Vol 439|19 January 2006|**doi:10.1038/nature04296**

Oncogenic pathway signatures in human cancers as a guide to targeted therapies

Andrea H. Bild[1,2], Guang Yao[1,2], Jeffrey T. Chang[1,2], Quanli Wang[1], Anil Potti[1,4], Dawn Chasse[1,2], Mary-Beth Joshi[3], David Harpole[3], Johnathan M. Lancaster[7], Andrew Berchuck[5], John A. Olson Jr[1,3], Jeffrey R. Marks[3], Holly K. Dressman[1,2], Mike West[6] & Joseph R. Nevins[1,2]

The development of an oncogenic state is a complex process involving the accumulation of multiple independent mutations that lead to deregulation of cell signalling pathways central to the control of cell growth and cell fate[1–3]. The ability to define cancer subtypes, recurrence of disease and response to specific therapies using DNA microarray-based gene expression signatures has been demonstrated in multiple studies[4]. Various studies have also demonstrated the potential for using gene expression profiles for the analysis of oncogenic pathways[5–11]. Here we show that gene expression signatures can be identified that reflect the activation status of several oncogenic pathways. When evaluated in several large collections of human cancers, these gene expression signatures identify patterns of pathway deregulation in tumours and clinically relevant associations with disease outcomes. Combining signature-based predictions across several pathways identifies coordinated patterns of pathway deregulation that distinguish between specific cancers and tumour subtypes. Clustering tumours based on pathway signatures further defines prognosis in respective patient subsets, demonstrating that patterns of oncogenic pathway deregulation underlie the development of the oncogenic phenotype and reflect the biology and outcome of specific cancers. Predictions of pathway deregulation in cancer cell lines are also shown to predict the sensitivity to therapeutic agents that target components of the pathway. Linking pathway deregulation with sensitivity to therapeutics that target components of the pathway provides an opportunity to make use of these oncogenic pathway signatures to guide the use of targeted therapeutics.

We used human primary mammary epithelial cell cultures (HMECs) to develop a series of pathway signatures. Recombinant adenoviruses were used to express various oncogenic activities in an otherwise quiescent cell, thereby specifically isolating the subsequent events as defined by the activation/deregulation of that single pathway. Various biochemical measures demonstrate pathway activation (Supplementary Fig. 1). RNA from multiple independent infections was collected for DNA microarray analysis using Affymetrix Human Genome U133 Plus 2.0 Array. Gene expression signatures that reflect the activity of a given pathway are identified using supervised classification methods of analysis previously described[12]. The analysis selects a set of genes for which the expression levels are most highly correlated with the classification of HMEC samples into oncogene-activated/deregulated versus control (green fluorescent protein, GFP). The dominant principal components from such a set of genes then defines a relevant phenotype-related metagene, and regression models assign the relative probability of pathway deregulation in tumour or cell line samples.

It is clear from Fig. 1a that the various signatures distinguish cells expressing the oncogenic activity from control cells. Given the potential for overlap in the pathways, we also examined the extent to which the signatures distinguish one pathway from another. Use of the first three principal components from each signature, evaluated across all experimental samples, demonstrates that the patterns of expression in each signature are specific to each pathway; the gene expression patterns accurately distinguish the individual oncogenic effects despite overlapping downstream consequences (Fig. 1b). The genes identified as comprising each signature are listed in Supplementary Table 1. To evaluate more formally the predictive validity and robustness of the pathway signatures, a leave-one-out cross validation study was applied to the set of pathway predictors. This analysis demonstrates that these signatures of oncogenic pathways can accurately predict the cells expressing the oncogenic activity from the control cells (Supplementary Fig. 2). The analysis clearly distinguishes and predicts the state of an oncogenic pathway.

