Survival prediction of stage I lung adenocarcinomas by expression of 10 genes

Fabrizio Bianchi, … , Francesco Nicassio, Pier Paolo Di Fiore

Research Article **Oncology**

Adenocarcinoma is the predominant histological subtype of lung cancer, the leading cause of cancer deaths in the world. At stage I, the tumor is cured by surgery alone in about 60% of cases. Markers are needed to stratify patients by prognostic outcomes and may help in devising more effective therapies for poor prognosis patients. To achieve this goal, we used an integrated strategy combining meta-analysis of published lung cancer microarray data with expression profiling from an experimental model. The resulting 80-gene model was tested on an independent cohort of patients using RT-PCR, resulting in a 10-gene predictive model that exhibited a prognostic accuracy of approximately 75% in stage I lung adenocarcinoma when tested on 2 additional independent cohorts. Thus, we have identified a predictive signature of limited size that can be analyzed by RT-PCR, a technology that is easy to implement in clinical laboratories.

Find the latest version:

https://jci.me/32007/pdf

Survival prediction of stage I lung adenocarcinomas by expression of 10 genes

Fabrizio Bianchi,[1] Paolo Nuciforo,[1] Manuela Vecchi,[1] Loris Bernard,[1] Laura Tizzoni,[1]
Antonio Marchetti,[2] Fiamma Buttitta,[2] Lara Felicioni,[2]
Francesco Nicassio,[1] and Pier Paolo Di Fiore[1,3,4]

[1]IFOM, Fondazione Istituto FIRC di Oncologia Molecolare, Milan, Italy. [2]Center of Excellence on Aging, University Foundation, Chieti, Italy.
[3]Istituto Europeo di Oncologia, Milan, Italy. [4]Dipartimento di Medicina, Chirurgia ed Odontoiatria, Università degli Studi di Milano, Milan, Italy.

Adenocarcinoma is the predominant histological subtype of lung cancer, the leading cause of cancer deaths in the world. At stage I, the tumor is cured by surgery alone in about 60% of cases. Markers are needed to stratify patients by prognostic outcomes and may help in devising more effective therapies for poor prognosis patients. To achieve this goal, we used an integrated strategy combining meta-analysis of published lung cancer microarray data with expression profiling from an experimental model. The resulting 80-gene model was tested on an independent cohort of patients using RT-PCR, resulting in a 10-gene predictive model that exhibited a prognostic accuracy of approximately 75% in stage I lung adenocarcinoma when tested on 2 additional independent cohorts. Thus, we have identified a predictive signature of limited size that can be analyzed by RT-PCR, a technology that is easy to implement in clinical laboratories.

Introduction

Lung cancer is the leading cause of cancer deaths in the world (1). The prognosis of non–small cell lung carcinoma (NSCLC) largely depends on tumor stage; indeed, the overall low survival rate (about 15% at 5 years; ref. 2) is primarily due to the high frequency of late diagnosis, when the tumor has become unresectable. Conversely, early-stage NSCLC patients (stage I–II) have a significantly better prognosis (30%–60% survival at 5 years; ref. 3).

One important issue in stage I NSCLC is that current diagnostic tools do not allow precise prognostic evaluation. In turn, this limits the power of clinical trials aimed at ameliorating prognosis through multimodality therapy. A case in point is represented by adjuvant chemotherapy, on which conflicting results in stage IB have been reported (4–6). Consequently, there is presently no indication for adjuvant treatment in stage I NSCLC (7, 8). Evidently enough, the availability of accurate prognostic markers might change this picture by allowing the selection, for clinical trials, of only those patients with a high risk of relapse. Thus, there is need for reliable prognostic indicators, both for diagnostic and prognostic purposes and for the design of clinical trials.

Microarray gene expression profiling has been used to identify molecular subtypes of lung cancer associated with different prognostic outcomes (9–20). Moreover, a proteomic-based approach allowed Yanagisawa et al. to distinguish histological subtypes of NSCLC as well as patients with resected tumors who had poor prognosis (21). One problem with these unbiased approaches, which is particularly evident in transcriptome analysis, is the high individual genetic noise associated with each profile, which causes relative instability of the resulting signatures when these are applied to independent datasets. In addition, these signatures tend to contain a high number of genes, and the methodology used is not directly transferable to the clinical setting. Thus, there

is need to develop strategies aimed at the identification of small signatures that can be easily analyzed in the clinical laboratory.

As an alternative to unbiased tumor profiling, some groups have developed approaches based on the profiling of experimental models that mimic specific oncogenic events (22–25). These "biased" approaches allowed the identification of signatures, which were subsequently validated in real human cancers, that might otherwise have been lost within the genetic noise of an unbiased profiling experiment.

We reasoned that a combination of the 2 strategies held potential for better insights into the mechanisms of lung tumorigenesis and for the definition of more reliable prognostic markers. Here, we describe an approach that integrates patterns derived from microarray lung cancer profiling from an experimental model and from known individual prognostic genes. Through this strategy, we identified a 10-gene prognostic signature in stage I lung adenocarcinoma, the predominant histological subtype of NSCLC. This signature, when tested by real-time PCR, a technology that can be rapidly implemented in a clinical setting, displayed excellent predictive power.

Results

Strategy of the integrated approach. The general strategy of our approach is illustrated in Figure 1. Initially, we performed meta-analyses on 2 published expression datasets of lung adenocarcinomas, totaling 170 patients, from studies by Beer et al. (ref. 9; henceforth the Michigan cohort) and by Bhattacharjee et al. (ref. 10; henceforth the Harvard cohort). Patients (Supplemental Table 1; supplemental material available online with this article; doi:10.1172/JCI32007DS1) were divided into good- and poor-prognosis groups according to their clinical outcomes (see Methods). A number of patients, who did not fit the established prognostic criteria, were therefore excluded from the meta-analysis (see Methods). We refer to the datasets from the initial 170 patients as *original* datasets (Michigan, n = 86; Harvard, n = 84) and to those of the selected patients as *reduced* datasets or cohorts (Michigan, n = 41; Harvard, n = 60); each of these datasets included patients
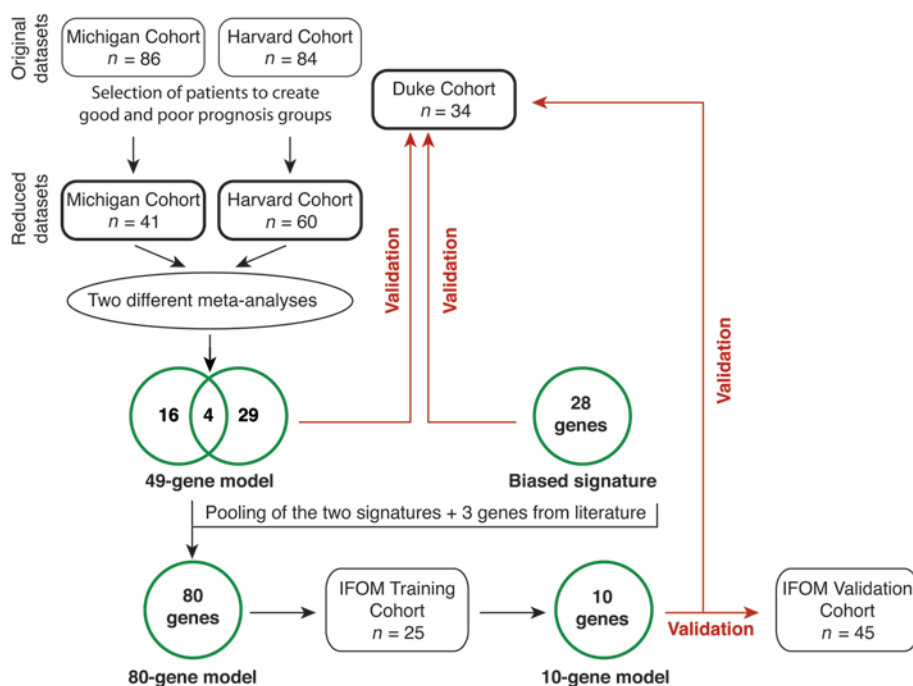
Figure 1
Strategy of the study. Validation of models as good predictors of prognosis in the Duke and IFOM validation cohorts is indicated. For details, see Results and Methods.

