**Research letters**

# Identification of high risk breast-cancer patients by gene expression profiling

*André Ahr, Thomas Karn, Christine Solbach, Tanja Seiter, Klaus Strebhardt, Uwe Holtrich, Manfred Kaufmann*

**We previously used DNA array analyses in the molecular profiling of breast cancers. By cluster analysis of 55 patients, we identified a subpopulation of breast cancers—designated class A—that contained a high number of nodal-positive tumours and that had frequently developed distant metastases at the time of diagnosis. We have now analysed follow-up data from these patients. We found that, despite a median of only 23·5 months of follow-up, 11 of 22 patients in class A progressed to metastatic disease, and three of five patients classified as having a nodal status of N0 in this subpopulation developed distant metastases. Our analysis identifies breast-cancer patients with a high risk of disease recurrence, and could act as a first step towards improved patient-adapted therapy.**

*Lancet* 2002; **359:** 131–32

Many tumours can be divided into pathological subclasses which need to be identified by use of genetic markers if therapy and follow-up strategies are to be optimised. DNA chip technology allows parallel expression profiling of several thousand genes, enabling the characterisation of complex cellular transcriptional activities. These DNA arrays have been used to identify gene expression patterns in various types of cancer tissues. One major aim is to use DNA arrays to classify tumours into categories on the basis of shared gene expression patterns.[1,2] Studies on mammary carcinomas can already categorise several subtypes of breast cancers.[3,4] However, these studies lack 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.

In patients with breast cancer, assessment of axillary lymph nodes and status with regard to steroid hormone receptors are the most important prognostic factors, since they can be used to predict disease-free and overall survival and to direct adjuvant systemic therapy. Concerning the outcome of an individual patient, the currently available prognostic factors are associated with a broad range of risk of recurrence. 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.

We have previously done cDNA array expression analyses in an attempt to establish a molecular profile for breast cancers.[5] Candidate marker genes for expression profiling were collected by array analyses of 15 ductal and two lobular mammary carcinomas.[5] Selection criteria for 41 marker genes included a cutoff of 10-fold expression difference among carcinomas, thereby ensuring each marker was relatively robust. Genes only sporadically altered were withdrawn. Also added to the panel were genes reported in published studies to be useful expression markers for mammary carcinomas. We took tumour samples from 55 consecutive patients attending our clinic for treatment between June, 1997, and October, 1998, and grouped them according to their expression of the 41 marker genes by hierarchical clustering with the Pearson correlation using the program CLUSTER (Stanford, CA, USA).[2,3] The result of the hierarchical clustering was verified by self-organising maps using the program GENECLUSTER (Cambridge, MA, USA), which forced the creation of two clusters: class A and non-class-A. The consistency of prediction of the class A samples was cross validated by the class prediction method,[1] which resulted in a median prediction strength of 0·73 for class A; median values seen in 500 random iterations were in the range of 0·1–0·4.

| Characteristic | Class A (n=20) | Non-class-A (n=27) |
|---|---|---|
| **Tumour stage*** | | |
| T1 | 4 | 10 |
| T2 | 14 | 16 |
| T3 | 2 | 1 |
| **Nodal status** | | |
| N0 | 5 | 15 |
| N1 | 15 | 12 |
| **Recurrences** | | |
| Total | 9 | 3 |
| N0 patients only | 3/5 | 0/15 |

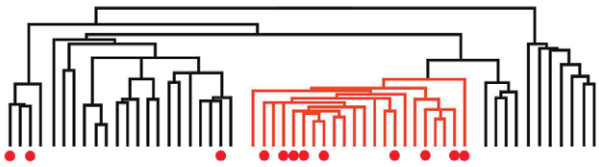
Median follow-up time 27 months, IQR 10·75 (class A), 24 months IQR 12·0 (non-class A) *Patients with T4 tumours (n=8) were omitted.

**Characteristics of patients**

The class A subgroup had a high proportion of patients with nodal-positive tumours (17 of 22) and with distant metastases at the time of diagnosis (23% in this subgroup compared with 4% among the rest of the patients).[5] These findings prompted us to examine whether class A tumours could predict a high risk of relapse in patients with breast cancer.

