2000, **406**:747–752.
3. Eisen MB, Spellman PT, Brown PO, Botstein D: **Cluster analysis and display of genome-wide expression patterns.** *Proc Natl Acad Sci USA* 1998, **95**:14863–14868.
4. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J, Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM: **Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling.** *Nature* 2000, **403**:503–511.
5. Bittner M, Meltzer P, Chen Y, Jiang Y, Seftor E, Hendrix M, Radmacher M, Simon R, Yakhini Z, Ben-Dor A, Sampas N, Dougherty E, Wang E, Marincola F, Gooden C, Lueders J, Glatfelter A, Pollock P, Carpten J, Gillanders E, Leja D, Dietrich K, Beaudry C, Berens M, Alberts D, Sondak V: **Molecular classification of cutaneous malignant melanoma by gene expression profiling.** *Nature* 2000, **406**:536–540.
6. Malzahn K, Mitze M, Thoenes M, Moll R: **Biological and prognostic significance of stratified epithelial cytokeratins in infiltrating ductal breast carcinomas.** *Virchows Arch* 1998, **433**:119–129.
7. Hastie T, Tibshirani R, Eisen MB, Alizadeh A, Levy R, Staudt L, Chan WC, Botstein D, Brown P: **'Gene shaving' as a method for identifying distinct sets of genes with similar expression patterns.** *Genome Biol* 2000, **1**:1–21.
8. Kononen J, Bubendorf L, Kallioniemi A, Barlund M, Schraml P, Leighton S, Torhorst J, Mihatsch MJ, Sauter G, Kallioniemi OP: **Tissue microarrays for high-throughput molecular profiling of tumor specimens.** *Nat Med* 1998, **4**:844–847.
9. Wang E, Miller LD, Ohnmacht GA, Liu ET, Marincola FM: **High-fidelity mRNA amplification for gene profiling.** *Nat Biotechnol* 2000, **18**:457–459.
10. Pinkel D, Segraves R, Sudar D, Clark S, Poole I, Kowbel D, Collins C, Kuo WL, Chen C, Zhai Y, Dairkee SH, Ljung BM, Gray JW, Albertson DG: **High resolution analysis of DNA copy number variation using comparative genomic hybridization to microarrays.** *Nat Genet* 1998, **20**:207–211.
11. Pollack JR, Perou CM, Alizadeh AA, Eisen MB, Pergamenschikov A, Williams CF, Jeffrey SS, Botstein D, Brown PO: **Genome-wide analysis of DNA copy-number changes using cDNA microarrays.** *Nat Genet* 1999, **23**:41–46.
12. Heiskanen MA, Bittner ML, Chen Y, Khan J, Adler KE, Trent JM, Meltzer PS: **Detection of gene amplification by genomic hybridization to cDNA microarrays.** *Cancer Res* 2000, **60**:799–802.
13. Lee ML, Kuo FC, Whitmore GA, Sklar J: **Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations.** *Proc Natl Acad Sci USA* 2000, **97**:9834–9839.
14. Forozan F, Mahlamaki EH, Monni O, Chen Y, Veldman R, Jiang Y, Gooden GC, Ethier SP, Kallioniemi A, Kallioniemi OP: **Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data.** *Cancer Res* 2000, **60**:4519–4525.