concomitantly found in independent analyses of the reduced Michigan and Harvard datasets (Figure 1). Therefore, we performed a class comparison test and identified 361 unique differentially expressed genes ($P < 0.05$, parametric Student's t test) in the reduced Michigan cohort and 429 unique differentially expressed genes ($P < 0.05$, parametric Student's t test) in the reduced Harvard cohort. Twenty genes were shared between the 2 lists ($P < 0.05$, parametric Student's t test). The modest overlap could be explained, at least in part, by the fact that the tumors from the 2 cohorts were analyzed on different array platforms — carrying a substantially different number of genes (Harvard, 9,096; Michigan, 5,588; of which 5,249 common genes were present) — with different protocols. Moreover, individual genetic differences can have an enormous impact on genetic signatures. Thus, the inherent imbalance between conditions (hundreds of patients) and variables (thousands of genes) may generate different signatures. Indeed, a recent study proposed that, to reach an overlap of 50% between 2 lists of prognostic genes, expression profiling studies would need several thousands of patients (29).

In a complementary approach, we assumed that a reliable list of genes should not necessarily be shared by the 2 independent analyses (Figure 1). Hence, we searched for the most stably differentially expressed genes in each dataset, using a stringent P value cutoff ($P < 0.001$, parametric Student's t test). We found 21 unique

with stage I, II, and III tumors. The reduced Michigan and Harvard datasets were then analyzed to obtain lists of genes that were differentially expressed between good- and poor-prognosis patients. This led to the identification of a 49-gene prognostic model that exhibited good prognostic value on the Michigan and Harvard cohorts. More importantly, the 49-gene model was a good predictor of prognosis (Figure 1) in a third independent cohort, composed of 34 stage I lung adenocarcinomas (ref. 23; henceforth the Duke cohort).

To improve the model, we used a biased cancer signature of 28 genes derived from an experimental model that mimics important cancer-related pathways (25). This signature was by itself predictive of prognosis in the Duke cohort (Figure 1).

Finally, we combined genes from the 2 models. We also added 3 genes identified in the literature as individual prognostic markers for stage I lung adenocarcinoma (26–28). The resulting 80-gene model was tested in a real-time PCR–based approach on a fourth cohort of patients (henceforth the IFOM training cohort) to define a predictive model using a limited number of genes as well as a readily accessible technical platform (Figure 1). By the leave-one-out validation method, we refined the model to a final 10-gene model. The 10-gene model was tested on a fifth independent cohort of patients (henceforth the IFOM validation cohort) and on the Duke cohort (Figure 1).

Meta-analysis of 2 lung adenocarcinoma expression profile datasets. As a first approach, we assumed that a reliable list of genes that are differentially regulated in the good- versus poor-prognosis groups should be

Table 1
The 10-gene model

Gene[A]	Name	Unigene	t value	Fold change (P)[B]
NUDCD1	NudC domain containing 1	Hs.558577	−3.95	1.80 (<0.001)
E2F1	E2F transcription factor 1	Hs.654393	−2.63	1.78 (0.016)
HOXB7	Homeobox B7	Hs.436181	−2.59	2.53 (0.017)
MCM6	Minichromosome maintenance complex component 6	Hs.444118	−2.57	1.80 (0.019)
SERPINB5	Serpin peptidase inhibitor, clade B (ovalbumin), member 5	Hs.55279	−2.53	3.95 (0.019)
E2F4	E2F transcription factor 4, p107/p130-binding	Hs.108371	−2.51	1.37 (0.019)
HSPG2	Heparan sulfate proteoglycan 2	Hs.651231	−2.38	1.58 (0.026)
SF3B1	Splicing factor 3b, subunit 1, 155 kDa	Hs.632554	−2.32	1.24 (0.029)
RRM2	Ribonucleotide reductase M2 polypeptide	Hs.226390	−2.16	2.56 (0.041)
SCGB3A1	Secretoglobin, family 3A, member 1	Hs.62492	5.54	0.01 (<0.001)

[A]Sorted by t value. [B]Mean fold change in poor- versus good-prognosis group.
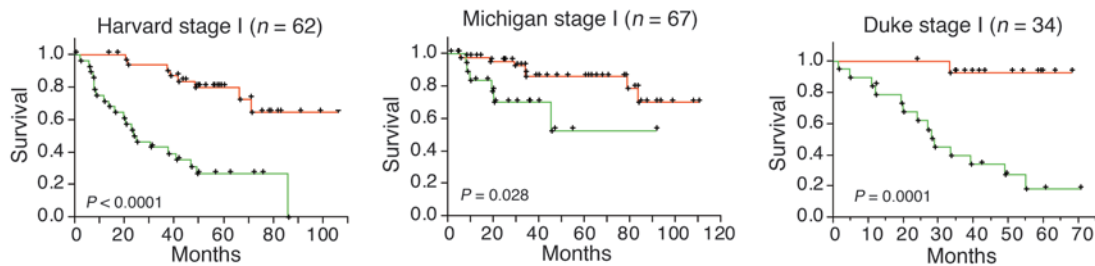
Figure 2

The 49-gene model predicts overall survival. The 49-gene model was used to predict overall survival in the stage I subset of lung adenocarcinomas from the Michigan ($n = 67$ of the original dataset's 86), Harvard ($n = 62$ of the original dataset's 84), and Duke ($n = 34$) cohorts. Data are shown as the probability of survival, in a Kaplan-Meier plot, as a function of a favorable (red line) or unfavorable (green line) signature.

genes in the reduced Michigan cohort and 12 unique genes in the reduced Harvard cohort, for a total of 33 unique genes. In total, by combining the 2 approaches, we identified 49 unique genes (including 4 genes in common between the approaches), which we referred to as the *49-gene model*.

