

24. Chorvatova, A., Gendron, L., Bilodeau, L., Gallo-Payet, N. & Payet, M. D. A Ras-dependent chloride current activated by adrenocorticotropin in rat adrenal zona glomerulosa cells. *Endocrinology* **141**, 684–692 (2000).
25. Tong, J. *et al.* NF1-regulated adenylyl cyclase pathway. *Soc. Neurosci. Abstr.* abstract no. 345.9 (Society for Neuroscience, New Orleans, 2000).
26. Ingram, D. A. *et al.* Hyperactivation of p21(ras) and the hematopoietic-specific Rho GTPase, Rac2, cooperate to alter the proliferation of neurofibromin-deficient mast cells in vivo and in vitro. *J. Exp. Med.* **194**, 57–69 (2001).
27. Jacks, T. *et al.* Tumour predisposition in mice heterozygous for a targeted mutation in *Nf1*. *Nature Genet.* **7**, 353–361 (1994).
28. Umanoff, H., Edelman, W., Pellicer, A. & Kucherlapati, R. The murine *N-ras* gene is not essential for growth and development. *Proc. Natl Acad. Sci. USA* **92**, 1709–1713 (1995).
29. Voikar, V., Koks, S., Vasar, E. & Rauvala, H. Strain and gender differences in the behaviour of mouse lines commonly used in transgenic studies. *Physiol. Behav.* **72**, 271–281 (2001).
30. Blanton, M. G., Lo Turco, J. J. & Kriegstein, A. R. Whole cell recording from neurons in slices of reptilian and mammalian cerebral cortex. *J. Neurosci. Methods* **30**, 203–210 (1989).

## Acknowledgements

We thank V. Manne for the BMS191563, and E. Friedman for technical assistance in earlier experiments. We are grateful to M. Barad, D. Buonomano, T. Cannon, J. Colicelli, P. Frankland, L. Kaczmarek, A. Matynia, M. Sanders and D. Smith for discussions, and to C. Brannan and S. Schlussel for encouragement. R.M.C. received support from the Graduated Program in Basic and Applied Biology (GABBA) of the University of Oporto, the Portuguese Foundation for Science and Technology (FCT) and the National Neurofibromatosis Foundation (NNF). This work was also supported by a generous donation from K. M. Spivak, and by grants from the NIH (R01 NS38480), Neurofibromatosis Inc. (National, Illinois, Mass Bay Area, Minnesota, Arizona, Kansas and Central Plains, Mid-Atlantic, and Texas chapters), the Merck and the NNF foundations to A.J.S.

## Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.J.S. (e-mail: Silvaa@mednet.ucla.edu).

## Gene expression profiling predicts clinical outcome of breast cancer

Laura J. van 't Veer<sup>\*†</sup>, Hongyue Dai<sup>‡</sup>, Marc J. van de Vijver<sup>\*†</sup>, Yudong D. He<sup>‡</sup>, Augustinus A. M. Hart<sup>\*</sup>, Mao Mao<sup>‡</sup>, Hans L. Peterse<sup>\*</sup>, Karin van der Kooy<sup>\*</sup>, Matthew J. Marton<sup>‡</sup>, Anke T. Witteveen<sup>\*</sup>, George J. Schreiber<sup>‡</sup>, Ron M. Kerkhoven<sup>\*</sup>, Chris Roberts<sup>‡</sup>, Peter S. Linsley<sup>‡</sup>, René Bernards<sup>\*</sup> & Stephen H. Friend<sup>‡</sup>

<sup>\*</sup> Divisions of Diagnostic Oncology, Radiotherapy and Molecular Carcinogenesis and Center for Biomedical Genetics, The Netherlands Cancer Institute, 121 Plesmanlaan, 1066 CX Amsterdam, The Netherlands

<sup>‡</sup> Rosetta Inpharmatics, 12040 115th Avenue NE, Kirkland, Washington 98034, USA

<sup>†</sup> These authors contributed equally to this work

Breast cancer patients with the same stage of disease can have markedly different treatment responses and overall outcome. The strongest predictors for metastases (for example, lymph node status and histological grade) fail to classify accurately breast tumours according to their clinical behaviour<sup>1–3</sup>. Chemotherapy or hormonal therapy reduces the risk of distant metastases by approximately one-third; however, 70–80% of patients receiving this treatment would have survived without it<sup>4,5</sup>. None of the signatures of breast cancer gene expression reported to date<sup>6–12</sup> allow for patient-tailored therapy strategies. Here we used DNA microarray analysis on primary breast tumours of 117 young patients, and applied supervised classification to identify a gene expression signature strongly predictive of a short interval to distant metastases ('poor prognosis' signature) in patients without tumour cells in local lymph nodes at diagnosis (lymph node negative). In addition, we established a signature that identifies tumours of *BRCA1* carriers. The poor prognosis signature consists of genes regulating cell cycle, invasion, metastasis and

angiogenesis. This gene expression profile will outperform all currently used clinical parameters in predicting disease outcome. Our findings provide a strategy to select patients who would benefit from adjuvant therapy.

