Correction and Retraction

CORRECTION

MEDICAL SCIENCES


The authors note, “We wish to bring to your attention an issue regarding our PNAS publication referenced above. The focus of the paper was the characterization of the signaling pathways of TTF-1, NKX2-8, and PAX9 in lung cancer using a transcriptome-based approach. As such, gene expression signatures representative of the activation of individual transcription factors (TFs), TTF-1, NKX2-8, or PAX9 were developed and were found to have prognostic implication in a data set of clinically annotated lung cancer specimens (Figs. 1–3).

“A second validation was then performed in an independent data set of lung cancer specimens. However, the source publication of the gene expression dataset (1) was retracting the paper’s authors in March 2011, calling into question the validity of the results using this dataset as described in Fig. 4, and with a subset of the data in Fig. S7. In addition, we have not been able to reproduce results reported in Fig. S6, which contains a subset of the data shown in Fig. 4. Given this finding, we believe that the affected figure in the paper (Fig. 4), as well as the data in Figs. S6 and S7 that report an analysis of a subset of the data in Fig. 4, are no longer valid.

“Finally, we also note that a specific portion of analysis (i.e., cisplatin sensitivity prediction) presented in Fig. 5 can no longer be supported, because the publication describing the cisplatin signature has been retracted (2). However, this issue does not alter the validity of the Fig. 5 data describing the unsupervised hierarchical clustering of 56 lung cancer cell lines using the three TF signatures (Fig. 5A Upper) and the experimental determination of cisplatin sensitivity (Fig. 5B Right). More importantly, the final conclusion of Fig. 5, which is that activation of the TF is correlated with sensitivity to cisplatin (Fig. 5B Right), remains unaltered. Importantly, we do not believe that this correction alters the primary conclusions of the paper.”


www.pnas.org/cgi/doi/10.1073/pnas.111138110

APPLIED PHYSICAL SCIENCES, BIOPHYSICS AND COMPUTATIONAL BIOLOGY


The authors wish to note the following: “We have determined that the conclusions stated in our paper are an artifact of errors of analysis. Our methods involved determining the rheological properties of lipid bilayers from analysis of the trajectories of lipid-anchored particles. Errors related to the particle-tracking methods we used generated an artifactual elastic-like signature.

“The imprecision in tracking resulted from determining the position of each particle by fitting the logarithm of the particle’s intensity profile to a quadratic function. Directly fitting the intensity to a Gaussian form is considerably less sensitive to noise and signal intensity. This improved fitting reduces positional uncertainty in test images from around 30 nm to a few nm and, applied to our membrane data, completely eliminates the ‘elastic’ response. Moreover, adding 30 nm of random error to precisely tracked images reproduces the reported elastic signature. Therefore, our data indicate that lipid membranes show a purely viscous character over the entire frequency range examined. Our initial assessments of particle tracking algorithms were not sufficiently stringent, leading to an overestimate of our tracking precision and the use of inadequate fitting methods.

“The artifactual elastic response we reported showed a feature at the chain ordering transition temperature of the lipid membranes examined. We are not certain what aspect of the images, when imprecisely tracked, led to this feature. We suspect that motions of the particles perpendicular to the membrane plane may be responsible.

“The conclusions of the paper are wrong and, regrettably, we must retract the work. We sincerely apologize for any confusion that our report may have caused.”

Christopher W. Harland
Miranda J. Bradley
Raghuveer Parthasarathy

www.pnas.org/cgi/doi/10.1073/pnas.1111961108
Characterizing the developmental pathways TTF-1, NKX2–8, and PAX9 in lung cancer

David S. Hsu,a,b Chaitanya R. Acharyaa,b Bala S. Balakumaran,a Richard F. Riedela,b Mickey K. Kim,a
Marvaretta Stevensonb,c Sascha Tuchmana,b Sayan Mukherjeea,c William Barrys,c Holly K. Dressma,a
Joseph R. Nevinsa,b Scott Powersd,e David Muif,1,2 and Anil Pottia,b,1,2

aInstitute for Genome Sciences and Policy, bDepartment of Medicine, and cInstitute for Statistics and Decision Sciences, Duke University, Durham, NC 27710; dCold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; eCancer Genome Center, Cold Spring Harbor Laboratory, Woodbury, NY 11724; and fDepartment of Pathology, Penn State University College of Medicine, Hershey, PA 17033

