nature medicine

A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer

Pierre Farmer[1,2,18], Hervé Bonnefoi[3–6], Pascale Anderle[1,7], David Cameron[8,18], Pratyakasha Wirapati[2], Véronique Becette[9,18], Sylvie André[1], Martine Piccart[10], Mario Campone[11], Etienne Brain[9], Gaëtan MacGrogan[3], Thierry Petit[12], Jacek Jassem[13], Frédéric Bibeau[14], Emmanuel Blot[15], Jan Bogaerts[6], Michel Aguet[1], Jonas Bergh[16], Richard Iggo[1,3,17] & Mauro Delorenzi[1,2]

To better understand the relationship between tumor-host interactions and the efficacy of chemotherapy, we have developed an analytical approach to quantify several biological processes observed in gene expression data sets. We tested the approach on tumor biopsies from individuals with estrogen receptor–negative breast cancer treated with chemotherapy. We report that increased stromal gene expression predicts resistance to preoperative chemotherapy with 5-fluorouracil, epirubicin and cyclophosphamide (FEC) in subjects in the EORTC 10994/BIG 00-01 trial. The predictive value of the stromal signature was successfully validated in two independent cohorts of subjects who received chemotherapy but not in an untreated control group, indicating that the signature is predictive rather than prognostic. The genes in the signature are expressed in reactive stroma, according to reanalysis of data from microdissected breast tumor samples. These findings identify a previously undescribed resistance mechanism to FEC treatment and suggest that antistromal agents may offer new ways to overcome resistance to chemotherapy.

Preoperative chemotherapy leads to the disappearance of the primary tumor (pathological complete response; pCR) in less than 10% of estrogen receptor–positive breast tumors and in 20–30% of estrogen receptor–negative tumors[1–3]. Several clinical studies have derived gene signatures that predict response to neoadjuvant therapy of breast cancer[4–8], but none of those signatures has yet been validated in an external data set. Recently, predictive signatures based on the *in vitro* response of cell lines to chemotherapy were proposed[9,10]. However, the use of cell lines has the potential disadvantage of ignoring the influence of the tumor microenvironment on drug response.

This study was designed to search for a gene signature predicting the response of breast tumors to FEC, a widely used anthracycline-containing chemotherapy regimen. To avoid confounding response-specific genes with tumor type-specific genes, the study was restricted to estrogen receptor–negative tumors. To isolate robust signatures, we developed a new bioinformatics method that decomposes the gene expression signal originating from a mixture of tumor cells and stromal cells into multiple independent signatures associated with distinct biological processes.

We report for the first time in the context of a randomized clinical trial a major contribution of stromal genes to drug sensitivity.

RESULTS

Chemotherapy response prediction from biological processes

Pretreatment biopsies of 63 estrogen receptor–negative tumors were included in this substudy of the EORTC 10994 neoadjuvant trial (clinical and pathological data are given in **Supplementary Tables 1** and **2** online). We were unable to predict the response of these tumors

[1]Swiss National Centre of Competence in Research Molecular Oncology, Swiss Institute for Experimental Cancer Research, Ecole Polytechnique Fédérale de Lausanne, School of Life Sciences, 155, Chemin des Boveresses, 1066 Epalinges s/Lausanne, Switzerland. [2]Swiss Institute of Bioinformatics, Génopode Building, Quartier Sorge, CH-1015 Lausanne, Switzerland. [3]Institut Bergonié and Bordeaux 2 University, Institute National de la Santé et de la Recherche Médicale unit U916, 229 Cours de l'Argonne, 33076 Bordeaux Cedex, France. [4]Hôpitaux Universitaires de Genève, 24 rue Micheli-du-Crest CH-1211 Geneva, Switzerland. [5]Swiss Group for Clinical Cancer Research, Effingerstrasse 40 CH-3008 Bern, Switzerland. [6]European Organisation for Research and Treatment of Cancer (EORTC), Avenue E. Mounier, 83, 1200 Brussels, Belgium. [7]Laboratory of Experimental Oncology, Oncology Institute of Southern Switzerland, Via Vincenza Vela 6, 6500 Bellinzona, Switzerland. [8]for the Anglo-Celtic Cooperative Oncology Group, University Department of Oncology, Western General Hospital, Crewe Road South, Edinburgh EH4 2XU, Scotland. [9]Centre René Huguenin, 35, rue Dailly, 92210 Saint-Cloud, France. [10]Institut Jules Bordet, Boulevard de Waterloo 125, B-1000 Brussels, Belgium. [11]Institut du Cancer Nantes Atlantique, Centre de Lutte Contre le Cancer René Gauducheau, Centre de Cancérologie, Institute National de la Santé et de la Recherche Médicale 892, Boulevard Jacques Monod, 44 805 Nantes Cedex, Saint Herbalin, France. [12]Centre Paul Strauss, 3, rue de la Porte de l'Hôpital 67065 Strasbourg Cedex, France. [13]Medical University, 7, Debinki Street, 80-211 Gdansk, Poland. [14]Centre Val d'Aurelle - Paul Lamarque, 208 rue des Apothicaires, 34298 Montpellier Cedex 5, France. [15]Centre Henri Becquerel, rue d'Amiens, 76038 Rouen Cedex, France. [16]Swedish Breast Cancer Group, Karolinska Institutet and Karolinska University Hospital, Radiumhemmet, 171 76 Stockholm, Sweden. [17]Bute Medical School, Westburn Lane, University of St. Andrews, Fife KY16 9TS, Scotland. [18]Present addresses: Merck Serono International S.A. Switzerland, 9, chemin des Mines, 1202 Geneva, Switzerland (P.F.), UK National Cancer Research Network Coordinating Centre, 24 Hyde Terrace, Leeds LS2 9LN, UK (D.C.), Centre François Baclesse, 3 avenue du Général Harris, BP 5026, 14076 CAEN Cedex 5, France (V.B.). Correspondence should be addressed to M.D. (mauro.delorenzi@isb-sib.ch).

