# Constructing Molecular Classifiers for the Accurate Prognosis of Lung Adenocarcinoma

Lan Guo,[1] Yan Ma,[2] Rebecca Ward,[3] Vince Castranova,[4] Xianglin Shi,[4] and Yong Qian[4]

**Abstract** **Purpose:** Individualized therapy of lung adenocarcinoma depends on the accurate classification of patients into subgroups of poor and good prognosis, which reflects a different probability of disease recurrence and survival following therapy. However, it is currently impossible to reliably identify specific high-risk patients. Here, we propose a computational model system which accurately predicts the clinical outcome of individual patients based on their gene expression profiles.

**Experimental Design:** Gene signatures were selected using feature selection algorithms *random forests, correlation-based feature selection*, and *gain ratio attribute selection*. Prediction models were built using *random committee* and *Bayesian belief networks*. The prognostic power of the survival predictors was also evaluated using hierarchical cluster analysis and Kaplan-Meier analysis.

**Results:** The predictive accuracy of an identified 37-gene survival signature is 0.96 as measured by the area under the time-dependent receiver operating curves. The cluster analysis, using the 37-gene signature, aggregates the patient samples into three groups with distinct prognoses (Kaplan-Meier analysis, $P < 0.0005$, log-rank test). All patients in cluster 1 were in stage I, with $N_0$ lymph node status (no metastasis) and smaller tumor size ($T_1$ or $T_2$). Additionally, a 12-gene signature correctly predicts the stage of 94.2% of patients.

**Conclusions:** Our results show that the prediction models based on the expression levels of a small number of marker genes could accurately predict patient outcome for individualized therapy of lung adenocarcinoma. Such an individualized treatment may significantly increase survival due to the optimization of treatment procedures and improve lung cancer survival every year through the 5-year checkpoint.

Lung cancer is one of the most aggressive cancer types and is the leading cause of cancer-related deaths in industrialized countries. There are an estimated 172,570 new cases and 163,510 deaths from lung cancer (small cell and non–small cell combined) in the U.S. in 2005 (1). About 25% to 30% of patients with non–small cell lung cancer have stage I disease and receive surgical intervention alone. However, 35% to 50% of patients with stage I non–small cell lung cancers will relapse within 5 years (2–4). It is currently impossible to accurately identify specific high-risk patients for individualized therapy.

Recent advances in our knowledge of human genomics and proteomics, as well as artificial intelligence, have revolutionized the ways in which researchers are able to identify new prognostic factors of human cancer. It has become increasingly promising to use molecular profiles to predict the outcome of lung cancer diseases (5–17). However, most approaches use statistical tests, such as *t* test or Cox analysis, to identify biomarkers. Such approaches ignore the intrinsic interactions among genes or proteins. In addition, most approaches use only Kaplan-Meier analysis based on historical population data to evaluate the prognostic power of identified biomarkers. This leads to weak predictive power for individual patients. It remains a challenge to construct molecular classifiers that achieve high prediction accuracy for individual patients.

Previously, Beer et al. (5) evaluated a group of the top 50 survival marker genes (and top 100 survival genes) to identify high-risk patients with lung adenocarcinoma. Bhattacharjee et al. (6) identified the top 175 genes for lung adenocarcinoma subclassification. Jiang et al. (18) identified 16 survival marker genes based on the data sets from the previous publications (5, 6). However, the prognostic power of these gene signatures

**Authors' Affiliations:** Mary Babb Randolph Cancer Center, Departments of [1]Community Medicine and [2]Statistics, [3]Biomedical Sciences Graduate Program, Health Science Center, West Virginia University, and [4]The Pathology and Physiology Research Branch, Health Effects Laboratory Division, National Institute for Occupational Safety and Health, Morgantown, West Virginia

Received 10/27/05; revised 3/22/06; accepted 3/27/06.

**Grant support:** NIH/NCRR P20 RR16440-03 (Dr. L. Guo) and NIH/NCI 1R01CA119028-01 (Dr. X. Shi).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (http://clincancerres.aacrjournals.org/).

The findings and conclusions in this report are those of the author(s) and do not necessarily represent the views of the National Institute for Occupational Safety and Health.

**Requests for reprints:** Lan Guo, 1814 HSS, Mary Babb Randolph Cancer Center, P.O. Box 9300, Morgantown, WV 26506-9300. Phone: 304-293-6455; Fax: 304-293-4667; E-mail: lguo@hsc.wvu.edu, or Yong Qian, Pathology and Physiology Research Branch, Health Effects Laboratory Division, National Institute for Occupational Safety and Health, 1095 Willowdale Road, Morgantown, WV 26505-2888. Phone: 304-285-6286; Fax: 304-285-5938; E-mail: yaq2@cdc.gov.

©2006 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-05-2336



Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

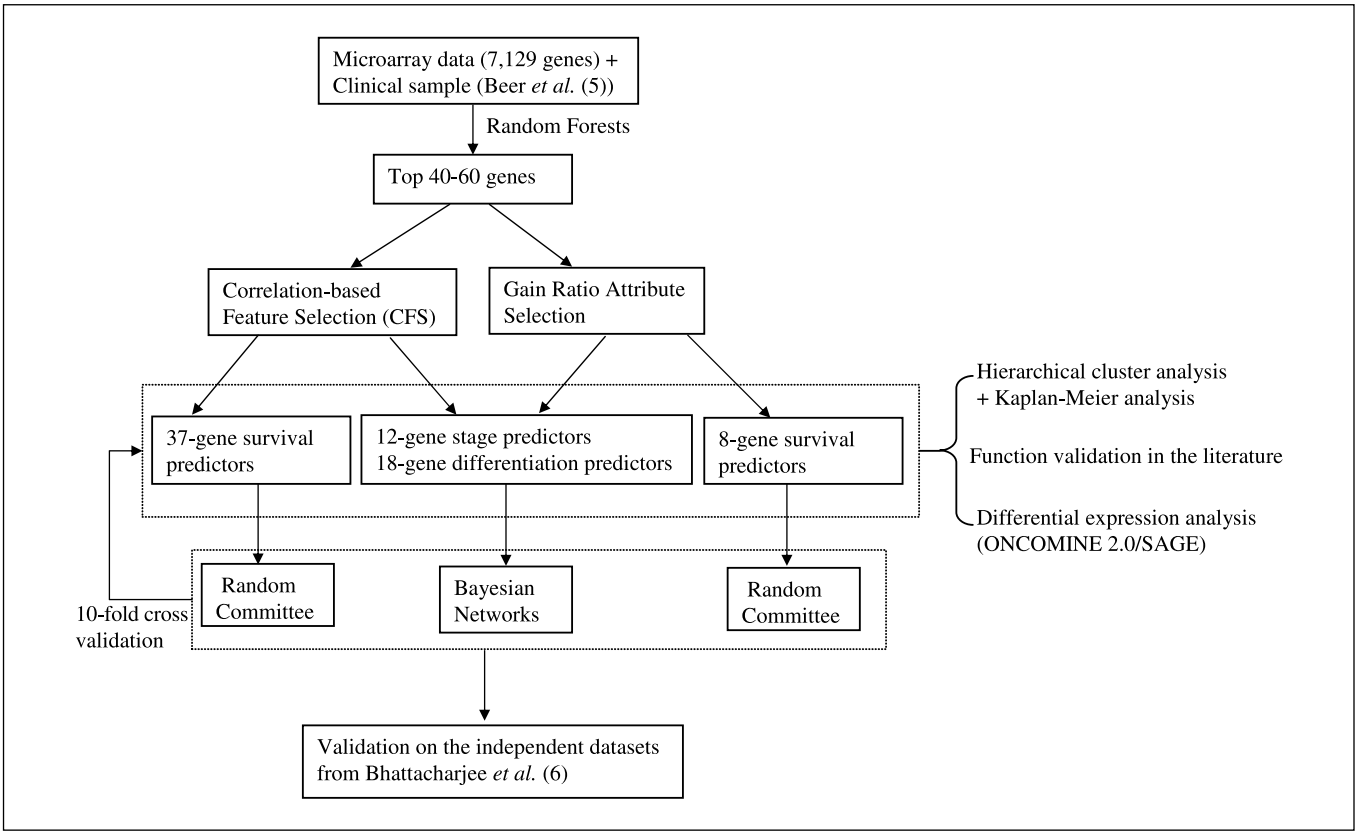


**Fig. 1.** The proposed model system for constructing molecular classifiers to identify important marker genes and predict clinical outcome for patients with lung adenocarcinoma.

for individual patients was not reported in their studies (5, 6, 18). Therefore, their reports are not directly applicable to devising individualized therapeutic strategies for specific high-risk patients.