Further verification of the capacity of oncogenic pathway signatures to predict accurately the status of pathways made use of tumour samples derived from various mouse cancer models. Pathway signatures were regenerated from the genes common to both human and mouse data sets; the analysis was trained on the HMEC-derived signatures and then used to predict the pathway status of all tumours. These studies were carried out using three of the pathway signatures for which we had matching mouse models that could be used for validation: Myc, Ras and E2F3. Across the set of mouse tumours, this analysis evaluates the relative probability of pathway deregulation of each tumour—that is, the predicted status of the pathway in each mouse tumour based only on the signatures developed in HMECs. These predictions are displayed as a colour map: red indicates a high probability of pathway deregulation and blue indicates a low probability, with predictions sorted by the relative probability of pathway deregulation. As shown in Fig. 2a, the pathway predictions exhibit close correlation with the molecular basis for tumour induction. For instance, the five mouse mammary tumour virus (MMTV)-*MYC* tumours exhibit the highest probability of Myc pathway deregulation, whereas the six Rb null tumours exhibit the highest probability of E2F3 deregulation. The probability of Ras pathway activation was highest in the MMTV-*HRAS* animals and MMTV-*MYC* tumours; this indication of Ras pathway activation in the MMTV-*MYC* tumours is consistent with past results demonstrating a selection for Ras mutations in these tumours[6,13]. Further substantiation and validation was obtained from a series of tumours in which Ras activity was spontaneously activated by homologous recombination in adult animals, more closely

[1]Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina 27708, USA. [2]Department of Molecular Genetics and Microbiology, [3]Department of Surgery, [4]Department of Medicine, [5]Department of Obstetrics & Gynecology, Duke University Medical Center, Durham, North Carolina 27710, USA. [6]Institute of Statistics and Decision Sciences, Duke University, Durham, North Carolina 27708, USA. [7]H. Lee Moffitt Cancer Center & Research Institute, University of South Florida, Tampa, Florida 33612, USA.

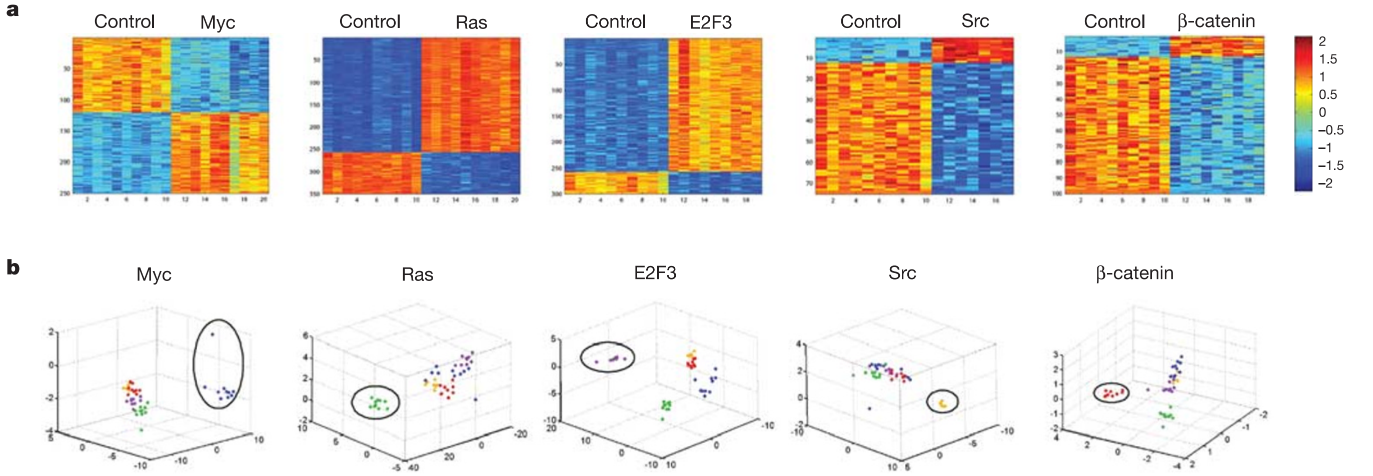

Figure 1 | Gene expression patterns that predict oncogenic pathway deregulation. **a**, Image intensity display of the expression levels of genes most highly weighted in the predictor differentiating GFP-expressing control cells from cells expressing the indicated oncogenic activity. Expression levels are standardized to zero mean and unit variance across samples, displayed with genes as rows and samples as columns, and colour coded to indicate high (red) or low (blue) expression levels. **b**, Scatter plots depicting the classification of samples based on the first three principal components (expression patterns) derived from each signature, as shown in **a**. The gene expression values for each signature were extracted from all experimental samples and mean centred, then single value decomposition analysis was applied across all samples. Colour coding for samples is as follows: Myc, blue; Ras, green; E2F3, purple; Src, yellow; β-catenin, red. Samples representing the specific pathway being examined are circled.

mimicking pathway deregulation in human tumours[14]. There was a consistent prediction of Ras pathway deregulation within these tumours when compared to the set of samples from control lung tissue (Fig. 2b). Taken together, these results strongly support the conclusion that the various oncogenic pathway signatures do reliably reflect pathway status under a variety of circumstances, and thus can serve as useful tools to probe the status of these pathways.