Next, we analyzed the prognostic predictive accuracy of the 49-gene model. Predictive accuracy for the reduced datasets was 90% and 72% in the Michigan and Harvard cohorts, respectively (Supplemental Table 2). For the original datasets, predictive accuracy was 69% and 71% in the Michigan and Harvard cohorts, respectively (Supplemental Table 2). Of note, the 49-gene model performed well when compared with the 2 signatures derived by Beer et al. (9) from the Michigan cohort (Supplemental Table 2).

Finally, the performance of the 49-gene model was tested by Kaplan-Meier analysis on stage I adenocarcinomas (Figure 2). The 49-gene model was very effective in predicting overall survival in the stage I patients from both the Michigan and the Harvard cohorts (Michigan, $n = 67$; Harvard, $n = 62$; Figure 2 and Supplemental Figure 1A). In addition, when we tested a dataset from a third independent expression profile study, the Duke cohort (23), the 49-gene model proved remarkably effective in predicting prognosis (Supplemental Table 2) and overall survival (Figure 2 and Supplemental Figure 1B).

Analysis of an in vitro–derived transcriptional signature. We have previously shown that a biased approach to cancer transcriptomes can lead to the identification of cancer signatures (25). In particular, a 28-gene biased signature was identified (Figure 1) by profiling terminally differentiated myotubes forced to reenter the cell cycle by the viral oncoprotein early region 1A (E1A). The expression of genes from this signature was frequently found to be altered in human neoplasia (25). Thus, we investigated whether the expression of these genes had predictive value in patients with stage I lung adenocarcinomas. We used the dataset from the Duke cohort, because it was the only one for which the expression data for all 28 genes was available. As shown in Figure 3A, the biased signature effectively predicted overall survival, further confirming that a

biased approach can lead to the discovery of cancer-relevant signatures (see Supplemental Table 3).

A 10-gene prognostic model in stage I lung adenocarcinomas. The next step in our experimental approach was to integrate models derived from unbiased and biased screenings. Thus, we combined the 49-gene model and the 28-gene biased signature. We also added 3 genes (*SCGB3A1*, *TERT*, and *EIF3S6*; see Methods and Figure 1) identified in the literature as individual prognostic markers for stage I lung adenocarcinoma (26–28). This set of 80 genes demonstrated excellent predictive power for overall patient survival in Kaplan-Meier analysis of the Duke cohort, the only one for which expression data for all 80 genes were available (Figure 3B and Supplemental Table 4).

The major goal of our efforts, however, was to identify a small number of genes, amenable to analysis using readily available technology (such as real-time PCR), that constitute a prognostic model that can be rapidly transferred to the clinical laboratory. Thus, we used TaqMan Low-Density Arrays (Applied Biosystems) to profile the IFOM training cohort, a set of 25 patients with stage I lung adenocarcinomas (Supplemental Table 1). At the time of our analysis, TaqMan Low-Density Arrays were available for 53 of the 80 genes (Supplemental Table 4). The results are summarized in
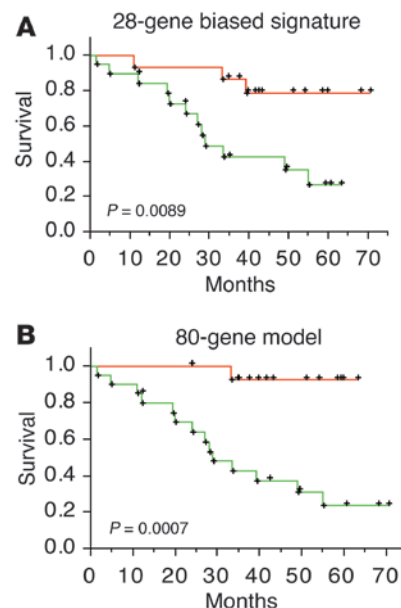
Figure 3

The 28-gene biased signature and the 80-gene model predict overall survival. The 28-gene biased signature (**A**) and the 80-gene model (**B**) were used to predict overall survival in stage I lung adenocarcinomas of the Duke cohort ($n = 34$). Data are shown as the probability of survival, in a Kaplan-Meier plot, as a function of a favorable (red line) or unfavorable (green line) signature.

Table 2
Univariate and multivariate analysis of various biological and biochemical parameters.

Variable (subset)	IFOM validation cohort (*n* = 45)		IFOM combined cohorts (*n* = 70)		IFOM mutational analysis cases (*n* = 39)[A]	
	OD (95% CI)	*P*	OD (95% CI)	*P*	OD (95% CI)	*P*
Univariate analysis						
Age (≥64, <64)	0.96 (0.29–3.16)	0.946	0.33 (0.24–1.61)	0.336		
Differentiation (well or moderate, poor)	0.64 (0.14–2.54)	0.534	0.78 (0.29–2.07)	0.621		
Stage (IA, IB)	0.44 (0.10–1.66)	0.234	1.14 (0.41–3.23)	0.794		
KRAS (mutant, WT)					2.00 (0.49–8.24)	0.327
10-gene model (poor, good)	3.94 (1.17–14.4)	0.026[B]	7.22 (2.60–21.7)	0.0002[B]	4.00 (1.06–16.89)	0.041[B]
Multivariate analysis						
Age (≥64, <64)	1.00 (0.26–3.90)	0.995	0.51 (0.16–1.50)	0.232		
Differentiation (well or moderate, poor)	0.65 (0.13–2.92)	0.584	0.79 (0.24–2.40)	0.679		
Stage (IA, IB)	0.38 (0.08–1.68)	0.222	0.94 (0.28–3.05)	0.922		
KRAS (mutant, WT)					9.07 (1.27–187.9)	0.058
10-gene model (poor, good)	3.98 (1.13–15.5)	0.036[B]	7.78 (2.72–24.8)	0.0002[B]	13.19 (2.06–261.4)	0.021[B]

The 10-gene model was tested for prediction of survival in the indicated cohorts of patients, in comparison to other biological or biochemical parameters, in univariate and multivariate analysis. OD, odds ratio; CI, confidence interval. [A]Cases of the combined IFOM training and validation cohorts for which mutational analysis of the *KRAS* gene was available. [B]Statistically significant value.

Supplemental Table 4. From these results, we excluded a number of genes that did not show variability between the good- and poor-prognosis groups; 16 genes were therefore selected for further analysis using cutoff values of $P \leq 0.05$ or fold change greater than 1.5 (Supplemental Table 4).

The final prognostic model was obtained by the leave-one-out cross-validation procedure, with independent gene selection ($P < 0.05$ as cutoff; parametric Student's *t* test). We found that on the IFOM training cohort, a 10-gene model (Table 1) displayed a predictive accuracy of 84% (sensitivity, 90%; specificity, 80%) and a *P* value of 0.004 after 2,000 random permutations of class labels.

To confirm the robustness of this new prognostic model, we used it on the IFOM validation cohort, an independent cohort of 45 stage I lung adenocarcinomas (Supplemental Table 1). Univariate and multivariate analysis showed that the 10-gene model predicted survival of patients more accurately than did tumor stage (IA versus IB), grading, age, sex, or presence of mutated *KRAS* (Table 2). The 10-gene model was also independent of tumor histological subtype (bronchoalveolar cell carcinoma versus adenocarcinoma; Supplemental Table 5). Kaplan-Meier survival curves showed a significant difference in the survival rate of patients stratified according to the 10-gene prognostic model ($P = 0.008$, log-rank test; Figure 4). It is also of note that our 10-gene model showed very good predictive power in Kaplan-Meier analysis and multivariate analysis of the Duke cohort, for which microarray expression data for all 10 genes were available (Figure 4, Supplemental Figure 2, and Supplemental Table 6).