Communicated by Aziz Sancar, University of North Carolina, Chapel Hill, NC, January 26, 2009 (received for review November 17, 2008)

We investigated the clinical implications of lung developmental transcription factors (TTF-1, NKX2–8, and PAX9) that we recently discovered as cooperating oncogenes activated by way of gene amplification at chromosome 14q13 in lung cancer. Using stable transfectants of human bronchial epithelial cells, RNA expression profiles (signatures) representing activation of the biological pathways defined by each of the 3 genes were determined and used to risk stratify a non-small-cell lung cancer (NSCLC) clinical data set consisting of 91 early stage tumors. Coactivation of the TTF-1 and NKX2–8 pathways identified a cluster of patients with poor survival, representing ~20% of patients with early stage NSCLC, whereas activation of individual pathways did not reveal significant prognostic power. Importantly, the poor prognosis associated with coactivation of TTF-1 and NKX2–8 was validated in 2 other independent clinical data sets. Furthermore, lung cancer cell lines showing coactivation of the TTF-1 and NKX2–8 pathways were shown to exhibit resistance to cisplatin, the standard of care for the treatment of NSCLC. This suggests that the cohort of patients with coactivation of TTF-1 and NKX2–8 pathways appears to be resistant to standard cisplatin therapy, suggesting the need for alternative therapies in this cohort of high-risk patients.

gene expression | lung cancer | transcription factors

The complexity of the oncogenic process involving the somatic acquisition of large numbers of mutations, coupled with the varied host genetic constitution, produces a disease of enormous complexity. The ability to dissect this complexity by better understanding models of carcinogenesis, lineage dependence, and oncogene addiction is critical for developing new effective therapeutic strategies.

There is strong evidence for the involvement of developmental and cell lineage-dependent processes in the development of human tumors. For example, multiple genes (e.g., AML1, PML-RAR fusion gene, and PAX5) that are critical to hematopoietic progenitor cell development are frequently altered in hematologic malignancies (1). Similar examples exist in cell proliferation and differentiation events that lead to melanoma (2). In contrast to the oncogene addiction model that invokes a tumor-specific gain-of-function event (3), the lineage-survival oncogene model (4) proposes that tumors may become highly dependent on survival mechanisms that are built into the lineage precursor cells during development and that alteration of the genes involved in these survival mechanisms could provide selective advantage for evolving tumor cells.

Despite the large number of transcription factors involved in mammalian lung development, the role of these factors in lung cancer progression, particularly non-small-cell lung cancer (NSCLC), and their clinical relevance have not been comprehensively evaluated. Recently, we discovered and functionally characterized a novel, recurrent amplicon located at 14q13.3 that appears to harbor lung cell lineage-specific oncogenes due to its preferential occurrence in lung cancer and the role that these 3 genes play in lung development (5). High-resolution quantitative PCR mapping of the commonly amplified region revealed genes encoding 3 transcription factors in the 413-kb core region of the amplicon: TTF-1, NKX2–8, and PAX9. Both TTF-1 and NKX2–8 are known to mediate lung development and maturation (5, 6). Similarly, PAX9 is also likely involved in lung development as evident by its expression in fetal lung tissue and the diverse negative impact on organogenesis in knockout animals (7). Whereas overexpression of individual members of this 3-gene set in the core region of the amplicon was ineffective in stimulating proliferation, combinatorial overexpression of the genes led to significant proliferation of premalignant human bronchial epithelial cells, implicating oncogene cooperation for the 3 amplified lung development regulators. Furthermore, our reported RNAi knockdown experiments in concert with other reports establish that each of these 3 genes can have essential tumor maintenance function (5, 8, 9).

