A Genomic Strategy to Refine Prognosis in Early-Stage Non–Small-Cell Lung Cancer

Anil Potti, M.D., Sayan Mukherjee, Ph.D., Rebecca Petersen, M.D.,
Holly K. Dressman, Ph.D., Andrea Bild, Ph.D., Jason Koontz, M.D.,
Robert Kratzke, M.D., Mark A. Watson, M.D., Ph.D., Michael Kelley, M.D.,
Geoffrey S. Ginsburg, M.D., Ph.D., Mike West, Ph.D., David H. Harpole, Jr., M.D.,
and Joseph R. Nevins, Ph.D.

ABSTRACT

From the Institute for Genome Sciences and Policy (A.P., S.M., H.K.D., A.B., J.K., G.S.G., M.W., J.R.N.) and the Institute of Statistics and Decision Sciences (S.M., M.W.), Duke University; and the Departments of Medicine (A.P., J.K., M.K., G.S.G.), Surgery (R.P., D.H.H.), and Molecular Genetics and Microbiology (H.K.D., A.B., J.R.N.), Duke University Medical Center — both in Durham, N.C.; the Department of Medicine, University of Minnesota, Minneapolis (R.K.); and the Department of Pathology and Immunology, Washington University School of Medicine, St. Louis (M.A.W.). Address reprint requests to Dr. Nevins at the Duke Institute for Genome Sciences and Policy, Duke University, 101 Science Dr., Box 3382, Durham, NC 27708, or at nevin001@mc.duke.edu.

BACKGROUND

Clinical trials have indicated a benefit of adjuvant chemotherapy for patients with stage IB, II, or IIIA — but not stage IA — non–small-cell lung cancer (NSCLC). This classification scheme is probably an imprecise predictor of the prognosis of an individual patient. Indeed, approximately 25 percent of patients with stage IA disease have a recurrence after surgery, suggesting the need to identify patients in this subgroup for more effective therapy.

METHODS

We identified gene-expression profiles that predicted the risk of recurrence in a cohort of 89 patients with early-stage NSCLC (the lung metagene model). We evaluated the predictor in two independent groups of 25 patients from the American College of Surgeons Oncology Group (ACOSOG) Z0030 study and 84 patients from the Cancer and Leukemia Group B (CALGB) 9761 study.

RESULTS

The lung metagene model predicted recurrence for individual patients significantly better than did clinical prognostic factors and was consistent across all early stages of NSCLC. Applied to the cohorts from the ACOSOG Z0030 trial and the CALGB 9761 trial, the lung metagene model had an overall predictive accuracy of 72 percent and 79 percent, respectively. The predictor also identified a subgroup of patients with stage IA disease who were at high risk for recurrence and who might be best treated by adjuvant chemotherapy.

CONCLUSIONS

The lung metagene model provides a potential mechanism to refine the estimation of a patient's risk of disease recurrence and, in principle, to alter decisions regarding the use of adjuvant chemotherapy in early-stage NSCLC.

LUNG CANCER IS THE LEADING CAUSE OF death from cancer among both men and women in the United States, and non–small-cell lung cancer (NSCLC) accounts for almost 80 percent of such deaths.[1,2] The clinical staging system has been the standard for determining lung-cancer prognosis.[3-5] Although other clinical and biochemical markers have prognostic significance,[6,7] none are more accurate than the clinico-pathological stage.[8]

The current standard of treatment for patients with stage I NSCLC is surgical resection, despite the observation that nearly 30 to 35 percent will relapse after the initial surgery and thus have a poor prognosis,[2,4] indicating that a subgroup of these patients might benefit from adjuvant chemotherapy. Similarly, as a population, patients with clinical stage IB, IIA or IIB, or IIIA NSCLC receive adjuvant chemotherapy,[9-13] but some may receive potentially toxic chemotherapy unnecessarily. Thus, the ability to identify subgroups of patients more accurately may improve health outcomes across the spectrum of disease.

Previous studies have described the development of gene-expression, protein, and messenger RNA profiles that are associated in some cases with the outcome of lung cancer.[14-24] However, the extent to which these profiles can be used to refine the clinical prognosis and the context in which improved prognostic capability could be used to alter a clinical treatment decision were not clear. Thus, we evaluated the use of gene-expression patterns as a means of stratifying risk and treatment in NSCLC.

METHODS

PATIENTS AND TUMOR SAMPLES

We analyzed 198 tumor samples from three cohorts of patients with NSCLC. The training cohort consisted of 89 patients enrolled through the Duke Lung Cancer Prognostic Laboratory. The independent validation cohorts included patients in two multicenter cooperative group trials: 25 patients from the American College of Surgeons Oncology Group (ACOSOG) Z0030 study and 84 from the prospective Cancer and Leukemia Group B (CALGB) 9761 trial. Table 1 lists the clinical and demographic characteristics of the patients in each cohort and their tumors, and complete details are listed in Table 1 of the Supplementary Appendix, available with the full text of this article at www.nejm.org. All patients were enrolled according to protocols approved by the institutional review board of Duke University, after written informed consent had been obtained.

HISTOPATHOLOGICAL EVALUATION

For each cohort, a single pathologist reviewed all slides to determine whether they met the histopathological criteria for NSCLC of the World Health Organization, including the subtype of adenocarcinoma and the degrees of differentiation, lymphatic invasion, and vascular invasion. Only samples with a tumor-cell content of more than 50 percent were used in the analysis.

