Chapter 10

Identification of high risk of recurrence in breast-cancer patients by DNA chip technology

Thomas Karn, Andre Ahr, Uwe Holtrich and Manfred Kaufmann

Introduction

Breast cancer is a major cause of death among women aged 35–55 years, affecting about 10% in Western countries. Despite important advances in therapy, still more than half of the affected patients suffer from relapses (Ries *et al*. 2000). This is in part due to the highly heterogeneous nature of the disease; the various pathological breast cancer subclasses have markedly different clinical courses and treatment responses. In patients with breast cancer, assessment of axillary lymph nodes and status of steroid hormone receptors are the most important prognostic factors, because they can be used to predict disease-free and overall survival, and to direct adjuvant systemic therapy. At the moment, most patients with lymph-node-negative disease (i.e., with no evidence that cancer cells have spread beyond the primary tumour) can be effectively treated with surgery and local radiation. Patients with more aggressive disease can benefit from adjuvant chemotherapy or hormone therapy and are currently identified according to a combination of criteria (Eifel *et al*. 2000; Goldhirsch *et al*. 2001): age, the size of the tumour, axillary-node status, the histologic type and pathological grade of cancer, and hormone-receptor status as depicted in Table 10.1. For the outcome of an individual patient, the currently available prognostic factors are associated with a broad range of risk of recurrence. Thus, the ability of these criteria to predict individual disease progression and clinical outcome is imperfect. This uncertainty in forecasting outcome means that some patients who need adjuvant treatment do not receive it, whereas others are unnecessarily treated and as a result are exposed to the risk of side effects without good reason. Improved tools are clearly needed for the assessment of prognosis in breast cancer. A major goal, therefore, is the development of an individual risk-profile system with high accuracy and reproducibility to estimate patients' prognosis and best treatment.

Table 10.1 Prognostic factors and guidelines for adjuvant therapy

Prognostic factor	*Low risk*	*High risk*	
Lymph node status	Negative	negative and	positive or
Tumour stage	T < 1 cm	T 1–2 cm or	T > 2 cm or
Grading	1	2–3 and	—
Hormone receptor	Positive	positive	negative or
Age	> 35 yrs	> 35 yrs	< 35 yrs
Adjuvant therapy (St Gallen 2001)	nil or tamoxifen	tamoxifen +/- chemotherapy	chemotherapy + tamoxifen (if ER- or PR-positive)

DNA chip technology and expression profiling

The completion of the Human Genome Project and the development of new, high-performance screening techniques have revolutionised the ways in which researchers can study the pathogenesis of disease. Analysis of the levels of expression of thousands of genes in parallel with the use of DNA chips has shown distinct patterns in different kinds of tumour. Because the expression of the genes is measured, such analysis is mostly referred as expression profiling. We can use these patterns to classify histologically similar tumours into specific subtypes (Alizadeh *et al*. 2001), a process that provides clinically relevant information. Studies on mammary carcinomas could already categorise several subtypes of breast cancer (Alizadeh *et al* 2000; Perou *et al*. 2000). However, these studies mostly lacked correlation with classic clinical variables and follow-up data. Global determination of cellular transcriptional activity is expected to identify gene expression signatures that predict clinical behaviour of tumours.

Figure 10.1 gives a representation of this new technology: a so-called DNA chip or microarray basically represents a glass slide, which carries probes for all human genes. The probes are spotted as a systematic array and the location of each probe is known to the investigator. An RNA sample from the tumour is now labelled and applied to this chip. Afterwards the chip is analysed with special equipment and the activity of each gene can be deduced from the intensity of the corresponding spot. The final result of a DNA chip analysis tells us the activity of all human genes in the cells of the corresponding sample. We can imagine this as if you are looking inside the cells and monitor the cellular program that is running there. The large amount of data from those analyses has to be analysed further by using mathematical methods to identify marker sets of interest. In addition, these multidimensional data mostly do not make immediate sense to the human eye. However, computational analyses can identify relationships in the data from different tumours and present them for example as tree-like structures or as scatter plots scaled in three dimensions.