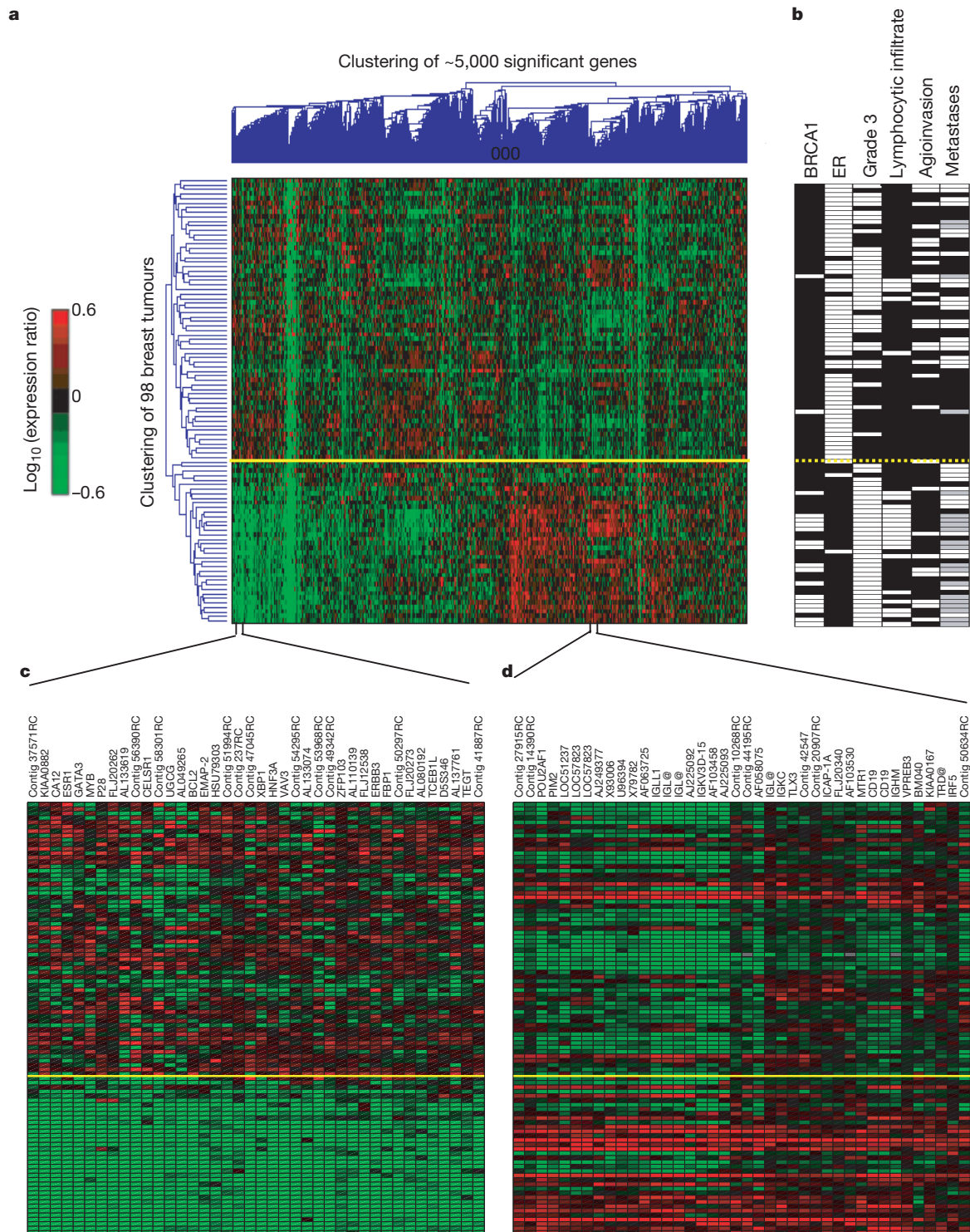
We selected 98 primary breast cancers: 34 from patients who developed distant metastases within 5 years, 44 from patients who continued to be disease-free after a period of at least 5 years, 18 from patients with *BRCA1* germline mutations, and 2 from *BRCA2* carriers. All 'sporadic' patients were lymph node negative, and under 55 years of age at diagnosis. From each patient, 5 µg total RNA was isolated from snap-frozen tumour material and used to derive complementary RNA (cRNA). A reference cRNA pool was made by pooling equal amounts of cRNA from each of the sporadic carcinomas. Two hybridizations were carried out for each tumour using a fluorescent dye reversal technique on microarrays containing approximately 25,000 human genes synthesized by inkjet technology<sup>13</sup>. Fluorescence intensities of scanned images were quantified, normalized and corrected to yield the transcript abundance of a gene as an intensity ratio with respect to that of the signal of the reference pool<sup>14</sup>. Some 5,000 genes were significantly regulated across the group of samples (that is, at least a twofold difference and a *P*-value of less than 0.01 in more than five tumours).

An unsupervised, hierarchical clustering algorithm allowed us to cluster the 98 tumours on the basis of their similarities measured over these approximately 5,000 significant genes. Similarly, the ~5,000 genes were clustered on the basis of their similarities measured over the group of 98 tumours (Fig. 1a). In the dendrograms shown in Fig. 1a (left and top), the length and the subdivision of the branches displays the relatedness of the breast tumours (left) and the expression of the genes (top). Two distinct groups of tumours are the dominant feature in this two-dimensional display (top and bottom of plot, representing 62 and 36 tumours, respectively), suggesting that the tumours can be divided into two types on the basis of this set of ~5,000 significant genes. Notably, in the upper group only 34% of the sporadic patients were from the group who developed distant metastases within 5 years, whereas in the lower group 70% of the sporadic patients had progressive disease (Fig. 1b). Thus, using unsupervised clustering we can already, to some extent, distinguish between 'good prognosis' and 'poor prognosis' tumours.

To gain insight into the genes of the dominant expression signatures, we associated them with histopathological data; for example, oestrogen receptor (ER)- $\alpha$  expression as determined by immunohistochemical (IHC) staining (Fig. 1b). Out of 39 IHC-stained tumours negative for ER- $\alpha$  expression (ER negative), 34 clustered together in the bottom branch of the tumour dendrogram. In the enlargement shown in Fig. 1c, a group of downregulated genes is represented containing both the ER- $\alpha$  gene (*ESR1*) and genes that are apparently co-regulated with ER, some of which are known ER target genes. A second dominant gene cluster is associated with lymphocytic infiltrate and includes several genes expressed primarily by B and T cells (Fig. 1d).

Sixteen out of eighteen tumours of *BRCA1* carriers are found in the bottom branch intermingled with sporadic tumours. This is consistent with the idea that most *BRCA1* mutant tumours are ER negative and manifest a higher amount of lymphocytic infiltrate<sup>15</sup>. The two tumours of *BRCA2* carriers are part of the upper cluster of tumours and do not show similarity with *BRCA1* tumours. Neither high histological grade nor angioinvasion is a specific feature of either of the clusters (Fig. 1b). We conclude that unsupervised clustering detects two subgroups of breast cancers, which differ in ER status and lymphocytic infiltration. A similar conclusion has also been reported previously<sup>7,16</sup>.