Received 13 August 2008; accepted 25 November 2008; published online 4 January 2009; doi:10.1038/nm.1908

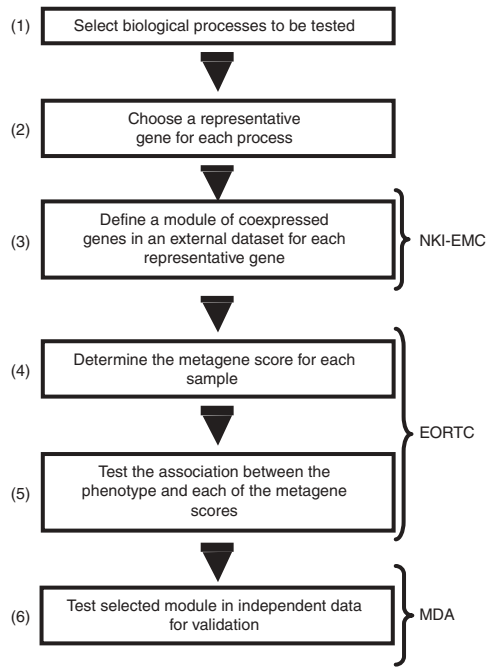

Figure 1 Flow chart showing how the metagenes were created and tested. The brackets indicate the data set used at each step.

to chemotherapy with conventional techniques, which selected genes with heterogeneous expression profiles in the classifiers (data not shown). To circumvent this difficulty, we have developed a new strategy that tests, by design, only coherent groups of genes.

In a previous study of gene expression and tumor subtypes in breast carcinoma[11], we identified nine main clusters of coexpressed genes. These clusters are related to tumor cell type (luminal, basal-like and erythroblastic leukemia viral oncogene homolog-2 (*ERBB2*)–molecular apocrine), cell physiology (proliferation, hypoxia and interferon signaling) and the tumor microenvironment (T and B cells, adipocytes and stroma) and have been described by others[12–15] (**Fig. 1** (1)). We aimed to identify groups of genes that robustly define these nine clusters by an automated and objective procedure. To do this, we first selected nine 'representative genes' on the basis that each is found at the heart of a prominent gene cluster in multiple published studies and has a known connection to the biological process linked to the cluster (**Fig. 1** (2)). The nine representative genes were included as explanatory variables in a multivariate regression model in order to identify groups of genes related to each biological process (the output of the model is given in **Supplementary Table 3** online). The model was not fitted to our own gene expression data but to an external data set, which includes profiles performed at the Nederlands Kanker Instituut (NKI) and at the Erasmus Medical Center (EMC) (NKI-EMC; see Methods), comprising 581 tumors hybridized to two different types of microarray (**Fig. 1** (3)). Notably, no information regarding response to therapy was used in this process. For each sample, the expression values in the EORTC data set of the 50 genes most strongly associated with each representative gene in the NKI-EMC data set were averaged to generate a single summary value per biological process. We refer to this value as a metagene (**Fig. 1** (4)). The ability of the linear model to identify groups of functionally related genes was verified by gene set enrichment analysis (GSEA; **Supplementary Fig. 1** online).

Table 1 Prediction of pathological complete response using metagene signatures

(**a**) EORTC ($n = 63$)

Biological process	Representative gene	HUGO name	AUC [95% CI]	*P* value (FDR)
Luminal-basal	*ESR1*	Estrogen receptor-α	0.53 [0.37–0.68]	0.40
Apocrine	*CLCA2*	Chloride-associated calcium channel-2	0.46 [0.32–0.59]	0.70
Stroma	*DCN*	Decorin	0.68 [0.54–0.80]	0.03
T Cell	*GZMA*	Granzyme A	0.62 [0.48–0.75]	0.14
B Cell	*CD83*	CD83 antigen	0.58 [0.44–0.71]	0.23
Adipocyte	*FABP4*	Fatty acid–binding protein-4	0.54 [0.38–0.68]	0.40
Proliferation	*TPX2*	TPX2 microtubule-associated homolog	0.55 [0.39–0.69]	0.40
Interferon	*MX1*	Myxovirus resistance gene-1	0.72 [0.59–0.84]	0.01
Hypoxia	*ADM*	Adrenomedulin	0.59 [0.44–0.73]	0.23

(**b**) MDA ($n = 51$)

Biological process	Representative gene	HUGO name	AUC [95% CI]	*P* value
Stroma	*DCN*	Decorin	0.70 [0.52–0.85]	0.01
Interferon	*MX1*	Myxovirus resistance gene-1	0.50 [0.33–0.67]	0.60

(**c**) Data set	ACC	PPV	SENS	NPV	SPEC	OR	*P* value
EORTC	0.65	0.57	0.86	0.81	0.49	5.51	0.01
MDA	0.65	0.64	0.78	0.67	0.50	3.41	0.05

(**a**) Prediction of pCR with metagenes in the EORTC data set (28 pCR cases among 63 cases). The 95% confidence intervals (CI) were calculated by the bootstrap method. *P* values were adjusted for multiple testing by computing the false discovery rate (FDR). (**b**) Validation of the interferon and stromal metagenes in the MDA data set (27 pCR cases among 51 cases). *P* values not adjusted. HUGO, Human Genome Organization. (**c**) Prediction of pCR by the stromal metagene using a logistic regression model trained on the EORTC data set and tested on the MDA data set. The accuracy (ACC), sensitivity (SENS), specificity (SPEC), positive (PPV) and negative (NPV) predictive value and odds ratio (OR) were determined at the threshold that maximizes the Youden index (SENS + SPEC – 1) in the EORTC data set. The *P* values were calculated by Fisher's exact test.

A stromal metagene is associated with chemoresistance

The ability of the nine metagenes to predict pathological complete response was tested by measuring the area under receiver operating characteristic curves (AUC; **Table 1** and **Fig. 1** (5)). A significant AUC for prediction of response to FEC was observed only for the interferon and stromal metagenes (**Table 1a** and **Supplementary Fig. 2a** online). The predictor showed high sensitivity but not specificity for detection of responders at a selected threshold (**Table 1c**). Varying the threshold resulted in other combinations of sensitivity, specificity and the other classification performance metrics (**Supplementary Fig. 2b**). The interferon and stromal metagenes were then tested in an independent cohort of estrogen receptor–negative tumors from the M. D. Anderson Medical Center (MDA data set) included in a recent study of response to neoadjuvant chemotherapy with paclitaxel, 5-fluorouracil, doxorubicin and cyclophosphamide (T-FAC)[6] (**Fig. 1** (6)). The stromal metagene was again significantly associated with response (AUC = 0.70; $P = 0.01$), whereas the interferon metagene was not (**Table 1b,c** and **Supplementary Fig. 2c,d**). The constituent genes in the stromal metagene have a coherent expression pattern in both the EORTC and the MDA studies; high stromal gene expression is associated with resistance to chemotherapy (**Fig. 2**).