Previous work has linked Ras activation with the development of adenocarcinomas of the lung[15,16]. We made use of a set of non-small cell lung carcinoma (NSCLC) samples to predict the pathway status and then sorted according to predicted Ras activity. As shown in Fig. 2c, Ras pathway status very clearly correlates with the histological subtype—most of the adenocarcinoma samples exhibit a high probability of Ras deregulation relative to the squamous cell carcinoma samples. Prediction of the status of the other pathways revealed a less distinct pattern, although each tended to be more active in the squamous cell carcinoma samples (Supplementary Fig. 3). This pattern becomes more evident in the analysis shown in Fig. 3. An examination of Ras mutation identified 11 samples with K-Ras mutations, all confined to the adenocarcinomas (indicated by an asterisk in the figure) (Supplementary Table 2). Overall, 14% of NSCLC tumours and 29% of the adenocarcinomas had K-Ras mutations in codon 12. Because nearly all of the adenocarcinomas exhibited Ras pathway deregulation, it seems that deregulation of the Ras pathway is indeed a characteristic of development of adenocarcinoma of the lung, and that this can occur as a result of Ras mutations as well as following other events that deregulate the pathway.

Whereas the analysis of pathway deregulation as shown in Fig. 2c depicts the status of an individual pathway, the real power in this approach is the ability to identify patterns of pathway deregulation, using hierarchical clustering, much the same as identifying patterns of gene expression. We started with an analysis of the lung cancer samples (Fig. 3a, left panel). This analysis distinguished adenocarcinomas from squamous cell carcinomas, driven in part by the Ras pathway distinction. It is also evident that the tumours predicted as exhibiting relatively low Ras activity are generally predicted at higher levels of Myc, E2F3, β-catenin and Src activity (clusters 1–3). Conversely, the tumours with relatively elevated Ras activity exhibited relatively lower levels of these other pathways (clusters 4–7). Independent of the tumour histopathology, concerted deregulation

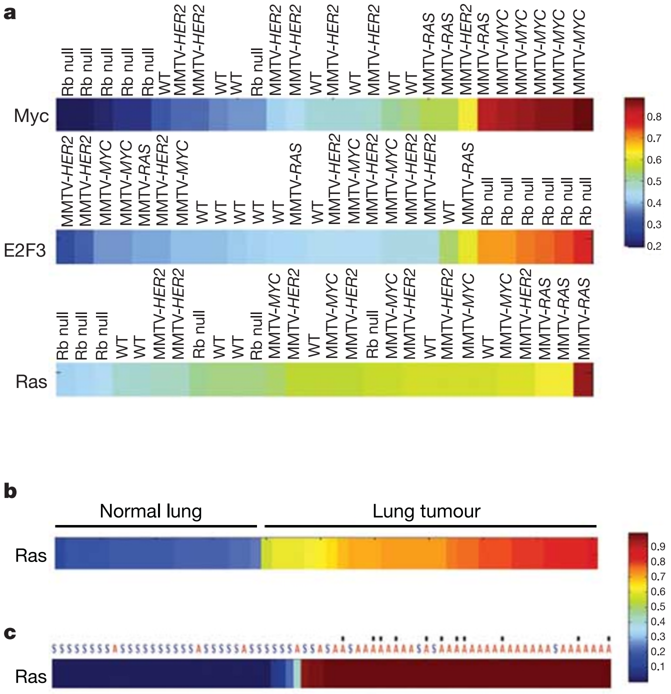

Figure 2 | Validation of pathway predictions in tumours. **a**, Mouse mammary tumours derived from mice transgenic for the MMTV-*MYC* (five samples), MMTV-*HRAS* (three samples) or MMTV-*HER2* (seven samples) oncogenes, tumours dependent on loss of Rb (six samples), or seven samples of normal mammary tissue were used to verify accuracy and specificity of our signatures. The predicted probability of Myc, E2F3 and Ras activity in mouse tumours was sorted from low (blue) to high (red), and displayed as a colour bar. **b**, Prediction of pathway status in mouse lung cancer model. A set of previously published mouse Affymetrix expression data comparing normal and tumour lung tissue with spontaneous activating *KRAS* mutations[14] was used to validate the predictive capacity of the Ras pathway signature. The predicted probability of Ras activity in the normal and tumour tissue was sorted from low (blue) to high (red), and displayed as a colour bar. **c**, Relationship of Ras pathway status in NSCLC samples to cell type of tumour origin. The corresponding tumour cell type is indicated as either squamous (S) or adenocarcinoma (A). Ras mutation status is indicated by an asterisk.