The 78 sporadic lymph-node-negative patients were selected specifically to search for a prognostic signature in their gene expression profiles. Forty-four patients remained free of disease

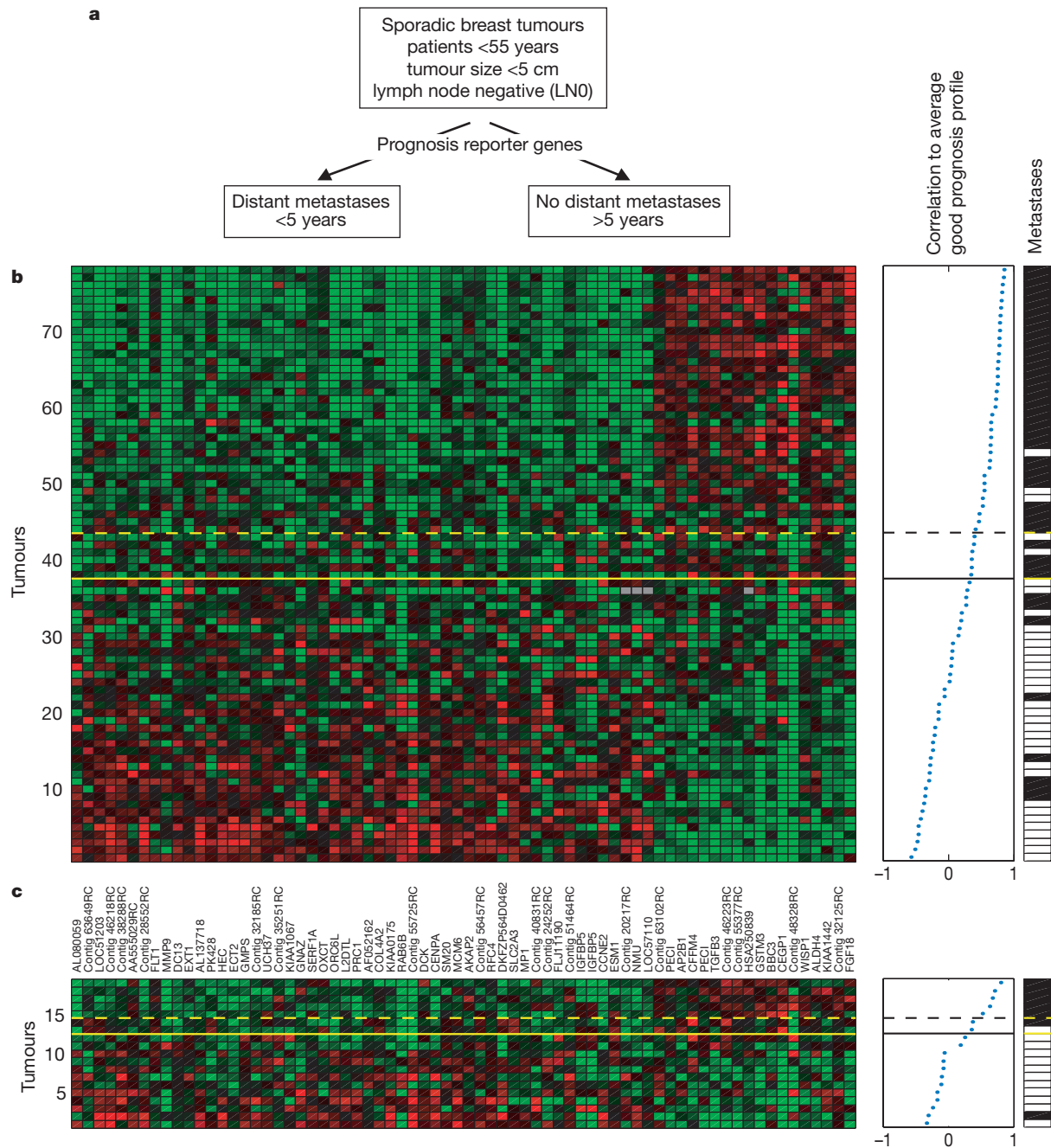


**Figure 1** Unsupervised two-dimensional cluster analysis of 98 breast tumours. **a**, Two-dimensional presentation of transcript ratios for 98 breast tumours. There were 4,968 significant genes across the group. Each row represents a tumour and each column a single gene. As shown in the colour bar, red indicates upregulation, green downregulation, black no change, and grey no data available. The yellow line marks the subdivision into two dominant tumour clusters. **b**, Selected clinical data for the 98 patients in **a**: *BRCA1* germline mutation carrier (or sporadic patient), ER expression, tumour grade 3 (versus grade 1 and 2), lymphocytic infiltrate, angioinvasion, and metastasis status. White indicates positive, black negative and grey denotes tumours derived from *BRCA1*

germline carriers who were excluded from the metastasis evaluation. The cluster below the yellow line consists of 36 tumours, of which 34 are ER negative (total 39 ER-negative) and 16 are carriers of the *BRCA1* mutation (total 18). **c**, Enlarged portion from **a** containing a group of genes that co-regulate with the ER- $\alpha$  gene (*ESR1*). Each gene is labelled by its gene name or accession number from GenBank. Contig ESTs ending with RC are reverse-complementary of the named contig EST. **d**, Enlarged portion from **a** containing a group of co-regulated genes that are the molecular reflection of extensive lymphocytic infiltrate, and comprise a set of genes expressed in T and B cells. (Gene annotation as in **c**.)

after their initial diagnosis for an interval of at least 5 years (good prognosis group, mean follow-up of 8.7 years), and 34 patients had developed distant metastases within 5 years (poor prognosis group, mean time to metastases 2.5 years) (Fig. 2a). To identify reliably good and poor prognostic tumours, we used a powerful three-step supervised classification method, similar to those used previously<sup>8,17,18</sup>. In brief, approximately 5,000 genes (significantly regulated in more than 3 tumours out of 78) were selected from the

25,000 genes on the microarray. The correlation coefficient of the expression for each gene with disease outcome was calculated and 231 genes were found to be significantly associated with disease outcome (correlation coefficient  $< -0.3$  or  $> 0.3$ ). In the second step, these 231 genes were rank-ordered on the basis of the magnitude of the correlation coefficient. Third, the number of genes in the 'prognosis classifier' was optimized by sequentially adding subsets of 5 genes from the top of this rank-ordered list and



**Figure 2** Supervised classification on prognosis signatures. **a**, Use of prognostic reporter genes to identify optimally two types of disease outcome from 78 sporadic breast tumours into a poor prognosis and good prognosis group (for patient data see Supplementary Information Table S1). **b**, Expression data matrix of 70 prognostic marker genes from tumours of 78 breast cancer patients (left panel). Each row represents a tumour and each column a gene, whose name is labelled between **b** and **c**. Genes are ordered according to their correlation coefficient with the two prognostic groups. Tumours are ordered by their correlation to the average profile of the good prognosis group (middle panel). Solid line,