In this article, we present our model system to identify important marker genes, which could improve the prognosis for individual patients with lung adenocarcinoma. We used several standard feature selection algorithms, random forests (19), correlation-based feature selection, and gain ratio attribute evaluation (20, 21), to identify novel molecular signatures with regard to the interactions among genes. We identified a 37-gene signature (38 probes), with a 5-year survival prediction accuracy of 0.96 as measured by the area under curve (AUC) of the time-dependent receiver operating curve (ROC) on the data sets from Beer et al. (5). The cluster analysis, using the 37-gene signature, aggregated the patient samples into three groups with distinct prognoses using Kaplan-Meier analysis ($P < 0.0005$, log-rank test). All patients in cluster 1 were in stage I, with $N_0$ lymph node status (no metastasis) and smaller tumor size ($T_1$ or $T_2$). In addition, we defined a 12-gene signature, which correctly predicted the stage of lung adenocarcinoma for 94.2% of patients. Furthermore, an identified 18-gene signature accurately predicted tumor differentiation (poor, moderate, and well) for 83.7% of lung adenocarcinoma patients. Our prediction models accurately predicted the clinical outcome for individual patients with lung adenocarcinoma. The differential expression analysis of the identified marker genes across cancer microarray data of various tissues and cell lines implies that the identified gene signatures are potentially novel biomarkers and therapeutics targets (12).

## Materials and Methods

***Clinical samples.*** Two independent data sets of clinical samples were used for building and validating the prediction models. The original gene expression profiles of patient samples were reported in previous publications (5, 6). The clinical information of patient samples and associated gene expression data files can be found online for the training set of 86 lung adenocarcinomas (5)[5] and the validation set of 84 adenocarcinomas (6).[6]

***Feature selection algorithms in gene shaving.*** Feature selection algorithms, random forests (19) in software package R,[7] correlation-based feature selection (20) and gain ratio attribute selection (22) in software package WEKA 3.4 (20, 21),[8] were used for signature discovery. A random forest is an ensemble of hundreds or thousands of classification trees. The final decision of the random forest is obtained using majority voting based on the results from these classification trees. The random forest algorithm ranks the variable importance in terms of the contribution to the predictive accuracy. The correlation-based feature selection algorithm identifies a good feature subset that contains features highly correlated with the class but is uncorrelated with each other. The gain ratio algorithm ranks a feature

[5] http://dot.ped.med.umich.edu:2000/ourimage/pub/Lung/index.html.
[6] http://dot.ped.med.umich.edu:2000/ourimage/microarrays/Lung_Affy/Harvard/idex.html.
[7] http://www.r-project.org/.
[8] http://www.cs.waikato.ac.nz/ml/weka/.



Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

based on the gained information by using this attribute in a classification rule. The random forest algorithm was used on the original training data set (5) to select the top 40 to 60 genes. The correlation-based feature selection and gain ratio algorithms were used to further refine the gene signatures.

*Machine-learning classifiers.* Two well-known machine-learning algorithms in software package WEKA 3.4 were employed to build our prediction models and molecular classifiers. The random committee algorithm is a diverse ensemble of random tree classifiers. In the case of classification, the random committee algorithm generates predictions by averaging probability estimates over these classification trees. The Bayesian Belief Networks (BBNs) are computational structures of acyclic graph. Nodes in the network structure represent propositions interrelated by links signifying causal relationships among the nodes. The BBNs are based on a sound mathematical theory of Bayesian probability. Specifically, the random committee algorithm was used to construct survival prediction models, and the BBNs were used to predict tumor stage and differentiation.

*Time-dependent ROC curves and AUC.* To evaluate the predictive performance of the proposed survival gene signatures, we employed time-dependent ROC analysis for censored data and AUC as our criteria to assess the 5-year survival predictions (23). The time-dependent sensitivity and specificity functions are defined as:

$$\text{sensitivity}(c, t) = P\{X > c \mid D(t) = 1\}$$

$$\text{specificity}(c, t) = P\{X \leq c \mid D(t) = 0\}$$

The corresponding ROC($t$) curve for any time $t$ is defined as the plot of {sensitivity($c$, $t$)} versus {1 − specificity($c$, $t$)}, with cutoff point $c$

**Table 1.** Information for the 37 survival genes grouped into two clusters

| Genes | Probe set | Cluster ID | Function (Unigene comment) | Classification |
|---|---|---|---|---|
| *CREB3* | AF009368_at | 1 | Transcriptional factor | Transcription |
| ***NULL*** | **AFFX.BioDn.3_at** | 1 | | |
| *NULL* | AFFX.DapX.5_at | 1 | | |
| *RER1* | AJ001421_at | 1 | Endoplasmic reticulum membrane proteins | Structure |
| *FCN2* | D63160_at | 1 | Innate immunity | Immunity |
| *NP220* | D83032_at | 1 | DNA binding protein: packaging, transferring, or processing transcripts | Transcription |
| *FBRNP* | HG1078.HT1078_at | 1 | Pseudogene | |
| *NULL* | HG180.HT180_at | 1 | | |
| *LBC* | HG2167.HT2237_at | 1 | Scaffolding protein for rho and PKA signaling | Signaling transduction |
| ***MSX2*** | **HG3729.HT3999_f_at** | 1 | Transformation suppressor genes | Signaling transduction |
| *TAL2* | HG4068.HT4338_at | 1 | Tcell leukemogenesis, brain development | Oncogene |
| *GHRHR* | L01406_at | 1 | Growth factor receptor, cancer development | Oncogene |
| *MT3* | M93311_at | 1 | Bind to heavy metals | Oncogene |
| *TNFSF9* | U03398_at | 1 | Tumor necrosis factor family | Oncogene |
| ***NULL*** | **U82313_at** | 1 | | |
| *ARHGDIG* | U82532_at | 1 | Cell signaling protein | Signaling transduction |
| ***TUBA3*** | **X01703_at** | 1 | Encode microtubules | Structure |
| *EGF* | X04571_at | 1 | Growth factor | Oncogene |
| *HFL3* | X64877_s.at | 1 | Unknown | |
| *FUT7* | X78031_at | 1 | Glycosylation | Metabolism |
| *GUCA2B* | Z70295_at | 1 | Endogenous activator of intestinal guanylate cyclase | Metabolism |
| *ATP5A1* | D14710_at | 2 | ATP synthesis | Metabolism |
| *EZFIT* | HG3565.HT3768_r_at | 2 | Regulate transcriptional control | Transcription |
| *UBE1* | M58028_at | 2 | Ubiquitin-activating protein | Protein degradation |
| ***ATRX*** | **U09820_s_at** | 2 | Transcriptional regulator | Transcription |
| ***ILF3*** | **U10324_at** | 2 | Transcriptional factor | Transcription |
| *E2F4* | U15641_s_at | 2 | Transcriptional factor, cell cycle apoptosis | Transcription, oncogene |
| *TAX1BP2* | U25801_at | 2 | Cellular transformation, gene activation | Oncogene |
| *UBE2I* | U45328_s_at | 2 | Ubiquitin-activating protein | Protein degradation |
| *ATRX* | U72935_cds3_s_at | 2 | Transcriptional regulator | Transcription |
| *OGT* | U77413_at | 2 | Glycosylation | Metabolism |
| *NULL* | U79256_at | 2 | | |
| *INSR* | X02160.at | 2 | Growth factor receptor: insulin receptor | Oncogene |
| *GNB1* | X04526_at | 2 | Cell signaling transduction | Signaling transduction |
| *NULL* | X57809_s_at | 2 | | |
| *CHD4* | X86691_at | 2 | Transcription regulator | Transcription |
| ***EMK1*** | **X97630_a_t** | 2 | Protein kinase | Signaling transduction |
| *HRMT1L2* | Y10807_s_at | 2 | Histone methyltransferase | Metabolism |

NOTE: Genes in boldface are also included in the eight-gene signature.



Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

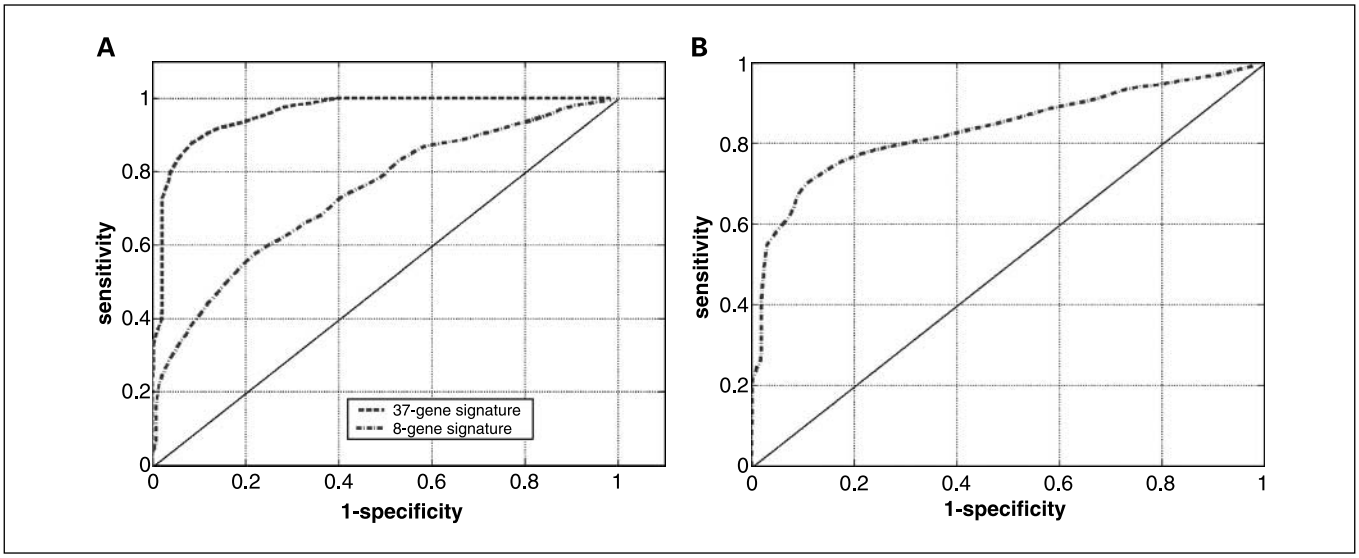


**Fig. 2.** Time-dependent ROC curves evaluating accuracy of survival prediction models. *A,* time-dependent ROC curves (5 years) for the 37-gene signature and the 8-gene signature on the data from Beer et al. (5). AUC (5 years) of the 37- and the 8-gene signature is 0.96 and 0.74, respectively. *B,* time-dependent ROC curves (5 years) for the 37-gene signature on the data from Bhattacharjee et al. (6). AUC of the 37-gene signature on the independent validation set is 0.835.

varying. $X$ is the covariate and $D(t)$ is the event indicator (here, death) at time $t$. The area under the curve, AUC($t$), is defined as the area under the ROC($t$) curve. A nearest neighbor estimator for the bivariate distribution function is used for estimating these conditional probabilities accounting for possible censoring (24). AUC can be used as an accuracy measure of the diagnostic marker; the larger the AUC, the better the prediction model. AUC = 0.5 indicates no predictive power, whereas AUC = 1 represents perfect predictive performance. The analysis was done using the software package $R$.

***Hierarchical cluster analysis.*** Hierarchical two-dimensional cluster analysis was done using identified survival marker genes on the 86 Michigan patient samples (5), using software package $R$. Similarity metrics were centered correlation, and the cluster method was complete linkage. The Silhouette validation method (25) implemented in $R$ was used to evaluate clustering validity and determine the number of clusters. A heat map was generated using Java Tree View.[9]

***Kaplan-Meier survival analysis.*** The Kaplan-Meier survival analysis was carried out using software package $R$. The statistical significance of the difference between the survival curves for different groups of patients was assessed, using likelihood ratio tests and log-rank tests.

***Differential expression analysis.*** ONCOMINE 2.0[10] (26, 27) is a cancer microarray database and a web-based data mining platform for cancer genomics research. The differential expression analysis module of ONCOMINE 2.0 was used to evaluate the expression patterns of identified marker genes across the lung cancer microarray data sets in the database. Results with statistical significance ($P < 0.05$, $t$ statistics) were included in the report.

The SAGE database[11] allows one to compare gene expression between normal tissues and cancer tissues (including cell lines), and between tumors of different histologic origin (28–33). We first used the SAGE Anatomic Viewer to identify the tissues in which a selected marker gene is differentially expressed between cancer and normal types. We then used the Digital Gene Expression Display tool to quantitatively analyze the differential expression of the marker gene in the cancer tissue and the normal tissue (including cell lines).

***Evaluating classifier accuracy.*** To assess the significance of individual classifier performance, we computed the probability of the observed prediction accuracy occurring by chance (random prediction using a fair coin flip). The probability of doing at least as well as our prediction models by chance was calculated using binomial distribution functions in software package $R$.

## Results and Discussion

***Architecture of the computational model system.*** The aim of the study was to identify important gene signatures with regard to the intrinsic complex interactions of the genes in lung adenocarcinoma. To provide an accurate prognosis for an individual patient, it is important to build prediction models using machine-learning algorithms, which are stable to the perturbations in the learning data. Based on these considerations, we constructed the model system consisting of three state-of-the-art supervised feature selection algorithms for gene signature discovery, two machine-learning classifiers for lung adenocarcinoma diagnosis and prognosis, hierarchical cluster analysis, and Kaplan-Meier analysis for the further validation of survival marker genes, and differential expression analysis for the elucidation of gene expression patterns (Fig. 1).

To identify important marker genes, we first used the random forests to select the top ranked 40 to 60 genes from the original microarray data sets, because the random forest algorithm is excellent and robust in processing large data sets which have a lot of noises (19). Then, the correlation-based feature selection (20) and the gain ratio attribute selection algorithms (22) were used to analyze the trimmed gene lists to further refine the gene signatures (Fig. 1). The correlation-based feature selection algorithm evaluates subsets of attributes instead of individual attributes. Thus, it is able to identify important attributes under moderate levels of interaction. When applied to gene signature discovery, the correlation-based feature selection algorithm could therefore reveal important marker genes with regard to their intrinsic interactions in lung adenocarcinoma. Gain ratio

[9] http://sourceforge.net/projects/jtreeview/.
[10] http://www.oncomine.org/main/index.jsp.
[11] http://cgap.nci.nih.gov/SAGE.



Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

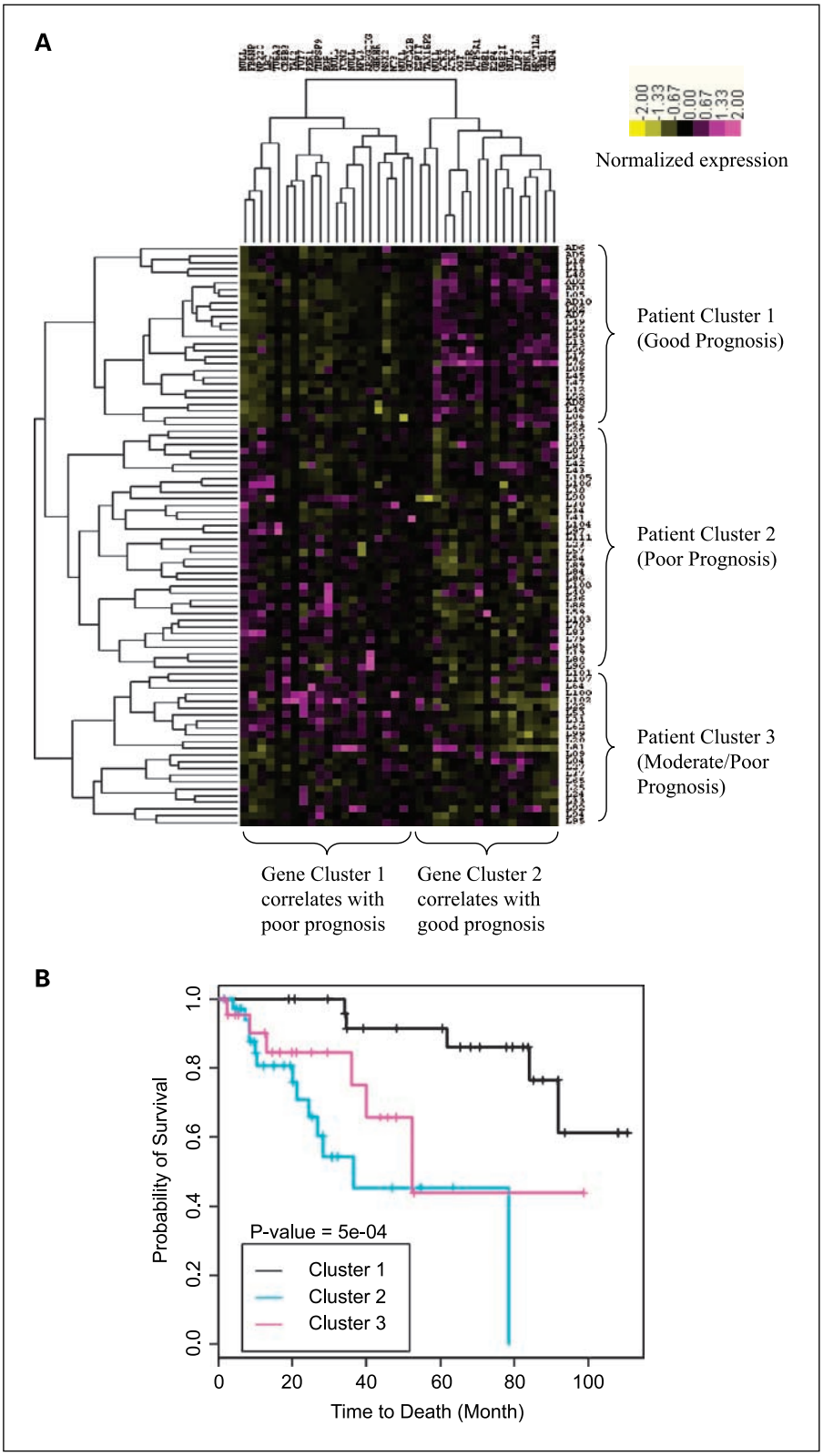


**Fig. 3.** Unsupervised hierarchical clustering analysis of the 37-gene signature on lung adenocarcinomas. *A,* heat map of gene expression patterns of the 37 survival marker genes determined using hierarchical clustering of the 86 lung adenocarcinomas from Beer et al. (5). *B,* Kaplan-Meier survival analysis of three clusters of patients. Average survival time of patients in cluster 1, 66.9 months; average survival time of patients in cluster 2, 22.4 months; average survival time of patients in cluster 3, 27.6 months ($P < 0.0005$, log-rank test).

attribute selection can rank the importance of individual attributes in the classification based on the measurement of the information requirements (22). In the selection of survival marker genes, the correlation-based feature selection and gain ratio algorithms were used separately to identify a 37- and 8-gene survival signature, respectively. The 8-gene signature is largely a subset of the 37-gene signature. Both the tumor stage predictors and tumor differentiation predictors were selected based on the overlapping results using the correlation-based feature selection and gain ratio algorithms (Fig. 1).



Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

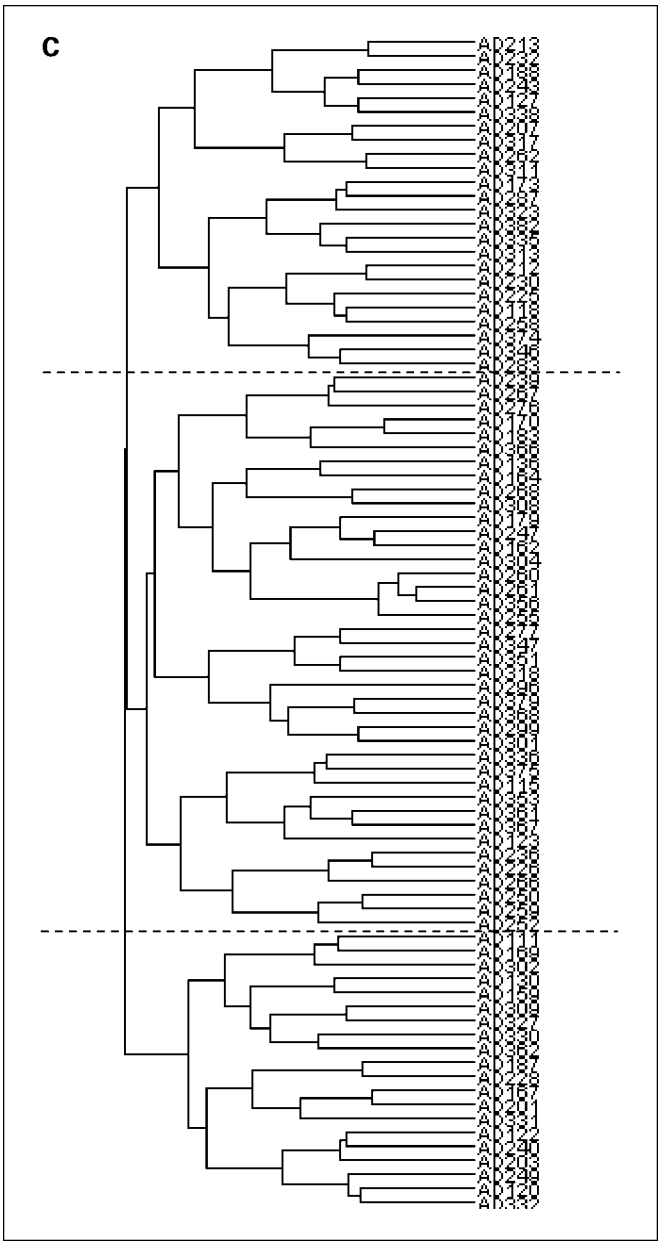


**Fig. 3** *Continued. C,* unsupervised cluster analysis of the 37-gene signature on the 84 lung adenocarcinomas from Bhattacharjee et al. (6). The tumor samples were aggregated into three clusters.

The supervised prediction models were built based on the Bayesian belief networks and the random committee algorithm (Fig. 1), which are accurate and stable. The random committee algorithm was used to build the survival prediction model, whereas the Bayesian belief networks were used to predict tumor stage and differentiation of lung adenocarcinoma. Ten-fold cross-validation was used to evaluate the prediction performance on the training set (5). We used 10-fold cross-validation to evaluate the prediction models, because the estimation accuracy by this validation method has been proven to have the lowest bias and variance among all validation methods, including the leave-one-out method (34). Thus, it provides an objective evaluation of the performance of our prediction models. The predictive power of the survival marker genes was further validated on the independent data sets from Bhattacharjee et al. (ref. 6; Fig. 1).