We have now obtained follow-up data on the 55 patients with primary breast cancer from our collective. Analysis of these follow up data revealed that despite a short observation time (median 23·5 months [IQR 14·25]), 11 of the 22 patients in class A progressed to metastatic disease. The high risk of recurrence of tumours in class A was further highlighted when we analysed only patients with tumour stage T1–T3 (T4 cases were excluded from the analysis because they were at high a priori risk of recurrence). Nine of 20 cases in class A had recurrences, compared with only three of 27 cases outside of this group (p=0·016) (table, figure). This association persisted after a possible confounding effect of nodal status was taken into account (p=0·023, Mantel-Haenszel test). Three of five patients classified as N0 in class A developed distant metastases during follow-up (p=0·009, table).

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 the time of diagnosis. Second, although class A and non-class-A contained similar numbers



**Class discovery of primary breast cancers by cluster analysis**
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.

of nodal-positive tumours, progression was limited mainly to class A. Finally, we saw progression of nodal-negative tumours only in class A. Taken together, our cluster analysis identifies breast-cancer patients with a high risk of recurrences, 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.

*Contributors*

A Ahr, M Kaufmann, T Karn and U Holtrich had the original idea and designed the overall study. The molecular analyses were developed and done by U Holtrich and T Karn. Tumour biopsies were collected by C Solbach and K Strebhardt. T Seiters and A Ahr handled the patient data and analysed the association between patient classification and survival data. A Ahr, M Kaufmann, T Karn and U Holtrich analysed the results and wrote the manuscipt.

*Conflict of interest statement*

None declared

*Acknowlegments*

We thank Silke Deckert and Katherina Kourtis for technical assistance. This work was supported by a grant from the Deutsche Krebshilfe (10-1478-Ka2). The sponsors had no role in the study design; data collection, analysis or interpretation; or in the writing of the paper. The Genbank accession numbers of the 41 marker genes are as follows: AJ131693, U66879, U43746, L33930, L11315, U24166, X52541, M11730, N54493, AA029434, R85813, T89015, AL050276, X03635, AB006589, AF123659, X59932, M60974, X75208, U07695, M35410, M65062, M62402, NM_002417, D26512, AF076622, X17620, U41745, NM_005030, X51730, X78817, L25081, U33920, M97935, U47686, L12350, X74764, M33294, NM_005204, X51602, and AF041259.

1 Golub TR, Slonim DK, Tamayo P, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999; **286:** 531–37.

2 Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 2000; **403:** 503–11.

3 Perou CM, Sorlie T, Eisen MB, et al. Molecular portraits of human breast tumours. *Nature* 2000; **406:** 747–52.

4 Hedenfalk I, Duggan D, Chen Y, et al. Gene-expression profiles in hereditary breast cancer. *N Engl J Med* 2001; **344:** 539–48.

5 Ahr A, Holtrich U, Solbach C, et al. Molecular classification of breast cancer patients by gene expression profiling. *J Pathol* 2001; **195:** 312–20.

**Department of Obstetrics and Gynecology, J W Goethe University, Theodor-Stern-Kai 7, D-60590 Frankfurt, Germany** (A Ahr MD, T Karn PhD, C Solbach MD, T Seiter MD, K Strebhardt PhD, U Holtrich PhD, Prof M Kaufmann MD)

**Correspondence to:** Dr André Ahr (e-mail: ahr@em.uni-frankfurt.de)

# A novel mechanism for thalassaemia intermedia

*C Badens, M G Mattei, A M Imbert, C Lapouméroulie, N Martini, G Michel, D Lena-Russo*

**Thalassaemia intermedia is a moderate form of thalassaemia resulting from various genetic defects. We report an undescribed mechanism leading to this condition: a somatic deletion of the β-globin gene in the haemopoietic lineage of a heterozygous β-thalassaemic patient. We did molecular studies and haemoglobin analysis of the patient and his parents. We found that the deletion gives rise to a mosaic of cells with either one or no functional β-globin gene and it extends to a region of frequent loss of heterozygosity called LOH11A, which is located close to the β-globin locus. Thus, loss of heterozygosity can be a cause of non-malignant genetic disease.**

