

Journal of Biology, Biological Systems and Bioinformatics

Volume 3, Issue 1, 2009

Application of the Gene-Ratio Method in the Diagnosis and Prognosis of a Variety of Human Diseases

Nathan Gaines, undergraduate student, Wheaton College, IL, nathandgaines@gmail.com

Christine O'Rourke, undergraduate student, Wheaton College, IL, christine.e.orourke@gmail.com

Pattle Pun, Professor of Biology, Wheaton College, IL, pattle.p.pun@wheaton.edu

Gavin Gordon, Co-Director, Thoracic Surgery Oncology Laboratory, Assistant Professor of Surgery, Harvard Medical School, ggordon@partners.org

Abstract:

A simple, highly accurate gene expression ratio-based methodology has been shown to be useful in the diagnosis and prognosis of cancer. This "Gene-Ratio Method" affords many advantages over similarly accurate mRNA-based assays that make it more suitable for clinical applications. It was determined that the gene ratio technique could be generally applied to other publicly available gene expression data representing a variety of normal and diseased human tissue types. This technique was able to distinguish in a statistically significant manner between normal and tumor tissues as well as among different tumor types and benign diseases using data from both cDNA and oligonucleotide microarray platforms.

Introduction:

The last decade has held the (relatively unfulfilled) promise for substantial improvements in existing cancer diagnosis and prognosis by analyzing global gene expression patterns. The most frequently used tools for these studies are microarrays (or similar devices) and sophisticated computer algorithms. Commercially available microarrays, small substrates to which a large number of genes or oligonucleotides (portions of genes) are bound as distinct spots, are capable of measuring the expression of as many as 40,000 genes and may be custom-made with relative ease. These devices are similar in principle to computer chips and quantify gene expression levels using selective hybridization of fluorescently dyed nucleic acid derivatives from a given tissue. Microarray-based technology has been used in proof-of-principle studies to classify human cancers and link gene expression patterns to patient clinical outcome (3, 4, 9, 11, 18, 23, 29, 38, 45, 46).

Interpretation of data generated from gene profiling with microarrays has so far required sophisticated computer algorithms and complex bioinformatics tools, leaving the practical use of expression profiling data beyond the scope of many scientists and clinicians (1, 24). To-date, no comprehensive method has been proposed to translate the results of tumor profiling to the analysis of a single tissue at a time. The crux of the problem is that current techniques of analysis allow classification of tumors only in groups and with inadequate reference to previous experiments. Also, the nature of bioinformatics tools (i.e. software) currently being used to analyze microarray data necessitates the use of a cohort of samples for adequate training, rendering the interpretation of data from any additional single sample impossible if its expression data was not acquired using the same platform.

Thus, despite approximately 1,500 peer-reviewed published scientific studies that use gene signatures in the diagnosis and prognosis of diseases¹, there are a relatively small number of candidate predictive gene panels that have been validated in independent cohorts. Among these studies only two are in clinical use, both for predicting recurrence in patients with resected breast cancer (40, 44, 45). In addition, a number of test-specific technical barriers serve to impede the translation of such prospective clinical tests from discovery to clinical implementation (40). These include between-platform variability, complex bioinformatics-based algorithms that are not easily reproducible, lack of sufficient frozen specimens of a quality adequate for test validation, and the necessity of lengthy patient follow-up for cancer prognosis validation, among others.

Recent publications describe a novel method that allows the translation of complex gene expression profiling into relatively simple clinical tests that overcome many of the limitations preventing the routine use of mRNA-based assays in the clinical management of cancer patients (5, 10, 12-17). This algorithm (i.e., the "Gene-Ratio" Method) identifies genes that are differentially expressed between two clinically distinct conditions and calculates the ratio of gene expression levels for a pair of two genes that can predict the condition alone or in combination with additional

¹PubMed searches conducted on July 23, 2009 using key word phrases "diagnostic gene signature" and "prognostic gene signature". (<http://www.pubmed.gov>)

gene ratios. In contrast to traditional approaches, both genes in such a gene pair ratio are informative, and by its use of a unitless ratio, this technique has the major advantages of platform-independence and facile applicability to individual specimens. In this report, we tested the hypothesis that this classification technique could be extended to multiple human diseases (primarily cancers) using previously published and freely available microarray data associated with both oligonucleotide and cDNA platforms. These data represent independent studies performed by a wide variety of investigators in multiple tissue types and using multiple platforms.