Identification of high-risk patients

In a previous study, we applied DNA chip analyses to identify differentially expressed genes and to evaluate transcriptional diversity among human breast cancers (Ahr *et*

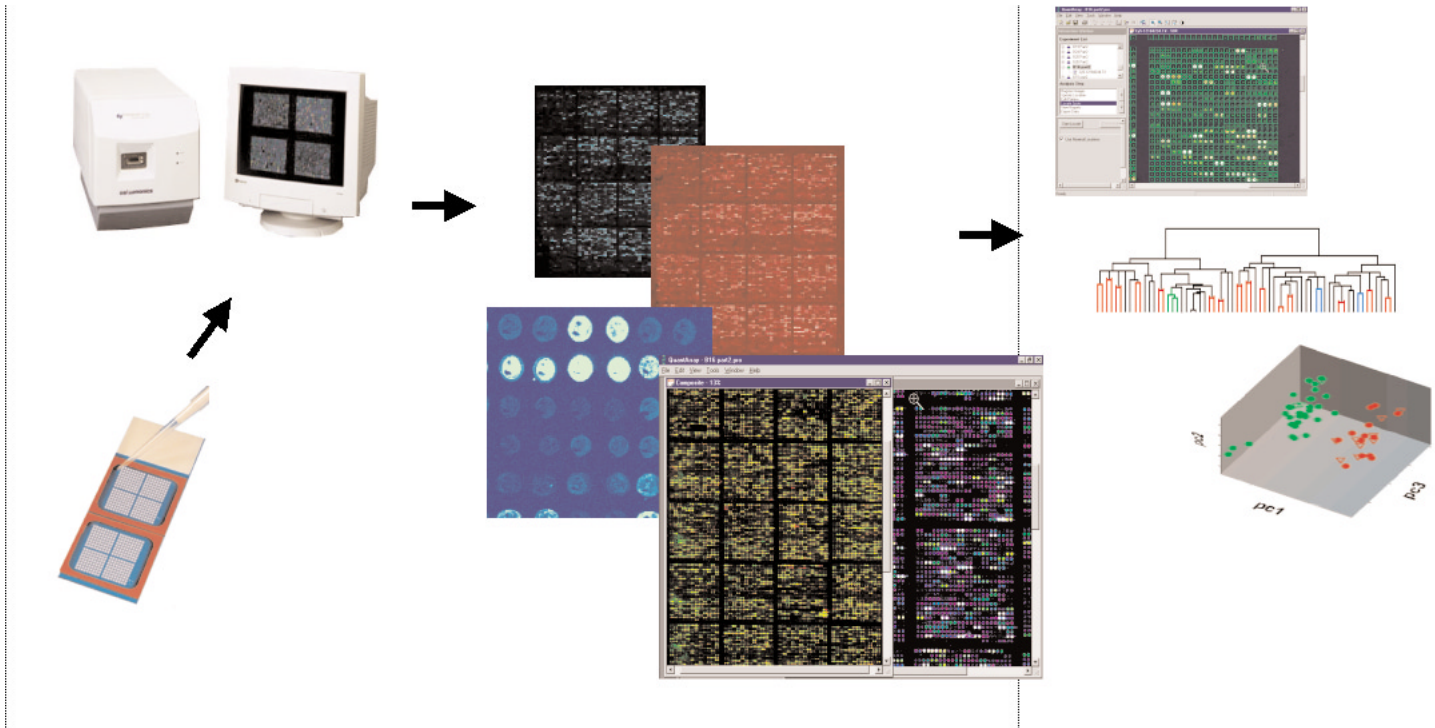


Figure 10.1 Principle of the DNA chip technology. Left: labelled RNA from tumour samples is applied to the DNA chip and analysis is performed in a special fluorescence laser scanner. Middle: on the obtained electronic images each spot represents a probe for a individual human gene (note the microscopic spots in magnification) and the intensity of the spot correlates with the activity of this specific gene in the sample. The colour visualisation represents an array of thousands of those spots measuring the activity of thousands of different genes in the sample. Right: data are analysed by mathematical and statistical methods using special software; phenotypic relationships of different tumours can be graphically visualised.

al. 2001). The detected differentially expressed transcripts include several genes known from the literature, as well as previously unrecognised transcripts. We showed that class discovery analysis based on our gene expression profiling of 82 breast tissue specimens as well as some reference samples like cell lines identifies four main sample groups. A correlation of the cluster data with classical clinicopathological parameters revealed that one subgroup was characterised by a remarkably high number of node-positive tumours and a disproportionate number of patients who had already developed distant metastases at the time of diagnosis. These cluster analysis data suggested that the technique could help to define patients with an early onset of disease progression, providing a first step towards improved patient-adapted therapy.