GENE-EXPRESSION ARRAYS

Total RNA was extracted from the tumor tissue with RNeasy Kits (Qiagen). The RNA quality was assessed with the use of a bioanalyzer (model 2100, Agilent). Hybridization targets were prepared from the total RNA according to standard Affymetrix protocols (described in detail in the Supplementary Appendix, along with the methods involved in the scanning of the arrays and the normalization of the resulting data). The microarray assays were carried out with Affymetrix GeneChips (U133 Plus2). All raw data and data transformed with the use of the robust multiarray average expression measure for the Duke, ACOSOG, and CALGB data sets are available elsewhere (accession number GSE3593 in the Gene Expression Omnibus database at www.ncbi.nlm.nih.gov/geo).

STATISTICAL ANALYSIS

We performed statistical analyses using the metagene construction and binary prediction tree analysis, as described previously[25-29] and in detail in the Supplementary Appendix. The metagene for a cluster of genes is the dominant singular factor (principal component), as computed with the use of a singular value decomposition of gene-expression levels in the gene cluster in all samples. The metagene represents the dominant average pattern of expression of the gene cluster across the tumor samples.[25]

We then used the set of metagenes and the clinical variables previously shown to be of prognostic value (age, sex, tumor diameter, stage of

Table 1. Characteristics of Patients and Tumors.*

Characteristic	Duke Training Cohort (N = 89)	ACOSOG Z0030 Validation Cohort (N = 25)	CALGB 9761 Validation Cohort (N = 84)
Age — yr			
Median	67	67	66
Range	32–83	41–80	33–82
Mean	65±10	65±5	65±10
Sex — %			
Male†	56 (63)	16 (64)	56 (67)
Female	33 (37)	9 (36)	28 (33)
Race — %‡		—	—
White	78 (88)		
Black	8 (9)		
Other	3 (3)		
Tobacco history — no. (%)		—	—
Cigarette smoking			
None	7 (8)		
≤20 Yr	10 (11)		
21–49 Yr	36 (40)		
≥50 Yr	34 (38)		
Heavy cigar smoking, ≥40 Yr	2 (2)		
Cancer-cell type — no. (%)			
Adenocarcinoma	45 (51)	11 (44)	84 (100)
Squamous	44 (49)	14 (56)	0
Stage — no. (%)			
IA	39 (44)	5 (20)	24 (29)
IB	30 (34)	13 (52)	28 (33)
IIA	4 (5)	2 (8)	7 (8)
IIB	10 (11)	5 (20)	8 (10)
IIIA	6 (7)	—	9 (11)
IIIB	—	—	8 (10)
Tumor diameter — cm	4±2	3±2	3±2
Tumor stage — no. (%)			
1	37 (42)	10 (40)	33 (39)
2	49 (55)	14 (56)	38 (45)
3	3 (3)	1 (4)	5 (6)
4	—	—	8 (10)
Nodal status — no. (%)			
Negative	66 (74)	19 (76)	60 (71)
Positive	23 (26)	6 (24)	24 (29)
Accuracy of the lung metagene model — %§	93	72	79

* Plus–minus values are means ±SD. Percentages may not total 100, because of rounding.
† There were more men in the study cohorts, since one of the principal sites involved was a Veterans Affairs medical center.
‡ Race was self-reported.
§ For the ACOSOG and CALGB data sets, the accuracy was predicted with the use of the Duke cohort as the training cohort. Recurrence was defined with the use of a probability of 0.5 as a cutoff.

disease, histologic subtype, and smoking history) in a binary classification-tree analysis to partition the samples recursively into smaller subgroups. Within these subgroups, predictions of recurrence (with 0 representing 5-year disease-free survival and 1 representing death within 2.5 years after the initial diagnosis of NSCLC) were made in terms of the estimated relative probabilities.[26,30,31] In the analysis, many classification trees were computed, weighed, and integrated to provide overall risk predictions for each patient. The dominant metagenes that constituted the final model are described in the Supplementary Appendix.

To compare the prognostic efficacy of the metagene and clinical strategies, the clinical variables were treated as factors or principal components (similar to the treatment of metagenes in the lung metagene model) in a classification-tree analysis to generate a clinical model. The end result was the probability of recurrence, which represents the conglomerate prognostic value of the individual clinical variables. Using GraphPad software, we computed a C statistic (comparable to the area under the curve in a receiver-operating-characteristic curve in the prediction of binary outcomes) for the model that included just the clinical variables, a C statistic for a model that included just the metagenes, and a C statistic for a model that included both the clinical and genomic variables.

The accuracy of each model was defined with the use of a probability of 0.5 as a cutoff. An estimated probability of recurrence of more than 0.5 was classified as a high risk of recurrence; an estimated probability of recurrence of 0.5 or less was classified as a low risk of recurrence.

