Discrimination of Human Lung Neoplasm from Normal Lung by Two Target Genes

Hans-Stefan Hofmann, Gesine Hansen, Stefan Burdach, Babett Bartling, Rolf-Edgar Silber, and Andreas Simm

Department of Cardiothoracic Surgery, Department of Pediatrics, and Children's Cancer Research Center, Martin Luther University Halle-Wittenberg, Halle; and Department of Pediatrics, Technical University of Munich, Munich, Germany

Simple tools for discrimination of lung tissues can be useful in a fast machine-aided diagnosis, for example, by tumor-specific microarrays. We demonstrate that an easy ratio technique, based on the expression levels of only two genes differentially expressed in lung tumor and normal lung samples, allows discrimination of normal and neoplastic lung with a sensitivity of 100% and specificity of 90.5%. DNA microarray analysis of 99 lung tumor samples and 15 normal lung tissues revealed that receptor for advanced glycation end products (RAGE) mRNA is reduced fourfold ($p = 7.8 \times 10^{-11}$) and cyclin-B2 mRNA is upregulated twofold ($p = 5.9 \times 10^{-18}$) in lung carcinoma compared with normal lung. The microarray-calculated expression ratio of RAGE to cyclin-B2 was used in polymerase chain reaction analysis of 84 independent blinded samples to discriminate tumor and corresponding normal lung tissues. In 94.7% of the samples this quotient correctly distinguished non–small cell lung cancer from normal lung tissue, suggesting the RAGE/cyclin-B2 quotient as a potential means for diagnosis of lung cancer.

Keywords: cyclin-B2; lung cancer; lung metastases; receptor for advanced glycation end products; tissue discrimination

Lung cancer is the leading cause of death worldwide among patients with cancer, with non–small cell lung cancer (NSCLC) accounting for about 80% of newly diagnosed cases (1). Current 5-year lung cancer survival rates are estimated at 14% (2). A patient's survival depends on histology and cancer stage. The current lung cancer classification is based on clinicopathological features. However, these methods are insufficient to reflect the complicated underlying molecular events that drive the neoplastic process (3). After sequencing the human genome, the development of high-throughput methods such as DNA microarrays offers new insights into disease mechanisms, target identification, and function as a possible diagnostic tool.

Currently, this technology is based on the analysis of an array, per centimeter squared, of up to 100,000 oligonucleotides or 10,000 polymerase chain reaction (PCR) products. However, in cancer studies only a small number of genes or expressed sequence tags are differentially expressed. Many bioinformatic tools under development and testing are quite complex and/or rely on these significantly expressed genes to establish a diagnosis for unknown samples. The concept of typical tissue-specific genes implies the existence of a few meaningful genes (tumor markers) for diagnosis of NSCLC. The minimal number of predictor genes for NSCLC is not known.

We have therefore generated a DNA microarray–based list of differentially expressed genes and identified lung- or lung tumor–specific target genes. We then tested the feasibility of using ratios of gene expression levels and chosen thresholds to accurately discriminate between lung tumor and normal lung tissue. These diagnostic marker genes should allow reproducible tissue determination and an extension to routine clinical application.

METHODS

Tumor Samples

Tissue specimens from tumor and noninvolved lung of consecutive patients with NSCLC (adenocarcinoma and squamous cell carcinoma) or lung metastases, who underwent pulmonary resection surgery between 1999 and 2001, were included. One hundred ninety-nine snap-frozen lung tumors (n = 133), lung metastases (n = 8), and normal lung tissues (n = 58) were used to create two data sets. Fifty-six squamous cell carcinomas, 43 adenocarcinomas, and 15 noninvolved lung samples were investigated by microarray analysis (Data Set A). Results of the microarray analysis were proven by reverse transcription (RT)-PCR on a second independent data set (Data Set B) consisting of samples from 17 squamous cell carcinomas, 17 adenocarcinomas, 8 lung metastases of different primary tumors, and the corresponding noninvolved paired lung tissues. Tumor histology was classified according to the World Health Organization classification system (4).

The use of human tissues was approved by the local ethics committee and the patients gave informed consent.