The ability of the stromal metagene, histological grade, node status, tumor size and *ERBB2* status to predict pCR was tested by logistic regression. The stromal metagene was the only significant variable in univariate analysis in both data sets (**Table 2**). Multivariate analysis confirmed that the stromal metagene was an independent predictive factor in both data sets (**Table 2**). The *ERBB2*–molecular apocrine tumor class was too small to test, but the stromal metagene was significantly associated with response in the basal-like class in both the EORTC (AUC = 0.69 [0.51–0.84]; $P = 0.02$; $n = 39$) and MDA data sets (AUC = 0.73 [0.54–0.92]; $P = 0.01$; $n = 27$).

The selection of decorin (*DCN*) as the representative gene to derive the stromal metagene was arbitrary. Two approaches were used to test whether this choice might have led to overfitting of the data. The stromal metagene is based on 50 genes, but the regression ranked all of the genes and identified many hundreds that are significantly associated with *DCN* expression ($P < 0.05$). Rather than using 50 genes, we constructed a series of metagenes from nonoverlapping groups of 15 genes associated with decorin expression. Nine of the first 12 metagenes, involving a total of 180 genes, gave an AUC significantly greater than 0.5 (**Supplementary Fig. 3** online). To test whether the use of *DCN* itself as the representative gene may have led to unintended bias, we repeated the regression procedure with other stroma-related genes. We iteratively replaced *DCN* in the regression model with the 49 other genes in the original stromal metagene and constructed new metagenes for each new representative gene (**Supplementary Table 4** online). The results with all of the metagenes were similar, in terms of both the genes identified and the AUC for prediction of response in the EORTC and MDA data sets (**Supplementary Table 4**). This was not surprising to us, because the mean pairwise correlation of genes in the original stromal metagene in the EORTC data set was high (0.55). We conclude that neither the choice of *DCN* as the stromal representative gene nor the inclusion of exactly 50 genes in the metagene led to overestimation of the predictive performance.

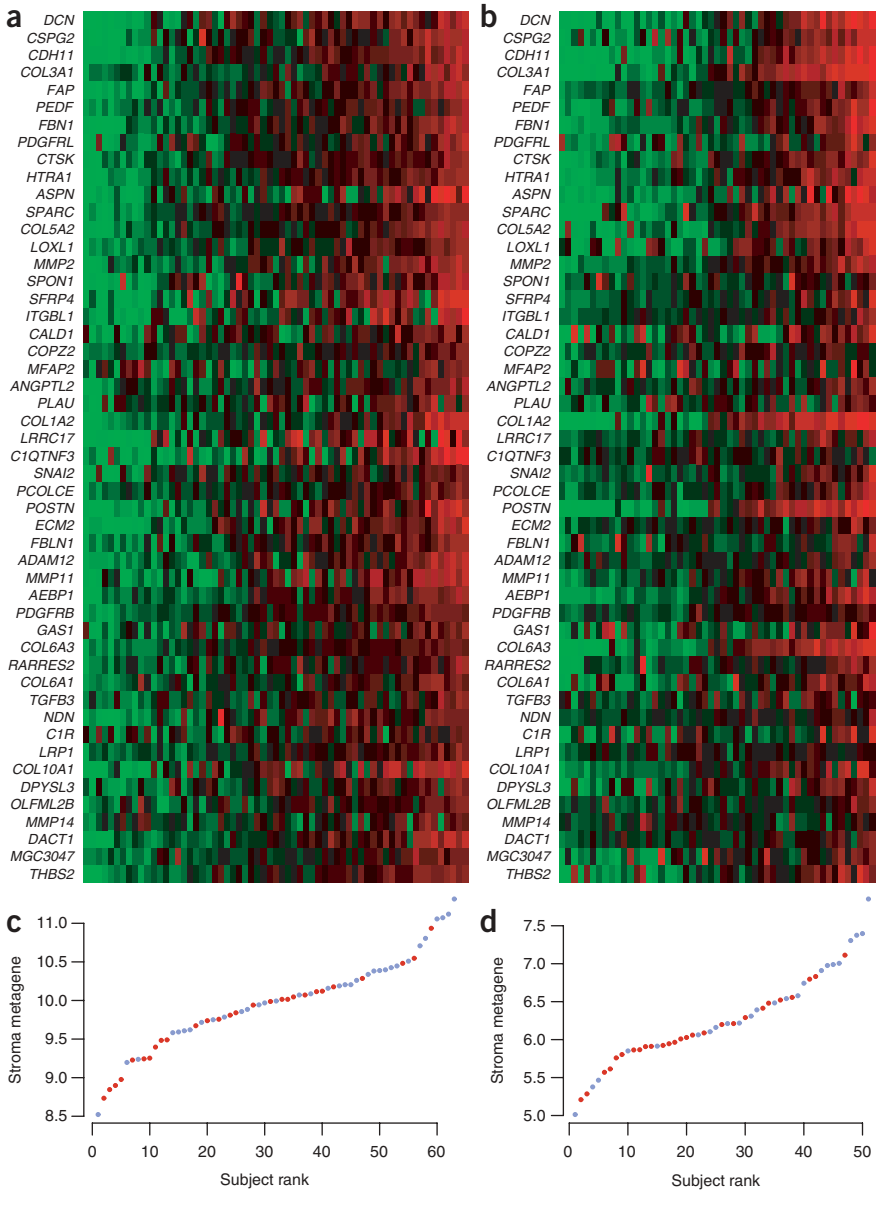

Figure 2 Stromal gene expression and metagene scores. (**a**,**b**) Heat map showing the individual genes included in the stromal metagene for the EORTC data set (**a**) and the MDA data set (**b**) (see also **Supplementary Table 4**). (**c**,**d**) Metagene scores for each subject in the EORTC data set (**c**) and the MDA data set (**d**) (pCR, red; non-pCR, blue). The samples in **a** and **b** are ordered horizontally as they are in **c** and **d**.