Simple univariate and multivariate logistic regressions for recurrence (with and without the metagene-based assessment of the risk) were also computed to assess the baseline prognostic value of each clinical variable (age, sex, tumor diameter, stage of disease, histologic subtype, and smoking history) in the cohorts. We also calculated the sensitivity, specificity, and positive and negative predictive values using a probability of recurrence of 0.5 as the cutoff value. Standard Kaplan–Meier survival curves were generated for the high-risk and low-risk groups of patients with the use of GraphPad software; the survival curves were compared with the use of the log-rank test. This test generates a two-tailed P value that tests the null hypothesis, which was that the survival curves were identical among the cohorts.
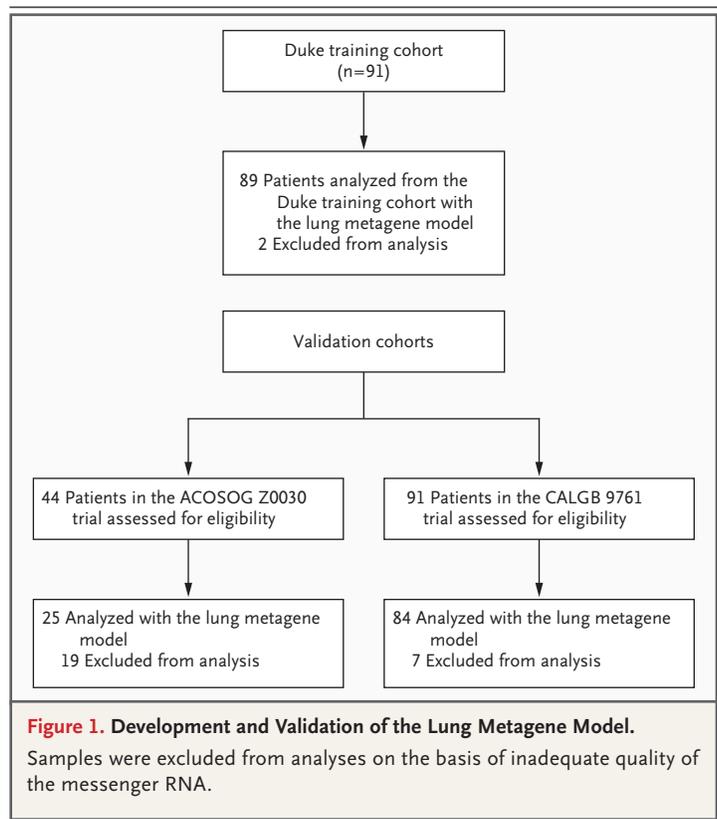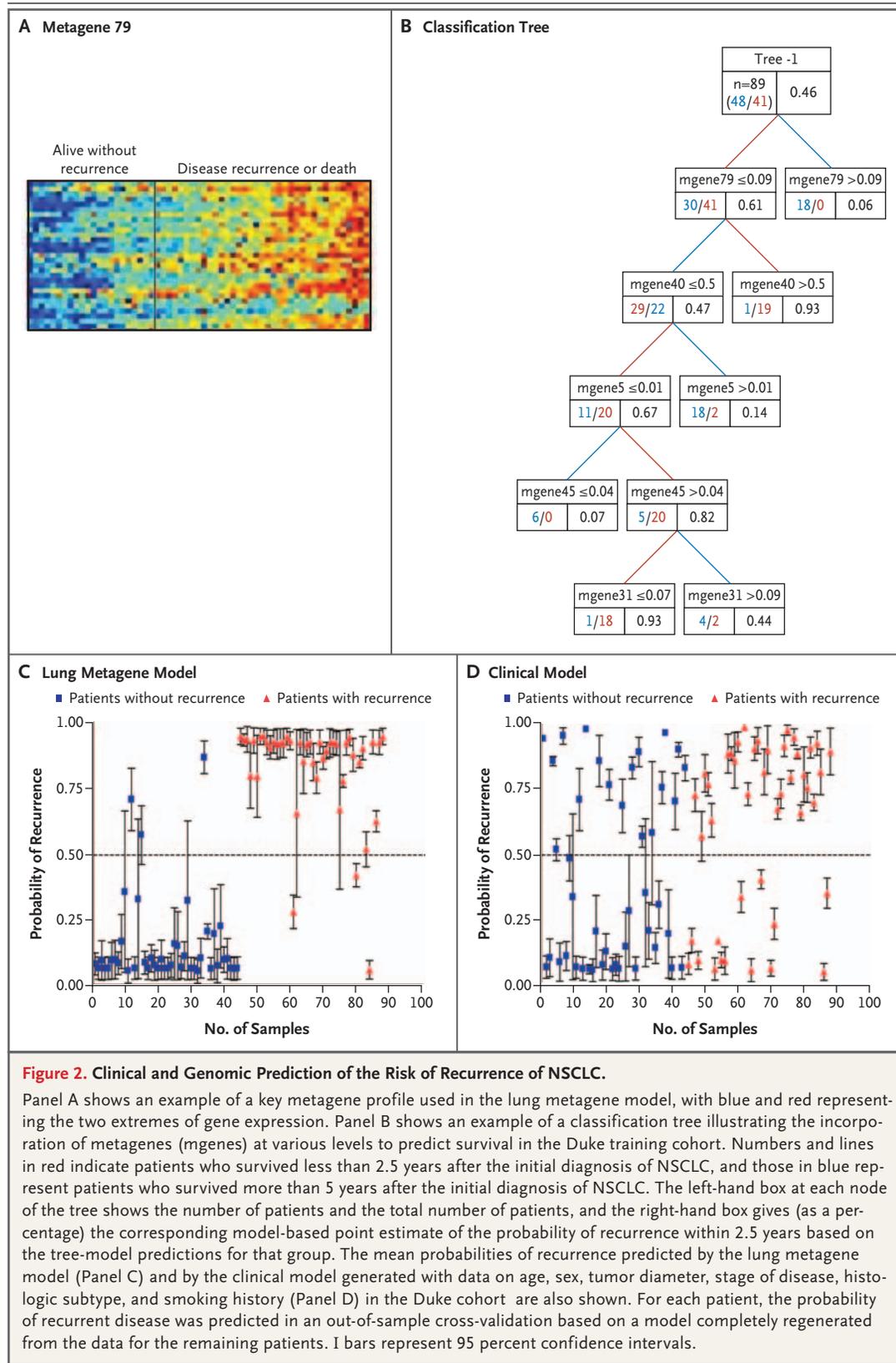
RESULTS

PATIENT CHARACTERISTICS

Table 1 lists the demographic and clinical characteristics of the patients (and their tumors) used to develop and test the prognostic model (Fig. 1).