*Lancet* 2002; **359:** 132–33

Thalassaemia intermedia is a clinical term used to describe patients who have mild thalassaemia with no regular transfusion requirements. The phenotype is highly heterogeneous, ranging in severity from severe anaemia with hepatosplenomegaly and thalassaemia-like bone modifications to moderate microcytic hypochromic anaemia. In terms of genotype, thalassaemia intermedia is also heterogeneous, and has been shown to arise in four distinct ways:[1] inheritance of mild β-thalassaemia alleles; part rescue of β-thalassaemia major by hereditary persistence of fetal haemoglobin; exacerbation of the heterozygous state by coinheritance of an extra α-globin gene, which worsens imbalance of the globin chains; or inheritance of a dominant β-thalassaemia mutation. We report a novel mechanism by which severity of the heterozygous state is increased leading to thalassaemia intermedia.

The propositus was first examined at age 10 years for anaemia, hepatosplenomegaly, and growth failure. He had a typical thalassaemic face, and skull radiographs showed a typical hair-on-end appearance. Analysis of his haemoglobin showed HbA 71%, HbF 25%, and $HbA_2$ 4%, and in-vitro globin chain synthesis gave an α/β globin-chain ratio of 3·9. Concentration of ferritin was 367 μg/L and of bilirubin, 60 μmol/L. The individual is now 18 years old and has never had a blood transfusion. His haemoglobin concentration has remained between 70 and 80 g/L.

The patient's father had a β-thalassaemic trait, with $HbA_2$ 5·1% and microcytosis, whereas his mother had a normal red-cell index, haemoglobin electrophoretic profile, and α/β globin-chain ratio. Results of molecular studies showed that the patient's father has the Mediterranean β-thalassaemia nonsense mutation at codon 39 (C→T), which the patient inherited, whereas his mother does not have any common β-thalassaemia mutations.

We screened the maternally inherited allele of the propositus for a rare or de novo mutation by cloning and sequencing a PCR fragment spanning the entire maternally-derived β gene (from –240 to +1665 nucleotide relative to the cap site). We did not identify a mutation. With PCR, we also showed that neither the 3·7 kb deletion nor the anti-3·7 kb triplication was present at the α gene locus.

After digestion of the patient's leucocyte DNA with the restriction enzyme *Mae* I to detect the β-thalassaemia mutation at codon 39, we noted that the band

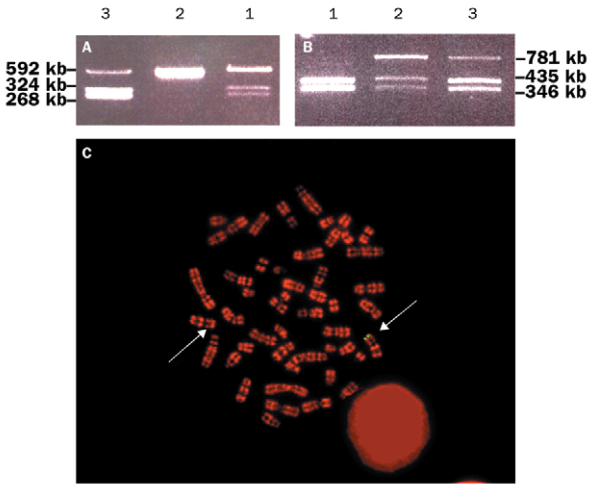



Figure 1: **Restriction enzyme digestion and FISH analysis**
PCR fragments of (a) exon 2 of the β-globin gene digested by Mae I and (b) IVS 2 of the γ-globin gene digested by Hind III. Lane 1=father; lane 2=mother, lane 3=propositus. (c) FISH on metaphases from lymphocytes of the propositus with a probe specific for the β-globin gene extending from 2 kb upstream of the δ-globin gene to the 3′ end of the β-globin gene