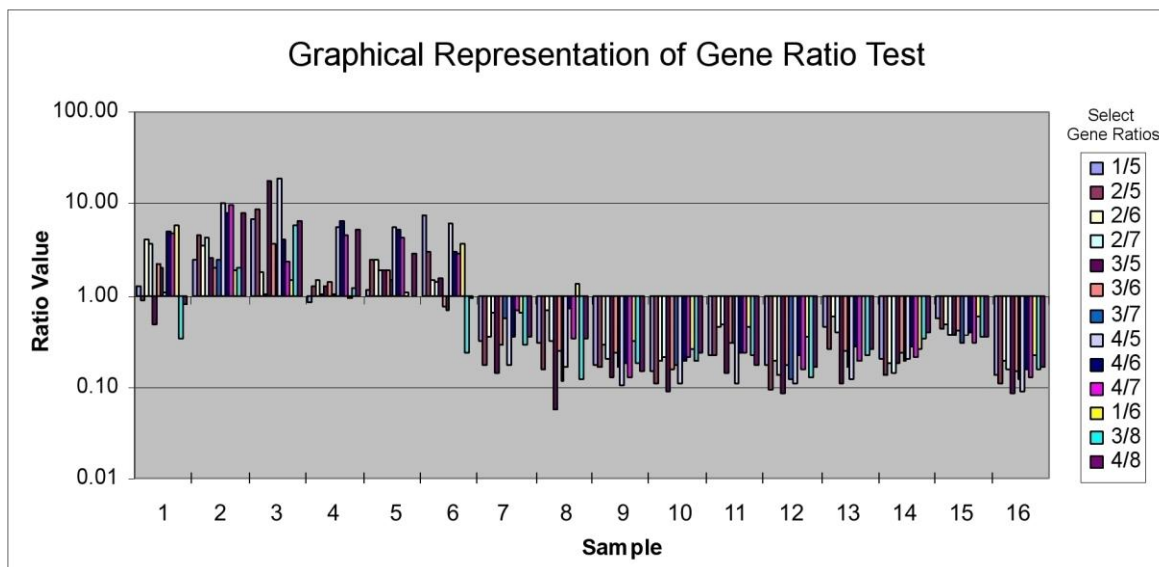
Materials and Methods:

Microarray Data. Previously published microarray data (1, 2, 4, 6-8, 11, 19-22, 24-27, 29, 30, 32-39, 41-43, 45, 47, 48) were downloaded from the following repositories: the Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>), the European Bioinformatics Institute (ArrayExpress, <http://www.ebi.ac.uk/microarray-as/ae/>), the Stanford Microarray Database (SMD, <http://genome-www5.stanford.edu/>), The Broad Institute (<http://www.broad.mit.edu/>), and the Oncomine Cancer Microarray Database (<http://www.oncomine.org/>). Each microarray data set contained two tissue types that were differentiated by some clinically relevant diagnostic or prognostic classification. See Table 1 for details. Each tissue type was arbitrarily divided into two equally sized sample sets: a “training set” for development of gene ratio tests and a “test set” for evaluation of the optimal gene ratio test (“Trial 1”). In a related analysis, the roles of the training and test sets were reversed to gauge the robustness of the gene ratio technique (“Trial 2”).

Selection of Predictor Genes. The selection of predictor genes (i.e. genes that are differentially expressed in the tissue types being analyzed) was performed, in effect, as previously described using standard statistical techniques (5, 12, 13-15). Essentially, we used a two-sided Student’s (parametric) *t*-test on the training set for pair-wise comparisons of average gene expression levels of all genes represented on the microarray to identify those genes differentially expressed in a statistically significant manner between the two tissue types. Cut-off *P* values were optimized on a case-by-case basis to minimize the false-discovery rate (<10%) and take into account multiple hypothesis testing. To further reduce the risk of false positives and increase the signal-to-noise metric of predictor genes, we chose for additional consideration those genes with a 2-fold difference in average expression levels and an average gene expression level >X in at least one of the two subsets where “X” represented an empirically determined average expression level that resulted in a small number of genes with an average expression level highly above the background noise on the chip. This number was optimized in each case by incremental increases starting from the target intensity of the chip (for Affymetrix oligonucleotide arrays) or median expression level (for cDNA arrays) until no fewer than 10 candidate genes resulted.

Generation and Evaluation of Gene Ratio Tests. The four most statistically significant known genes (i.e., no ESTs) overexpressed in both tissue types (i.e., 8 genes total) from each individual training set were used to determine whether expression ratios could accurately classify the samples used to train the model. We calculated a total of 16 possible expression ratios per sample by dividing the expression value of each of the 4 genes expressed at relatively higher levels in tissue type “X” by the expression value of each of the 4 genes expressed at relatively higher levels in tissue type “Y”. Samples with ratio values >1 were predicted to be “X” and those with ratio values <1 were predicted to be “Y”. The most accurate individual gene pair ratios were chosen for further study. To incorporate the predictive accuracy of multiple ratios, we calculated the geometric mean, $(R_1R_2R_3)^{1/3}$ where R_i represents a single ratio value, of multiple ratios similar to previous studies (12, 13, 15, 16). This is the mathematical equivalent to the average of $[\log_2(R_1), \log_2(R_2), \log_2(R_3)]$, thereby giving equal weight to ratio fold-changes of identical magnitude but opposite direction. (However, unlike these cited studies, we only used the geometric mean of multiple ratios when two or more ratios shared the highest classification accuracy. Accordingly, we expect that our classification accuracies would improve in some cases if ratios were combined in a case specific manner as they were in the cited studies.) This ratio test was then used to classify the test set samples using the same technique (Figure 1).