USE OF GENE-EXPRESSION PROFILES TO IMPROVE PROGNOSIS

Lung cancer is a heterogeneous disease resulting from the acquisition of multiple somatic mutations; given this complexity, it would be surprising if a single gene-expression pattern could effectively describe and ultimately predict the clinical course of the disease for all patients. Recognizing the importance of addressing this complexity, we have previously described methods to integrate various forms of data, including clinical variables and multiple gene-expression profiles, to build robust predictive models for the individual patient.[25,26] There are two critical components of this methodologic approach. First, we generated a collection of gene-expression profiles, termed "metagenes" (an example is given in Fig. 2A), that provide the basis for the predictive models. Second, we used clas-

Figure 1. Development and Validation of the Lung Metagene Model.

Samples were excluded from analyses on the basis of inadequate quality of the messenger RNA.

A Metagene 79

Alive without
recurrence   Disease recurrence or death

B Classification Tree

Tree -1

n=89   0.46
(48/41)

mgene79 ≤0.09   mgene79 >0.09
30/41   0.61   18/0   0.06

mgene40 ≤0.5   mgene40 >0.5
29/22   0.47   1/19   0.93

mgene5 ≤0.01   mgene5 >0.01
11/20   0.67   18/2   0.14

mgene45 ≤0.04   mgene45 >0.04
6/0   0.07   5/20   0.82

mgene31 ≤0.07   mgene31 >0.09
1/18   0.93   4/2   0.44

C Lung Metagene Model

■ Patients without recurrence   ▲ Patients with recurrence

Probability of Recurrence / No. of Samples

D Clinical Model

■ Patients without recurrence   ▲ Patients with recurrence

Probability of Recurrence / No. of Samples

Figure 2. Clinical and Genomic Prediction of the Risk of Recurrence of NSCLC.

Panel A shows an example of a key metagene profile used in the lung metagene model, with blue and red representing the two extremes of gene expression. Panel B shows an example of a classification tree illustrating the incorporation of metagenes (mgenes) at various levels to predict survival in the Duke training cohort. Numbers and lines in red indicate patients who survived less than 2.5 years after the initial diagnosis of NSCLC, and those in blue represent patients who survived more than 5 years after the initial diagnosis of NSCLC. The left-hand box at each node of the tree shows the number of patients and the total number of patients, and the right-hand box gives (as a percentage) the corresponding model-based point estimate of the probability of recurrence within 2.5 years based on the tree-model predictions for that group. The mean probabilities of recurrence predicted by the lung metagene model (Panel C) and by the clinical model generated with data on age, sex, tumor diameter, stage of disease, histologic subtype, and smoking history (Panel D) in the Duke cohort are also shown. For each patient, the probability of recurrent disease was predicted in an out-of-sample cross-validation based on a model completely regenerated from the data for the remaining patients. I bars represent 95 percent confidence intervals.

sification- and regression-tree analysis to sample these metagenes and build prognostic models; this approach mines the collection of profiles to predict the clinical outcome best. An example tree (one of many generated in the analysis) is depicted in Figure 2B.

The predictive accuracy of each model was initially assessed with the use of leave-one-out cross-validation, in which the analysis is performed repeatedly, one sample is removed each time, and the probability of recurrence is predicted for that sample. Because the entire model-building process is repeated for each prediction, the reproducibility of the approach is also evaluated. As a measure of model stability, we generated multiple iterations of randomly split training and validation sets from within the Duke cohort; the resulting accuracy of prognostic capability exceeded 85 percent (data not shown).

The lung metagene model for the prediction of recurrence was superior to a predictive model generated with the same methods but that included clinical data alone (including age, sex, tumor diameter, stage of disease, histologic subtype, and smoking history). In the Duke cohort, the lung metagene model predicted disease recurrence with an overall accuracy of 93 percent (Fig. 2C). The model built with clinical data had an accuracy of only 64 percent (Fig. 2D). Inclusion of the clinical data with the genomic data did not further improve the accuracy of the prediction of recurrence over that of the genomic data alone.

The outperformance of the clinical model by the lung metagene model in identifying patients at risk for recurrence was also supported by the results of Kaplan–Meier analyses. The lung metagene model identified two distinct groups of patients with respect to survival (Fig. 3A). In contrast, the distinction was less clear for each of the models based on clinical predictions (one that combined the clinical variables in a manner similar to the lung metagene model, and another that was based on individual clinical prognostic factors [tumor diameter and stage of disease are shown]) (Fig. 3B). Univariate and multivariate analyses (with and without the genome-based assessment of the risk of recurrence) to assess the relative prognostic value of the individual clinical variables and the lung metagene model showed that the lung metagene model performed significantly better ($P<0.001$ by multivariate analysis)

than stage of disease, tumor diameter, nodal status, age, sex, histologic subtype, or smoking history (Table 3 in the Supplementary Appendix).

Finally, further confirmation that the lung metagene model represents the biology of the tumor was provided by the finding that the metagenes with the greatest discriminatory capability in the model included genes that have previously been shown to have clinical relevance in NSCLC. In some instances, a metagene represented a single molecular process such as angiogenesis (metagene 19), which is a proven target for therapy in NSCLC. Other key metagenes, such as metagene 41, represented a combination of biologic processes — for example, the BRAF, phosphatidylinositol 3′ kinase, TP53, and MYC signaling pathways.