Figure 10.2 shows a colour representation of the activity of 41 marker genes in 94 samples and as already mentioned, it is confusing for the human eye and difficult to find a structure in such data. However, computers can identify relationships between the tissue samples as represented by the tree-structure on the right. So we can analyse the resulting groups of patients according to their clinical data. Much easier to look at is the visualisation method in Figure 10.3, a principal component analysis (PCA) of the data from Figure 10.2, reducing the amount of information. From this we can clearly discriminate between two major tumour groups designated class A (shown in red) and non-A (green) using this method. When one looks now at the clinical data of the patients, class A is characterised by many node-positive cases, suggesting a high risk of relapse for patients in this subgroup. This observation forced us to analyse the follow-up of the patients. And we found that nearly half of the patients in this group did suffer from a relapse in less than 2 years in contrast to only 11% in the other group (Ahr *et al*. 2002). Figure 10.4 shows the data for patients with available follow-up once more visualised by the tree like structure: The patients with a relapse are marked by red dots here; the clustering of those patients in class A, which is represented by red branches in this tree, is obvious. This accumulation cannot be explained by the number of lymph-node-positive patients, because these numbers are comparable in both groups. Strikingly, the high risk of patients in class A is further highlighted by the observation that three of the five node-negative patients in class A had a relapse in contrast to none of the 15 node-negative patients in class non-A.

Although validation studies with larger numbers need to be done, several lines of evidence support the suggestion that tumours of class A represent cancers with a high risk of recurrence. First, our initial clustering of the sample collective revealed an accumulation of tumours that had already developed distant metastases at diagnosis. Second, although class A and non-class-A contained similar numbers of node-positive tumours, progression was limited mainly to class A. Finally, we saw progression of node-negative tumours only in class A. Taken together, our cluster analysis identifies breast-cancer patients with a high risk of recurrence, and is a step towards the establishment of an individual risk-profile system. Future directions should combine these molecular methods with the standard tumour classification system to obtain improved patient-tailored therapies.