of Ras with β-catenin, Src and Myc (cluster 8) identified a population of patients with poor survival—a median survival of 19.7 months versus 51.3 months for all other clusters (Fig. 3a, right panel). Furthermore, this subpopulation of patients exhibited worse survival than any of the groups of patients identified based on the status of any single pathway deregulation (Supplementary Fig. 4). This analysis demonstrates the ability of integrated pathway analysis, based on multiple signatures of component pathway deregulation, to define improved categorization of lung cancer patients.

Two additional examples made use of large sets of breast cancer samples (Fig. 3b) and ovarian cancer samples (Fig. 3c). Again, there were evident patterns of pathway deregulation, distinct from that seen in the lung samples, which characterized the breast and ovarian tumours. For breast cancer, there were two clusters of patients with good prognosis (clusters 2 and 4), and two clusters with poor prognosis (clusters 1 and 3). Furthermore, clusters 2 and 3, which both contain oestrogen receptor (ER)-positive tumours (and no discernable differences in HER2 status or other clinical parameters), show distinct survival rates (P-value = 0.07). Patients defined by cluster 5 (in which higher than average β-catenin and Myc activities were predicted, and E2F3 activity was lower than average) exhibited very poor survival, again illustrating the importance of co-deregulation of multiple oncogenic pathways as a determinant of clinical outcome. A final analysis made use of an advanced stage (III or IV) ovarian cancer data set. The ovarian samples exhibited a dominant pattern of β-catenin and Src deregulation, either elevated (cluster 1 and 2) or diminished (clusters 3-6). Notably, the co-deregulation of Src and β-catenin defined by clusters 1 and 2 identifies a population of patients with very poor survival compared to other pathway clusters (median survival: 29.0 months versus 91.0 months) (Fig. 3c, right panel). Once again, for these cases, individual pathway status did not stratify patient subgroups as effectively as patterns of multiple pathway deregulation (Supplementary Fig. 4).

Given the capacity of the gene expression signatures to predict deregulation of oncogenic signalling pathways, we have also addressed the extent to which this could predict sensitivity to a therapeutic agent that targets that pathway. To explore this, we predicted pathway deregulation in a series of breast cancer cell lines to be screened against potential therapeutic drugs. The results using the set of five pathway predictors, together with an initial collection of breast cancer cell lines, are shown in Fig. 4a. Biochemical characteristics of the cell lines relevant for pathway analysis are summarized in Supplementary Table 3 and Supplementary Fig. 5. In each case, the relative probabilities of pathway activation are predicted from the signature in a manner completely analogous to the prediction of pathway status in tumours. In most cases, there is a good correlation between biochemical measures of pathway activation and prediction based on gene expression signatures. An exception is with Ras, where there is not a significant correlation between the biochemical measure of pathway activation and pathway prediction, presumably reflecting additional events not measured in the biochemical assay. Clearly, the critical issue is whether the gene expression signature predicts drug sensitivity—this point is addressed by the dose–response assays in Fig. 4b.

In parallel with mapping the pathway status, the cell lines were assayed with drugs known to target specific activities within given