VALIDATION OF THE METAGENE PROGNOSTIC MODEL

Validation across Early Stages and Subtypes of NSCLC

The samples used to devise the prognostic model represented both the major histologic subtypes of NSCLC (adenocarcinoma and squamous-cell carcinoma) and all the early stages of disease. To assess the general robustness of the prognostic model in the Duke cohort, we examined the predictions of risk as a function of these variables. The lung metagene model was consistently accurate across all the early stages of NSCLC (Fig. 1 in the Supplementary Appendix) and between the major histologic subtypes (Fig. 2 in the Supplementary Appendix), not only in the estimated risk of recurrence but also in the results of the Kaplan–Meier survival analysis for each stage or subtype.
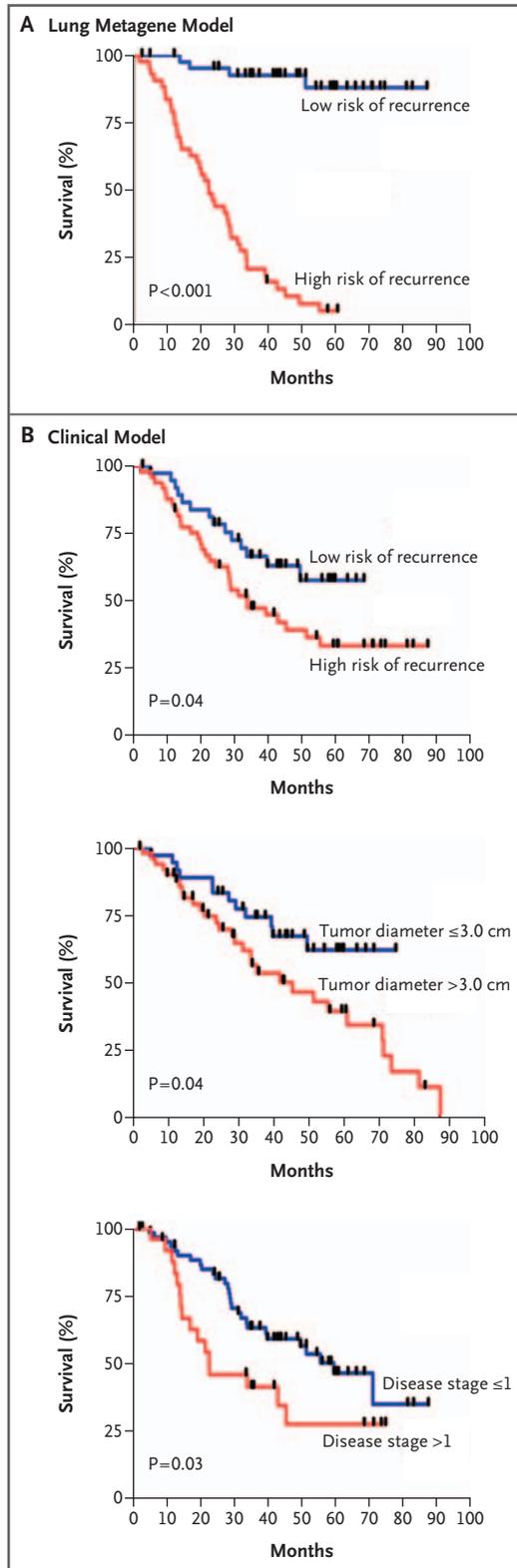
Validation across Data from Two Multicenter Studies

For a new prognostic model that assesses the risk of recurrence to be used to inform the decision of whether to administer adjuvant chemotherapy, the model must be shown to be robust when applied to independent, heterogeneous populations of patients and conditions of sample acquisition. We therefore evaluated the ability of the metagene model generated from the Duke training cohort to predict the risk of recurrence by using samples from two multicenter, cooperative group studies (ACOSOG Z0030 and CALGB 9761) (Fig. 1). These sets of samples represented the full spectrum of clinical outcomes; the samples were not selected with respect to the duration of survival.
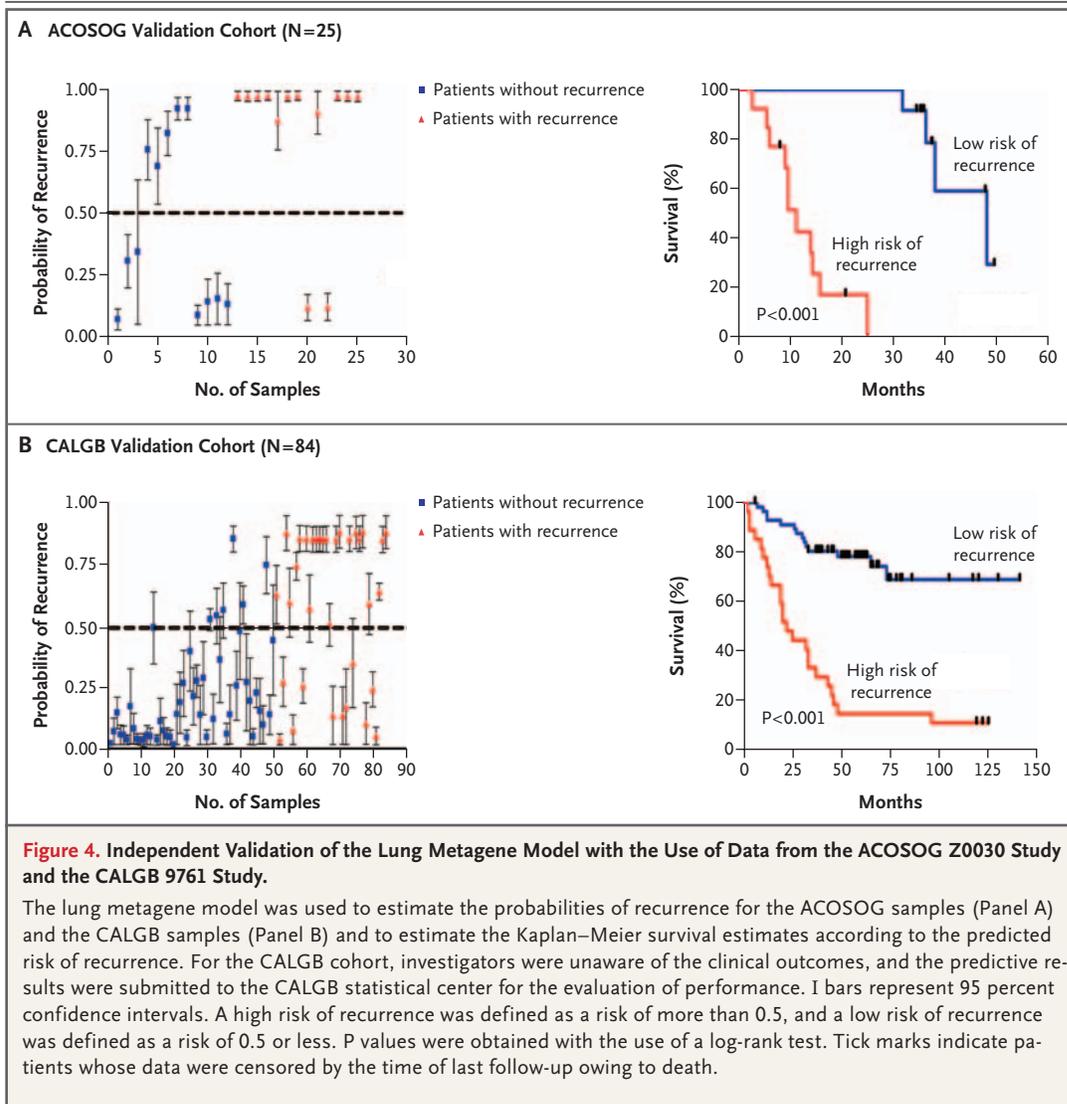
Figure 3. Kaplan–Meier Survival Estimates for the Duke Training Cohort.