Table 2 Predictive factors for pCR

EORTC	Univariate[b]		Multivariate[b]	
Variable[a]	Coefficient	*P* value	Coefficient	*P* value
Clinical node (N0 versus N1 and N2)	–0.099	0.862	0.014	0.982
Clinical size (T1 and T2 versus T3)	–0.076	0.895	–0.222	0.736
Grade (grade 1 and 2 versus 3)	1.030	0.080	0.923	0.141
ERBB2 (low versus high)	–0.588	0.362	0.114	0.882
Stromal metagene score (low versus high)	1.658	0.022	1.673	0.036
MDA	**Univariate[b]**		**Multivariate[b]**	
Variable[a]	Coefficient	*P* value	Coefficient	*P* value
Clinical node (N0 versus N1 and N2)	0.789	0.231	1.844	0.044
Clinical size (T1 and T2 versus T3)	–0.357	0.54	–1.698	0.040
Grade (grade 1 and 2 versus 3)	1.191	0.181	1.009	0.333
ERBB2 (negative versus positive)	0.577	0.336	0.489	0.477
Stromal metagene score (low versus high)	1.217	0.043	1.605	0.039

[a]The node and size determination followed the tumor, nodes and metastasis (TNM) staging system[35]. [b]Univariate and multivariate logistic regression models were used to test the predictive power of the stromal signature and conventional variables.

The stromal metagene is predictive but not prognostic

To test whether the stromal metagene is detecting bad prognosis tumors that are intrinsically more aggressive, regardless of therapy, we examined the impact of the stromal metagene on relapse-free survival in three cohorts of subjects, one treated with adjuvant chemotherapy[16], the other two not. Subjects in the NKI and EMC studies who did not receive either chemotherapy or hormonal therapy were used as the reference untreated population. A proliferation metagene known to be associated with high tumor grade and poor survival[14] was used as a positive control. Subjects were split into two equally sized groups on the basis of the value of their metagenes. As expected, high proliferation identified subjects with poor prognosis in all three data sets (**Fig. 3a–c**). The stromal metagene was unrelated to survival in the untreated subjects (**Fig. 3d,e**). In subjects who received adjuvant chemotherapy[16], higher expression of the stromal metagene was associated with significantly shorter relapse-free survival times (**Fig. 3f**). This indicates that the stromal metagene is predictive of treatment response rather than prognostic of survival and supports the hypothesis that the main association is between high stromal content and resistance to chemotherapy.

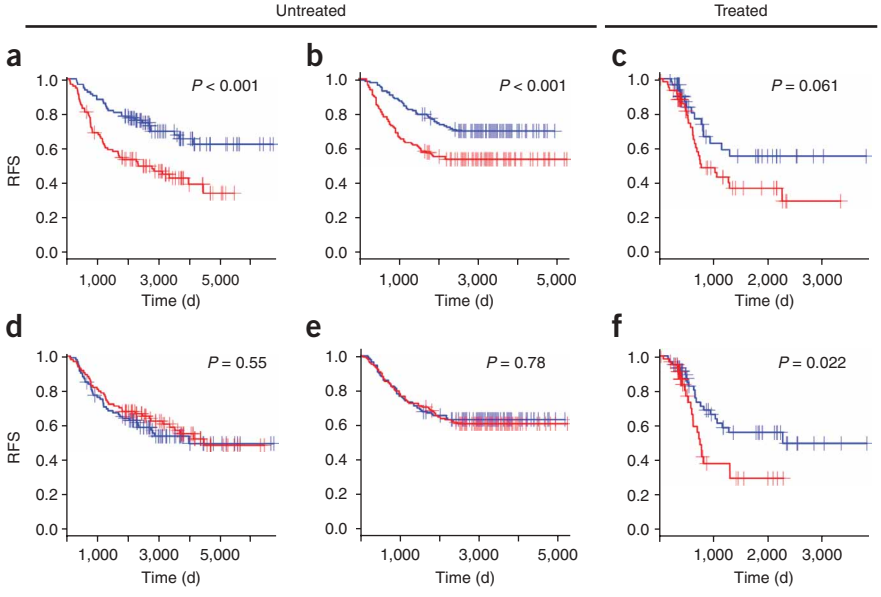

Figure 3 Prognostic versus predictive value of the proliferation and stromal metagenes. (**a–c**) Relapse-free survival (RFS) in the proliferation metagene for the NKI (left), EMC (center) and Duke (right) data sets. (**d–f**) Relapse-free survival in the stromal metagene for the NKI (left), EMC (center) and Duke (right) data sets. In panels **a**, **b**, **d** and **e**, the subjects did not receive chemotherapy. In panels **c** and **f**, subjects received adjuvant chemotherapy. Blue lines show subjects with low metagene scores. Red lines show subjects with high metagene scores. Significance was calculated by log-rank test.

The stromal signature is characteristic of reactive stroma

A key question about the stromal metagene is whether it measures an intrinsic property of the tumor or merely the amount of normal tissue in the biopsy. To determine which pathological property of the tumor is captured by the stromal metagene, we performed a blinded examination of the EORTC biopsy sections. For each sample, the proportions of tumor and nontumor tissue compartments were evaluated. As expected, most biopsies contained only tumor tissue (**Supplementary Fig. 4a** online). We use the term reactive stroma to describe areas rich in cells having fibroblast-like morphologies within the tumor compartment. The fibroblasts themselves are commonly referred to as activated fibroblasts or cancer-associated fibroblasts. There were noteworthy differences between tumors in the amount of reactive stroma (**Fig. 4a,b**). Comparison of the histological and microarray data showed a significant association of the stromal metagene with the amount of reactive stroma (**Fig. 4c**, $P = 0.009$, Wilcoxon test) but not with the proportion of nontumor tissue in the biopsy (**Supplementary Fig. 4b**). This means we can confidently exclude a major contribution of normal mammary tissue, such as lobules, ducts, adipose tissue or normal mammary interstitial tissue, to the stromal metagene.