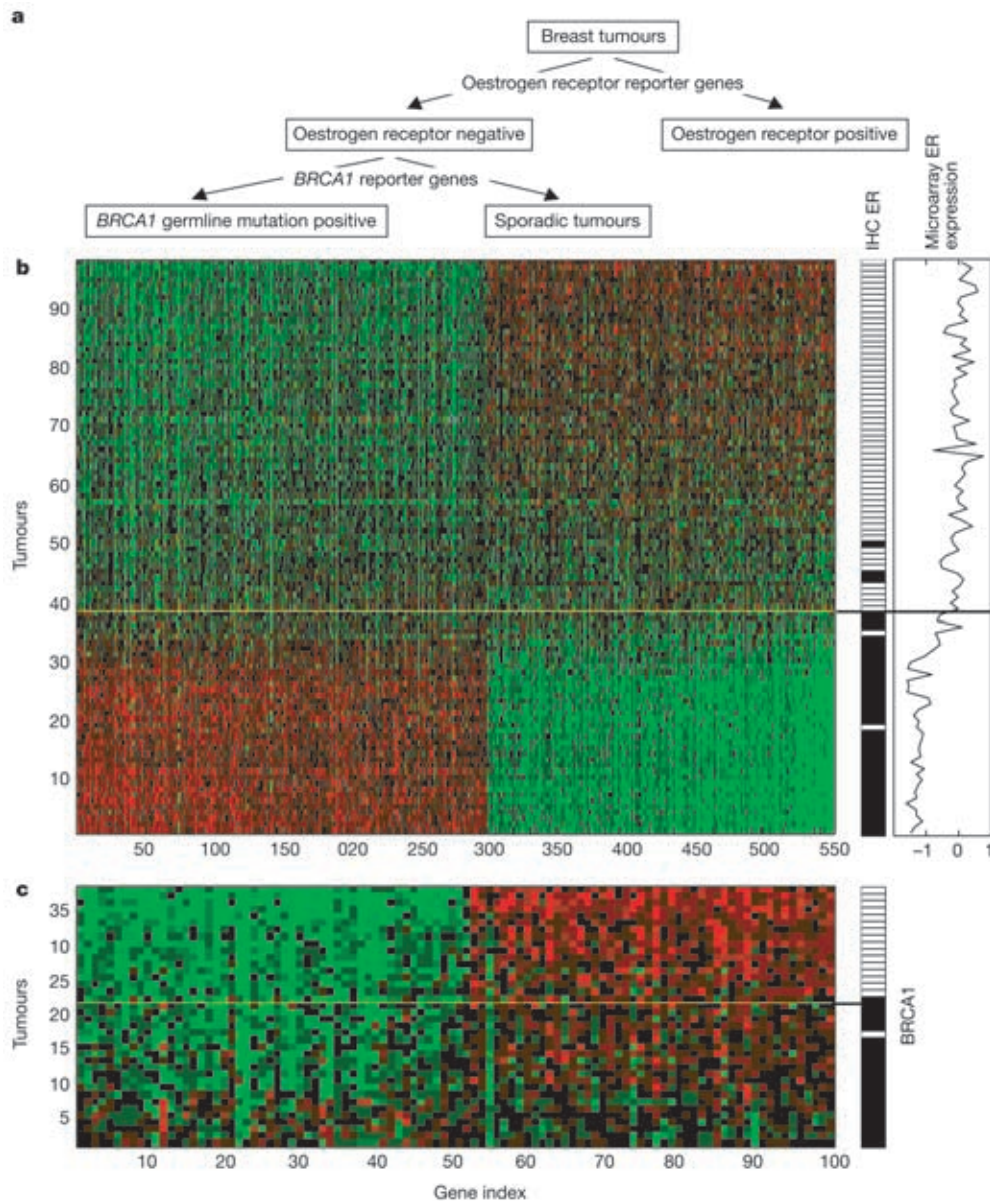
prognostic classifier with optimal accuracy; dashed line, with optimized sensitivity. Above the dashed line patients have a good prognosis signature, below the dashed line the prognosis signature is poor. The metastasis status for each patient is shown in the right panel: white indicates patients who developed distant metastases within 5 years after the primary diagnosis; black indicates patients who continued to be disease-free for at least 5 years. **c**, Same as for **b**, but the expression data matrix is for tumours of 19 additional breast cancer patients using the same 70 optimal prognostic marker genes. Thresholds in the classifier (solid and dashed line) are the same as **b**. (See Fig. 1 for colour scheme.)



evaluating its power for correct classification using the 'leave-one-out' method for cross-validation (see Supplementary Information). Classification was made on the basis of the correlations of the expression profile of the 'leave-one-out' sample with the mean expression levels of the remaining samples from the good and the poor prognosis patients, respectively. The accuracy improved until the optimal number of marker genes was reached (70 genes).

The expression pattern of the 70 genes in the 78 samples is shown in the colour plot of Fig. 2b (left panel), where tumours were ordered by rank according to their correlation coefficients with the average good prognosis profile (Fig. 2b, middle panel). The classifier predicted correctly the actual outcome of disease for 65 out of the 78 patients (83%), with respectively 5 poor prognosis and 8 good

prognosis patients assigned to the opposite category (Fig. 2b, threshold 'optimal accuracy', solid line). However, for the selection of patients eligible for adjuvant systemic therapy, a lower number of poor prognosis patients assigned to the good prognosis category should be attained. For this purpose, we set a threshold that resulted in misclassification of no more than 10% of the poor prognosis patients (3 patients out of 34 of the poor prognosis group). This optimized sensitivity threshold resulted in a total of 15 misclassifications: 3 poor prognosis tumours were classified as good prognosis, and 12 good prognosis tumours were classified as poor prognosis (Fig. 2b, dashed line). We classified tumours having a gene expression profile with a correlation coefficient above the 'optimized sensitivity' threshold (dashed line) as a good prognosis



**Figure 3** Supervised classification on ER and *BRCA1* signatures. **a**, Outline of a two-level classification system: 98 breast tumours are first classified into an ER-positive group and an ER-negative group, which is further divided into *BRCA1* mutation and sporadic tumours. **b**, Expression data matrix of the 98 sporadic tumours across 550 optimal ER reporter genes. The contrasting patterns discriminate between tumours with an ER-negative signature (below solid line) and an ER-positive signature (above solid line). The reporter genes were ordered on the basis of their level of contribution to the classifiers. Tumours are arranged according to the leave-one-out correlation coefficients to the

average signatures of the classifier. The ER status, as determined by IHC and microarray, are indicated in the two right panels. **c**, Expression data matrix of 38 ER-negative tumours defined by the ER classifier over the 100 optimal *BRCA1* reporter genes. The degree of the patterns divides the tumours in the ER-negative group into two subgroups: *BRCA1*-like and sporadic-like. Patients above the solid line are characterized by a *BRCA1* signature. The classification for each tumour was based on the leave-one-out procedure. The *BRCA1* germline mutation status is indicated in the right panel (white indicates mutation). (See Fig. 1 for colour scheme.)

signature, and below this threshold as a poor prognosis signature. Even small primary tumours without lymph node metastases can display the poor prognosis signature, indicating that they are already programmed for this metastatic phenotype.