Estimates based on predictions from the lung metagene model demonstrate the value of that approach (Panel A). Panel B shows the estimates based on the clinical model of prognosis, as well as those based on individual clinical characteristics — here, tumor diameter and stage of disease. A high risk of recurrence was defined as a probability of recurrence of more than 0.5, and a low risk of recurrence was defined as a risk of 0.5 or less. P values were obtained with the use of a log-rank test. Tick marks indicate patients whose data were censored by the time of last follow-up owing to death.

A **Lung Metagene Model**

B **Clinical Model**

We analyzed 25 samples from the ACOSOG Z0030 trial to validate the performance of the predictive model of recurrence based on the Duke training cohort. As was the case with the Duke cohort, for the ACOSOG Z0030 cohort, univariate and multivariate analyses showed that the metagene model was a significantly more accurate predictor ($P<0.001$ by multivariate analysis) than stage of disease, tumor diameter, nodal status, age, sex, histologic subtype, or smoking history (Table 3 in the Supplementary Appendix). The accuracy of the prediction of recurrence in the ACOSOG samples was approximately 72 percent (sensitivity, 85 percent; specificity, 58 percent; positive predictive value, 69 percent; and negative predictive value, 78 percent) (Fig. 4A). The level of accuracy provides an assessment of the robustness of the risk predictions and is substantial, particularly given the heterogeneity of the cohort and the fact that the clinical outcomes among the patients in the ACOSOG cohort are prospective. The Kaplan–Meier survival curves, stratified according to the risk predictions based on the lung metagene model, provide strong evidence of the reliability of those predictions (Fig. 4A). In addition, a multivariate analysis showed that in this cohort, the patients predicted by the lung metagene model to have a probability of recurrence of more than 0.5 were more likely to have a recurrence than those with a predicted probability of recurrence of 0.5 or less (adjusted odds ratio, 35.9; 95 percent confidence interval, 2.8 to 46.3).

We analyzed 84 samples from the CALGB 9761 trial as a second independent validation cohort. The investigators applying the predictive model were unaware of the outcomes among these patients; thus, the genome-based predictions of re-

Figure 4. Independent Validation of the Lung Metagene Model with the Use of Data from the ACOSOG Z0030 Study and the CALGB 9761 Study.

The lung metagene model was used to estimate the probabilities of recurrence for the ACOSOG samples (Panel A) and the CALGB samples (Panel B) and to estimate the Kaplan–Meier survival estimates according to the predicted risk of recurrence. For the CALGB cohort, investigators were unaware of the clinical outcomes, and the predictive results were submitted to the CALGB statistical center for the evaluation of performance. I bars represent 95 percent confidence intervals. A high risk of recurrence was defined as a risk of more than 0.5, and a low risk of recurrence was defined as a risk of 0.5 or less. P values were obtained with the use of a log-rank test. Tick marks indicate patients whose data were censored by the time of last follow-up owing to death.

currence were submitted to a CALGB statistician for comparison with the true outcomes. Once again, univariate and multivariate analyses showed that the lung metagene model predicted outcome significantly better (P<0.001 by multivariate analysis) than the stage of disease, tumor diameter, nodal status, age, sex, histologic subtype, or smoking history (Table 3 in the Supplementary Appendix). The overall predictive accuracy of the model for the CALGB samples was 79 percent (sensitivity, 68 percent; specificity, 88 percent; positive predictive value, 79 percent; and negative predictive value, 80 percent) (Fig. 4A). Again, the Kaplan–Meier analysis showed a significant difference in the survival rates of patients with a

probability of recurrence of greater than 0.5 as compared with 0.5 or less, according to the lung metagene model (Fig. 4B). Similar to the results seen for the Duke and ACOSOG data, the adjusted odds ratio for disease recurrence in the CALGB cohort was 16.6 (95 percent confidence interval, 4.4 to 62.8) when the model estimate for recurrence was greater than 0.5 (Table 3 in the Supplementary Appendix).

We also applied the lung metagene model to another cohort of 15 patients with surgically resected stage I squamous-cell lung cancer. Using the lung metagene model, we were able to predict the outcome accurately in all 5 patients with recurrence and in 7 of 10 patients without recur-

rence, for an overall accuracy of 80 percent (Fig. 3 in the Supplementary Appendix).