Figure 1. Graphical representation of select gene expression ratios with the highest classification accuracies from Analysis 6, Trial 1. Samples 1 through 6 were clinically diagnosed as lung adenocarcinoma, samples 7 through 16 were normal lung tissues. In this case, ratio values >1 predicted a diagnosis of adenocarcinoma, while ratio values <1 predicted normal lung tissue. The select gene ratios are thirteen of the sixteen possible gene ratios generated through combination of the eight predictor genes (see Materials and Methods sections Selection of Predictor Genes and Generation and Evaluation of Gene Ratio Tests). (The classification accuracy found in Table 1 for Analysis 6 Trial 1 represents only those gene ratios with classification accuracies of 100%.)



Statistical Analysis. The classification accuracy of the gene ratio tests in the test sets were independently assessed using two methods. In one case, we used an exact one-sample binomial test. The P -values are reported under the null hypothesis of classification randomly assigned with equal probability of 0.5 based on one-sided tests in order to reject lower levels of accuracy. These tests were performed online at Simple Interactive Statistical Analysis (SISA, <http://home.clara.net/sisa/binomial.htm>). Second, a more conservative Goodman-Kruskal tau statistic and Chi-square P -value approximation to rule out chance events was calculated using SPSS 13.0 (©1989-LEAD Technologies, Inc.). Since our method is not case specific for a given human disease, adjustment of the P -values using a Bonferroni procedure or any other is not necessary.

Results

We performed a total of 36 independent analyses with 36 different tissue types using microarrays containing between ~1,500 and ~40,000 genes (Table 1). Two different microarray platforms were examined; approximately two-thirds of the microarrays were cDNA arrays with the remainder consisting of high-density oligonucleotide arrays. The majority (85%) of analyses involved cancer and were associated either with diagnosis or prognosis.

In total, 81% (29/36) of the analyses resulted in the identification of a gene ratio test associated with statistically significant ($P < 0.05$, binomial test) predictions in Trial 1 and Trial 2 thus attesting to the general robustness of the gene ratio technique. Only a single analysis (Analysis 7) did not identify a statistically significant gene ratio test in either Trial 1 or Trial 2. Similar results were obtained using the more conservative Chi-Square analysis. No obvious trends in classification accuracy were apparent between analyses conducted using oligonucleotide or cDNA microarrays.

Table 1. All analyses with tissue types, sample sizes, total number of genes on microarray, classification accuracy, and statistical results. Both cDNA and oligonucleotide microarrays were used; cDNA expression values were converted to absolute values to mimic oligonucleotide array. Sample Size shows the total number of samples for the analysis with the sample sizes of the individual tissue types in parentheses. Classification Accuracy shows the percentage of samples in each trial for which the gene ratio test correctly predicted the clinical diagnosis. *P*-values were calculated using both an exact one-sample binomial test and a Chi-Square *P*-value approximation. Significant *P*-values (<0.05) calculated from the binomial test are in bold font.