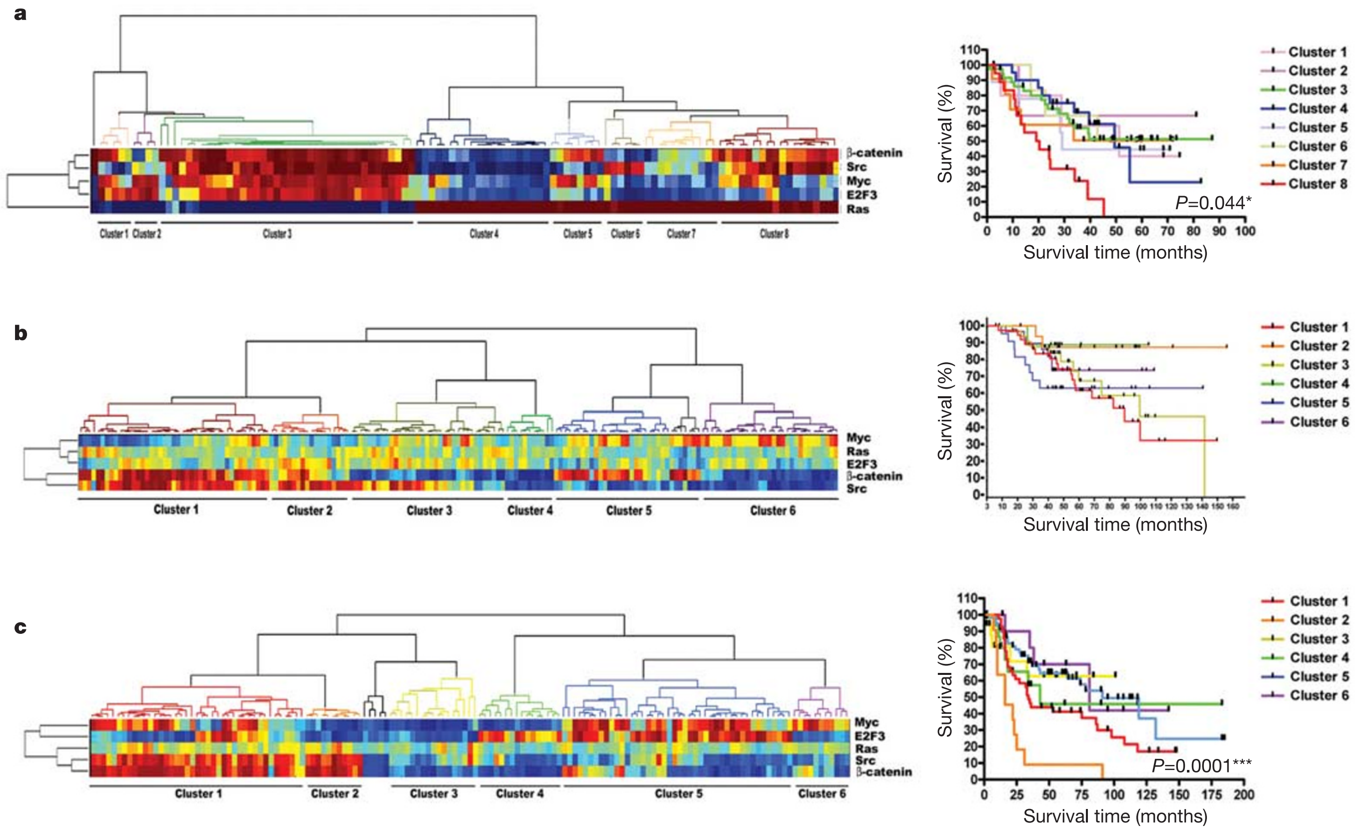

Figure 3 | Patterns of pathway deregulation in human cancers. **a**, Hierarchical clustering of predictions of pathway deregulation in samples of human lung tumours (left panel). Prediction of Ras, Myc, E2F3, β-catenin and Src pathway status for each tumour sample was independently determined using supervised binary regression analysis, as described. Red indicates high probability of pathway activation, with blue indicating a low probability. Patterns in the tumour pathway predictions were identified by hierarchical clustering, and separate clusters are indicated by coloured dendograms. Kaplan–Meier survival analysis for lung cancer patients based on pathway clusters (right panel). Patient clusters with correlative pathway deregulation shown in the left panel correspond to clusters comprising each independent survival curve. Black tick marks represent censored patients. **b**, Breast cancer. Same as for **a**. **c**, Ovarian cancer. Same as for **a**.

oncogenic pathways. The assays involve growth inhibition measurements using standard colorimetric assays[17,18]. The result of testing the sensitivity of the cell lines to inhibitors of the Ras pathway using both a farnesyl transferase inhibitor (L-744,832) and a farnesylthiosalicylic acid (FTS) is shown in Fig. 4b. In addition, a Src inhibitor (SU6656) was also used for these assays. In each case, the results show a close concordance and correlation between the probability of Ras and Src pathway deregulation based on the gene expression prediction, and the extent of cell proliferation inhibition by the respective drugs (Fig. 4b). Furthermore, comparison of the drug inhibition results with predictions of other pathways failed to demonstrate a significant correlation (Supplementary Fig. 6). These results confirm the ability of the defined 'pathway deregulation signatures' to also predict sensitivity to therapeutic agents that target the corresponding pathways.