The histological appearance of the biopsies in this study, as of breast tumors in general, strongly suggested that the amount of reactive stroma is a regular feature of individual tumors, making it an intrinsic property of the tumors. The classic approach to identify genes intrinsic to tumors is to compare between-tumor to within-tumor gene expression variance[12,17]. Genes intrinsic to tumors should have low variance within tumors but high variance between tumors. The variance ratios (intrinsic scores) for the individual genes in the stromal metagene can be used to define intrinsic genes. Applying a previously used threshold[17], we found 20 out of 46 testable genes of the stromal metagene to be intrinsic (**Fig. 4d**). This is significantly more than would be expected by chance (namely, 4 genes; $P = 0.00005$, Fisher's exact test).

To test the stromal metagene as a whole, we calculated the mean of the intrinsic scores of all of the genes in the stromal metagene (**Fig. 4e**). To compute an expected distribution for this parameter, we broke the link between biopsies and tumors in the previous data[17]; specifically, the pairing of biopsies used to calculate the within tumor variance was randomly permuted. Taking the 97.5% upper boundary

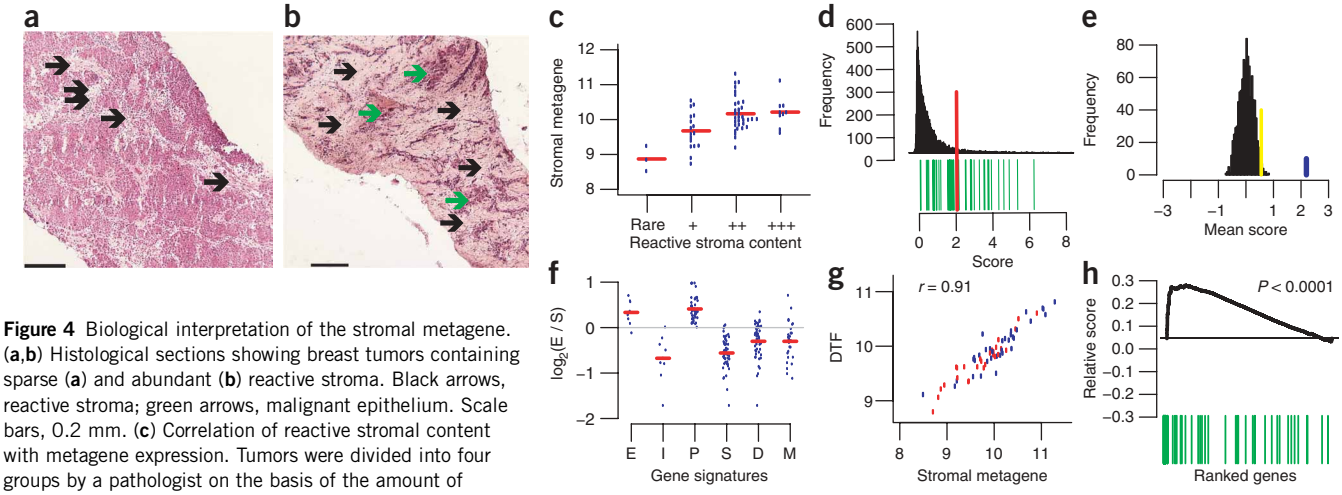

Figure 4 Biological interpretation of the stromal metagene. (**a**,**b**) Histological sections showing breast tumors containing sparse (**a**) and abundant (**b**) reactive stroma. Black arrows, reactive stroma; green arrows, malignant epithelium. Scale bars, 0.2 mm. (**c**) Correlation of reactive stromal content with metagene expression. Tumors were divided into four groups by a pathologist on the basis of the amount of reactive stroma. $P < 0.05$ for the + versus ++ and + versus +++ comparisons. (**d**,**e**) Intrinsic gene analysis. The intrinsic score (*x* axis) is the log ratio of the between-tumor to within-tumor variance for paired biopsies; red bar, cutoff previously used to define intrinsic genes[17]. Black bars in **d**, all genes; green bars, genes included in the stromal metagene. Blue bar in **e**, mean of the intrinsic scores of all the stromal genes, used to assess the full stromal metagene; yellow bar, upper 97.5th percentile of the distribution of the mean intrinsic scores (black bars) under random pairing of the biopsies. (**f**) Gene expression in microdissected breast tissue. Blue dots, genes belonging to the indicated signature; E / S score, ratio of the expression values in microdissected epithelial and stromal tissue[18]. Genes below the horizontal gray line are preferentially expressed in stroma. E (epithelium) and I (stroma), control gene lists taken from another microdissection study[36]; P (proliferation) and S (stroma), metagenes described in this study; D (desmoid-type fibromatosis[22]) and M (mammosphere[19]), gene signatures related to Wnt signaling in fibroblasts and EMT in mammary epithelial cells, respectively. Red horizontal bars, mean score for each signature. (**g**) Correlation of the stromal metagene with a DTF gene expression signature[22]. Each dot is a single tumor (pCR, red; non-pCR, blue); *r*, Pearson correlation coefficient. (**h**) GSEA testing whether the ranks (*x* axis) of EMT signature genes (green bars) are uniformly distributed when genes are sorted by decreasing *t* statistic for *DCN* in the linear regression model.

of the expected distribution as reference, we concluded that, by the conventional definition, the stromal metagene is intrinsic to the tumor (**Fig. 4e**).

To further confirm that the genes in the stromal metagene are expressed by stromal cells in the tumor compartment, we analyzed data generated by microdissection of breast tumors[18]. To determine whether a gene was expressed preferentially by epithelial or stromal cells, we calculated an epithelial (E) to stromal (S) ratio for each gene. The stromal metagene (**Fig. 4f**) showed a low E / S ratio, consistent with the constituent genes being expressed primarily in the stroma ($P = 0.001$, Sign test). To confirm that this result is not confined to a single study or even a single tumor type, we performed the same analysis with data from another breast microdissection study and from a colon cancer study comparing tumor epithelial cells with colon cancer–associated fibroblasts (**Supplementary Fig. 5a–c** online).