Analysis #	Tissue Types Compared	Reference	Sample Size	Genes on Microarray	Classification Accuracy		P-Value (Binomial Test)		P-Value (Chi-Square)	
					Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
1	Enrometrioid vs Mucinous Ovarian Adenocarcinoma	43	21(11:10)	10,239	80%	91%	0.055	5.86E-03	0.720	0.008
2	Colorectal vs Gastric Adenocarcinoma	43	33(20:13)	10,239	88%	94%	2.09E-03	1.37E-04	0.003	<0.001
3	Crohn's Disease vs Ulcerative Colitis	7	20(10:10)	33,792	90%	80%	0.011	0.172	0.014	0.134
4	Normal Tissue vs Crohn's Disease	7	21(11:10)	33,792	90%	73%	0.011	5.90E-03	0.014	0.009
5	Normal Tissue vs Ulcerative Colitis	7	21(11:10)	33,792	90%	100%	0.011	4.88E-04	0.014	0.002
6	Normal Lung vs Lung Adenocarcinoma	22	31(19:12)	40,368	100%	100%	3.00E-05	1.50E-05	<0.001	<0.001
7	Neuroendocrine vs Small Cell Lung Carcinoma	22	23(15:8)	40,368	55%	75%	0.500	0.073	0.572	0.176
8	B-Cell Type vs T-Cell Type Acute Lymphoblastic Leukemia	8	30(20:10)	9,984	100%	100%	6.10E-05	3.00E-05	<0.001	<0.001
9	Head and Neck Squamous Cell Carcinoma; Lymph Node Metastasis vs No Metastasis	6	47(31:16)	15,389	61%	79%	0.047	3.31E-03	0.101	0.009
10	Normal Breast vs Luminal like ER+ Breast Cancer	29	43(30:13)	9,216	100%	95%	4.70E-07	5.00E-06	<0.001	<0.001
11	Normal Breast vs Basal-like Breast Cancer	29	23(13:10)	9,216	91%	92%	5.90E-03	3.20E-03	0.009	0.005
12	Luminal like ER+ vs Basal-like Breast Cancer	29	40(30:10)	9,216	100%	100%	9.50E-07	9.50E-07	<0.001	<0.001
13	Breast Cancer Response to Tamoxifen; Disease Free vs Cancer Recurrence	26	60(32:28)	22,575	53%	73%	0.428	8.06E-03	0.551	0.012
14	Leukemia Response to Chemotherapy; Good Response v. Poor Response	19	33(22:11)	41,665	75%	71%	0.038	0.072	0.155	0.224
15	Normal Breast Tissues; Luminal vs Myoepithelial Types	21	52(26:26)	9,984	92%	100%	5.00E-06	1.00E-08	<0.001	<0.001
16	Benign Gastric Mucosa vs Primary Gastric Cancer	25	112(90:22)	6,688	77%	73%	1.00E-05	3.42E-04	<0.001	0.001
17	Breast Carcinoma; Solitary Fibrous Tumour vs Desmoid-type Fibromatosis	47	23(13:10)	45,207	100%	100%	4.88E-04	2.44E-04	0.002	0.001
18	Serous Ovarian Cancer; Primary vs Metastasis	35	36(28:8)	39,726	100%	100%	3.00E-06	3.00E-06	<0.001	<0.001
19	Prostate Cancer; Normal vs Tumor	24	103(61:42)	5,153	96%	94%	1.00E-08	1.00E-08	<0.001	<0.001
20	Lymphoma; Diffuse Large B Cell vs Activated/Germinal Center B Cell	33	64(38:26)	12,598	90%	90%	7.00E-06	8.00E-07	<0.001	<0.001
21	Breast Cancer Overall Survival; <5 years vs >5 years	45	78(44:34)	24,481	72%	79%	1.69E-03	1.47E-04	0.003	<0.001
22	Large Cell vs Small Cell Lung Cancer	22	27(15:12)	40,368	86%	92%	6.47E-03	1.71E-03	0.011	0.003
23	Acute Lymphoblastic Leukemia vs Lymphoblastic Leukemias with MLL translocations	1	44(24:20)	1,427	100%	100%	2.30E-07	2.30E-07	<0.001	<0.001
24	Acute Myeloid Leukemia vs Acute Lymphoblastic Leukemia	11	72(47:25)	6,787	100%	95%	1.00E-08	1.00E-08	<0.001	<0.001
25	Papillary Thyroid Carcinoma vs Normal Thyroid	20	16(8:8)	9,658	88%	88%	0.035	0.035	0.040	0.040
26	Large B-cell Lymphoma; Mediastinal vs Diffuse	34	90(56:34)	44,928	86%	68%	1.00E-08	0.061	<0.001	0.091
27	Colon Cancer v. Normal Colon	27	36(18:18)	7,464	100%	100%	3.81E-06	3.81E-06	<0.001	<0.001
28	Medulloblastomas vs Other Brain Tumours	30	34(14:10)	7,128	71%	72%	0.072	0.048	0.160	0.470
29	Renal Cancer vs Other Uteral Tumours	32	21(11:10)	10,406	100%	100%	0.031	9.76E-04	0.046	0.003
30	Soft Tissue Sarcoma vs Clear Cell Sarcoma	37	52(47:5)	12,559	96%	93%	1.00E-06	2.40E-05	0.001	1
31	Diffuse Large B-cell Lymphoma vs Follicular Lymphoma	38	77(58:19)	5,669	79%	89%	1.47E-04	2.60E-07	0.056	<0.001
32	Normal Lung vs Pulmonary Carcinoid	4	37(20:17)	12,600	100%	100%	3.81E-06	1.90E-06	<0.001	<0.001
33	Ewing's Sarcoma vs Pediatric Rhabdomyosarcoma	2	20(11:9)	12,559	89%	100%	0.020	4.88E-04	0.025	0.002
34	Oxidatively Elicited Macular Degeneration Patients vs Healthy Controls	42	36(18:18)	12,625	100%	94%	3.81E-06	7.20E-05	<0.001	<0.001
35	No or Mild Emphysema vs Severe Emphysema	41	30(18:12)	22,283	73%	80%	3.69E-03	0.018	0.072	0.022
36	Prostate Cancer vs Normal Prostate	39	100(50:50)	12,600	84%	82%	4.70E-07	2.00E-06	<0.001	<0.001

Discussion:

Our results suggest that the gene ratio methodology is useful for discriminating among a wide variety of normal and diseased human tissues with gene expression data acquired on multiple types of microarray platforms from multiple manufacturers. The classification accuracies (and associated *P*-values) were encouraging within this limited proof-of-principle study since no additional attempt was made to further optimize the selection of predictor genes or gene ratio tests for any specific comparison. It is likely that this additional refinement would result in even greater statistically significant classification accuracy.