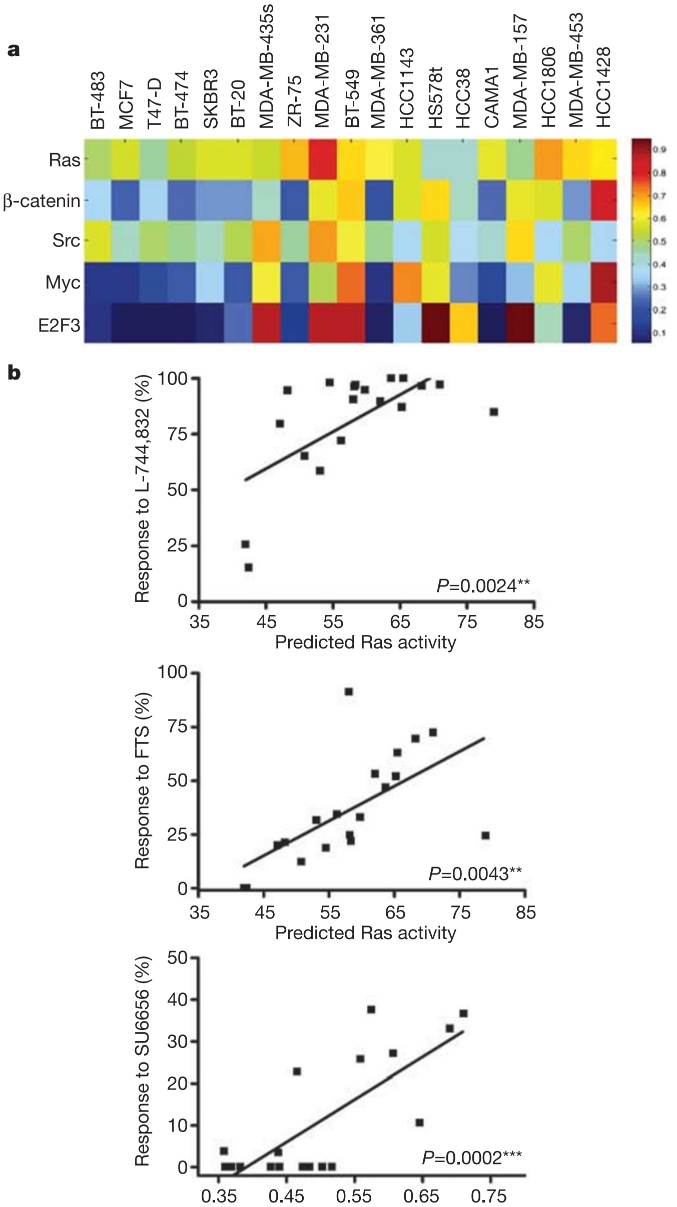

Figure 4 | Pathway deregulation in breast cancer cell lines predicts drug sensitivity. a, Pathway predictions in breast cancer cell lines. The results plotted show the predicted probability of pathway activation (red indicates high probability; blue indicates low probability). **b**, Sensitivity to pathway-specific drugs. Top: cells were treated with 3.75 μM of farnesyltransferase inhibitor (L-744,832) for 96 h. Proliferation was assayed using a standard MTS tetrazolium colorimetric method. The degree of proliferation inhibition was plotted as a function of probability of Ras pathway activation as determined in **a**. Middle: same as for the top panel but using FTS (200 μM). Bottom: same as for the top panel but using the Src pathway inhibitor SU6656 (1.5 μM), and with the degree of proliferation inhibition plotted as a function of Src pathway activation.

In most instances, the consequence of mutations in proto-oncogenesor inactivation of tumour suppressor genes is the deregulation of cellular signalling pathways, which ultimately affects the expression of a variety of genes. Our use of gene expression signatures that reflect the action of oncogenic pathway deregulation provides a strategy for measuring the functional consequence of these events. Undoubtedly, an ability to distinguish the deregulation of additional subpathways, as well as pathways reflective of additional aspects of tumorigenesis (apoptosis, DNA repair, and so on), will help to categorize further and understand the complexity of tumour development and the oncogenic process.

Although the development of targeted biological agents holds the promise of a more precise matching of therapy with disease mechanism, it is nevertheless true that the success rate of single agents as well as the selection of combination therapies could be improved. The ability to predict the deregulation of various oncogenic pathways through gene expression analysis offers an opportunity to identify new therapeutic options for patients by providing a potential basis for guiding the use of pathway-specific drugs. The major value of this approach may be the capacity to direct combinations of therapies—multiple drugs that target multiple pathways—based on information that specifies the activation state of the pathways.