To explore the biology underlying the stromal metagene, we compared it to several potentially relevant gene lists. Because the epithelial-to-mesenchymal transition (EMT) is a potential explanation for stromal gene expression by tumor cells, we tested a gene list derived from undifferentiated epithelial cells cultivated as floating mammospheres[19,20]. These structures contain normal mammary epithelial cell progenitors that produce an abundant extracellular matrix containing decorin[19]. Mammospheres and EMT signature genes have a pattern of expression similar to that of the stromal metagene (**Fig. 4f** and **Supplementary Fig. 6a**,**b** online). Thus, EMT of tumor epithelial cells could account for increased stromal metagene expression in some tumors. As EMT was reported to generate cells with properties of stem cells[20], chemoresistance observed in this study could be related to the widely held opinion that tumor stem cells and progenitors are more resistant to chemotherapy[21]. We also analyzed two gene signatures that are reported to distinguish between different types of fibroblastic tumor, solitary fibrous tumors and desmoid-type fibromatosis (DTF)[22]. Only the DTF signature showed a strong correlation with the stromal metagene (**Fig. 4f**,**g** and **Supplementary Fig. 6c**). Desmoid tumors arise in fibroblasts with germline mutations in the *APC* gene that activates the Wnt signaling pathway. To further explore a potential role for EMT and Wnt signaling in the biology of the stromal metagene, we ranked genes by similarity to *DCN* expression and performed GSEA with a Wnt gene set, an EMT gene set and a transforming growth factor-β (TGF-β) gene set; the last was tested because of its known role in EMT. A GSEA peak skewed to the left (as in **Fig. 4h**) indicates that a high proportion of EMT signature genes have an expression pattern similar to that of the genes in the stromal metagene. GSEA supports a potential role for all three processes in reactive stroma (**Fig. 4h** and **Supplementary Fig. 7a**,**b** online). We conclude that reactive stroma is intrinsic to tumors that are resistant to FEC, and this may arise from activation of TGF-β or Wnt signaling.

DISCUSSION

The two main contributions of this work are the development of a new technique for the analysis of gene expression data and the identification of stromal gene expression as a marker for resistance to chemotherapy. The cell state represented by a high stromal signature was associated with resistance of estrogen receptor–negative breast tumors to neoadjuvant therapy with two different anthracycline-based regimens, FEC (EORTC) and T-FAC (MDA[6]).

The fact that the stromal signature was unable to predict survival in subjects who did not receive chemotherapy (**Fig. 3**) indicates that it does not merely detect a difference in the innate risk of recurrence of breast cancers. Most prognostic signatures currently studied in breast cancer are dominated by proliferation genes[23]. Together with a large body of historical data on classical markers, they highlight the

importance of proliferation in the innate risk of relapse or aggressiveness of tumors[23]. Notably, despite including topoisomerase II α (*TOP2A*), the target of anthracyclines, the proliferation metagene was not associated with response. Taken together, these findings strongly suggest that the stromal metagene is predictive rather than prognostic.

To identify the cellular origin of the stromal signature, we analyzed microarray data from microdissected breast tumors[18]. This showed that the dominant source of stromal gene expression is the tumor stroma. Although this result might appear to settle the issue, it is crucial to recognize that mammary epithelial cells can adopt a stromal gene expression pattern almost indistinguishable from that of reactive stroma if grown in conditions that promote EMT[20]. Our analysis of EMT and mammosphere signatures fully supports this possibility. Nevertheless, genetic analysis of the tumor stroma rarely finds identical changes in tumor epithelial cells and stromal fibroblasts, and few pathologists would seriously contemplate EMT as a general explanation for the existence of reactive stroma in tumors[24].

Recently, another group[25] described a new stroma-derived prognostic predictor (SDPP) based on genes expressed by microdissected tumor-associated stroma. They showed that stromal samples naturally fall into distinct groups associated with very different outcomes. The poor prognosis group has high expression of hypoxia and angiogenesis genes and low expression of type I immune response genes. Despite being derived from microdissected stroma, the SDPP signature is prognostic even in whole-tumor samples comprising tumor epithelium and stroma. In that study, fibroblasts and other normal cell types, such as immune cells and endothelial cells, were analyzed in the microdissected stroma. In contrast, when we refer to reactive stroma, we mean primarily the fibroblasts within the stroma. The key difference between the two studies is that our signature assesses only a single aspect of gene expression in the stroma, whereas those researchers used a composite signature that takes into account multiple cell types and processes occurring in the stroma. Not surprisingly, the conclusions reached by the two studies are different: the SDPP captures information about the innate aggressiveness of the tumor, whereas the *DCN* signature captures information about the responsiveness of the tumor to chemotherapy.

There are many precedents for tumor-stroma interactions to modulate the response to chemotherapy. Integrin activation renders mammary epithelial cells resistant to apoptosis induction by a wide range of different treatments[26], hyaluronic acid promotes resistance of breast cancer cell lines to doxorubicin[27] and adhesion of multiple myeloma cells and monocytic leukemia cells to fibronectin confers a survival advantage in the presence of doxorubicin[28,29]. In experimental models, osteopontin secretion by tumor cells has recently been linked to mobilization of bone marrow cells that contribute to the formation of reactive stroma in tumors[30]. Release of paracrine survival factors by stromal cells recruited by the tumor cells is one possible explanation of our results: the reactive stromal signature could be a marker for a form of cross-talk between tumor epithelial and reactive stromal cells that makes the tumor cells resistant to FEC chemotherapy.

In conclusion, we have developed a 50-gene signature that predicts poor response to anthracycline-based neoadjuvant chemotherapy in two independent data sets. This predictor reflects the activation state of the tumor stroma. Future and larger studies will be required to test whether this signature can be combined with other variables—for example, from gene dosage or mutational analysis—to construct multivariable classifiers that can predict response even better. Our findings provide new insights into the contribution of the tumor microenvironment to breast cancer biology and may lead to the development of new therapies for breast cancer.

METHODS

Subject selection and sample processing. This study was performed in the context of a prospective trial of neoadjuvant chemotherapy (EORTC 10994/BIG 00-01). Ethical approval for the clinical trial and associated translational projects was obtained from the institutional review boards of Institut Bergonié, Bordeaux; Hôpitaux Universitaires de Genève; Western General Hospital, Edinburgh; Centre René Huguenin, Saint-Cloud; Institut Jules Bordet, Brussels; Institut du Cancer Nantes Atlantique, Nantes; Centre Paul Strauss, Strasbourg; Medical University Gdansk; Centre Val d'Aurelle, Montpellier; Centre Henri Becquerel, Rouen; and Karolinska Institutet, Stockholm. Subjects gave signed informed consent for both the clinical and the translational studies. Inclusion criteria for this substudy are described in the **Supplementary Methods** online. We gave the subjects six cycles of 500 mg m^{-2} 5-fluorouracil, 100 mg m^{-2} epirubicin and 500 mg m^{-2} cyclophosphamide (FEC) or of a modified tailored FEC (Swedish patients; see **Supplementary Methods**) followed by breast-conserving surgery or mastectomy. We defined pCR as disappearance of the invasive component of the primary tumor after treatment, with, at most, scattered tumor cells detected by the pathologist in the resection specimen. There were 28 pCR cases among 63 cases included in this study. Complete details of the methods used for RNA extraction, hybridization and data preprocessing and cross-platform mapping are described in the **Supplementary Methods**.