Finally, to evaluate the extent to which the metagene model could increase the ability of clinicians to estimate prognosis, we computed a C statistic as a measure of the capacity of the clinical or genomic information to identify patients according to the risk of recurrence. For the ACOSOG cohort, the C statistic based on clinical variables alone was 0.67; this value was increased to 0.84 by the inclusion of genomic data. For the CALGB cohort, inclusion of the genomic data increased the value from 0.73 to 0.87. Clearly, the genomic data transformed a limited clinical-based prognosis to one with substantial capacity to identify patients who were likely to have disease recurrence.

APPLICATION OF THE REFINED PROGNOSIS

Previous studies have shown that 25 percent of patients with stage IA NSCLC will have disease recurrence within five years. Thus, some patients with stage IA NSCLC might be more appropriately categorized as being at higher risk than others and might be candidates for adjuvant chemother-apy. We therefore focused on the 68 patients from the Duke, ACOSOG, and CALGB cohorts who were classified clinically as having stage IA disease. Kaplan–Meier survival curves were generated for the group as a whole, as well as for the subgroups predicted to be at high or low risk for recurrence by the lung metagene model. Although the survival rate for the group was approximately 70 percent at four years, the survival rate for those predicted to be at high risk was less than 10 percent (Fig. 5A), thus identifying the subgroup of patients with stage IA NSCLC at risk for recurrence.

DISCUSSION

Although gene-expression profiles that can classify patients with cancer according to their risk of recurrence have been described in many instances, the prognostic tool we devised could be used to change a clinical decision. In particular, the guidelines for the treatment of patients with stage I NSCLC provide an opportunity to use an improved prognostic model to refine the currently imprecise assessment of risk and the decision re-
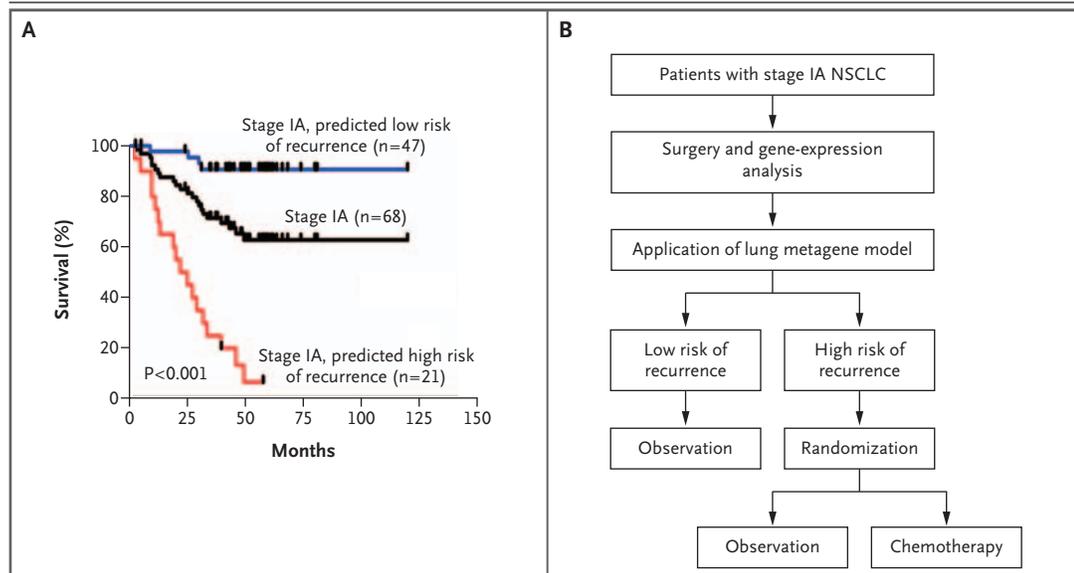
Figure 5. Application of the Lung Metagene Model to Refine the Assessment of Risk and Guide the Use of Adjuvant Chemotherapy in Stage IA NSCLC.

Panel A shows the Kaplan–Meier survival estimates for a group of patients with stage IA disease from the Duke, ACOSOG, and CALGB cohorts and the subgroups predicted to have either a high probability (>0.5) or a low probability (≤0.5) of recurrence. P values were obtained with the use of a log-rank test. Tick marks indicate patients whose data were censored by the time of last follow-up owing to death. Panel B illustrates the possible design of a planned prospective, phase 3 clinical trial involving patients with stage IA NSCLC to evaluate the performance of the metagene model.

garding whom to treat, and thus potentially leading to more personalized cancer treatment. In this case, the refinement of prognosis with the use of the metagene model provides the opportunity for a prospective, randomized, phase 3 clinical trial that would evaluate the benefit of the identification of a subgroup of patients with stage IA disease estimated to be at high risk for recurrence (Fig. 5B). Patients initially classified as having clinical stage IA disease would undergo surgery, and the metagene model would then be applied to identify the patients predicted to be at high risk for recurrence. Patients at high risk would then be randomly assigned to observation (the current standard of care for stage IA disease) or adjuvant chemotherapy, in order to evaluate the extent to which the use of genomic reclassification improves survival. Our study is a critical first step in the use of genomic tools as a strategy to refine the prognosis and improve the selection of patients appropriate for adjuvant chemotherapy.