Importantly, the 10-gene model retained excellent predictive power also when patients with stage IA and IB disease were considered separately (Supplemental Figure 3). In particular, it was able to accurately predict prognosis in stage IA patients from both the IFOM and Duke cohorts. This is relevant, because the 5-year survival rate of stage IA NSCLC patients ranges from 67% to 77% (30–32) after surgery alone. Thus, in this group, molecular tools for prognostic prediction and patient stratification are greatly needed. On the other hand, the 10-gene model did not show pre-

dictive power on stage II–III adenocarcinomas (Supplemental Figure 2), possibly suggesting the existence of additional molecular mechanisms, occurring in more advanced lung carcinomas, that might influence the natural history of the tumor. Thus, the sum of our findings indicates that we have identified a prognostic signature specific for stage I lung adenocarcinoma.

One final question concerned the performance of our 10-gene model with respect to other prognostic models in NSCLC. Three prognostic models are described in the literature: a 5-gene model described by Chen et al. (33), and 50- and 100-gene models described by Beer et al. (9). These models were challenged against our 10-gene model on the independent Duke cohort of patients with stage I disease. While all of the models displayed good predictive accuracy, the 10-gene model displayed an overall better performance in terms of accuracy, sensitivity, specificity, and positive and negative predictive values (Table 3).

A fourth model composed of 134 genes was previously described by Potti et al. (16). This 134-gene "lung metagenes" model is somewhat different from our 10-gene model and from the other models described above. It is composed of several metagenes that are used to partition the samples recursively into smaller groups and predict recurrence through binary classification-tree analysis (16). Consequently, a direct comparison of our 10-gene model with the 134-metagene model on an independent cohort was unfeasible. However, when we compared the prognostic power of the 134-metagene model (Table 3) as described by the authors (16), we found that its overall performance was similar to that of our model, which uses only 10 genes.

Discussion
When we embarked upon the analysis of stage I lung adenocarcinomas with our integrated approach, we did so with the prospect of identifying more stable signatures than those obtained by unbiased profiling alone. As a consequence, we hoped to isolate a prognostic predictor that was of limited size and amenable to reduction into practice with technologies commonly available in
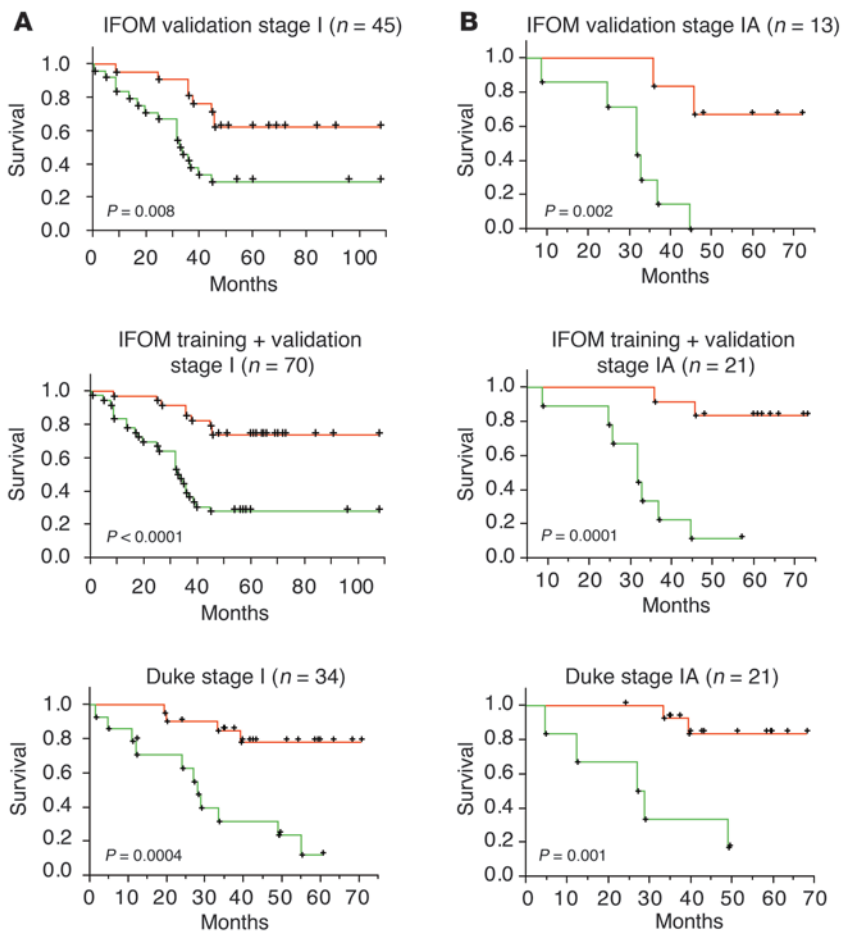
Figure 4
The 10-gene model predicts overall survival. The 10-gene model was tested to predict overall survival in the indicated cohorts of stage I (**A**) and stage IA (**B**) lung adenocarcinomas. Data are shown as the probability of survival, in a Kaplan-Meier plot, as a function of a favorable (red line) or unfavorable (green line) signature.

power of our biased approach in revealing gene expression programs not otherwise easily identifiable. Finally, the composition of the 10-gene prognostic model confirmed the hypothesized efficacy of our integrated approach, as it consisted of 5 genes derived from the E1A signature (*NUDCD1*, *E2F1*, *MCM6*, *RRM2*, and *SF3B1*), 4 genes from the meta-analysis of microarray data (*HOXB7*, *SERPINB5*, *E2F4*, and *HSPG2*), and 1 gene (*SCGB3A1*) from the literature.

We note that the 10-gene model, when tested by real-time PCR or on Affymetrix datasets (IFOM and Duke cohorts, respectively) in a total of 104 patients, displayed a predictive power of about 75% (79% for the Duke cohort, 71% for the IFOM cohorts). This indicates the existence of a subgroup of patients for whom our model is not predictive. One possible explanation for this is that the existence of genetically distinct subtypes of lung adenocarcinoma (10, 12) might have exerted a limited negative effect on the accuracy of our prognostic model, as well as of other published models (9, 16, 33).

An alternative explanation is that, because of limitations in the availability of TaqMan Low-Density Arrays at the time we performed the experiments, we could only analyze 53 of 80 genes of our 80-gene model to derive the final model. Thus, it will be interesting to extend our validation to the remaining genes in an attempt to achieve an even higher level of accuracy.