METHODS

Cell and RNA preparation. Human mammary epithelial cells from a breast reduction surgery at Duke University were isolated and cultured according to previously published protocols[19]. These cells were a gift from G. Huper. Cells are brought to quiescence, and then infected with adenovirus expressing either human c-Myc, activated H-Ras, human c-Src, human E2F3, or activated β-catenin (we thank J. Kitajewski, W. El-Deiry, Z. German and D. Kuppuswamy for DNA constructs and adenovirus). Eighteen hours after infection, cells were collected and expression of oncogenes and their secondary targets was determined by a standard western blotting protocol (see Supplementary Information). Activation status of kinase pathways for the breast cancer cell lines was determined for growing cells using the following methods. Ras activation is measured using a Ras activation assay kit (Upstate Biotechnology). c-Src activation was determined by western blotting using a phospho-Tyr 416 Src antibody. E2F3, Myc and β-catenin activity was measured by isolating nuclear extracts from cells as previously described, and performing western blotting analysis using antibodies specific for each oncoprotein. Total RNA was extracted for cell lines using the Qiashredder and Qiagen RNeasy Mini kit. Quality of the RNA was checked by an Agilent 2100 Bioanalyser.

Tumour analyses. Tumour tissue from breast, ovarian and lung cancer patients was >60% tumour, and was selected for by stage and histology. Total RNA was extracted as previously described[20].

DNA microarray analysis. Samples were prepared according to the manufacturer's instructions and as previously published[21,22]. Experiments to generate signatures use Human U133 2.0 Plus GeneChips. Breast tumours were hybridized to Hu95Av2 arrays, ovarian tumours to Hu133A arrays, and lung tumours to Human U133 2.0 plus arrays (Affymetrix). All microarray data are available at http://data.cgt.duke.edu/oncogene.php and on GEO.

Cross-platform Affymetrix GeneChip comparison. To map the probe sets across various generations of Affymetrix GeneChip arrays, we used an in-house program: Chip Comparer (http://tenero.duhs.duke.edu/genearray/perl/chip/chipcomparer.pl). Details are in Supplementary Information.

Statistical analysis methods. Analysis of expression data is as previously described[12]. Briefly, before statistical modelling, gene expression data are filtered to exclude probe sets with signals present at background noise levels, and probe sets that do not vary significantly across samples. Each signature summarizes its constituent genes as a single expression profile, and is here derived as the top principal components of that set of genes. When predicting the pathway activation of cancer cell lines or tumour samples, gene selection and identification is based on the training data, and then metagene values are computed using the principal components of the training data and additional cell line or tumour expression data. Bayesian fitting of binary probit regression models to the training data then permits an assessment of the relevance of the metagene signatures in within-sample classification, and estimation and uncertainty assessments for the binary regression weights mapping metagenes to probabilities of relative pathway status. Predictions of the relative pathway status of the validation cell lines or tumour samples are then evaluated, producing estimated relative probabilities—and associated measures of uncertainty—of activation/deregulation across the validation samples. Hierarchical clustering of tumour predictions was performed using Gene Cluster 3.0 (ref. 23). Genes and tumours were clustered using average linkage with the uncentred correlation similarity metric. Standard Kaplan–Meier mortality curves and their significance levels were generated for clusters of patients with similar patterns of oncogenic pathway deregulation using GraphPad software. For the Kaplan–Meier survival analyses, the survival curves are compared using the logrank test. For full details, see Supplementary Information.

Cell proliferation assays. Growth curves and dosing ranges for the breast cancer cell lines were carried out as described in Supplementary Information. Sensitivity to a farnesyl transferase inhibitor (L-744,832), FTS and an Src inhibitor (SU6656) was determined by quantifying the percentage reduction in growth (versus DMSO controls) at 96 h using a standard MTS (3-(4,5-dimethylthiazol-2-yl)-5-(3-carboxymethoxyphenyl)-2-(4-sulphophenyl)-2H-tetrazolium) colorimetric assay (Promega). Concentrations used were from 100 nM to 10 μM L-744,832, 10 to 200 μM FTS, and 300 nM to 10 μM SU6656. All experiments were repeated at least three times.

K-Ras mutation assay. K-Ras mutation status was determined using restriction fragment length polymorphism and sequencing as previously described[24]. See Supplementary Information for additional details.

Received 11 August; accepted 3 October 2005.
Published online 6 November 2005.