The functional annotation for the genes provides insight into the underlying biological mechanism leading to rapid metastases. Genes involved in cell cycle, invasion and metastasis, angiogenesis, and signal transduction are significantly upregulated in the poor prognosis signature (for example cyclin E2, *MCM6*, metalloproteinases *MMP9* and *MP1*, *RAB6B*, *PK428*, *ESM1*, and the VEGF receptor *FLT1*; see Fig. 2b). If we evaluate all 231 prognostic reporter genes, more genes belonging to these functional categories become apparent (for example, *RAD21*, cyclin B2, *PCTAIRE*, *CDC25B*, *CENPE*, *VEGF*, *PGK1*, *MAD2*, *CKS2*, *BUB1*) (for a complete list, see Supplementary Information Table S2).

Many clinical studies have correlated alterations in expression of individual genes with breast cancer disease outcome, often with contradictory results. Examples include cyclin D1, ER- $\alpha$ , *UPA*, *PAI-1*, *HER2/neu* and *c-myc*<sup>19–22</sup>. Surprisingly, none of these genes are present in our set of 70 marker genes. This could be due to the fact that here we determine gene expression at the level of transcription, whereas most previous studies measured protein levels. However, it is more likely that these genes in isolation have only limited predictive power, which highlights the need for an approach based on many genes.

To validate the prognosis classifier, an additional independent set of primary tumours from 19 young, lymph-node-negative breast cancer patients was selected. This group consisted of 7 patients who remained metastasis free for at least five years, and 12 patients who developed distant metastases within five years. The disease outcome was predicted by the 70-gene classifier and resulted in 2 out of 19 incorrect classifications using both the optimal accuracy threshold (Fig. 2c, solid line) and the optimized sensitivity threshold (Fig. 2c, dashed line). Thus, the classifier showed a comparable performance on the validation set of 19 independent sporadic tumours and confirmed the predictive power and robustness of prognosis classification using the 70 optimal marker genes (Fisher's exact test for association  $P = 0.0018$ ).

The prediction of the classifier presented in Fig. 2b would indicate that women under 55 years of age who are diagnosed with lymph-node-negative breast cancer that has a poor prognosis signature have a 28-fold odds ratio (OR) (95% confidence interval, CI 7–107,  $P = 1.0 \times 10^{-8}$ ) to develop a distant metastasis within 5 years compared with those that have the good prognosis signature (see Methods for odds ratio definition). This estimate, however, is based on the same series of patients that the classifier was derived from, and therefore this odds ratio represents an upper limit. A performance cross-validation procedure, in which the leave-one-out sample is not involved in selecting the prognosis reporter genes and the number of reporter genes is not optimized, results in an odds ratio of 15 for a short interval to metastases (95% CI 4–56,  $P = 4.1 \times 10^{-6}$ ) (see Supplementary Information). This cross-validated predictive value of our classifier is superior to the currently available clinical and histopathological prognostic factors: high grade (odds ratio, OR = 6.4 (95% CI 2.1–19),  $P = 0.0008$ ), tumour size greater than 2 cm (OR = 4.4 (95% CI 1.7–11),  $P = 0.0028$ ), angioinvasion (OR = 4.2 (95% CI 1.5–12),  $P = 0.01$ ), age  $\leq 40$  (OR = 3.7 (95% CI 1.3–11),  $P = 0.02$ ), and ER negative (OR = 2.4 (95% CI 0.9–6.6),  $P = 0.13$ ). Furthermore, the evaluation of the cross-validated classifier in a multivariate model that includes all classical prognostic factors indicates that it is an independent factor in predicting outcome of disease (logistic regression OR = 18 (3.3–94),  $P$ -value of likelihood ratio test  $1.4 \times 10^{-4}$ ). Studying a large and unselected cohort of breast cancer patients is required to provide a more accurate estimate of the metastatic risk associated with the prognosis signature.

Unsupervised cluster analysis distinguishes between ER-positive

and ER-negative tumours (Fig. 1a). To investigate the expression patterns associated with the immunohistochemical staining of ER and to explore the differences between the sporadic and *BRCA1* tumours that fall into the ER-negative cluster (Fig. 1a), a supervised two-layer classification was performed (Fig. 3a). Figure 3b shows that 550 genes optimally report the dominant pattern associated with ER status, including genes such as keratin 18, *BCL2*, *ERBB3* and *ERBB4* (see Supplementary Information Table S3). The leave-one-out analysis shows that only two ER-positive and three ER-negative tumours (as determined by IHC) were classified in the opposite gene expression group (95% correct classification, Fig. 3b, middle panel). However, in all five discordant cases, the abundance of ER messenger RNA measured by the microarray agrees with the classification (Fig. 3b, right panel). An ER status reporter signature was also determined by others using a similar classification method<sup>8</sup>, and their ER signature gene set overlaps with ours (21 out of their 50 ER status reporter genes are present in our set of 550 ER reporters). Our observation in the unsupervised analysis that ER clustering has predictive power for prognosis is also valid for the ER supervised classification, although it does not reach the level of significance of the prognosis classifier (ER signature prediction for prognosis, OR = 3.7 (95% CI 1.3–11)  $P = 0.02$ ; data not shown).

Figure 3c shows the leave-one-out classification of the 38 ER-negative tumours into sporadic cases and *BRCA1*-associated cases based on an optimal set of 100 genes. This set is enriched in lymphocyte-specific genes (see Supplementary Information Table S4). The classification into sporadic and *BRCA1* tumours was caused mainly by the differences in levels of gene expression (amplitude), in concordance with recent findings that *BRCA1* mediates ligand-independent transcriptional repression of the ER<sup>23</sup> (95% accuracy, 2/38 misclassified, Fig. 3c). The one sporadic tumour that was classified as a *BRCA1* tumour was shown to contain methylation of the *BRCA1* promoter, indicating an epigenetic modification of *BRCA1*<sup>24</sup> (data not shown). Notably, the discordant *BRCA1* tumour is from a patient where the germline mutation has only altered the last 29 amino acids of the *BRCA1* protein (*BRCA1* mutation 5,622del62), which abolishes transcriptional activation by *BRCA1*<sup>25</sup>. One previous study defined a gene expression signature associated with *BRCA1* germline mutations using a panel of seven tumours<sup>26</sup>; however, the study was unable to appreciate the overlap in signatures between the ER-negative and *BRCA1* tumours. Furthermore, the nine *BRCA1* status reporter genes<sup>26</sup> were not present in our set of 100 optimal reporter genes. The two-layer cluster analysis that we have used and the larger number of tumours we analysed may account for these differences.