Analysis 30 demonstrated the importance of using a relatively high number of samples for both tissue types in an analysis or having roughly equal sample sizes for both tissue types. While Trial 2 in Analysis 30 had a classification accuracy of 93% and a *P*-value of 2.4×10^{-5} calculated using the binomial test, the *P*-value from the Chi-Square approximation was 1. Because there were only five clear cell sarcoma samples compared to forty-seven soft tissue sarcoma samples, the ratio test produced the deceptively high classification accuracy by selecting for gene ratios that were not discriminatory but rather classified all samples as soft tissue sarcoma. This case also demonstrated the utility of the *P*-value from the Chi-Square approximation which, unlike the binomial test, takes into account the size of each sample subset.

Sub-optimal classification accuracy for Analysis 7 may also be due to relatively small sample sizes. Interestingly, the original analysis of these data (22) reaches the same conclusion; that these two subclasses of tumors should not be classified as distinct entities but instead as a single class (high grade neuroendocrine carcinoma).

In summary, we have shown that the gene ratio technique is likely to be widely applicable to a variety of human diseases and as such may facilitate increased usage of mRNA-based clinical assays for the identification of tissue type, most notably in the context of cancer.

Acknowledgment:

We thank Ryan M. Wilke and the bioinformatics class at Wheaton College (2006) for gathering preliminary data for this project and the Wheaton College Student Summer Researcher in Residence program for supporting Nathan Gaines and Christine O'Rourke. Preliminary data of this project were presented at Experimental Biology 2007.

References:

1. Armstrong, S. A., Staunton, J. E., Silverman, L. B., Pieters, R., den Boer, M. L., Minden, M. D., Sallan, S. E., Lander, E. S., Golub, T. R., and Korsmeyer, S. J. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. *Nature Genetics*, 30: 41-47, 2002.
2. Baer, C., Nees, M., Breit, S., Selle, B., Kulozik, A. E., Schaefer, K. L., Braun, Y., Wai, D., and Poremba, C. Profiling and functional annotation of mRNA gene expression in pediatric rhabdomyosarcoma and Ewing's sarcoma. *International Journal of Cancer*, 110: 687-694, 2004.
3. Beer, D. G., Kardia, S. L. R., Huang, C.-C., Giordana, T. J., Levin, A. M., Misek, D. E., Lin, L., Chen, G., Gharib, T. G., Thomas, D. G., Lizyness, M. L., Kuick, R., Hayasaka, S., Taylor, J. M. G., Iannettoni, M. D., Orringer, M. B., and Hanash, S. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Medicine*, 8: 816-824, 2002.
4. Bhattacharjee, A., Richards, W. G., Staunton, J., Li, C., Monti, S., Vasa, P., Ladd, C., Beheshti, J., Bueno, R., Gillette, M., Loda, M., Weber, G., Mark, E. J., Lander, E. S., Wong, W., Johnson, B. E., Golub, T. R., Sugarbaker, D. J., and Meyerson, M. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of the National Academy of Sciences of the United States of America*, 98: 13790-13795, 2001.
5. Bueno, R., Loughlin, K. R., Powell, M. H., and Gordon, G. J. A diagnostic test for prostate cancer from gene expression profiling data. *Journal of Urology*, 171: 903-906, 2004.
6. Chung, C. H., Parker, J. S., Karaca, G., Wu, J., Funkhouser, W. K., Moorem, D., Butterfoss, D., Xiang, D., Zanation, A., Yin, X., Shockley, W. W., Weissler, M. C., Dressler, L. G., Shores, C. G., Yarbrough, W. G., and Perou, C. M. Molecular classification of head and neck squamous cell carcinomas using patterns of gene expression. *Cancer Cell*, 5: 489-500, 2004.