There are potential clinical implications of our findings. On the one hand, they provide prognostic tools. On the other, they might identify targets for molecular therapies or help to direct the choice of therapeutic regime. With respect to the first issue, one possible application of our model is patient stratification for the experimental testing of multimodality therapies. For instance, the disappointing results of adjuvant chemotherapy in stage I NSCLC might derive, at least in part, from the lack of reliable criteria for patient stratification, given the rather high rate of cure by surgery alone. Our 10-gene model might provide such a tool. While this will require large-scale studies, the feasibility of such studies should be greatly enhanced by the characteristics of our signature and our methodology of analysis. We note that a similar approach has been recently proposed by Chen et al., who identified a 5-gene real-time PCR signature predictive of clinical outcome in NSCLC patients (33). We compared the performance of our signature with that of Chen et al. on the Duke cohort, which constitutes an independent dataset for both signatures, and found that our signature performed significantly better. We ascribe this to the fact that we specifically focused on adenocarcinomas of the lung, both in our meta-analysis and in subsequent validations of the integrated

the clinical laboratory. Indeed, we identified a 10-gene signature that, by real-time PCR technology, predicted prognosis and overall survival of patients with stage I lung adenocarcinomas. Furthermore, the 10-gene model appeared robust enough to withstand validation across different technological platforms, as shown by its predictive power on the Duke dataset, which was generated by Affymetrix technology.

The major difference between our approach and the widely employed, more traditional ones (9, 11, 13, 14, 18, 20) was our merging biased and unbiased signatures. A biased approach to cancer transcriptomes relies on the assumption that a limited number of altered signaling pathways leads to the development of a malignant state. Thus, molecular tools that cause the transformation of genetically uniform cells in vitro might be used to circumvent the problems connected with unbiased transcriptome analysis, such as high individual genetic noise. We previously validated such an approach by showing that a signature obtained in myotubes that were forced to reenter the cell cycle by the E1A oncogene contained genes overexpressed in human tumors that could predict an unfavorable prognosis in breast cancer (25). In the present study, we showed that the same biased signature could, by itself, predict clinical outcome in patients with stage I lung adenocarcinomas. It is of note that the 28-gene biased signature did not emerge from the analysis of lung cancer microarray datasets, either in this study or in other studies (9, 11, 13, 14, 18, 20, 33) — the only exception is *SYNCRIP*, reported in a recent study (16). This confirms the

Table 3

Comparison of the prognostic predictive accuracy of several prognostic models

Model	Source	Acc[A]	Sens[B]	Spec[C]	PPV[D]	NPV[E]
Duke cohort (*n* = 34)						
10-gene	IFOM	79	73	84	79	80
5-gene	(33)	65	67	63	59	71
50-gene	(9)	74	80	68	67	81
100-gene	(9)	76	80	74	71	82
ACOSOG cohort (*n* = 25)						
134-gene	(16)	72	85	58	69	78
CALGB cohort (*n* = 84)						
134-gene	(16)	79	68	88	79	80
MAYO cohort (*n* = 15)						
134-gene	(16)	80	NA	NA	NA	NA

The 10-gene model was tested for prognostic predictive accuracy and compared with 3 other models of 5, 50 and 100 genes reported previously (9, 33). The models were tested on the Duke cohort, because it represents an independent dataset for all models and because all the genes of the 4 models were present on the microarray platform used to generate the datasets of the Duke cohort. A fifth, previously reported prognostic model (134 genes; ref. 16) is shown for comparison. Because immediate comparison was impossible, the predictive accuracy of the model is reported as described in the original paper on 3 different cohorts of patients. NA, not available. [A]Accuracy (percentage of correctly predicted patients). [B]Sensitivity (probability that a patient with poor prognosis will be predicted as having a poor prognosis). [C]Specificity (probability that a patient with good prognosis will be predicted as having a good prognosis). [D]Positive predictive value (probability that a sample predicted as poor prognosis actually belongs to the poor prognosis group). [E]Negative predictive value (probability that a sample predicted as good prognosis actually belongs to the good prognosis group).

model, while Chen et al. derived their signature from a global analysis of major subtypes of NSCLC (both squamous carcinomas and adenocarcinomas). Squamous carcinomas and adenocarcinomas are distinct disease entities (34–36) with different gene expression patterns (10, 12). Therefore, although unique prognostic signatures for NSCLC are more attractive because of their wider applicability, using independent prognostic signatures for squamous carcinomas and adenocarcinomas may be more biologically significant and less influenced by genetic heterogeneity.

An additional implication of our findings concerns their potential exploitation for the identification of novel therapeutic targets; this issue is directly linked to whether genes of our 10-gene model are relevant to lung carcinogenesis. Of course, no immediate biological implications can be derived from expression profile analysis. Nevertheless, we found it remarkable that, for many genes of our signature, there is ample literature indicating their association with cancer. For instance, the E2F family of genes, of which *E2F1* and *E2F4* are present in the signature, are crucial regulators of cell-cycle progression and have been implicated in numerous types of cancer (37, 38). *MCM6* is part of a family of proteins essential for the formation of the prereplicative complex in G1 and for DNA replication in S phase. Deregulation of the MCM complex results in chromosomal defects and is frequently detected in cancer (39, 40). *SERPINB5* (*MASPIN*) is a serine protease inhibitor, with multifaceted functions in the regulation of cell adhesion, motility, apoptosis, angiogenesis, and development, that is being actively studied for its potential usefulness as a diag-

nostic cancer marker and a therapeutic target (41). Other genes of our signature, such as *SCGB3A1* (*HIN-1*) or *NUDCD1* (*CML-66*), have been reported to be altered in cancer (28, 42). While a comprehensive review of the relevant literature will be impossible here, it is clear that our signature identifies a set of genes whose involvement in cancer is already established or strongly suspected and that might constitute attractive targets for molecular therapies.

In this context, one of the most interesting genes is probably *RRM2*, which encodes the small subunit of the ribonucleotide reductase holoenzyme (RNR), an essential enzyme for DNA synthesis. The other subunit of RNR is encoded by the *RRM1* gene, whose mRNA levels correlate with shorter survival in gemcitabine/cisplatin-treated advanced NSCLC patients (43–46). Interestingly, *RRM2* is also associated with resistance to a variety of chemotherapeutic agents, including gemcitabine (47, 48). In pancreatic adenocarcinomas, direct targeting of *RRM2* by siRNA enhanced chemosensitivity to gemcitabine both in vitro and in vivo (49). In NSCLC, gemcitabine is one of the most active chemotherapeutic agents (50). Our present findings establish that *RRM2* is overexpressed in stage I lung adenocarcinomas and is part of a poor-prognosis signature in these tumors. Thus, *RRM2* might constitute an appealing target for molecular therapies, also with the perspective of increasing responsiveness to certain drugs. At minimum, our findings should caution against the use of drugs such as gemcitabine in clinical trials in stage I lung adenocarcinoma exhibiting the poor-prognosis signature described here.

Methods

Meta-analysis of expression datasets from microarray analysis. The original datasets included 86 adenocarcinomas analyzed by Beer et al. (Michigan cohort; ref. 9) and 84 adenocarcinomas analyzed by Bhattacharjee et al. (Harvard cohort; ref. 10). The 84 adenocarcinomas of the Harvard cohort correspond to those of the Michigan cohort, for the validation of their prognostic signature. Microarray expression datasets of the Michigan and Harvard cohorts (obtained on the HU6800 and HU95Av2 Affymetrix chips, respectively) and details of patient selection criteria and methods for data normalization are shown by Beer et al. (ref. 9 and http://dot.ped.med.umich.edu:2000/ourimage/pub/Lung/index.html).