Statistical analyses. The data sets and the statistical methods used are described in detail in the **Supplementary Materials**. Briefly, the aim of the linear regression model was to condense the information contained in a cluster of functionally related genes into a single value that we refer to as a metagene. For each of the nine functionally related gene clusters included in the model, a representative gene was defined a priori. The representative genes are genes regularly found grouped with other functionally related genes when performing clustering. We used a multivariate linear regression model to identify the genes showing similar expression to the representative genes. Linear models provide the framework to allow easy adjustment for potential confounding effects and integration of data coming from different technological platforms. In all cases, we selected the genes belonging to a metagene with external data (NKI-EMC data set) only; hence, no pCR outcome information was used for gene selection. We fitted the linear model separately for each study, and we combined the associated t statistics with the fixed-effect meta-analytical method[31]. We estimated P values by random permutation according to the method of Westfall[32]. **Supplementary Table 3** lists the t statistics and the corresponding P values for all genes and model coefficients. We fixed the number of genes used per metagene at 50. The number of genes significantly associated ($P < 0.05$) with the representative gene was always greater that 50 genes. The approach used to test the impact of changing the number of genes in the stromal metagene is described in the **Supplementary Methods**. The heat maps used to visualize the expression of single genes in the metagenes are color-coded representations of the mean centered gene expression matrix.

We assessed the ability of the metagenes to classify the samples by their pCR status by analyzing the area under receiver operating characteristic curves. We estimated the 95% confidence intervals for this AUC by bootstrapping samples (1,000 iterations). We defined sensitivity as the proportion of pCR cases detected. We adjusted P values for multiple testing by the false discovery rate method[33]. We analyzed survival with the R programming environment[34], using the library 'survival'; we assessed significance with log-rank tests. In the survival analysis, the cases with chemotherapy were taken from a study at Duke University[16] (Duke data set).

We performed the intrinsic gene analysis as previously described[17]. We performed GSEA and molecular subtype classification as previously described[11]. The epithelial, stromal, DTF, solitary fibrous tumors, mammosphere, EMT, TGF-β and Wnt signature genes used for the figures are described in the **Supplementary Methods** and are listed in **Supplementary Table 5** online.

Accession codes. Gene Expression Omnibus: Minimum Information About a Microarray Experiment–compliant data have been deposited with accession code GSE4779. Microdissection data[18], GSE5847.

Note: Supplementary information is available on the Nature Medicine website.

ACKNOWLEDGMENTS

We thank the subjects, doctors and nurses involved in the EORTC 10994/BIG01 study for their generous participation. We thank the staff of the EORTC data center (M. de Vos, S. Lejeune, I. Delmotte and M. Karina) and the technician in the Iggo laboratory (A.-L. Nicoulaz) for assistance with data management and sample processing. We thank I. Xenarios, V. Praz and T. Sengstag for support with bioinformatics. We thank H. Chebab (Pathology Institute, University Lausanne) for supplying human colon tumors. We thank M. Zahn for critical reading of the manuscript. We thank the University of Lausanne DNA array facility and the Swiss Institute for Bioinformatics Vital-IT project for infrastructure support. We thank the Fondation Medic, Fondation Widmer, Oncosuisse, Swiss National Science Foundation and Swiss National Center for Competence in Research (NCCR) Molecular Oncology, EORTC Translational Research Fund, Swedish Cancer Society, King Gustav the Fifth Jubilee Fund and Swedish Research Council for financial support.

AUTHOR CONTRIBUTIONS

J.B., F.B., E. Blot, H.B., D.C., M.C., J.J., G.M., T.P., M.P. and E. Brain supplied tumor tissues and collected the clinical follow-up data. V.B. did the central pathology review. R.I. supervised the laboratory experiments. S.A. performed laboratory experiments. P.F., P.W. and M.D. developed the statistical models. P.F., M.D. and R.I. analyzed the data. J.B. performed additional statistical analyses. P.A. and M.A. performed and supervised additional laboratory experiments. H.B., R.I., P.F. and M.D. designed the study. P.F., M.D., H.B., P.A., J.B., M.A., D.C. and R.I. wrote the report. All investigators contributed to and reviewed the final report.