Microarray Expression Analysis

A total of 10 μg of RNA from each sample was used to prepare biotinylated target cRNA as previously described (5–7). A detailed protocol is available at www.affymetrix.com. Samples were hybridized to a custom expression monitoring DNA microarray designed by Eos Biotechnology (South San Francisco, CA), using Affymetrix GeneChip technology (Affymetrix, Santa Clara, CA) (8), that contained essentially all expressed human genes in the public domain at the time of design (EOS-K). Sequences included on the array were derived from human genomic, expressed mRNA, and expressed sequence tag databases in GenBank (9). Consensus sequences representing human expressed sequences were generated with Clustering and Alignment Tools software (DoubleTwist, Oakland, CA), and prediction of the expressed genome from the human genome sequence was done by *ab initio* exon prediction (10). The 59,000 probe sets on this microarray represent about 45,000 mRNA and expressed sequence tag clusters and 6,200 predicted exons. Data were used after γ distribution normalization.

Gene Chip Analysis

For analysis of gene expression total RNA was extracted with TRIzol (GIBCO, Karlsruhe, Germany) and biotinylated cRNA was prepared by *in vitro* transcription after synthesis of double-stranded cDNA by standard protocols. After cRNA fragmentation and hybridization with microarrays (EOS-K chips), signals were detected with streptavidin–phycoerythrin. Signal enhancement was performed with biotinylated goat anti-streptavidin antibodies. Arrays were washed and stained with the GeneChip Fluidics Station 400 (Affymetrix) and scanned with a GeneArray scanner (Agilent Technologies, Palo Alto, CA). Primary image analysis was performed by using Microarray Suite 5.0 (Affymetrix). Images were scaled to an average hybridization intensity of 200. All expression values below 60 were set to 60.

(*Received in original form January 29, 2004; accepted in final form June 1, 2004*)

Supported by grants from EOS Biotechnology, DFG SFB 610 TPB1, SFB 598 TP A5, Deutsche Krebshilfe (70-2787-Bu 3), and BMBF (01-ZZ0109). The sponsors of the study had no role in data interpretation or writing of the report.

Correspondence and requests for reprints should be addressed to Hans-Stefan Hofmann, M.D., Department of Cardiothoracic Surgery, Martin Luther Universität Halle-Wittenberg, Ernst Grube Strasse 4, 06097 Halle, Germany. E-mail: stefan.hofmann@medizin.uni-halle.de

Am J Respir Crit Care Med Vol 170. pp 516–519, 2004
Originally Published in Press as DOI: 10.1164/rccm.200401-127OC on June 7, 2004
Internet address: www.atsjournals.org

To identify specific genes that were differentially expressed in tumors as compared with normal lung tissue we used a criterion that marks differential gene expression at an approximate significance level (determined by the Bonferroni method) of 8.0×10^{-7}, using a Student *t* test, and a fold change cutoff of 2.0 (for upregulated genes) or 0.5 (for downregulated genes). Calculation of fold change was performed by dividing the mean expression level of a gene in tumor samples by the mean expression level of the same gene in normal lung samples.

RT-PCR

In a reverse transcription reaction, cDNA was synthesized from 200 ng of total RNA with 100 units of SuperScript II reverse transcriptase (Invitrogen, Karlsruhe, Germany). A one-fifth volume of cDNA reaction was used for semiquantitative PCR of RAGE and cyclin-B2 per 18S rRNA amplification for intersample correction. Each PCR contained 1.5 mM $MgCl_2$, gene-specific primers (5 pmol each), dNTPs (10 μM each), and 1 unit of recombinant *Taq* DNA polymerase (Promega, Mannheim, Germany). PCR primers were 5′-TGA ACA CAG GCC GGA CAG AAG-3′ (sense) and 5′-CCC ATC CAA GTG CCA GCT AAG-3′ (antisense) for RAGE, 5′-AGC TGC TTC CTG CTT GTC TC-3′ (sense) and 5′-GCA CAA TGA AGC ACA CAT CC-3′ (antisense) for cyclin-B2, and 5′-GTT GGT GGA GCG ATT TGT CTG G-3′ (sense) and 5′-AGGGCAGGGACTTAATCAACGC-3′ (antisense) for 18S rRNA. PCR was performed after initial denaturation at 95°C for 2 minutes and 30 seconds at 95°C, 20 seconds at 58°C for primer annealing, and 40 seconds at 72°C, with 35 PCR cycles for RAGE, 37 cycles for cyclin-B2, and 12 cycles for 18S rRNA. After gel electrophoretic separation, the intensity of PCR products was densitometrically evaluated with AIDA 2.0 software (Raytest, Straubenhardt, Germany). In a first step the PCR analyses were optimized with a few samples to obtain the same expression ratio as when using the DNA chip assay. For internal signal correction between different PCR processes we always used loading samples. Subsequently, the ratio between RAGE and cyclin-B2 amplification was calculated.