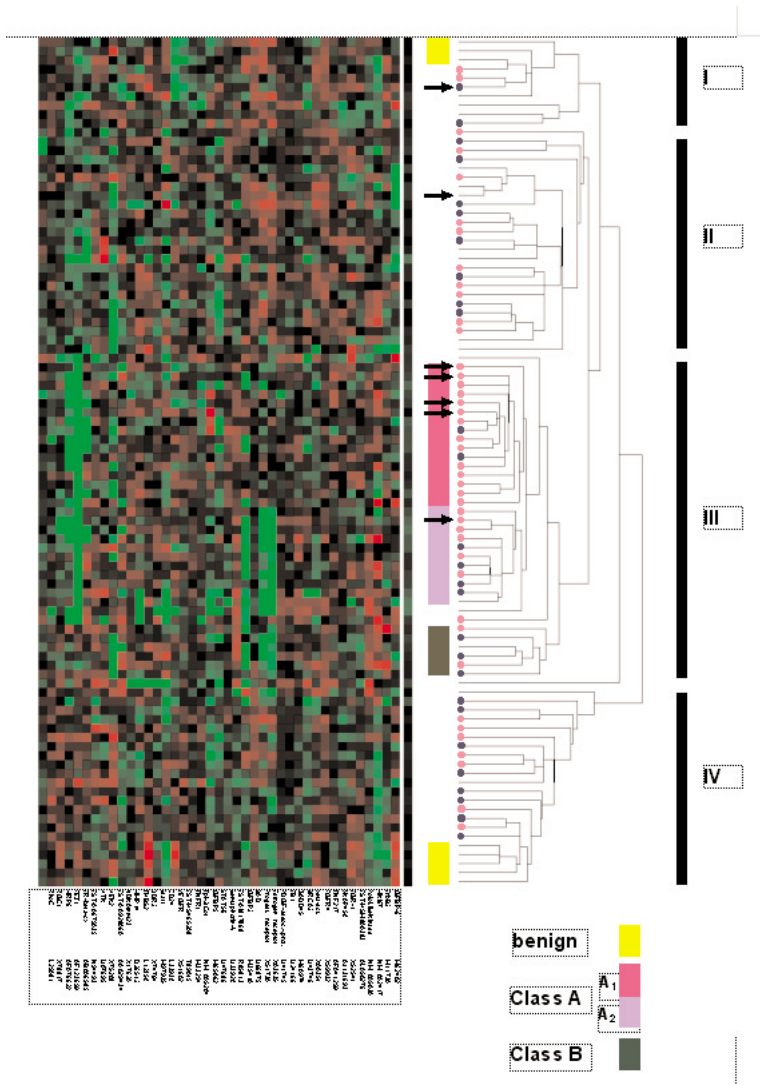
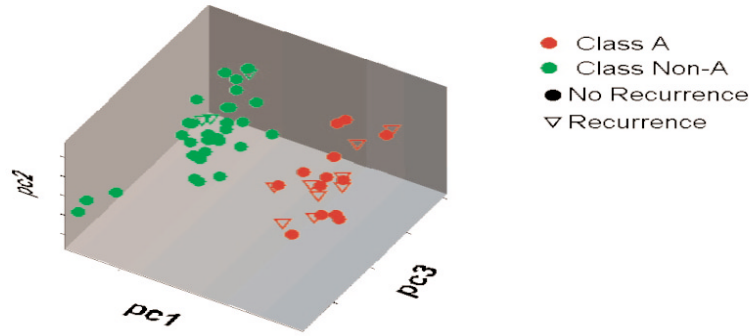


Figure 10.2 Molecular classification of breast cancers. Genes repeatedly found to be differentially expressed in array analyses were applied in real-time polymerase chain reaction assays (TaqMan-technology) to perform a molecular tumour similarity classification of 82 normal and malignant breast specimens as well as reference samples. Red indicates expression levels above median, green below. Each sample is represented by a horizontal line of the matrix. The columns refer to the analysed marker genes. The corresponding unrooted tree, where branch lengths represent distances (1, Pearson correlation coefficient) of samples as judged by their expression patterns, is depicted on the right. The four main sample groups (I–IV) are indicated by vertical bars on the right. For further information see Ahr *et al.* (2001) and online resources at http://www.kgu.de/zfg/dnachip.



follow up 23.5 months	class A	class Non-A	
relapses (total)	9 (45%)	3 (11%)	p = 0.016
relapses (N0 only)	3 of 5	0 of 15	p = 0.009

Lancet 2002; 359: 131–32

Figure 10.3 Discrimination of carcinoma classes. Fifty-five mammary carcinoma samples were characterised by expression profiling and the classification visualised by a principal component analysis. Tumours can be separated in two classes by using these methods: class A (red symbols) and class non-A (green symbols). The incidence of relapses is more than four times higher in class A, and recurrences in node-negative patients were only seen in class A. For further information see Ahr *et al.* (2002) and online resources at http://www.kgu.de/zfg/dnachip.