Published online at http://www.nature.com/naturemedicine/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Colleoni, M. *et al.* Chemotherapy is more effective in patients with breast cancer not expressing steroid hormone receptors: a study of preoperative treatment. *Clin. Cancer Res.* **10**, 6622–6628 (2004).
2. Guarneri, V. *et al.* Prognostic value of pathologic complete response after primary chemotherapy in relation to hormone receptor status and other factors. *J. Clin. Oncol.* **24**, 1037–1044 (2006).
3. Fisher, E.R. *et al.* Pathobiology of preoperative chemotherapy: findings from the National Surgical Adjuvant Breast and Bowel (NSABP) protocol B-18. *Cancer* **95**, 681–695 (2002).
4. Chang, J.C. *et al.* Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* **362**, 362–369 (2003).
5. Gianni, L. *et al.* Gene expression profiles in paraffin-embedded core biopsy tissue predict response to chemotherapy in women with locally advanced breast cancer. *J. Clin. Oncol.* **23**, 7265–7277 (2005).
6. Hess, K.R. *et al.* Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin and cyclophosphamide in breast cancer. *J. Clin. Oncol.* **24**, 4236–4244 (2006).
7. Thuerigen, O. *et al.* Gene expression signature predicting pathologic complete response with gemcitabine, epirubicin and docetaxel in primary breast cancer. *J. Clin. Oncol.* **24**, 1839–1845 (2006).
8. Hannemann, J. *et al.* Changes in gene expression associated with response to neoadjuvant chemotherapy in breast cancer. *J. Clin. Oncol.* **23**, 3331–3342 (2005).
9. Potti, A. *et al.* Genomic signatures to guide the use of chemotherapeutics. *Nat. Med.* **12**, 1294–1300 (2006).
10. Bonnefoi, H. *et al.* Validation of gene signatures that predict the response of breast cancer to neoadjuvant chemotherapy: a substudy of the EORTC 10994/BIG 00–01 clinical trial. *Lancet Oncol.* **8**, 1071–1078 (2007).
11. Farmer, P. *et al.* Identification of molecular apocrine breast tumours by microarray analysis. *Oncogene* **24**, 4660–4671 (2005).
12. Perou, C.M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
13. 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).
14. Sotiriou, C. *et al.* Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J. Natl. Cancer Inst.* **98**, 262–272 (2006).
15. van de Vijver, M.J. *et al.* A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347**, 1999–2009 (2002).
16. Bild, A.H., Potti, A. & Nevins, J.R. Linking oncogenic pathways with therapeutic opportunities. *Nat. Rev. Cancer* **6**, 735–741 (2006).
17. Hu, Z. *et al.* The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7**, 96 (2006).
18. Boersma, B.J. *et al.* A stromal gene signature associated with inflammatory breast cancer. *Int. J. Cancer* **122**, 1324–1332 (2008).
19. Dontu, G. *et al. In vitro* propagation and transcriptional profiling of human mammary stem/progenitor cells. *Genes Dev.* **17**, 1253–1270 (2003).
20. Mani, S.A. *et al.* The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**, 704–715 (2008).
21. Li, X. *et al.* Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. *J. Natl. Cancer Inst.* **100**, 672–679 (2008).
22. West, R.B. *et al.* Determination of stromal signatures in breast carcinoma. *PLoS Biol.* **3**, e187 (2005).
23. Wirapati, P. *et al.* Meta-analysis of gene-expression profiles in breast cancer: toward a unified understanding of breast cancer sub-typing and prognosis signatures. *Breast Cancer Res.* **10**, R65 (2008).
24. Qiu, W. *et al.* No evidence of clonal somatic genetic alterations in cancer-associated fibroblasts from human breast and ovarian carcinomas. *Nat. Genet.* **40**, 650–655 (2008).
25. Finak, G. *et al.* Stromal gene expression predicts clinical outcome in breast cancer. *Nat. Med.* **14**, 518–527 (2008).
26. Weaver, V.M. *et al.* β_4 integrin–dependent formation of polarized three-dimensional architecture confers resistance to apoptosis in normal and malignant mammary epithelium. *Cancer Cell* **2**, 205–216 (2002).
27. Misra, S., Ghatak, S. & Toole, B.P. Regulation of MDR1 expression and drug resistance by a positive feedback loop involving hyaluronan, phosphoinositide 3-kinase, and ErbB2. *J. Biol. Chem.* **280**, 20310–20315 (2005).
28. Damiano, J.S., Cress, A.E., Hazlehurst, L.A., Shtil, A.A. & Dalton, W.S. Cell adhesion mediated drug resistance (CAM-DR): role of integrins and resistance to apoptosis in human myeloma cell lines. *Blood* **93**, 1658–1667 (1999).
29. Hazlehurst, L.A. *et al.* Reduction in drug-induced DNA double-strand breaks associated with β_1 integrin–mediated adhesion correlates with drug resistance in U937 cells. *Blood* **98**, 1897–1903 (2001).
30. McAllister, S.S. *et al.* Systemic endocrine instigation of indolent tumor growth requires osteopontin. *Cell* **133**, 994–1005 (2008).
31. Hedges, L. & Olkin, I. *Statistical Methods for Meta-Analysis*. 39 (Academic Press, London, 1985).
32. Westfall, P. & Young, S. *Resampling-Based Multiple Testing: Examples And Methods For P-Values Adjustment*. Ch.2 (Wiley, New York, 1993).
33. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. [Ser B]* **57**, 289–300 (1995).
34. R Development Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, Vienna, 2005).
35. International Union Against Cancer (UICC). *TNM Classification of Malignant Tumours* 5th edn. (eds. Sobin, L.H. & Wittekind, C.) (Wiley, New York, 1997).
36. Finak, G. *et al.* Gene expression signatures of morphologically normal breast tissue identify basal-like tumors. *Breast Cancer Res.* **8**, R58 (2006).

Erratum: A stroma-related gene signature predicts resistance to neoadjuvant chemotherapy in breast cancer

Pierre Farmer, Hervé Bonnefoi, Pascale Anderle, David Cameron, Pratyaksha Wirapati, Véronique Becette, Sylvie André, Martine Piccart, Mario Campone, Etienne Brain, Gaëtan MacGrogan, Thierry Petit, Jacek Jassem, Frédéric Bibeau, Emmanuel Blot, Jan Bogaerts, Michel Aguet, Jonas Bergh, Richard Iggo & Mauro Delorenzi
Nat. Med. **15, 68–74 (2009); published online 4 January 2009; corrected after print 5 February 2009**

In the version of this article initially published, one of the authors' names was misspelled. Pratyakasha Wirapati should have been Pratyaksha Wirapati. The error has been corrected in the HTML and PDF versions of the article.

Erratum: Loss of ETHE1, a mitochondrial dioxygenase, causes fatal sulfide toxicity in ethylmalonic encephalopathy

Valeria Tiranti, Carlo Viscomi, Tatjana Hildebrandt, Ivano Di Meo, Rossana Mineri, Cecilia Tiveron, Michael D Levitt, Alessandro Prelle, Gigliola Fagiolari, Marco Rimoldi & Massimo Zeviani
Nat. Med.; **doi:10.1038/nm.1907; corrected online 20 January 2009**

In the version of this article initially published online, 'SDO' should have been 'SDH' in Figure 4a,b. The error has been corrected for the print, PDF and HTML versions of this article.

Corrigendum: Rescuing a failing heart: putting on the squeeze

David A Kass
Nat. Med. **15, 24–25 (2009); corrected after print February 5 2009**

In the version of this article initially published, the competing financial interests statement was missing. The error has been corrected in the HTML and PDF versions of the article.