Unsupervised hierarchical two-dimensional cluster analysis was also used to study the expression patterns of the identified 37 survival marker genes in the 86 patients with lung adenocarcinoma (Fig. 1). The validity of the clustering was determined using the Silhouette validation method (25). The resulting clusters of patient samples were further analyzed using Kaplan-Meier survival analysis (Fig. 1). The unsupervised cluster analysis reveals the intrinsic gene expression patterns and provides useful information for the prognosis of lung adenocarcinoma.

The functionality of the identified marker genes was evaluated in extensive literature survey (Fig. 1). In addition, the differentiation expression of the identified genes regarding cancer classification, metastasis, stage, and differentiation was further analyzed across the lung cancer microarray data sets stored in the ONCOMINE database (refs. 26, 27; Fig. 1). The expression of all the identified marker genes in different tissues (including cell lines) between normal and cancer types was also analyzed using the SAGE techniques (Fig. 1).

Our results have shown that the proposed model system (Fig. 1) is able to identify novel gene signatures, which could be used to build prediction models for accurate prognosis of outcome for individual lung adenocarcinoma patients. The different sources of information and techniques used in this study quantitatively validate the expression patterns of the identified marker genes. Such expression patterns provide meaningful insights of gene functionality to guide future hypothesis-driven experimentation.

***Identification of genes associated with survival.*** The previous studies (5, 6, 18) did not report the prognostic power of their survival gene signatures on individual patients. To identify specific patients at high risk, we first selected important survival marker genes and then used these genes to predict patient survival after therapy. Using the random forest algorithm, a total of 7,129 genes were ranked by the variable importance measure provided in the random forest toolset. In this process, the random forest algorithm was also used as a classifier to evaluate the prediction results based on the selected genes. The prediction accuracy increased with the noise and after irrelevant genes were removed from the prediction model until 59 genes remained as predictors. To further refine the gene signatures based on these 59 genes, we used the feature selection algorithm, correlation-based feature selection, and the classifier, random committee, to identify the 37-gene signature (38 probes; Table 1). To evaluate the predictive power of the 37-gene survival signature, we employed the time-dependent ROC curve (5 years) for censored data (ref. 23; Fig. 2A). The AUC is 0.96, indicating that the 5-year survival prediction is highly accurate ($P < 1 \times 10^{-7}$) by using the 37-gene signature.

To evaluate the survival predictive accuracy of the 37-gene signature on independent data sets, we also validated the 37-gene prediction model on the data sets from Bhattacharjee et al. (6). The prediction accuracy of the 37-gene survival signature reached 0.835 ($P < 1 \times 10^{-7}$) as measured by the AUC of the time-dependent ROC curve (5 years) on the independent validation set (Fig. 2B). Noticeably, the gene expressions in the original training set (5) and the independent validation data set (6) were measured across two generations of the Affymetrix



Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

Genechip, HumanFL, and HumanGenome_U95Av2, which have significant variations in the measurements and the hybridization signal intensity (5, 18, 35). Although the gene expression profiles of the independent validation data sets (6) have been normalized for reproducibility (5), the discrepancy of the gene expression patterns due to different platforms cannot be totally ignored. Our results suggest that the accuracy of the 37-gene survival prediction model was reproducible using gene expression profiles across different platforms.

Using the gain ratio attribute selection algorithm, we further reduced the list of 59 genes to the top 8 genes. Seven genes in the 8-gene signature were also present in the 37-gene signature (boldface in Table 1), and the remaining gene is an unknown gene (with Microarray ID S73149_at). The predictive accuracy of the eight-gene signature is 0.74 ($P < 0.008$) as measured by the AUC of the time-dependent ROC curve (5 years; Fig. 2A). None of the identified survival marker genes in our 37- or 8-gene signatures was included in the previous 50-gene signature (5) or in the 16-gene signature (18) based on the same data sets (*TUBA1* was identified as one of the 16 survival marker genes; ref. 18). The results show that a smaller amount of genes identified using our model system can achieve higher prediction accuracy for individualized prognosis than the previously identified survival marker genes (5, 6, 18).

Among these 37 genes (Table 1), 8 genes are oncogenes, which include *TAL2*, *MT3*, *TNFSF9*, *GHRHR*, *THFSF*, *TAX1BP2*, *INSF*, and *EGF*. Five genes encode cell signaling proteins. Further analysis found that the *LBC* (lymphoid blast crisis) gene encodes a protein that is one of the antigens most identified in lung cancer (36) and the *MT3* gene encodes a protein that plays an important role in the destruction of lung tissue (37). Our results also include two well-established lung cancer prognostic factors, *TNFSF* and *EGF*, indicating that our results are consistent with the published observations. Interestingly, 8 of 37 genes encode either transcription factors or the protein products related to transcription.

In the hierarchical cluster analysis, 86 lung adenocarcinoma patients in the Michigan data set were aggregated into three clusters, whereas the 37 survival genes were grouped into two clusters (Fig. 3A). Genes in cluster 1 are correlated with a poor prognosis of lung adenocarcinoma, whereas genes in cluster 2 are correlated with a good prognosis for adenocarcinoma (Fig. 3A). In Fig. 3A, low relative expression of the genes in cluster 1 (Table 1) is associated with good prognosis in patients with adenocarcinoma, whereas high relative expression of the genes in cluster 1 is associated with poor prognosis in patients with adenocarcinoma. On the other hand, high relative expression of the genes in cluster 2 is associated with good prognosis in patients with adenocarcinoma, whereas low relative expression of the genes in cluster 2 is associated with poor prognosis in patients with adenocarcinoma. For patient sample clustering, all patients stratified to cluster I were in stage I, with $N_0$ lymph node status (no metastasis) and smaller tumor size ($T_1$ or $T_2$; Supplementary Table S1). Our results show that the identified 37-gene signature and their intrinsic interactions are important for the accurate prognosis of patients with lung adenocarcinoma.

Kaplan-Meier survival analysis showed that the survival time after therapy was significantly different in three patient clusters ($P < 0.0005$, log-rank test; Fig. 3B). Cluster 1 is the good prognosis group with less aggressive lung adenocarcinomas [all in stage I, with smaller tumor size ($T_1$ or $T_2$), and no metastasis]. Two-thirds of patients in cluster 1 survived a 5-year interval (Fig. 3B). It shows that, using the identified 37-gene signature, the patients could be stratified into clinically meaningful groups for further prognosis.

Hierarchical clustering of the 37-gene signature aggregated the 84 adenocarcinomas in the independent validation set (6) into three clusters (Fig. 3C). We examined the relationships between cluster and tumor characteristics (Supplementary Table S2). There were marginal associations between cluster and K-*ras* mutation ($P = 0.07$) and between cluster and differentiation ($P = 0.10$). Cluster 3 contained the greatest percentage (58.3%) of K-*ras* mutation followed by cluster 2 (33.3%) and cluster 1 (8.3%). Cluster 2 contained the greatest percentage (53.3%) of poorly differentiated tumors, followed by cluster 3 (40%), and cluster 1 (6.7%). The survival probability among the clusters was not statistically different using the Kaplan-Meier analysis. As mentioned before, the gene expression data in the validation set were generated from a different gene chip and were normalized to comparable scales for analysis in this study. The discrepancy of the gene expression patterns due to different platforms cannot be totally ignored. In addition, 76.2% of the patients in the validation set had recurrences during the follow up period. However, the Kaplan-Meier analysis cannot take this factor into account.

***Differential expression analysis of five survival genes.*** The 8-gene signature is largely a subset of the 37-gene signature (Table 1). It suggests that these overlapping genes selected using both feature selection algorithms are important survival marker genes. Within the eight-gene signature, the function of five of these genes are known. Three of them, *TUBA3* (38–40), *MSX2* (41, 42), and *ATRX* (43), have been found to be directly related to cancer development. The other two genes, *ILF3* (44) and *EMK1* (45), may be involved in tumor formation. No reports have shown that their expression is related to tumor formation.