Our results indicate that breast cancer prognosis can already be derived from the gene expression profile of the primary tumour. Recent consensus conferences on treatment of breast cancer in Europe and the USA (St. Gallen<sup>2</sup> and NIH consensus<sup>3</sup>) have developed guidelines for the eligibility of adjuvant chemotherapy based on histological and clinical characteristics. Following these

**Table 1 Breast cancer patients eligible for adjuvant systemic therapy**

| Consensus          | Patient group                       |  |                                      |
|--------------------|-------------------------------------|--|--------------------------------------|
|                    | Total patient group<br>( $n = 78$ ) | Metastatic disease<br>at 5 yr ( $n = 34$ ) | Disease free<br>at 5 yr ( $n = 44$ ) |
| St Gallen          | 64/78 (82%)                         | 33/34 (97%)                                | 31/44 (70%)                          |
| NIH                | 72/78 (92%)                         | 32/34 (94%)                                | 40/44 (91%)                          |
| Prognosis profile* | 43/78 (55%)                         | 31/34 (91%)                                | 12/44 (27%)<br>(18/44 (41%)†)        |

The conventional consensus criteria are: tumour  $\geq 2$  cm, ER negative, grade 2–3, patient  $< 35$  yr (either one of these criteria; St Gallen consensus); tumour  $> 1$  cm (NIH consensus).

\* Number of tumours having a poor prognosis signature using our microarray profile, defined by the optimized sensitivity threshold in the 70-gene classifier (see Fig. 2b).

† Number of tumours with a poor prognosis signature in the group of disease-free patients, when the cross-validated classifier is applied.

guidelines, up to 90% of lymph-node-negative young breast cancer patients are candidates for adjuvant systemic treatment. As 70–80% of these patients would not have developed distant metastases without adjuvant treatment, these patients may not benefit from the treatment, and may potentially suffer from the side effects. We applied the St Gallen and NIH consensus criteria on our patient group to compare the efficacy of the microarray classifier for the selection of patients for adjuvant systemic treatment. Table 1 shows that the prognosis classifier selects just as effectively those high-risk patients that would benefit from adjuvant therapy, but significantly reduces the number of patients that receive unnecessary treatment. Thus, the prognostic profile potentially provides a powerful tool to tailor adjuvant systemic treatment that could greatly reduce the cost of breast cancer treatment, both in terms of adverse side effects and health care expenditure. Furthermore, the signature that defines ER status can be used to decide on adjuvant hormonal therapy, and the signature that reveals *BRCA1* status may further improve the diagnosis of hereditary breast cancer. Finally, genes that are over-expressed in tumours with a poor prognosis profile are potential targets for the rational development of new cancer drugs. Identification of such targets may improve the efficiency of developing therapeutics for many tumour types. □

## Methods

### Breast tumour selection criteria

The criteria for the sporadic patients ( $n = 97$ ) were: primary invasive breast carcinoma less than 5 cm (T1 or T2), no axillary metastases (N0), age at diagnosis less than 55 years, calendar year of diagnosis 1983–1996, no previous malignancies; all patients were treated by modified radical mastectomy ( $n = 35$ ) or breast-conserving treatment ( $n = 62$ ), including axillary lymph node dissection followed by radiotherapy. Five patients of the metastases group received adjuvant systemic therapy consisting of chemotherapy ( $n = 3$ ) or hormonal therapy ( $n = 2$ ), all other patients did not receive additional treatment. All patients were followed at least annually for a period of at least 5 years. The criteria for hereditary patients ( $n = 20$ ) were: carriers of a germline mutation in *BRCA1* or *BRCA2*, and primary invasive breast carcinoma; no other selection criterion was applied. This study was approved by the Medical Ethical Committee of the Netherlands Cancer Institute. For complete patient data, see Table S1 in Supplementary Information.

### Clinical parameters of breast tumours

Tumour material was snap-frozen in liquid nitrogen within 1 h after surgery. A haematoxylin and eosin stained section was prepared before and after cutting slides for RNA isolation for assessment of the percentage of tumour cells. Only samples with greater than 50% tumour cells were selected, mean 67% and median 70% for all groups studied. Formalin-fixed, paraffin-embedded tumour tissue was used to evaluate the following: tumour type (according to the World Health Organisation classification), histological grade (grade 1–3), and the presence of angioinvasive growth and extensive lymphocytic infiltrate. ER expression was determined by immunohistochemical staining (negative when less than 10% of the nuclei showed staining, all others ER positive).

### RNA isolation

We used 30 sections of 30- $\mu$ m thickness for total RNA isolation. Total RNA was isolated with RNeasyLysR, and finally dissolved in RNase-free H<sub>2</sub>O. Twenty-five micrograms of total RNA was treated with DNase using the Qiagen RNase-free DNase kit and RNeasy spin columns. Total RNA treated with DNase was dissolved in RNase-free H<sub>2</sub>O to a final concentration of 0.2  $\mu$ g  $\mu$ l<sup>-1</sup>.

### cRNA labelling

cRNA was generated by *in vitro* transcription using T7 RNA polymerase on 5  $\mu$ g total RNA and labelled with Cy3 or Cy5 (CyDye, Amersham Pharmacia Biotech)<sup>13</sup>. Five micrograms of Cy-labelled cRNA from one breast cancer tumour was mixed with the same amount of reverse colour Cy-labelled product from a pool, which consisted of an equal amount of cRNA from each individual sporadic patient.

### Expression profiling using microarray

Labelled cRNAs were fragmented to an average size of approximately 50–100 nucleotides by heating at 60 °C in the presence of 10 mM ZnCl<sub>2</sub>, added to a hybridization buffer containing 1 M NaCl, 0.5% sodium sarcosine, 50 mM MES, pH 6.5, and formamide to a final concentration of 30%, final volume 3 ml at 40 °C. Hu25K microarrays represented the 24,479 biological oligonucleotides plus 1,281 control probes. Sequences for microarrays were selected from RefSeq (a collection of non-redundant mRNA sequences; <http://www.ncbi.nlm.nih.gov/LocusLink/refseq.html>) and from expressed sequence tag (EST) contigs ([http://www.phrap.org/est\\_assembly/human/gene\\_number\\_methods.html](http://www.phrap.org/est_assembly/human/gene_number_methods.html)). Each mRNA or EST contig was represented on the Hu25K microarray by a single 60-polymer

oligonucleotide chosen by the oligonucleotide probe design programme<sup>13</sup>. After hybridization, slides were washed and scanned using a confocal laser scanner (Agilent Technologies). Fluorescence intensities on scanned images were quantified, corrected for background noise and normalized<sup>13</sup>. Microarray data are available at <http://www.rii.com/publications/default.htm>.