Statistical Analysis

Results of gene expression are given as medians. Box-and-whisker plots were constructed for illustration: the boundary of the box closest to zero indicates the 25th percentile centered about the mean, the line within the box marks the 50th percentile (median), and the boundary of the box farthest from zero indicates the 75th percentile. Whiskers above and below the box represent the 5th and 95th percentiles, respectively, and circles represent the 1st and 99th percentiles. Statistical significance was determined on the basis of patient expression data, using a Student *t* test.

RESULTS

Comparison of expression profiles determined by microarray analysis (Data Set A) of normal lung and NSCLC samples revealed significant changes in gene expression in a total of 344 genes (0.6%). We detected 0.3% upregulated and 0.3% downregulated genes in tumor compared with normal lung tissue. Among these differentially expressed genes we searched for stable diagnostic targets.

The classic marker cytokeratin (CK-6A) is specific for squamous cell carcinomas, whereas thyroid transcription factor-1 (TTF-1) is specific for adenocarcinomas. These two markers can be used to differentiate squamous cell carcinoma (gene chip expression for CK-6A, 851.6 ± 383.8 versus adenocarcinoma, 117.0 ± 146.1) and adenocarcinoma (TTF-1, 379.0 ± 203.4 versus squamous cell carcinoma, 94.5 ± 66.7). However, neither marker can be used to entirely differentiate NSCLC from normal lung tissue. In contrast, we found that the inverse expression, between normal lung and NSCLC, of the genes encoding RAGE (receptor for advanced glycation end products) and cyclin-B2 was highly diagnostic for nearly all NSCLC specimens (Figure 1).

RAGE is involved in the process of arteriosclerosis, diabetes, aging, and Alzheimer's disease (11). Cyclin-B2 is a member of the cyclin family and plays an essential role in the cell cycle machinery governing the transition from G_2 to M phase (12). The expression of RAGE mRNA alone was, in terms of median values, fourfold downregulated in NSCLC, whereas cyclin-B2 mRNA was twofold upregulated (Figure 2). Using these two genes identified in data set A, we calculated the expression ratio per sample by dividing the expression value of RAGE by that of cyclin-B2 (Figure 3A). On the basis of the median mRNA ratio of RAGE to cyclin-B2, samples with ratio values greater than 2 were identified as normal lung and those with ratio values

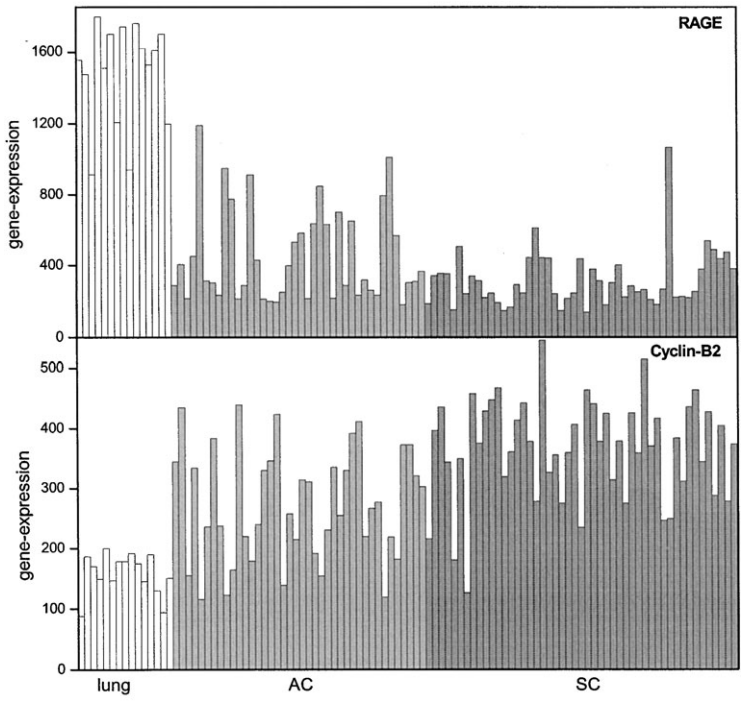

Figure 1. Gene expression for RAGE and cyclin-B2 in all normal lung, adenocarinoma (AC), and squamous cell carcinoma (SC) samples (Data Set A).