We further analyzed the differential expression of *EMK1* (*MARK2*), *ILF3*, *TUBA3*, *ATRX*, and *MSX2* across the lung cancer microarray data sets stored in the ONCOMINE database. The analytic results show that the identified survival marker genes, *ILF3*, *TUBA3*, and *ATRX* was overexpressed in females on the data sets from Bhattacharjee et al. (6), which is consistent with the fact that adenocarcinoma is the most common cause of lung cancer in women. The differential expression analysis indicate that *ILF3*, *EMK1* (*MARK2*), *MSX2*, and *ATRX* were strongly related with metastasis. Overall, these five known genes were all differentially expressed in lung cancer versus normal lung, and were differentially expressed in subclasses of lung cancer with statistical significance on the previous data sets (refs. 5, 6, 9, 13–16, 46; see Supplementary Materials for details). Using the differential gene expression analysis, the function of some identified marker genes have been validated as directly related to cancer development in the literature, whereas other identified marker genes might be involved in tumor formation as suggested in our analytic studies.

***Marker genes to predict tumor stage.*** It currently remains an open problem to determine the stage of lung adenocarcinoma using quantitative and standardized models based on molecular profiles. To identify a gene signature to predict tumor stage, we used the random forests to select 49 genes out of 7,129



Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

genes from the Michigan data sets. The 49-gene list was further reduced to 12 genes that overlap in the results from the analysis using the correlation-based feature selection and gain ratio algorithms (Table 2). Based on the identified 12-gene tumor stage predictors, the prediction model using the Bayesian belief networks accurately predicted the stage of 94.2% lung adenocarcinoma patients, with a prediction accuracy of 98.5% (66 out of 67) for stage I and 78.9% (15 out of 19) for stage III. The errors in the 10-fold cross-validation of the stage prediction model were plotted in Fig. 4A. The output probability for each variable was computed by the Bayesian inference methods, with 0.5 as the cutoff probability in the final classification. One misclassified sample is close to the cutoff with an output probability 0.413, whereas the remaining three with output probability <0.25.

The 12-gene signature (Table 2) does not overlap with the 37- or the 8-gene survival signature (Table 1). The 12-gene predictors were not included in the marker genes identified in the previous studies (5, 18) on the same data sets. The results indicate that, for the first time, the tumor stage of lung adenocarcinoma could be determined by standardized and quantified measurement of the expression profiles of these unique marker genes.

Interestingly, functional analysis found that 4 out 12 genes are directly related to the human immune system. Both *D12S2489E* and *ELA2* gene products mediate natural killer

**Table 2.** Marker genes to predict the tumor stage and differentiation of lung adenocarcinoma

| Gene | Probe set | Function (Unigene comments) | Classification |
|---|---|---|---|
| Tumor stage predictors | | | |
| *D12S2489E* | X54870_at | Mediate natural killer cell killing | Immune response |
| *CD8B1* | X13444_at | Mediate T cell killing | Immune response |
| *NULL* | AFFX.CreX.5_at | | |
| *L1CAM* | U52112_rna1_at | Cell adhesion | Structure |
| *PDK2* | L42451_at | Inhibits the mitochondrial pyruvate dehydrogenase complex | Metabolism |
| *GBP2* | M55543_at | Regulate IFN | Immune response |
| *ELA2* | Y00477_at | Mediate natural killer cells, monocytes, and granulocytes killing | Immune response |
| *DIO2* | U53506_at | Activate thyroid hormone | Metabolism |
| *P63* | X69910_at | Activate thyroid hormone | Metabolism |
| *LYL1* | M22638_at | Involve in T cell acute lymphoblastic leukemia | Oncogene |
| *GPR6* | U18549_at | Cell signaling protein | Signaling transduction |
| *PRKCE* | X65293_at | Protein kinase | Oncogene |
| Tumor differentiation predictors | | | |
| *LGALS4* | AB006781_s_at | May be involved in cell adhesion | Metabolism |
| *KIAA0101* | D14657_at | May be relative to follicular lymphoma | Unknown |
| *FCGBP* | D84239_at | May be relative to follicular adenoma and a follicular carcinoma | Unknown |
| *PTPN13* | HG3187.HT3366_s_at | Apoptosis, protein phosphatase | Oncogene |
| *CRYM* | L02950_at | Cell development, binds thyroid hormone | Metabolism |
| *ADH1* | M12963_s_at | Alcohol dehydrogenase | Metabolism |
| *CCNB1* | M25753_at | Cell cycle | Oncogene |
| *IDUA* | M74715_s_at | Hydrolyzes the terminal α-L-iduronic acid residues of two glycosaminoglycans, dermatan sulfate and heparan sulfate | Metabolism |
| *LOC55969* | S83364_at | Unknown | Unknown |
| *CSPG2* | U16306_at | Cell growth and differentiation | Oncogene |
| *RAB27B* | U57093_at | Cell signaling protein | Signaling transduction |
| *PLOD2* | U84573_at | The component of collagen | Structure |
| *P40 (EBNA1BP2)* | U86602_at | Cell signaling protein | Transcription |
| *MTHFD2* | X16396_at | Bifunctional enzyme with methylenetetrahydrofolate dehydrogenase and methenyltetrahydrofolate cyclohydrolase activities | Metabolism |
| *ADE2H1* | X53793_at | Purine biosynthesis | Metabolism |
| *FMO2* | Y09267_at | Catalyzes the N-oxidation of certain primary alkylamines to their oximes | Metabolism |
| *RPC* | Y11651_at | Catalyzes the conversion of 3′-phosphate to a 2′,3′-cyclic phosphodiester at the end of RNA | Metabolism |
| *COL1A1* | Z74615_at | The major component of type I collagen | Structure |



Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

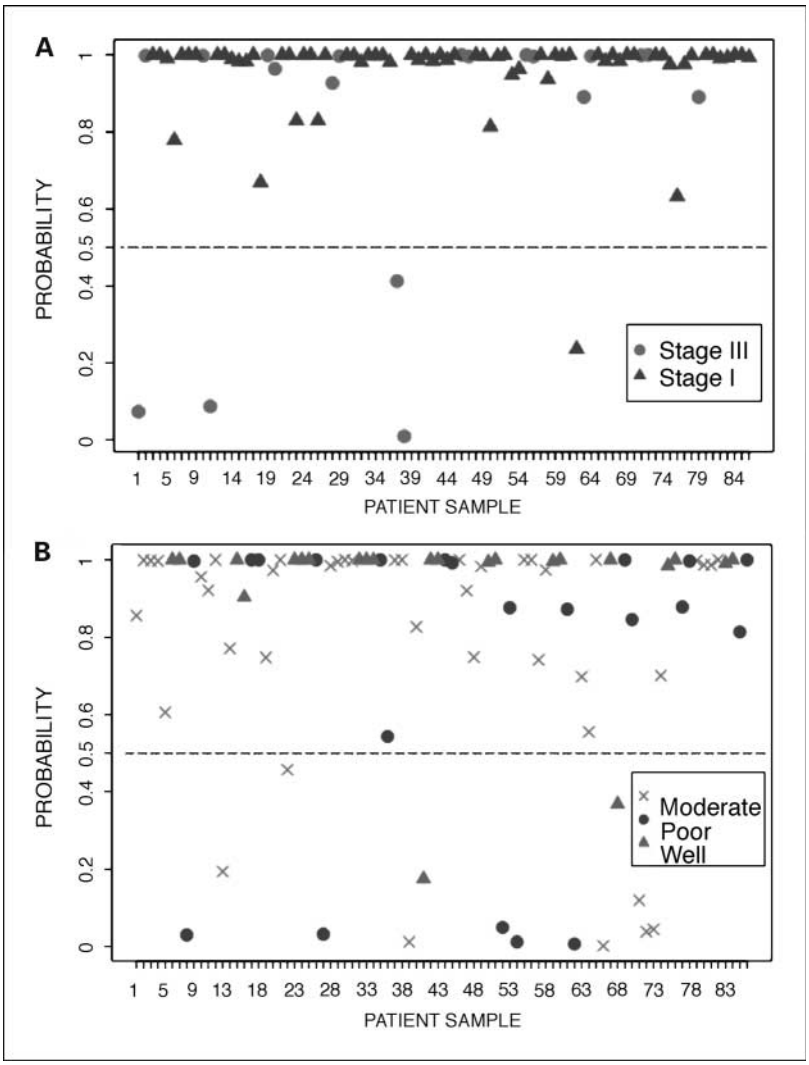


**Fig. 4.** Errors in 10-fold cross-validation of the prediction models. Misclassified samples are those with probability less than 0.5. *A,* errors in 10-fold cross-validation of the adenocarcinoma stage prediction model using the 12-gene signature. The total number of errors is 5 out of 86. *B,* errors in 10-fold cross-validation of the adenocarcinoma differentiation prediction model using the 18-gene signature. The total number of errors is 14 out of 86.

cells, *CD8B1* encodes proteins involved in mediating T cell killing, and GBP2 protein regulates IFN. The results indicate that the immune response system is critical in the progress of lung adenocarcinoma, which implies that the therapeutic strategies targeting the immune system could play an important role in altering lung adenocarcinoma development. Indeed, immunotherapy is currently undergoing clinical trials and may provide additional options for those patients with lung cancer who are resistant to current conventional therapies (47).