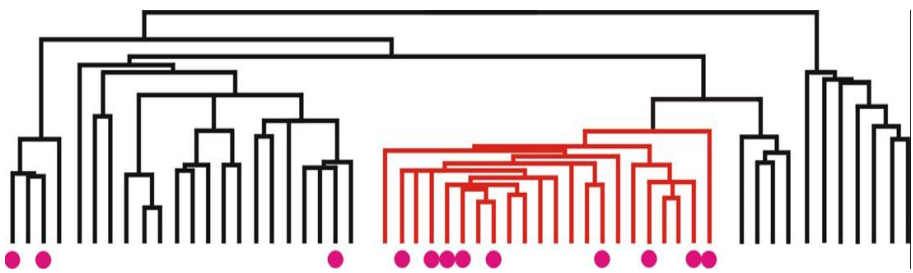


Figure 10.4 Detection of high-risk patients. The 55 mammary carcinoma samples were grouped according to marker gene activity by hierarchical clustering with the Pearson correlation using the program CLUSTER (Stanford University, California, USA) Branch length represents similarity distances of samples as judged by their expression patterns. Class A breast cancers are represented by red branches. Tumour samples (T1–T3) of patients with recurrences during follow-up are marked by red dots. For further information see Ahr *et al.* (2002) and online resources at http://www.kgu.de/zfg/dnachip.

Conclusions

Several studies on breast cancer have already been done using these novel techniques: In 2000, the pioneering microarray group from Stanford University categorised several new subtypes of breast cancer (Perou *et al*. 2000), and later on showed a correlation of these subtypes with survival data (Sorlie *et al*. 2001). In 2001, a group from the National Institutes of Health was able to distinguish hereditary cases of breast cancer (*BRCA1*, *BRCA2* and sporadic cases) according to their expression profiles (Hedenfalk *et al*. 2001). In January 2002, in addition to our results (Ahr *et al*. 2002), a group from The Netherlands Cancer Institute established a similar profiling system, allowing them the identification of those lower-risk patients who seem not to need an adjuvant treatment (van't Veer *et al*. 2002).

In summary, expression profiling can already distinguish known subtypes of breast cancer. In addition, however, novel subgroups have been identified using this technology, which differ in their clinical behaviour and which will have an impact on future treatment decisions. Finally, it is anticipated that further studies will supply us with profiles that allow the individual prediction of response to a specific therapy.

References

Ahr, A., Holtrich, U., Solbach, C., Scharl, A., Strebhardt, K., Karn, T. & Kaufmann, M. (2001). Molecular classification of breast cancer patients by gene expression profiling. *Journal of Pathology* **195**, 312–320.

Ahr A., Karn T., Solbach C., Seiter T., Strebhardt K., Holtrich U. & Kaufmann M. (2002). Identification of high risk breast-cancer patients by gene expression profiling. *The Lancet* **359**, 131–132.

Alizadeh, A. A., Eisen, M. B., Davis, R. E. *et al*. (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511.

Alizadeh, A. A., Ross, D. T., Perou, C. M. & van de Rijn, M. (2001). Towards a novel classification of human malignancies based on gene expression patterns. *Journal of Pathology* **195**, 41–52.

Eifel, P., Axelson, J. A. & Costa, J. *et al*. (2001). National Institutes of Health Consensus Development Conference Statement: adjuvant therapy for breast cancer, November 1–3, 2000. *Journal of the National Cancer Institute* **93**, 979–989.

Goldhirsch, A., Glick, J. H., Gelber, R. D., Coates, A. S., Senn, H. J. (2001). Meeting highlights: International Consensus Panel on the Treatment of Primary Breast Cancer: Seventh International Conference on Adjuvant Therapy of Primary Breast Cancer. *Journal of Clinical Oncology* **19**, 3817–3827.

Hedenfalk I, Duggan D, Chen Y, *et al*. (2001). Gene-expression profiles in hereditary breast cancer. *New England Journal of Medicine* **344**, 539–548.

Perou, C. M., Sorlie, T., Eisen, M. B. *et al*. (2000). Molecular portraits of human breast tumours. *Nature* **406**, 747–752.

Ries, L. A. G., Eisner, M. P., Kosary, C. L. *et al*. (eds) (2000). *SEER Cancer Statistics Review*, 1973–1997. Bethesda, Maryland: National Cancer Institute.

Sorlie, T., Perou, C. M., Tibshirani, R. *et al*. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 10,869–10,874.

van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A. M., Mao, M., Peterse, J. L., van der Kooy, K., Marton, M. J., Witteveen, A. T. *et al*. (2002). Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**, 530–536.

Online resources

http://www.kgu.de/zfg/dnachip/
http://www.gene-chips.com/