We gathered microarray expression datasets of a study by Bild et al. obtained on the HU133 2.0–plus Affymetrix chip (Duke cohort; ref. 23). Affymetrix CEL format files were processed using Microarray Suite version 5 software (MAS 5; Affymetrix).

Analyses of gene expression data were performed using BRB ArrayTools version 3.3.1 (http://linus.nci.nih.gov/BRB-ArrayTools.html). Microarray spot intensities below the minimum value of 10 (the BRB software default for Affymetrix array analysis) were excluded, and arrays were then normalized (centered) using the median value of the signal over the entire array. When we derived the 49-gene signature for the Michigan and Harvard reduced datasets, genes were excluded if less than 20% of their expression data across the patients had at least a 1.5-fold change in either direction from the gene's median value. Genes were also excluded if the percentage of data missing or filtered out exceeded 75%. All data were log-transformed (base 2). The 2-sample parametric Student's *t* test was used to select significant genes.

Class prediction analyses were performed with diagonal linear discriminant analysis. We estimated the prediction efficiency of the classifiers using the leave-one-out cross-validation, and the *P* value of each classifier was evaluated by 2,000 random permutations of the patient class labels.

Because data obtained with different microarray platforms were used in our study, the Affymetrix NetAffx web analysis tool was used to match probesets to identical genes.

Criteria for the selection of the reduced Harvard and Michigan datasets. To perform the meta-analysis on the Harvard and Michigan cohorts, we had to divide patients into good- and poor-prognosis groups according to their clinical outcome. We initially established the cutoff according to an international adjuvant lung cancer trial (IALT) that set the median disease-free survival of NSCLC patients at 30.5 months (51), which resulted in the exclusion of a number of patients. In addition, the analyses performed in this study stringently required size-balanced classes of patients. Indeed, imbalance in class size can increase the number of false positives and false negatives (52–55). Thus, we had to modify the cutoff in some cases, as described below.

In the Harvard cohort, we labeled as good prognosis those patients who were alive and had a follow-up of at least 30 months. By this criterion, the good-prognosis group of the Harvard reduced dataset contained 33 patients. Three patients who were labeled alive in the Harvard study were excluded because their follow-up was less than 30 months. We labeled as poor prognosis those patients who died before the 30-month cutoff. By this criterion, the poor-prognosis group of the Harvard reduced dataset contained 27 patients. Twenty-one patients who were labeled dead in the Harvard study were excluded because their death event occurred after 30 months. In this case it was impossible to include these patients in the poor-prognosis group because they died after the cutoff; at the same time, we reasoned that it was unwise to include them in the good-prognosis group because they died. By using these criteria, the Harvard reduced dataset contained 33 good-prognosis and 27 poor-prognosis patients, thus fulfilling the requirement of balance in class size.

In the Michigan cohort, by applying the above parameters, the 2 classes were remarkably unbalanced (48 good-prognosis versus 18 poor-prognosis patients). Thus, while retaining the cutoff for the poor-prognosis group (death event before the 30-month cutoff), we had to change the cutoff for the good-prognosis group. In this case, we labeled as good prognosis those patients who were alive and with a follow-up of at least 50 months. By doing this, the reduced Michigan dataset resulted in 23 good-prognosis patients and 18 poor-prognosis patients, thereby obtaining 2 balanced classes.

A relevant question is whether the criteria used for the cutoffs introduced bias to the analysis and/or somehow led to overfitting of the data. This is excluded by the facts that the 49-gene model displayed good predictive power when applied to all stage I adenocarcinomas for the Harvard and Michigan cohorts, without any selection of patients (Figure 2), as well as when applied to the original datasets (Supplemental Table 2 and Supplemental Figure 2). In addition, the 49-gene model had good prognostic value when applied to a third independent cohort (Duke cohort; Figure 2).

In addition, as shown in Supplemental Table 2, we performed an additional control by changing the cutoff values, so as to include more patients in the meta-analysis. This led to the identification of a 71-gene signature (which shared 23 genes with the 49-gene model); however, this did not perform better than the 49-gene model when applied to the original Michigan and Harvard datasets and to the Duke cohort.

Patients and the IFOM cohorts. Clinicopathological data for all patient groups in the present study are in Supplemental Table 1.

Patients within the IFOM cohorts were selected within a consecutive series of 391 stage I (T1-2N0M0) NSCLC patients surgically treated at the Division of Thoracic Surgery, University of Pisa (Pisa, Italy), between 1994 and 1999. Patient stage at the time of diagnosis was determined according to guidelines of the American Joint Committee on Cancer. The 70 patients selected for this study (25 and 45 for the training and validation cohorts, respectively) were selected solely on the basis of histotype (adenocarcinoma), availability of adequate tissue samples (>80% tumor cellularity), and complete follow-up data. Informed consent was obtained from all patients under study. Tumors were snap-frozen in liquid nitrogen within 10 minutes of excision and stored at –80°C. Total RNA was isolated with TRIZOL

(Invitrogen) according to the manufacturer's instructions, and its quality was evaluated by gel electrophoresis and by 2100 Bioanalyzer (Agilent).

Criteria for selection of candidate prognostic genes from the literature. The purpose of our study was to identify a signature, at the mRNA level, amenable to straightforward reduction into practice by a technology of easy access in the clinical laboratory. Thus, we selected *SCGB3A1* (*HIN-1*), *EIF3S6* (*INT-6*), and *hTERT* because these genes were studied at the mRNA level in stage I NSCLC by real-time PCR and proved to have prognostic value (26–28). Several other prognostic markers were proposed for NSCLC (56–59). However, most of these markers were proposed based on studies at the protein level (essentially by immunohistochemistry), which, for our purposes, would first require their independent validation as prognostic markers at the mRNA level and then their validation as potential members of a prognostic model on independent cohorts of patients. Thus, these markers were not considered for the purpose of this study. Similarly, we did not consider *EGFR*, which has been reported to be mutated in NSCLC, because the prognostic value of *EGFR* overexpression in untreated NSCLC remains controversial (60, 61).

TaqMan Low-Density Array analysis. TaqMan Low-Density Arrays were purchased from Applied Biosystems. Total RNA (0.5 μg) was reverse transcribed with 200 U Superscript II RT (Invitrogen) and 250 ng random hexamers according to the manufacturer's instructions. A reaction mix containing 75 ng of cDNA and 50 μl of 2× PCR Master Mix (Euregentec) in a final volume of 100 μl was then prepared and loaded in the array. PCR conditions were as follows: 2 min at 50°C, 10 min at 94.5°C, followed by 45 cycles at 97°C for 30 s and 59.7°C for 1 min, on an Applied Biosystems 7900HT PCR System.