***Marker genes to predict tumor differentiation.*** The previous studies (5–7, 9, 10, 15, 18) have not addressed the preoperative determination of tumor differentiation of lung adenocarcinoma using molecular profiles. We sought to identify important tumor differentiation marker genes and employ them to predict tumor differentiation (poor, moderate, and well) of lung adenocarcinoma. From the computational point of view, it is a multiclassification problem, which is intrinsically more difficult than the binary classification problems addressed thus far (48).

To predict tumor differentiation, we used the random forests to identify the top 50 genes out of 7,129 genes from the Michigan data sets. The 50-gene list was further reduced to 18 genes (Table 2) that overlap in the results from the analysis using the correlation-based feature selection and gain ratio algorithms. Based on the identified 18-gene tumor differentiation predictors, the prediction model using the Bayesian belief networks accurately predicted the differentiation for 83.7% of lung adenocarcinoma patients. The prediction accuracy of well-differentiated tumors was 91.3% (21 out of 23), moderate differentiation 83.3% (35 out of 42), and poor differentiation 76.2% (16 out of 21). Among the misclassified samples, no well-differentiated tumor samples were misclassified as poor differentiation and vise versa. There was no overlap between the tumor differentiation predictors and the survival predictors (Table 1) or the tumor stage predictors identified in this study (Table 2). The 18-gene predictors were not included in the marker genes identified in previous studies (5, 18) on the same data sets. The results show that our identified marker genes are unique and capable of accurately predicting the tumor differentiation of lung adenocarcinomas. Ten-fold cross-validation results for the tumor differentiation prediction model were depicted in Fig. 4B. The cutoff probability is 0.5 in the classification. One misclassified sample is close to the cutoff with an output probability 0.457, whereas the remaining 13 with output probability was <0.40.



Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

Noticeably, several genes from this group are directly involved in cell differentiation. *PTPN13* is a proapoptotic protein tyrosine phosphatase, which is overexpressed in most cancer cells, and is involved in the regulation of cell differentiation (49). The expression pattern of *CCNB1* is markedly different among differentiated lung cancers (11). Interestingly, *CSPG2* is a target gene of *p53* that is a major regulator of cell differentiation and growth. *CSPG2* was found to be selectively induced and overexpressed in lung cancer and the knockdown of *CSPG2* significantly inhibited lung tumor growth *in vivo* (50).

***Differential expression of the identified marker genes between cancer and normal tissues.*** We have identified a 37-gene signature and an 8-gene signature (Table 1) to predict lung adenocarcinoma patient survival after therapy, a 12-gene signature (Table 2) to predict the stage of lung adenocarcinoma, and an 18-gene signature (Table 2) to predict the differentiation of lung adenocarcinoma. The function of these identified marker genes has been searched in extensive literature survey. To further validate the gene expression patterns in cancer and normal tissues, we used the SAGE database and its associated tools (28–33) to quantitatively analyze the differential expression of our identified marker genes between cancer and normal tissues.

The SAGE database encompasses over 200 libraries including solid tumor specimens and cell lines. Using the SAGE Anatomic Viewer, we found that the identified marker genes are differentially expressed between cancer and normal types in 13 tissues (including cell lines): such as lung, brain, thyroid, breast, stomach, liver, kidney, pancreas, colon, peritoneum, prostate, ovary, and skin. Furthermore, to quantitatively analyze the gene's differential expression patterns, we used the SAGE Digital Gene Expression Display tool to identify the genes that overexpress or underexpress at least 2-fold ($P < 0.05$) between cancer and normal tissues including cell lines. In many cases, the overexpression ratios of 5- to 100-fold were found (see Supplementary Materials for details). Our results imply that the identified marker genes are related to cancer development from various histologic origins, which provides possible directions for future cancer research.

In this study, we constructed a comprehensive molecular classification system for the accurate prognosis of lung adenocarcinoma. Our system provides prediction models based on the identified gene expression profiles not only for the identification of specific high-risk lung adenocarcinoma patients, but also for the prediction of the stage as well as the differentiation of lung adenocarcinoma. Random prediction is unlikely to achieve our prediction accuracy. The prediction accuracy of 5-year survival is 0.96 ($P < 1 \times 10^{-7}$) using the identified 37-gene signature. The correct classification of stage I and stage III lung adenocarcinoma is 94.2% ($P < 2.9 \times 10^{-20}$) using the 12-gene stage predictors. The prediction accuracy of lung adenocarcinoma differentiation reaches 83.7% ($P < 3.7 \times 10^{-23}$) using the 18-gene signature in the studied cohort. Our results showed that, using the proposed prediction models and the identified marker genes, we can accurately predict the clinical outcome of lung adenocarcinoma patients. The differential expression analysis of the identified marker genes shows that most genes are either overexpressed or underexpressed in tumor tissues with a variety of histologic origins. It implies that our identified marker genes are related to cancer development. Although some genes have not gained wide recognition, their differential expression patterns have reached statistical significance over various sources of data, which provides directions for future hypothesis-driven experimentation. The knowledge in this study is extracted from the public domain data, which is therefore repeatable and provides a consistent platform for the evaluation of future prognostic models.

## Acknowledgments

We thank Dr. Daniel C. Flynn for his thoughtful comments on the work, and Dr. Lexin Li at North Carolina State University, who provided valuable help and the source code for generating time-dependent ROC curves.