1. Fearon, E. R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **17,** 671–674 (1990).
2. Hanahan, D. & Weinberg, R. A. The hallmarks of cancer. *Cell* **100,** 57–70 (2000).
3. Sherr, C. J. Cancer cell cycles. *Science* **274,** 1672–1677 (1996).
4. Ramaswamy, S. & Golub, T. R. DNA microarrays in clinical oncology. *J. Clin. Oncol.* **20,** 1932–1941 (2002).
5. Lamb, J. *et al.* A mechanism of cyclin D1 action encoded in the patterns of gene expression in human cancer. *Cell* **114,** 323–334 (2003).
6. Huang, E. *et al.* Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nature Genet.* **34,** 226–230 (2003).
7. Black, E. P. *et al.* Distinct gene expression phenotypes of cells lacking Rb and Rb family members. *Cancer Res.* **63,** 3716–3723 (2003).
8. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. *Nature Genet.* **36,** 1090–1098 (2004).
9. Rhodes, D. R. *et al.* Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. *Proc. Natl Acad. Sci. USA* **101,** 9309–9314 (2004).
10. Ramaswamy, S., Ross, K. N., Lander, E. S. & Golub, T. R. A molecular signature of metastasis in primary solid tumors. *Nature Genet.* **33,** 49–54 (2003).
11. Mootha, V. K. *et al.* PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature Genet.* **34,** 267–273 (2003).
12. 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).
13. D'Crus, C. M. *et al.* c-MYC induces mammary tumorigenesis by means of a preferred pathway involving spontaneous Kras2 mutations. *Nature Med.* **7,** 235–239 (2001).
14. Sweet-Cordero, A. *et al.* An oncogenic KRAS2 expression signature identified by cross-species gene expression analysis. *Nature Genet.* **37,** 48–54 (2005).
15. Rodenhuis, S. *et al.* Mutational activation of the K-ras oncogene and the effect of chemotherapy in advanced adenocarcinoma of the lung: a prospective study. *J. Clin. Oncol.* **15,** 285–291 (1997).
16. Salgia, R. & Skarin, A. T. Molecular abnormalitities in lung cancer. *J. Clin. Oncol.* **16,** 1207–1217 (1998).
17. Cory, A. H. Use of an aqueous soluble tetrazolium/formazan assay for cell growth assays in culture. *Cancer Commun.* **3,** 207–212 (1991).
18. Riss, T. L. & Moravec, R. A. Comparison of MTT, Xtt, and a novel tetrazolium compound for MTS for *in vitro* proliferation and chemosensitivity assays. *Mol. Biol. Cell* **3,** 184a (1992).
19. Stampfer, M. R. & Yaswen, P. Culture systems for study of human mammary epithelial cell proliferation, differentiation, and transformation. *Cancer Surv.* **18,** 7–34 (1993).
20. Huang, E. *et al.* Gene expression predictors of breast cancer outcomes. *Lancet* **361,** 1590–1596 (2003).
21. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* (in the press).
22. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalizaton methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19,** 185–193 (2003).
23. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci.* **95,** 14863–14868 (1998).
24. Mitsudomi, T. *et al.* Mutations of ras genes distinguish a subset of non-small-cell lung cancer cell lines from small-cell lung cancer cell lines. *Oncogene* **6,** 1353–1362 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to P. Bild for the inspiration to pursue this research. We also thank K. Shianna, J. Freedman, S. Mori, L. Jakoi and K. Culler for their assistance. A.H.B. has been supported by an AACR-Anna D. Barker Fellowship and an ACS Postdoctoral Fellowship. This work was supported by grants from the NIH (to J.R.N.).

Author Contributions A.H.B. was responsible for all experimental work and computational data analysis. A.H.B. and J.R.N. were responsible for project planning and data analysis. G.Y., J.T.C. and Q.W. were responsible for generation of specialized computer programs used in these studies. H.K.D. and A.P. provided intellectual input and data management support. D.C. and M.-B.J. provided technical support for experiments. M.W. was responsible for conception of the statistical approach and intellectual input. A.B., J.M.L., J.R.M., J.A.O. and D.H. were responsible for the development of clinical resources used in the study.

Author Information The GEO accession numbers for the datasets are: GSE3156, breast cancer cell lines; GSE3158, mouse tumour data set; GSE3151, oncogene signature data set; GSE3141, lung cancer data set; GSE3143, breast cancer data set; GSE3149, ovarian cancer data set. Reprints and permissions information is available at npg.nature.com/reprintsandpermissions. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.R.N. (j.nevins@duke.edu).