7. Costello, C. M., Mah, N., Häsler, R., Rosenstiel, P., Waetzig, G. H., Hahn, A., Lu, T., Gurbuz, Y., Nikolaus, S., Albrecht, M., Hampe, J., Lucius, R., Klöppel, G., Eickhoff, H., Lehrach, H., Lengauer, T., and Schreiber, S. Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays. *PLoS Medicine*, 2: e199, 2005.
8. De Pittà, C., Tombolan, L., Campo Dell'Orto, M., Accordi, B., te Kronnie, G., Romualdi, C., Vitulo, N., Basso, G., and G., L. A leukemia-enriched cDNA microarray platform identifies new transcripts with relevance to the biology of pediatric acute lymphoblastic leukemia. *Haematologica*, 90: 890-898, 2005.
9. Dhanasekaran, S. M., Barrette, T. R., Ghosh, D., Shah, R., Varambally, S., Kurachi, K., Pienta, K. J., Rubin, M. A., and Chinnaiyan, A. M. Delineation of prognostic biomarkers in prostate cancer. *Nature*, 412: 822-826, 2001.
10. Dong, L., Bard, A. J., Richards, W. G., Nitz, M. D., Theodorescu, D., Bueno, R., and Gordon, G. J. A gene expression ratio-based diagnostic test for bladder cancer. *Computational Biology and Chemistry*, In Press, 2008.
11. Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H., Loh, M. L., Downing, J. R., Caligiuri, M. A., Bloomfield, C. D., and Lander, E. S. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*, 286: 531-537, 1999.
12. Gordon, G. J., Deters, L. A., Nitz, M. D., Lieberman, B. C., Yeap, B. Y., and Bueno, R. Differential diagnosis of solitary lung nodules using gene expression ratios. *Journal of Thoracic and Cardiovascular Surgery*, 132: 621-627, 2006.
13. Gordon, G. J., Jensen, R. V., Hsiao, L. L., Gullans, S. R., Blumenstock, J. E., Richards, W. G., Jaklitsch, M. T., Sugarbaker, D. J., and Bueno, R. Using gene expression ratios to predict outcome among patients with mesothelioma. *Journal of the National Cancer Institute*, 95: 598-605, 2003.
14. Gordon, G. J., Jensen, R. V., Hsiao, L.-L., Gullans, S. R., Blumenstock, J. E., Ramaswami, S., Richards, W. G., Sugarbaker, D. J., and Bueno, R. Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. *Cancer Research*, 62: 4963-4967, 2002.
15. Gordon, G. J., Richards, W. G., Sugarbaker, D. J., Jaklitsch, M. T., and Bueno, R. A prognostic test for adenocarcinoma of the lung from gene expression profiling data. *Cancer Epidemiol Biomarkers Prev*, 12: 905-910, 2003.
16. Gordon, G. J., Rockwell, G. N., Godfrey, P. A., Jensen, R. V., Glickman, J. N., Yeap, B. W., Richards, W. G., Sugarbaker, D. J., and Bueno, R. Validation of genomics-based prognostic tests in malignant pleural mesothelioma. *Clinical Cancer Research*, 11: 4406-4414, 2005.
17. Gordon, G. J., Rockwell, G. N., Jensen, R. V., Rheinwald, J. G., Glickman, J. N., Aronson, J. P., Pottorf, B. J., Nitz, M. D., Richards, W. G., Sugarbaker, D. J., and Bueno, R. Identification of novel candidate oncogenes and tumor suppressors in malignant pleural mesothelioma using large-scale transcriptional profiling. *American Journal of Pathology*, 166: 1827-1840, 2005.
18. Hedenfalk, I., Duggan, D., Chen, Y., Radmacher, M., Bittner, M., Simon, R., Meltzer, P., Gusterson, B., Esteller, M., Kallioniemi, O.-P., Wilfond, B., Borg, A., and Trent, J. Gene expression profiles in hereditary breast cancer. *New England Journal of Medicine*, 344: 539-548, 2001.
19. Heuser, M., Wingen, L. U., Steinemann, D., Cario, G., von Neuhoff, N., Tauscher, M., Bullinger, L., Krauter, J., Heil, G., Döhner, H., Schlegelberger, B., and Ganser, A. Gene-expression profiles and their association with drug resistance in adult acute myeloid leukemia. *Haematologica*, 90: 1484-1492, 2005.
20. Huang, Y., Prasad, M., Lemon, W. J., Hampel, H., Wright, F. A., Kornacker, K., LiVolsi, V., Frankel, W., Kloos, R. T., Eng, C., Pellegata, N. S., and de la Chapelle, A. Gene expression in papillary thyroid carcinoma reveals highly consistent profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 98: 15044-15049, 2001.
21. Jones, C., Mackay, A., Grigoriadis, A., Cossu, A., Reis-Filho, J. S., Fulford, L., Dexter, T., Davies, S., Bulmer, K., Ford, E., Parry, S., Budroni, M., Palmieri, G., Neville, A. M., O'Hare, M. J., and Lakhani, S. R. Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer. *Cancer Research*, 64: 3037-3045, 2004.