## References

1. American Cancer Society. Cancer facts and figures 2005. Atlanta (GA): American Cancer Society; 2005.
2. Naruke T, Goya T, Tsuchiya R, Suemasu K. Prognosis and survival in resected lung carcinoma based on the new international staging system. J Thorac Cardiovasc Surg 1988;96:440–7.
3. Pairolero PC, Williams DE, Bergstralh EJ, Piehler JM, Bernatz PE, Payne WS. Postsurgical stage I bronchogenic carcinoma: morbid implications of recurrent disease. Ann Thorac Surg 1984;38:331–8.
4. Williams DE, Pairolero PC, Davis CS, et al. Survival of patients surgically treated for stage I lung cancer. J Thorac Cardiovasc Surg 1981;82:70–6.
5. Beer DG, Kardia SL, Huang CC, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. Nat Med 2002;8:816–24.
6. 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.
7. 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.
8. Franklin WA, Carbone DP. Molecular staging and pharmacogenomics. Clinical implications: from lab to patients and back. Lung Cancer 2003;41 Suppl 1: S147–54.
9. 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.
10. Gordon GJ, Jensen RV, Hsiao LL, et al. Translation of microarray data into clinically relevant cancer diagnostic tests using gene expression ratios in lung cancer and mesothelioma. Cancer Res 2002;62:4963–7.
11. Kettunen E, Anttila S, Seppanen JK, et al. Differentially expressed genes in nonsmall cell lung cancer: expression profiling of cancer-related genes in squamous cell lung cancer. Cancer Genet Cytogenet 2004;149:98–106.
12. Muller-Hagen G, Beinert T, Sommer A. Aspects of lung cancer gene expression profiling. Curr Opin Drug Discov Devel 2004;7:290–303.
13. Powell CA, Spira A, Derti A, et al. Gene expression in lung adenocarcinomas of smokers and nonsmokers. Am J Respir Cell Mol Biol 2003;29:157–62.
14. Ramaswamy S, Tamayo P, Rifkin R, et al. Multiclass cancer diagnosis using tumor gene expression signatures. Proc Natl Acad Sci U S A 2001;98:15149–54.
15. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet 2003;33:49–54.
16. Su AI, Welsh JB, Sapinoso LM, et al. Molecular classification of human carcinomas by use of gene expression signatures. Cancer Res 2001;61:7388–93.
17. 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.
18. Jiang H, Deng Y, Chen HS, et al. Joint analysis of two microarray gene-expression data sets to select lung adenocarcinoma marker genes. BMC Bioinformatics 2004;5:81.
19. Breiman L. Random Forests. Mach Learn 2001;45:5–32.
20. Hall MA, Holmes G. Benchmarking attribute selection techniques for discrete class data mining. IEEE Trans Knowl Data Eng 2003;15:1437–47.
21. Witten IH, Frank E. Data mining: practical machine learning tools and techniques. 2nd ed. Morgan Kaufmann; 2005.
22. Quinlan JR. Induction of decision tree. Mach Learn 1986;1:81–106.
23. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. Biometrics 2000;56:337–44.



Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

**24.** Akritas MG. Nearest neighbor estimation of a bivariate distribution under random censoring. Annu Stat 1994;22:1299–327.

**25.** Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math 1987;20:53–63.

**26.** Rhodes DR, Yu J, Shanker K, et al. ONCOMINE: a cancer microarray database and integrated data-mining platform. Neoplasia 2004;6:1–6.

**27.** Rhodes DR, Yu J, Shanker K, et al. Large-scale meta-analysis of cancer microarray data identifies common transcriptional profiles of neoplastic transformation and progression. Proc Natl Acad Sci U S A 2004;101:9309–14.

**28.** Lal A, Lash AE, Altschul SF, et al. A public database for gene expression in human cancers. Cancer Res 1999;59:5403–7.

**29.** Nacht M, Ferguson AT, Zhang W, et al. Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. Cancer Res 1999;59:5464–70.

**30.** Stein WD, Litman T, Fojo T, Bates SE. A Serial Analysis of Gene Expression (SAGE) database analysis of chemosensitivity: comparing solid tumors with cell lines and comparing solid tumors from different tissue origins. Cancer Res 2004;64:2805–16.

**31.** Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. Serial analysis of gene expression. Science 1995;270: 484–7.

**32.** Velculescu VE, Madden SL, Zhang L, et al. Analysis of human transcriptomes. Nat Genet 1999;23: 387–8.

**33.** Zhang W, Laborde PM, Coombes KR, Berry DA, Hamilton SR. Cancer genomics: promises and complexities. Clin Cancer Res 2001;7:2159–67.

**34.** Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. International Joint Conference on Artificial Intelligence (IJCAI) 1995;1137–43.

**35.** Nimgaonkar A, Sanoudou D, Butte AJ, et al. Reproducibility of gene expression across generations of Affymetrix microarrays. BMC Bioinformatics 2003;4:27.

**36.** Diesinger I, Bauer C, Brass N, et al. Toward a more complete recognition of immunoreactive antigens in squamous cell lung carcinoma. Int J Cancer 2002; 102:372–8.

**37.** Matsui K, Takeda K, Yu ZX, Travis WD, Moss J, Ferrans VJ. Role for activation of matrix metalloproteinases in the pathogenesis of pulmonary lymphangioleiomyomatosis. Arch Pathol Lab Med 2000;124: 267–75.

**38.** Loganzo F, Hari M, Annable T, et al. Cells resistant to HTI-286 do not overexpress P-glycoprotein but have reduced drug accumulation and a point mutation in α-tubulin. Mol Cancer Ther 2004;3: 1319–27.

**39.** Poruchynsky MS, Kim JH, Nogales E, et al. Tumor cells resistant to a microtubule-depolymerizing hemiasterlin analogue, HTI-286, have mutations in α- or β-tubulin and increased microtubule stability. Biochemistry 2004;43:13944–54.

**40.** Seve P, Isaac S, Tredan O, et al. Expression of class III {β}-tubulin is predictive of patient outcome in patients with non-small cell lung cancer receiving vinorelbine-based chemotherapy. Clin Cancer Res 2005;11:5481–6.

**41.** Malewski T, Milewicz T, Krzysiek J, Gregoraszczuk EL, Augustowska K. Regulation of Msx2 gene expression by steroid hormones in human nonmalignant and malignant breast cancer explants cultured *in vitro*. Cancer Invest 2005;23:222–8.

**42.** Satoh K, Ginsburg E, Vonderhaar BK. Msx-1 and Msx-2 in mammary gland development. J Mammary Gland Biol Neoplasia 2004;9:195–205.

**43.** Steensma DP, Gibbons RJ, Higgs DR. Acquired α-thalassemia in association with myelodysplastic syndrome and other hematologic malignancies. Blood 2005;105:443–52.

**44.** Zhao G, Shi L, Qiu D, Hu H, Kao PN. NF45/ILF2 tissue expression, promoter analysis, and interleukin-2 transactivating function. Exp Cell Res 2005;305: 312–23.

**45.** Hueso M, Beltran V, Moreso F, et al. Splicing alterations in human renal allografts: detection of a new splice variant of protein kinase Par1/Emk1 whose expression is associated with an increase of inflammation in protocol biopsies of transplanted patients. Biochim Biophys Acta 2004;1689:58–65.

**46.** Hsiao LL, Dangond F, Yoshida T, et al. A compendium of gene expression in normal human tissues. Physiol Genomics 2001;7:97–104.

**47.** Swisher SG, Roth JA, Carbone DP. Genetic and immunologic therapies for lung cancer. Semin Oncol 2002;29:95–101.

**48.** Rifkin R, Mukherjee S, Tamayo P, et al. An analytical method for multi-class molecular cancer classification. SIAM Rev 2003;45:706–23.

**49.** Nedachi T, Conti M. Potential role of protein tyrosine phosphatase nonreceptor type 13 in the control of oocyte meiotic maturation. Development 2004;131: 4987–98.

**50.** Yoon H, Liyanarachchi S, Wright FA, et al. Gene expression profiling of isogenic cells with different TP53 gene dosage reveals numerous genes that are affected by TP53 dosage and identifies CSPG2 as a direct target of p53. Proc Natl Acad Sci U S A 2002; 99:15632–7.




Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.

American Association for Cancer Research

Clinical Cancer Research

# Constructing Molecular Classifiers for the Accurate Prognosis of Lung Adenocarcinoma

Lan Guo, Yan Ma, Rebecca Ward, et al.

*Clin Cancer Res* 2006;12:3344-3354.

| | |
|---|---|
| **Updated version** | Access the most recent version of this article at: http://clincancerres.aacrjournals.org/content/12/11/3344 |
| **Supplementary Material** | Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2006/06/05/12.11.3344.DC1 |

| | |
|---|---|
| **Cited articles** | This article cites 47 articles, 18 of which you can access for free at: http://clincancerres.aacrjournals.org/content/12/11/3344.full#ref-list-1 |
| **Citing articles** | This article has been cited by 11 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/12/11/3344.full#related-urls |

| | |
|---|---|
| **E-mail alerts** | Sign up to receive free email-alerts related to this article or journal. |
| **Reprints and Subscriptions** | To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org. |
| **Permissions** | To request permission to re-use all or part of this article, use this link http://clincancerres.aacrjournals.org/content/12/11/3344. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site. |

Downloaded from clincancerres.aacrjournals.org on March 11, 2019. © 2006 American Association for Cancer Research.