The expression level of each gene was measured in triplicate, and a panel of 8 reference genes (*RPL14*, *RPL18*, *AGPAT1*, *ACTB*, *TBP*, *GUSB*, *PPIA*, and *18S*) was used. GeNorm software (62) was used to evaluate the expression stability of the reference genes. The average Ct value of each target gene was normalized against the geometric mean of the Ct values of the 8 reference genes. Universal Reference RNA (Stratagene) was used as calibrator for all the samples analyzed. The relative fold change of gene expression in lung cancer patients was calculated as $2^{-\Delta\Delta Ct}$, where $\Delta\Delta Ct$ represents $\Delta Ct_{sample} - \Delta Ct_{univ-ref}$.

Data were then analyzed using BRB ArrayTools version 3.3.1. Definition of the classifier was performed with the diagonal linear discriminant analysis and leave-one-out cross-validation. The *P* value was calculated by 2,000 random permutations of the class labels. The 45 patients of the IFOM validation cohort were labeled as "predict" in BRB Array-Tools to perform a completely blind classification of the class labels (good and poor outcome). TaqMan assay IDs (Applied Biosystems) were as follows: NUDCD1-Hs00292614_m1, CXCL6-Hs00237017_m1, E2F1-Hs00153451_m1, E2F4-Hs00608098_m1, GABPB2-Hs00242573_m1, HLA-DQB1-Hs00409790_m1, HOXB7-Hs00270131_m1, HSPG2-Hs00194179_m1, MCM4-Hs00381533_m1, MCM6-Hs00195504_m1, MCM7-Hs00428518_m1, RAFTLIN-Hs00412084_m1, RRM2-Hs00357247_g1, SCGB3A1-Hs00369360_g1, SERPINB5-Hs00184728_m1, and SF3B1-Hs00202782_m1.

Statistics. Univariate and multivariate analyses were performed using the Nominal Logistic Regression tool within JMP IN software (version 5.1; SAS). The *P* values were calculated with the likelihood-ratio χ^2 test. Kaplan-Meier survival curves were generated using JMP IN version 5.1 and were based on the diagonal linear discriminant analysis classification results. Kaplan-Meier associated *P* values were computed with the log-rank test. *P* values of less than or equal to 0.05 were considered significant.

1. Parkin, D.M., Bray, F., Ferlay, J., and Pisani, P. 2005. Global cancer statistics, 2002. *CA Cancer J. Clin.* **55**:74–108.

2. Jemal, A., Siegel, R., Ward, E., Murray, T., Xu, J., Smigal, C., and Thun, M.J. 2006. Cancer statistics, 2006. *CA Cancer J. Clin.* **56**:106–130.

3. Mountain, C.F. 1997. Revisions in the International System for Staging Lung Cancer. *Chest.* **111**:1710–1717.

4. Kato, H., et al. 2004. A randomized trial of adjuvant chemotherapy with uracil-tegafur for adenocarcinoma of the lung. *N. Engl. J. Med.* **350**:1713–1721.

5. Strauss, G.M., et al. 2006. Adjuvant chemotherapy in stage IB non-small cell lung cancer (NSCLC): Update of Cancer and Leukemia Group B (CALGB) protocol 9633. In *ASCO Annual Meeting Proceedings Part I.* June 20, 2006. *J. Clin. Oncol.* **24**(Suppl. 18S):7007.

6. Winton, T., et al. 2005. Vinorelbine plus cisplatin vs. observation in resected non-small-cell lung cancer. *N. Engl. J. Med.* **352**:2589–2597.

7. Betticher, D.C. 2005. Adjuvant and neoadjuvant chemotherapy in NSCLC: a paradigm shift. *Lung Cancer.* **50**(Suppl. 2):S9–S16.

8. Le Chevalier, T., Arriagada, R., Pignon, J.P., and Scagliotti, G.V. 2005. Should adjuvant chemotherapy become standard treatment in all patients with resected non-small-cell lung cancer? *Lancet Oncol.* **6**:182–184.

9. Beer, D.G., et al. 2002. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat. Med.* **8**:816–824.

10. Bhattacharjee, A., et al. 2001. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl. Acad. Sci. U. S. A.* **98**:13790–13795.

11. Endoh, H., et al. 2004. Prognostic model of pulmonary adenocarcinoma by expression profiling of eight genes as determined by quantitative real-time reverse transcriptase polymerase chain reaction. *J. Clin. Oncol.* **22**:811–819.

12. Garber, M.E., et al. 2001. Diversity of gene expression in adenocarcinoma of the lung. *Proc. Natl. Acad. Sci. U. S. A.* **98**:13784–13789.

13. Jiang, H., et al. 2004. Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. *BMC Bioinformatics.* **5**:81.

14. Lu, Y., et al. 2006. A gene expression signature predicts survival of patients with stage I non-small cell lung cancer. *PLoS Med.* **3**:e467.

15. Miura, K., et al. 2002. Laser capture microdissection and microarray expression analysis of lung adenocarcinoma reveals tobacco smoking- and prognosis-related molecular profiles. *Cancer Res.* **62**:3244–3250.

16. Potti, A., et al. 2006. A genomic strategy to refine prognosis in early-stage non-small-cell lung cancer. *N. Engl. J. Med.* **355**:570–580.

17. Ramaswamy, S., Ross, K.N., Lander, E.S., and Golub, T.R. 2003. A molecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33**:49–54.

18. Raponi, M., et al. 2006. Gene expression signatures for predicting prognosis of squamous cell and adenocarcinomas of the lung. *Cancer Res.* **66**:7466–7472.

19. Takeuchi, T., et al. 2006. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. *J. Clin. Oncol.* **24**:1679–1688.

20. Wigle, D.A., et al. 2002. Molecular profiling of non-small cell lung cancer and correlation with disease-free survival. *Cancer Res.* **62**:3005–3008.

21. Yanagisawa, K., et al. 2003. Proteomic patterns of tumour subsets in non-small-cell lung cancer. *Lancet.* **362**:433–439.

22. Huang, E., et al. 2003. Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat. Genet.* **34**:226–230.

23. Bild, A.H., et al. 2006. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature.* **439**:353–357.

24. Sweet-Cordero, A., et al. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat. Genet.* **37**:48–55.

25. Nicassio, F., et al. 2005. A cancer-specific transcriptional signature in human neoplasia. *J. Clin. Invest.* **115**:3015–3025. doi:10.1172/JCI24862.

26. Buttitta, F., et al. 2005. Int6 expression can predict survival in early-stage non-small cell lung cancer patients. *Clin. Cancer Res.* **11**:3198–3204.

27. Wang, L., et al. 2002. hTERT expression is a prognostic factor of survival in patients with stage I non-small cell lung cancer. *Clin. Cancer Res.* **8**:2883–2889.

28. Marchetti, A., et al. 2004. Down regulation of high in normal-1 (HIN-1) is a frequent event in stage I non-small cell lung cancer and correlates with poor clinical outcome. *Clin. Cancer Res.* **10**:1338–1343.

29. Ein-Dor, L., Zuk, O., and Domany, E. 2006. Thousands of samples are needed to generate a robust gene list for predicting outcome in cancer. *Proc. Natl. Acad. Sci. U. S. A.* **103**:5923–5928.