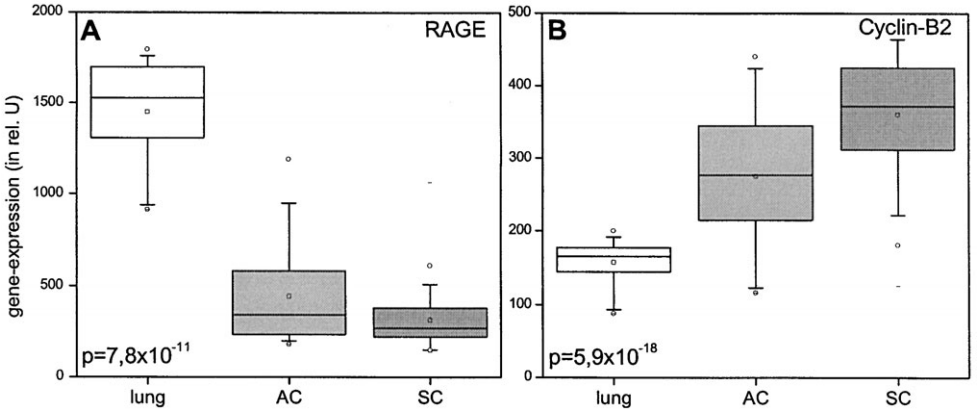

Figure 2. Box-and-whisker plots of gene expression of RAGE (*A*) and cyclin-B2 (*B*) for the samples of Data Set A. p Values of AC and SC versus normal lung group were determined by *t* test.

less than 2 were identified as lung tumor. The sensitivity for discrimination between NSCLC and normal lung tissue was determined as 87%, with a specificity of 100%.

Thereafter, we tested the accuracy of the RAGE:cyclin-B2 ratio in Data Set B. For this purpose we blinded all lung tumors, metastases, and corresponding normal lung samples (n = 84) and performed RT-PCR for expression analysis (Figure 4). We confirmed that the ratio of RAGE per cyclin-B2 correctly distinguishes between NSCLC and normal lung with 94.7% accuracy. All NSCLC tissues were classified correctly (sensitivity, 100%), and 4 of the 42 normal lung tissues were incorrectly defined (specificity, 90%).

The quotients of the overall RAGE and cyclin-B2 mRNA levels for lung cancer and normal lung tissue determined by RT-PCR (Data Set B) were comparable to the results of Data Set A from the gene chip analysis (Figure 3B). Moreover, the expression ratio of RAGE mRNA per cyclin-B2 mRNA also correlated positively in lung metastases with the ratios in adenocarcinomas and squamous cell carcinomas. Also, all blinded lung metastasis samples were correctly classified as tumor tissue.

DISCUSSION

Current gene expression profiling based on bioinformatics tools is highly accurate in the diagnosis and classification of cancer. The possibility of cancer classification based solely on gene expression monitoring by microarray analysis was first shown for human acute leukemias (6). The translation of gene expression data to potentially useful targets for molecular diagnosis depends largely on statistical analysis. Although some statistical tests may identify robust diagnostic target genes (13), the expression of these genes is often extremely variable from sample to sample. The expression ratio technique of two differentially expressed genes can reduce this interindividual variability. The physiologic function of these genes can be connected with tumorigenesis, but plays no role in tissue discrimination by the mathematical technique of expression ratios. We evaluated the ratio of RAGE per cyclin-B2 as a reliable marker to differentiate normal lung from NSCLC or lung metastases of different histologic origin.

Whereas the physiological function of cyclin-B2 seems to be clear as a regulator of the cell cycle (12), the functional role of RAGE in cancer needs to be evaluated. RAGE was previously detected in a differential display experiment to be downregulated in lung cancer in comparison with corresponding normal tissue (14). Data from the group of Huttunen showed a decreased invasive transendothelial migration *in vitro*, using human fibrosarcoma cells, and suppression of lung metastasis formation *in vivo* after treatment with the RAGE ligand amphoterin (15). Glioma cells overexpressing RAGE had increased tumor growth (16). These conflicting data may be explained by differences in the experimental design as well as by organ-specific differences. From our results it cannot be determined whether downregulation of RAGE is a consequence or cause of tumorigenesis. In the latter case, we would predict a higher rate of lung tumors in RAGE knockout mice, which has indeed been identified (A. Bierhaus, personal communication).