REFERENCES

1. Spira A, Ettinger DS. Multidisciplinary management of lung cancer. N Engl J Med 2004;350:379-92.
2. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. Lancet 2000;355:479-85. [Erratum, Lancet 2000;355:1280.]
3. Mountain CF. Revisions in the International System for Staging Lung Cancer. Chest 1997;111:1710-7.
4. Nesbitt JC, Putnam JB Jr, Walsh GL, Roth JA, Mountain CF. Survival in early-stage non-small cell lung cancer. Ann Thorac Surg 1995;60:466-72.
5. Mountain CF. The new International Staging System for Lung Cancer. Surg Clin North Am 1987;67:925-35.
6. D'Amico TA, Massey M, Herndon JE II, Moore MB, Harpole DH Jr. A biologic risk model for stage I lung cancer: immunohistochemical analysis of 408 patients with the use of ten molecular markers. J Thorac Cardiovasc Surg 1999;117:736-43.
7. Brundage MD, Davies D, Mackillop WJ. Prognostic factors in non-small cell lung cancer: a decade of progress. Chest 2002;122:1037-57.
8. Meyerson M, Carbone DP. Genomic and proteomic profiling of lung cancers: lung cancer classification in the age of targeted therapy. J Clin Oncol 2005;23:3219-26.
9. Arriagada R, Bergman B, Dunant A, et al. Cisplatin-based adjuvant chemotherapy in patients with completely resected non-small-cell lung cancer. N Engl J Med 2004;350:351-60.
10. Winton T, Livingston R, Johnson D, et al. Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. N Engl J Med 2005;352:2589-97.
11. Douillard J-Y, Rosell R, Delena M, Legroumellec A, Torres A, Carpagnano F. ANITA: phase III adjuvant vinorelbine (N) and cisplatin (P) versus observation (OBS) in completely resected (stage I-III) non-small-cell lung cancer (NSCLC) patients (pts): final results after 70-month median follow-up. J Clin Oncol 2005;23:Suppl:7013. abstract.
12. Kato H, Ichinose Y, Ohta M, et al. A randomized trial of adjuvant chemotherapy with uracil–tegafur for adenocarcinoma of the lung. N Engl J Med 2004;350:1713-21.
13. Strauss GM. Herndon JE II, Maddaus MA, et al. Randomized clinical trial of adjuvant chemotherapy with paclitaxel and carboplatin following resection in Stage 1B non-small cell lung cancer. J Clin Oncol 2004;22:7019. abstract.
14. Tonon G, Wong KK, Maulik G, et al. High-resolution genomic profiles of human lung cancer. Proc Natl Acad Sci U S A 2005;102:9625-30.
15. Schneider PM, Praeuer HW, Stoeltzing O, et al. Multiple molecular marker testing (p53, C-Ki-ras, c-erbB-2) improves estimation of prognosis in potentially curative resected non-small cell lung cancer. Br J Cancer 2000;83:473-9.
16. Berrar D, Sturgeon B, Bradbury I, Downes CS, Dubitzky W. Survival trees for analyzing clinical outcome in lung adenocarcinomas based on gene expression profiles: identification of neogenin and diacylglycerol kinase alpha expression as critical factors. J Comput Biol 2005;12:534-44.
17. Ju Z, Kapoor M, Newton K, et al. Global detection of molecular changes reveals concurrent alteration of several biological pathways in nonsmall cell lung cancer cells. Mol Genet Genomics 2005;274:141-54.
18. Beer DG, Kardia SLR, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816-24.
19. Chen G, Gharib TG, Wang H, et al. Protein profiles associated with survival in lung adenocarcinoma. Proc Natl Acad Sci U S A 2003;100:13537-42.
20. Bhattacharjee A, Richards WG, Staunton J, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. Proc Natl Acad Sci U S A 2001;98:13790-5.
21. Wigle DA, Jurisica I, Radulovich N, et al. Molecular profiling of non-small cell lung cancer and correlation with disease-free survival. Cancer Res 2002;62:3005-8.
22. Kikuchi T, Daigo Y, Katagiri T, et al. Expression profiles of non-small cell lung cancers on cDNA microarrays: identification of genes for prediction of lymph-node metastasis and sensitivity to anti-cancer drugs. Oncogene 2003;22:2192-205.
23. Garber ME, Troyanskaya OG, Schluens K, et al. Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci U S A 2001;98:13784-9.
24. Yanaihara N, Caplen N, Bowman E, et al. Unique microRNA molecular profiles in lung cancer diagnosis and prognosis. Cancer Cell 2006;9:189-98.
25. Pittman J, Huang E, Dressman H, et

al. Integrated modeling of clinical and gene expression information for personalized prediction of disease outcomes. Proc Natl Acad Sci U S A 2004;101:8431-6.

26. Pittman J, Huang E, Nevins JR, Wang Q, West M. Bayesian analysis of binary prediction tree models for retrospectively sampled outcomes. Biostatistics 2004;5:587-601.

27. Nevins JR, Huang ES, Dressman H,

Pittman J, Huang AT, West M. Towards integrated clinico-genomic models for personalized medicine: combining gene expression signatures and clinical factors in breast cancer outcomes prediction. Hum Mol Genet 2003;12:R153-R157.

28. Huang E, Cheng SH, Dressman H, et al. Gene expression predictors of breast cancer outcomes. Lancet 2003;361:1590-6.

29. West M, Blanchette C, Dressman H, et

al. Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci U S A 2001; 98:11462-7.

30. Denison DGT, Mallick BK, Smith AFM. A Bayesian CART algorithm. Biometrika 1998;85:363-77.

31. Breiman L. Statistical modeling: the two cultures. Stat Sci 2001;16:199-225.

Copyright © 2006 Massachusetts Medical Society.

JOURNAL EDITORIAL FELLOW

The *Journal*'s editorial office invites applications for a one-year research fellowship beginning in July 2007 from individuals at any stage of training. The editorial fellow will work on *Journal* projects and will participate in the day-to-day editorial activities of the *Journal* but is expected in addition to have his or her own independent projects. Please send curriculum vitae and research interests to the Editor-in-Chief, 10 Shattuck St., Boston, MA 02115 (fax, 617-739-9864), by October 1, 2006.