### Method of unsupervised two-dimensional clustering

In the two-dimensional cluster analysis, gene clustering and tumour clustering were performed independently using an agglomerative hierarchical clustering algorithm. For gene clustering, pairwise similarity metrics among genes are calculated on the basis of expression ratio measurements across all tumours. Similarly, for tumour clustering, pairwise similarity measures among tumours are calculated based on expression ratio measurements across all significant genes (for details see Supplementary Information).

### Method of supervised classification

We developed a method for classifying breast tumours into prognostic or diagnostic categories based on gene expression profiles. This method includes the following three steps: (1) selection of discriminating candidate genes by their correlation with the category; (2) determination of the optimal set of reporter genes using a leave-one-out cross validation procedure; (3) prognostic or diagnostic prediction based on the gene expression of the optimal set of reporter genes (for details see Supplementary Information).

### Statistical analysis

The odds ratio is the ratio of the odds in favour of developing distant metastases within 5 years for a patient in this study with a tumour characterized by the poor prognosis signature, to the odds in favour of developing metastases without this signature (2 × 2 table). *P*-values associated with odds ratios are calculated by Fisher's exact test. In the multivariate analysis a logistic model was applied with outcome of disease as the dependent variable, and the *P*-value for the relevant parameter is derived from the likelihood ratio test in the model (see Supplementary Information).

Received 24 August; accepted 22 November 2001.

- McGuire, W. L. Breast cancer prognostic factors: evaluation guidelines. *J. Natl Cancer Inst.* **83**, 154–155 (1991).
- Goldhirsch, A., Glick, J. H., Gelber, R. D. & Senn, H. J. Meeting highlights: international consensus panel on the treatment of primary breast cancer. *J. Natl Cancer Inst.* **90**, 1601–1608 (1998).
- Eifel, P. *et al.* National institutes of health consensus development conference statement: adjuvant therapy for breast cancer, November 1–3, 2000. *J. Natl Cancer Inst.* **93**, 979–989 (2001).
- Early Breast Cancer Trialists' Collaborative Group. Polychemotherapy for early breast cancer: an overview of the randomised trials. *Lancet* **352**, 930–942 (1998).
- Early Breast Cancer Trialists' Collaborative Group. Tamoxifen for early breast cancer: an overview of the randomised trials. *Lancet* **351**, 1451–1467 (1998).
- 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).
- Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Gruvberger, S. *et al.* Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res.* **61**, 5979–5984 (2001).
- Martin, K. J. *et al.* Linking gene expression patterns to therapeutic groups in breast cancer. *Cancer Res.* **60**, 2232–2238 (2000).
- Zajchowski, D. A. *et al.* Identification of gene expression profiles that predict the aggressive behavior of breast cancer cells. *Cancer Res.* **61**, 5168–5178 (2001).
- Sorlie, T. *et al.* Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl Acad. Sci. USA* **98**, 10869–10874 (2001).
- West, M. *et al.* Predicting the clinical status of human breast cancer by using gene expression profiles. *Proc. Natl Acad. Sci. USA* **98**, 11462–11467 (2001).
- Hughes, T. R. *et al.* Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nature Biotechnol.* **19**, 342–347 (2001).
- Roberts, C. J. *et al.* Signaling and circuitry of multiple MAPK pathways revealed by a matrix of global gene expression profiles. *Science* **287**, 873–880 (2000).
- Lakhani, S. R. *et al.* Multifactorial analysis of differences between sporadic breast cancers and cancers involving *BRCA1* and *BRCA2* mutations. *J. Natl Cancer Inst.* **90**, 1138–1145 (1998).
- Brenton, J. D., Aparicio, S. A. & Caldas, C. Molecular profiling of breast cancer: portraits but not physiognomy. *Breast Cancer Res.* **3**, 77–80 (2001).
- Khan, J. *et al.* Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Med.* **7**, 673–679 (2001).
- He, Y. D. & Friend, S. H. Microarrays—the 21st century divining rod? *Nature Med.* **7**, 658–659 (2001).
- Bieche, I. *et al.* Genetic alterations in breast cancer. *Genes Chromosomes Cancer* **14**, 227–251 (1995).
- Steege, P. S. & Zhou, Q. Cyclins and breast cancer. *Breast Cancer Res. Treat.* **52**, 17–28 (1998).
- Janicke, F. *et al.* Randomized adjuvant chemotherapy trial in high-risk, lymph node-negative breast cancer patients identified by urokinase-type plasminogen activator and plasminogen activator inhibitor type 1. *J. Natl Cancer Inst.* **93**, 913–920 (2001).
- van Diest, P. J. *et al.* Cyclin D1 expression in invasive breast cancer. Correlations and prognostic value. *Am. J. Pathol.* **150**, 705–711 (1997).
- Zheng, L., Annab, L. A., Afshari, C. A., Lee, W. H. & Boyer, T. G. *BRCA1* mediates ligand-independent transcriptional repression of the estrogen receptor. *Proc. Natl Acad. Sci. USA* **98**, 9587–9592 (2001).
- Esteller, M. *et al.* Promoter hypermethylation and *BRCA1* inactivation in sporadic breast and ovarian tumors. *J. Natl Cancer Inst.* **92**, 564–569 (2000).
- Chapman, M. S. & Verma, I. M. Transcriptional activation by *BRCA1*. *Nature* **382**, 678–679 (1996).
- Hedenfalk, I. *et al.* Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* **344**, 539–548 (2001).



Supplementary Information accompanies the paper on Nature's website (http://www.nature.com).

**Acknowledgements**

We thank D. Atsma and D. Majoor for assistance with the histological analyses and the preparation of tumour RNA; T. van der Velde, W. van Waardenburg and O. Dalesio for medical record data extraction; D. Slade, J. McDonald, J. Koch, T. Erkkila, M. Parrish and others at Rosetta's High Throughput Gene Expression Profiling Facility for microarray experiments; R. Stoughton, F. van Leeuwen, M. Rookus, P. Nederlof, F. Hogervorst and D. Voskuil for suggestions; and A. Berns, L. Hartwell, J. Radich and S. Rodenhuis for support and reading of the manuscript. This work was supported by a grant from the Center for Biomedical Genetics.

**Competing interests statement**

The authors declare competing financial interests: details accompany the paper on Nature's website (http://www.nature.com).