The expression ratio technique represents an effective method to translate the strengths of microarray expression profiling into

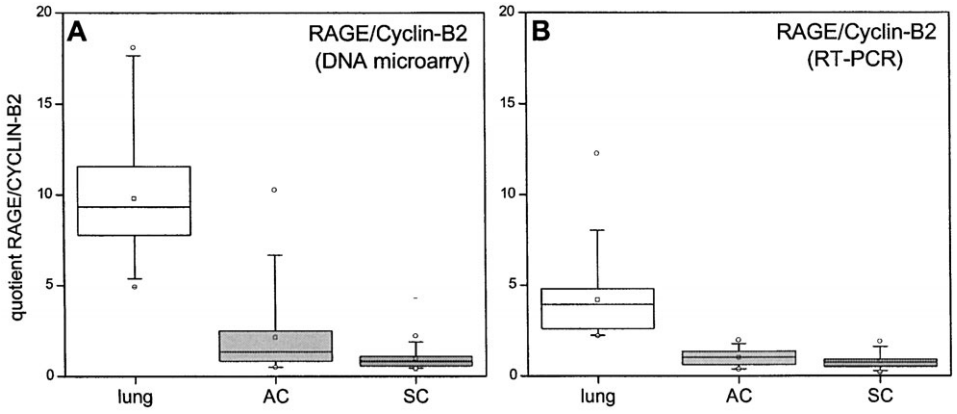

Figure 3. Box-and-whisker plots of the RAGE/cyclin-B2 expression quotient determined by DNA microarray analysis (*A*) for normal lung, adenocarcinomas (AC), and squamous cell carcinomas (SC)—Data Set A. (*B*) RAGE/cyclin-B2 quotient based on RT-PCR in adenocarcinomas (AC), squamous cell carcinoma (SC), and lung metastases compared with paired normal lung samples—Data Set B.

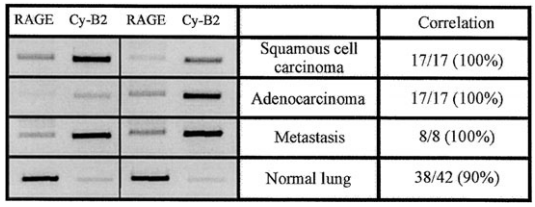

Figure 4. Correlation of the blinded differentiation between tumor tissue (adenocarcinomas, squamous cell carcinomas, and lung metastases) and paired normal lung samples and representative agarose gel electrophoresis after RT-PCR of amplified cDNA fragments for RAGE and cyclin-B2 for two different patient samples of each tissue type.

simple clinical tests. The technique is simple and effective with broad clinical use in diagnosis as well as prediction of prognosis in cancer. This ratio-based technique is independent from the expression measuring method, needs no gene expression reference (housekeeping gene as loading control), requires only small pieces of RNA, and does not require the coupling of transcription to translation for chosen genes (17). Gordon and coworkers were the first to test expression ratios in the discrimination of two different tissues (18). Using two or three expression ratios of two differentially expressed genes, they found that the differential diagnosis of mesothelioma and pulmonary adenocarcinoma was 95 and 99% accurate, respectively. With the geometric mean of the three most accurate individual ratios of four genes, Bueno and coworkers could distinguish, with 90% accuracy, normal prostate and prostate cancer samples obtained at surgery (19). This technique is not only highly precise in the discrimination of cancer tissues; it can be equally useful in additional clinical applications such as prediction of outcome. In patients with mesothelioma, treatment-related outcome was predicted by gene expression ratio-based analysis (17).

By using ratios of gene expression and rationally chosen thresholds we have demonstrated an alternative and simple approach to predict clinical parameters such as malignancy of tissue samples. This technique can be easily adapted and extended to routine clinical application without the need for additional sophisticated analyses.