22. Jones, M. H., Virtanen, C., Honjoh, D., Miyoshi, T., Satoh, Y., Okumura, S., Nakagawa, K., Nomura, H., and Ishikawa, Y. Two prognostically significant subtypes of high-grade lung neuroendocrine tumours independent of small-cell and large-cell neuroendocrine carcinomas identified by gene expression profiles. *Lancet*, 363: 775-781, 2004.
23. Khan, J., Wei, J. S., Ringner, M., Saal, L. H., Ladanyi, M., Westermann, F., Berthold, F., Schwab, M., Antonescu, C. R., Peterson, C., and Meltzer, P. S. Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nature Medicine*, 7: 673-679, 2001.
24. Lapointe, J., Li, C., Higgins, J. P., van de Rijn, M., Bair, E., Montgomery, K., Ferrari, M., Egevad, L., Rayford, W., Bergerheim, U., Ekman, P., DeMarzo, A. M., Tibshirani, R., Botstein, D., Brown, P. O., Brooks, J. D., and Pollack, J. R. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proceedings of the National Academy of Sciences of the United States of America*, 101: 811-816, 2004.
25. Leung, S. Y., Chen, X., Chu, K. M., Yuen, S. T., Mathy, J., Ji, J., Chan, A. S., Li, R., Law, S., Troyanskaya, O. G., Tu, I. P., Wong, J., So, S., Botstein, D., and Brown, P. O. Phospholipase A2 group IIA expression in gastric adenocarcinoma is associated with prolonged survival and less frequent metastasis. *Proceedings of the National Academy of Sciences of the United States of America*, 99: 16203-16208, 2002.
26. Ma, X. J., Wang, Z., Ryan, P. D., Isakoff, S. J., Barmettler, A., Fuller, A., Muir, B., Mohapatra, G., Salunga, R., Tuggle, J. T., Tran, Y., Tran, D., Tassin, A., Amon, P., Wang, W., Wang, W., Enright, E., Stecker, K., Estepa-Sabal, E., Smith, B., Younger, J., Balis, U., Michaelson, J., Bhan, A., Habin, K., Baer, T. M., Brugge, J., Haber, D. A., Erlander, M. G., and D.C., S. A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell*, 5: 607-616, 2004.
27. Notterman, D. A., Alon, U., Sierk, A. J., and Levine, A. J. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Research*, 61: 3124-3130, 2001.
28. Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F. L., Walker, M. G., Watson, D., Park, T., Hiller, W., Fisher, E. R., Wickerham, D. L., Bryant, J., and Wolmark, N. A Multigene Assay to Predict Recurrence of Tamoxifen-Treated, Node-Negative Breast Cancer. *N Engl J Med*, 351: 2817-2826, 2004.
29. Perou, C. M., Sørlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R., Ross, D. T., Johnsen, H., Akslen, L. A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S. X., Lønning, P. E., Børresen-Dale, A. L., Brown, P. O., and Botstein, D. Molecular portraits of human breast tumours. *Nature*, 406: 747-752, 2000.
30. Pomeroy, S. L., Tamayo, P., Gaasenbeek, M., Sturla, L. M., Angelo, M., McLaughlin, M. E., Kim, J. Y., Goumnerova, L. C., Black, P. M., Lau, C., Allen, J. C., Zagzag, D., Olson, J. M., Curran, T., Wetmore, C., Biegel, J. A., Poggio, T., Mukherjee, S., Rifkin, R., Califano, A., Stolovitzky, G., Louis, D. N., Mesirov, J. P., Lander, E. S., and Golub, T. R. Prediction of central nervous system embryonal tumour outcome based on gene expression. *Nature*, 415: 436-442, 2002.
31. Quackenbush, J. Computational analysis of microarray data. *Nature Reviews Genetics*, 2: 418-427, 2001.
32. Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C. H., Angelo, M., Ladd, C., Reich, M., Latulippe, E., Mesirov, J. P., Poggio, T., Gerald, W., Loda, M., Lander, E. S., and Golub, T. R. Multiclass cancer diagnosis using tumor gene expression signatures. *Proceedings of the National Academy of Sciences of the United States of America*, 98: 15149-15154, 2001.
33. Rosenwald, A., Wright, G., Leroy, K., Yu, X., Gaulard, P., Gascoyne, R. D., Chan, W. C., Zhao, T., Haioun, C., Greiner, T. C., Weisenburger, D. D., Lynch, J. C., Vose, J., Armitage, J. O., Smeland, E. B., Kvaloy, S., Holte, H., Delabie, J., Campo, E., Montserrat, E., Lopez-Guillermo, A., Ott, G., Muller-Hermelink, H. K., Connors, J. M., Braziel, R., Grogan, T. M., Fisher, R. I., Miller, T. P., LeBlanc, M., Chiorazzi, M., Zhao, H., Yang, L., Powell, J., Wilson, W. H., Jaffe, E. S., Simon, R., Klausner, R. D., and Staudt, L. M. Molecular diagnosis of primary mediastinal B cell lymphoma identifies a clinically favorable subgroup of diffuse large B cell lymphoma related to Hodgkin lymphoma. *The Journal of Experimental Medicine*, 198: 851-862, 2003.
34. Savage, K. J., Monti, S., Kutok, J. L., Cattoretti, G., Neuberger, D., De Leval, L., Kurtin, P., Dal Cin, P., Ladd, C., Feuerhake, F., Aguiar, R. C., Li, S., Salles, G., Berger, F., Jing, W., Pinkus, G. S., Habermann, T., Dalla-Favera, R., Harris, N. L., Aster, J. C., Golub, T. R., and Shipp, M. A. The molecular signature of mediastinal large B-cell

lymphoma differs from that of other diffuse large B-cell lymphomas and shares features with classical Hodgkin lymphoma. *Blood*, 102: 3871-3879, 2003.