Correspondence and requests for materials should be addressed to S.H.F. (e-mail: stephen\_friend@merck.com).

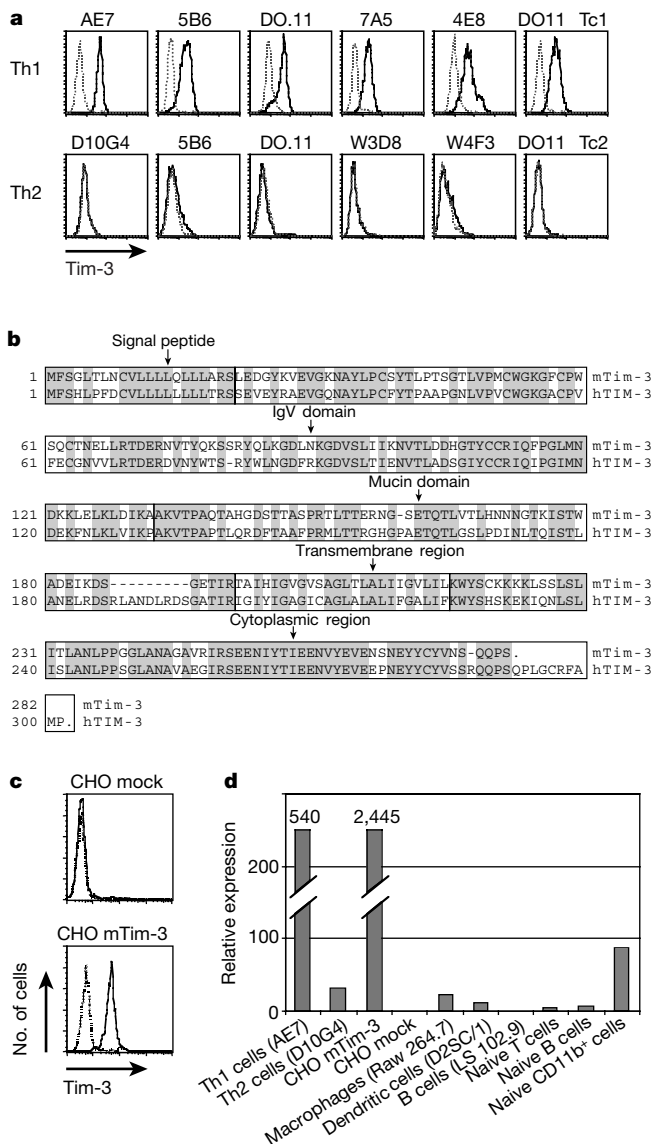
# Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease

Laurent Monney\*, Catherine A. Sabatos\*, Jason L. Gaglia\*, Akemi Ryu\*, Hanspeter Waldner\*, Tatyana Chernova†, Stephen Manning‡, Edward A. Greenfield\*†, Anthony J. Coyle‡, Raymond A. Sobel§, Gordon J. Freeman† & Vijay K. Kuchroo\*

\* Center for Neurologic Diseases, Department of Neurology, Brigham and Women's Hospital and Harvard Medical School; and † Department of Adult Oncology, Dana-Farber Cancer Institute and Department of Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA  
‡ Millennium Pharmaceuticals, 640 Memorial Drive, Cambridge, Massachusetts 02139, USA  
§ VA Health Care System, Palo Alto and Department of Pathology, Stanford University School of Medicine, Stanford, California 95305, USA

Activation of naive CD4<sup>+</sup> T-helper cells results in the development of at least two distinct effector populations, Th1 and Th2 cells<sup>1-3</sup>. Th1 cells produce cytokines (interferon (IFN)- $\gamma$ , interleukin (IL)-2, tumour-necrosis factor (TNF)- $\alpha$  and lymphotoxin) that are commonly associated with cell-mediated immune responses against intracellular pathogens, delayed-type hypersensitivity reactions<sup>4</sup>, and induction of organ-specific autoimmune diseases<sup>5</sup>. Th2 cells produce cytokines (IL-4, IL-10 and IL-13) that are crucial for control of extracellular helminthic infections and promote atopic and allergic diseases<sup>4</sup>. Although much is known about the functions of these two subsets of T-helper cells, there are few known surface molecules that distinguish between them<sup>6</sup>. We report here the identification and characterization of a transmembrane protein, Tim-3, which contains an immunoglobulin and a mucin-like domain and is expressed on differentiated Th1 cells. *In vivo* administration of antibody to Tim-3 enhances the clinical and pathological severity of experimental autoimmune encephalomyelitis (EAE), a Th1-dependent autoimmune disease, and increases the number and activation level of macrophages. Tim-3 may have an important role in the induction of autoimmune diseases by regulating macrophage activation and/or function.

In addition to their distinct roles in disease, Th1 and Th2 cells cross-regulate each other's expansion and functions. Thus, preferential induction of Th2 cells inhibits autoimmune diseases<sup>7,8</sup>, and



**Figure 1** Cloning of a Th1-specific cell surface protein, Tim-3. **a**, Th1, Th2, Tc1 and Tc2 cells were stained with monoclonal antibody to Tim-3 (solid line) or rat IgG isotype control (dotted line). **b**, Deduced amino-acid sequence of murine and human Tim-3. Shading indicates regions of homology. IgV, variable region of immunoglobulin. **c**, CHO cells transfected using either Tim-3 cDNA (CHO mTim-3) or vector alone (CHO mock). Stable puromycin-resistant cells were stained with monoclonal antibody to Tim-3 (solid line) or rat IgG isotype control (dotted line). **d**, Total RNA from various cell lines and cells purified from SJL mice was isolated and transcribed to cDNA by reverse transcription, and cDNA was used for Taqman PCR. The figure shows expression of Tim-3 RNA relative to control GAPDH expression.

predominant induction of Th1 cells can regulate induction of asthma, atopy and allergies<sup>9,10</sup>. Several groups have reported the association of chemokine and co-stimulatory receptors with Th1 (refs 11–14) and Th2 (refs 12, 13, 15–18) cells; however, the nature of the differences in expression of most of these molecules is quantitative.

To identify new Th1-specific cell surface proteins, we immunized Lewis and Lou/M rats with Th1 T-cell clones and lines, including the established Th1-specific clone AE7 and *in vitro* differentiated Th1 cell lines derived from 5B6 (ref. 19) and DO11.10 T-cell receptor (TCR) transgenic mice. A panel of approximately 20,000 monoclonal antibodies was generated and screened on Th1 and Th2 cells. Two of the monoclonal antibodies (8B.2C12 and 25F.1D6) that selectively stained Th1 cells were further characterized. These