30. Wada, H., Miyahara, R., Tanaka, F., and Hitomi, S. 1999. Postoperative adjuvant chemotherapy with PVM (Cisplatin + Vindesine + Mitomycin C) and UFT (Uracil + Tegaful) in resected stage I-II NSCLC (non-small cell lung cancer): a randomized clinical trial. West Japan Study Group for lung cancer surgery (WJSG). *Eur. J. Cardiothorac. Surg.* **15**:438–443.

31. Suzuki, K., et al. 1999. Prognostic factors in clinical stage I non-small cell lung cancer. *Ann. Thorac. Surg.* **67**:927–932.

32. Sakao, Y., Nakazono, T., Sakuragi, T., Natsuaki, M., and Itoh, T. 2004. Predictive factors for survival in surgically resected clinical IA peripheral adenocarcinoma of the lung. *Ann. Thorac. Surg.* **77**:1157–1161; discussion 1161–1152.

33. Chen, H.Y., et al. 2007. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N. Engl. J. Med.* **356**:11–20.

34. Travis, W.D., Colby, T.V., Corrin, B., Shimosato, Y., and Brambilla, E. 1999. *Histological typing of lung and pleural tumors.* Springer. Berlin, Germany. 156 pp.

35. Travis, W.D., Travis, L.B., and Devesa, S.S. 1995. Lung cancer. *Cancer.* **75**:191–202.

36. Yakut, T., et al. 2006. Assessment of molecular events in squamous and non-squamous cell lung carcinoma. *Lung Cancer.* **54**:293–301.

37. Attwooll, C., Lazzerini Denchi, E., and Helin, K. 2004. The E2F family: specific functions and overlapping interests. *EMBO J.* **23**:4709–4716.

38. Seville, L.L., Shah, N., Westwell, A.D., and Chan, W.C. 2005. Modulation of pRB/E2F functions in the regulation of cell cycle and in cancer. *Curr. Cancer Drug Targets.* **5**:159–170.

39. Blow, J.J., and Dutta, A. 2005. Preventing re-replication of chromosomal DNA. *Nat. Rev. Mol. Cell Biol.* **6**:476–486.

40. Lei, M. 2005. The MCM complex: its role in DNA replication and implications for cancer therapy. *Curr. Cancer Drug Targets.* **5**:365–380.

41. Khalkhali-Ellis, Z. 2006. Maspin: the new frontier. *Clin. Cancer Res.* **12**:7279–7283.

42. Yang, X.F., et al. 2001. CML66, a broadly immunogenic tumor antigen, elicits a humoral immune response associated with remission of chronic myelogenous leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **98**:7492–7497.

43. Bepler, G., et al. 2005. Ribonucleotide reductase M1 gene promoter activity, polymorphisms, population frequencies, and clinical relevance. *Lung Cancer.* **47**:183–192.

44. Rosell, R., et al. 2003. Targeted therapy in combination with gemcitabine in non-small cell lung cancer. *Semin. Oncol.* **30**:19–25.

45. Rosell, R., et al. 2004. Ribonucleotide reductase messenger RNA expression and survival in gemcitabine/cisplatin-treated advanced non-small cell lung cancer patients. *Clin. Cancer Res.* **10**:1318–1325.

46. Rosell, R., et al. 2003. Transcripts in pretreatment biopsies from a three-arm randomized trial in metastatic non-small-cell lung cancer. *Oncogene.* **22**:3548–3553.

47. Zhou, B.S., Hsu, N.Y., Pan, B.C., Doroshow, J.H., and Yen, Y. 1995. Overexpression of ribonucleotide reductase in transfected human KB cells increases their resistance to hydroxyurea: M2 but not M1 is sufficient to increase resistance to hydroxyurea in transfected cells. *Cancer Res.* **55**:1328–1333.

48. Goan, Y.G., Zhou, B., Hu, E., Mi, S., and Yen, Y. 1999. Overexpression of ribonucleotide reductase as a mechanism of resistance to 2,2-difluorodeoxycytidine in the human KB cancer cell line. *Cancer Res.* **59**:4204–4207.

49. Duxbury, M.S., Ito, H., Zinner, M.J., Ashley, S.W., and Whang, E.E. 2004. RNA interference targeting the M2 subunit of ribonucleotide reductase enhances pancreatic adenocarcinoma chemosensitivity to gemcitabine. *Oncogene.* **23**:1539–1548.

50. Natale, R. 2005. A ten-year review of progress in the treatment of non-small-cell lung cancer with gemcitabine. *Lung Cancer.* **50**(Suppl. 1):S2–S4.

51. Le Chevalier, T., and Lynch, T. 2004. Adjuvant treatment of lung cancer: current status and potential applications of new regimens. *Lung Cancer.* **46**(Suppl. 2):S33–S39.

52. Japkowicz, N. 2000. The class imbalance problem: significance and strategies. In *Proceedings of the 2000 International Conference on Artificial Intelligence.* June 26–29. Las Vegas, Nevada, USA. H. Arabnia, editor. CSREA Press. Las Vegas, Nevada, USA. 111–117.

53. Japkowicz, N., and Stephen, S. 2002. The class imbalance problem: a systematic study. *Intelligent Data Analysis.* **6**:429–450.

54. Yang, K., Li, J., and Gao, H. 2006. The impact of sample imbalance on identifying differentially expressed genes. *BMC Bioinformatics.* **7**(Suppl. 4):S8.

55. Pawitan, Y., Michiels, S., Koscielny, S., Gusnanto, A., and Ploner, A. 2005. False discovery rate, sensitivity and sample size for microarray studies. *Bioinformatics.* **21**:3017–3024.

56. Singhal, S., et al. 2005. Prognostic implications of cell cycle, apoptosis, and angiogenesis biomarkers in non-small cell lung cancer: a review. *Clin. Cancer*

Res. **11**:3974–3986.

57. Han, J.Y., Choi, B.G., Choi, J.Y., Lee, S.Y., and Ju, S.Y. 2006. The prognostic significance of pretreatment plasma levels of insulin-like growth factor (IGF)-1, IGF-2, and IGF binding protein-3 in patients with advanced non-small cell lung cancer. *Lung Cancer.* **54**:227–234.

58. Masuya, D., et al. 2004. The tumour-stromal interaction between intratumoral c-Met and stromal hepatocyte growth factor associated with tumour growth and prognosis in non-small-cell lung cancer patients. *Br. J. Cancer.* **90**:1555–1562.

59. Tokunou, M., et al. 2001. c-MET expression in myofibroblasts: role in autocrine activation and prognostic significance in lung adenocarcinoma. *Am. J. Pathol.* **158**:1451–1463.

60. Hirsch, F.R., et al. 2003. Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J. Clin. Oncol.* **21**:3798–3807.

61. Jeon, Y.K., et al. 2006. Clinicopathologic features and prognostic implications of epidermal growth factor receptor (EGFR) gene copy number and protein expression in non-small cell lung cancer. *Lung Cancer.* **54**:387–398.

62. Vandesompele, J., et al. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3**:research0034.1–research0034.11.