Diagnoses provided by the expression ratio-based technique on the basis of limited amounts of tissue (obtained by fine needle aspiration, pleural effusion), without traditional histology results, have been shown to correlate with response to chemotherapy, and hence may be useful in predicting clinical course (20). Our results, which demonstrate not only a simple characterization of NSCLC, could also be a first step in an automated diagnostic process. Further investigations will test the usefulness of the RAGE/cyclin-B2 quotient for differentiation of lung neoplasms from normal lung, and of other tumor types from their paired normal tissue samples.

Conflict of Interest Statement: H.-S.H. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; G.H. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; S.B. states that the results of this study do not impact on the past grant as the present shares; B.B. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; R.-E.S. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript; A.S. does not have a financial relationship with a commercial entity that has an interest in the subject of this manuscript.

References

1. Parkin DM, Pisani P, Ferlay J. Estimates of the worldwide incidence of eighteen major cancers in 1985. *Int J Cancer* 1993;54:594–606.
2. Greenlee RT, Hill-Harmon MB, Murray T, Thun M. Cancer statistics, 2001. *CA Cancer J Clin* 2001;51:15–36.
3. Liotta L, Petricoin E. Molecular profiling of human cancer. *Nat Rev Genet* 2000;1:48–56.
4. Travis WD, Colby TV, Corrin B, Shimosato Y, Brambilla E, editors. Histological typing of lung and pleural tumours, 3rd ed. Berlin: Springer-Verlag; 1999. p. 31–47.
5. Wodicka L, Dong H, Mittmann M, Ho MH, Lockhart DJ. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat Biotechnol* 1997;15:1359–1367.
6. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, *et al.* Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 1999;286:531–537.
7. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. *Proc Natl Acad Sci USA* 1999;96:2907–2912.
8. Lipshutz RJ, Fodor SP, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet* 1999;21(1 Suppl):20–24.
9. Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Rapp BA, Wheeler DL. GenBank. *Nucleic Acids Res* 2000;28:15–18.
10. Salamov AA, Solovyev VV. *Ab initio* gene finding in *Drosophila* genomic DNA. *Genome Res* 2000;10:516–522.
11. Thornalley PJ. Cell activation by glycated proteins: AGE receptors, receptor recognition factors and functional classification of AGEs. *Cell Mol Biol (Noisy-le-grand)* 1998;44:1013–1023.
12. Roberts JM. Evolving ideas about cyclins. *Cell* 1999;98:129–132.
13. Boer JM, Huber WK, Sultmann H, Wilmer F, von Heydebreck A, Haas S, Korn B, Gunawan B, Vente A, Fuzesi L, *et al.* Identification and classification of differentially expressed genes in renal cell carcinoma by expression profiling on a global human 31,500-element cDNA array. *Genome Res* 2001;11:1861–1870.
14. Schraml P, Bendik I, Ludwig CU. Differential messenger RNA and protein expression of the receptor for advanced glycosylated end products in normal lung and non–small cell lung carcinoma. *Cancer Res* 1997;57:3669–3671.
15. Huttunen HJ, Fages C, Kuja-Panula J, Ridley AJ, Rauvala H. Receptor for advanced glycation end products-binding COOH-terminal motif of amphoterin inhibits invasive migration and metastasis. *Cancer Res* 2002;62:4805–4811.
16. Taguchi A, Blood DC, del Toro G, Canet A, Lee DC, Qu W, Tanji N, Lu Y, Lalla E, Fu C, *et al.* Blockade of RAGE–amphoterin signalling suppresses tumour growth and metastases. *Nature* 2000;405:354–360.
17. Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Richards WG, Jaklitsch MT, Sugarbaker DJ, Bueno R. Using gene expressio ratios to predict outcome among patients with mesothelioma. *J Natl Cancer Inst* 2003;95:598–605.
18. Gordon GJ, Jensen RV, Hsiao LL, Gullans SR, Blumenstock JE, Ramaswamy S, Richards WG, Sugarbaker DJ, Bueno R. Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Res* 2002; 62:4963–4967.
19. Bueno R, Loughlin KR, Powell MH, Gordon GJ. A diagnostic test for prostate cancer from gene expression profiling data. *J Urol* 2004;171: 903–906.
20. Sotiriou C, Powles TJ, Dowsett M, Jazaeri AA, Feldman AL, Assersohn L, Gadisetti C, Libutti SK, Liu ET. Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. *Breast Cancer Res* 2002;4:R3.