35. Schaner, M. E., Davidson, B., Skrede, M., Reich, R., Flørenes, V. A., Risberg, B., Berner, A., Goldberg, I., Givant-Horwitz, V., Tropè, C. G., Kristensen, G. B., Nesland, J. M., and Børresen-Dale, A. L. Variation in gene expression patterns in effusions and primary tumors from serous ovarian cancer patients. *Molecular Cancer*, 4: 26, 2005.
36. Schwartz, D. R., Kardia, S. L., Shedden, K. A., Kuick, R., Michailidis, G., Taylor, J. M., Misek, D. E., Wu, R., Zhai, Y., Darrah, D. M., Reed, H., Ellenson, L. H., Giordano, T. J., Fearon, E. R., Hanash, S. M., and Cho, K. R. Gene expression in ovarian cancer reflects both morphology and biological behavior, distinguishing clear cell from other poor-prognosis ovarian carcinomas. *Cancer Research*, 62: 4722-4729, 2002.
37. Segal, N. H., Pavlidis, P., Noble, W. S., Antonescu, C. R., Viale, A., Wesley, U. V., Busam, K., Gallardo, H., DeSantis, D., Brennan, M. F., Cordon-Cardo, C., Wolchok, J. D., and Houghton, A. N. Classification of clear-cell sarcoma as a subtype of melanoma by genomic profiling. *Journal of Clinical Oncology*, 21: 1775-1781, 2003.
38. Shipp, M. A., Ross, K. A., Tamayo, P., Weng, A. P., Kutok, J. L., Aguiar, R. C. T., Gaasenbeek, M., Angelo, M., Reich, M., Pinkus, G. S., Ray, T. S., Koval, M. A., Last, K. M., Norton, A., Lister, T. A., Mesirov, J., Neuberg, D. S., Lander, E. S., Aster, J. C., and Golub, T. R. Diffuse large B-cell lymphoma outcome prediction by gene expression profiling and supervised machine learning. *Nature Medicine*, 8: 68-74, 2002.
39. Singh, D., Febbo, P. G., Ross, K., Jackson, D. G., Manola, J., Ladd, C., Tamayo, P., Renshaw, A. A., D'Amico, A. V., Richie, J. P., Lander, E. S., Loda, M., Kantoff, P. W., Golub, T. R., and Sellers, W. R. Gene expression correlates of clinical prostate cancer behavior. *Cancer Cell*, 1: 203-209, 2002.
40. Sotiriou, C. and Piccart, M. J. Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? *Nat Rev Cancer*, 7: 545-553, 2007.
41. Spira, A., Beane, J., Pinto-Plata, V., Kadar, A., Liu, G., Shah, V., Celli, B., and Brody, J. S. Gene expression profiling of human lung tissue from smokers with severe emphysema. *American Journal of Respiratory Cell and Molecular Biology*, 31: 601-610, 2004.
42. Strunnikova, N., Hilmer, S., Flippin, J., Robinson, M., Hoffman, E., and Csaky, K. G. Differences in gene expression profiles in dermal fibroblasts from control and patients with age-related macular degeneration elicited by oxidative injury. *Free Radical Biology & Medicine*, 39: 781-796, 2005.
43. Tothill, R. W., Kowalczyk, A., Rischin, D., Bousioutas, A., Haviv, I., van Laar, R. K., Waring, P. M., Zalberg, J., Ward, R., Biankin, A. V., Sutherland, R. L., Henshall, S. M., Fong, K., Pollack, J. R., Bowtell, D. D., and Holloway, A. J. An expression-based site of origin diagnostic method designed for clinical application to cancer of unknown origin. *Cancer Research*, 65: 4031-4040, 2005.
44. van de Vijver, M. J., He, Y. D., van't Veer, L. J., Dai, H., Hart, A. A., Voskuil, D. W., Schreiber, G. J., Peterse, J. L., Roberts, C., Marton, M. J., Parrish, M., Atsma, D., Witteveen, A., Glas, A., Delahaye, L., van der Velde, T., Bartelink, H., Rodenhuis, S., Rutgers, E. T., Friend, S. H., and Bernards, R. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med*, 347: 1999-2009, 2002.
45. van 't Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*, 415: 530-536, 2002.
46. Welsh, J. B., Sapinoso, L. M., Su, A. I., Kern, S. G., Wang-Rodriguez, J., Moskaluk, C. A., Frierson, H. F., and Hampton, G. M. Analysis of gene expression identifies candidate markers and pharmacological targets in prostate cancer. *Cancer Research*, 61: 5974-5978, 2001.
47. West, R. B., Nuyten, D. S., Subramanian, S., Nielsen, T. O., Corless, C. L., Rubin, B. P., Montgomery, K., Zhu, S., Patel, R., Hernandez-Boussard, T., Goldblum, J. R., Brown, P. O., van de Vijver, M., and van de Rijn, M. Determination of stromal signatures in breast carcinoma. *PLoS Biology*, 3: e187, 2005.
48. Ye, Q. H., Qin, L. X., Forgues, M., He, P., Kim, J. W., Peng, A. C., Simon, R., Li, Y., Robles, A. I., Chen, Y., Ma, Z. C., Wu, Z. Q., Ye, S. L., Liu, Y. K., Tang, Z. Y., and Wang, X. W. Predicting hepatitis B virus-positive

metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nature Medicine*, 9: 416-423, 2003.