To investigate the impact of the 3 collaborating lung cell lineage oncogenes on clinical outcome, we adapted a validated strategy that utilizes gene expression profiles as a measure of the consequence of an activated oncogenic pathway, irrespective of how the signaling pathway might have been altered. Thus, even if the known oncogene is not mutated, but rather another component of the pathway is altered, the expression profile will likely still detect the biological alteration. Building upon this approach, here we describe a significant correlation of lung tumors with dual activation of the biological pathways directed by both TTF-1 and NKX2–8 with poor survival. Furthermore, we demonstrate that lung cancer cells possessing coactivated TTF-1 and NKX2–8 pathways exhibit cisplatin resistance, suggesting the need for alternative therapies in this cohort of high-risk patients.

Results

Gene Expression Signatures That Reflect the Activity of TTF-1, NKX2–8, and PAX9. Using gene expression levels of TTF-1 (Probe IDs 207771 Lair, 207772 s.at), NKX2–8 (Probe ID 207451 a.at), and PAX9 (207059 a.at), we first explored the possibility that individual gene expression levels of the developmental transcription factors may reveal correlations with clinical parameters. To test this hypothesis, we chose survival in patients with surgically resected stage I NSCLC. Supporting information (SI) Fig. S1 and Fig. S2 depict an analysis of transcription factor status in a cohort of NSCLC tumors (n = 91) that included comparable numbers of squamous cell


The authors declare no conflict of interest.

1D.M. and A.P. contributed equally to this work.

This article contains supporting information online at www.pnas.org/cgi/content/full/0900827106/DCSupplemental.
carcinoma and adenocarcinoma (GSE3141). Importantly, none of these patients had received adjuvant chemotherapy. Kaplan–Meier survival analysis shows that no statistically significant difference was detected between high or low expression level of NKX2–8, TTF-1, and PAX9 (Fig. S1) or when combinations of high or low expression level of NKX2–8/TTF-1 were examined (Fig. S2), suggesting that gene expression levels of these transcription factors hold no definitive prognostic implications.

As gene expression levels of NKX2–8, TTF-1, and PAX9 could not be used to identify a clinically relevant phenotype, we hypothesized that signatures representative of coordinated activation of NKX2–8, TTF-1, and PAX9 pathways may be more informative. Using oncogene-immortalized, premalignant lung epithelial cells (BEAS-2B), pathway signatures were developed that embodied the activation of TTF-1, NKX2–8, or PAX9. Individual genes were retrovirally transduced into BEAS-2B cells for stable expression and uninfected cells were eliminated by drug selection. Quantitative assays evaluating overexpression of TTF-1, NKX2–8, and PAX9 confirmed the biochemical activation of each gene (Fig. 1A).

Gene expression signatures represented by metagenes, which represent groups of genes that together exhibit a consistent pattern of expression in a collection of samples, were identified and correlated to the observable phenotype of activation of a transcription factor (e.g., TTF-1, NKX2–8, or PAX9), as described previously (10). This analysis selects a set of genes whose expression levels are most highly correlated with the classification of cell lines into oncogene activated/deregulated vs. control (Fig. 1B Upper). The dominant principal component from such a set of genes then defines a relevant phenotype-related metagene, and regression models assign the probability of TTF-1, NKX2–8, or PAX9 activation in an independent validation set of tumor or cell line samples.

Fig. 1B (Upper) illustrates the gene expression patterns that reflect the activity of the 3 transcription factors (TTF-1, NKX2–8, and PAX9) coded for by 3 coamplified genes in the 14q13.3 amplicon. Each individual signature distinguishes cells expressing the transcription factors from control cells (Fig. 1B Lower). As such, this result suggests that these gene expression signatures represent tools to accurately profile the biological pathway status defined by each transcription factor. Importantly, gene annotation of the individual gene expression profiles using GATHER (11) further demonstrated that the TTF-1, NKX2–8, and PAX9 signatures are composed of genes/ pathways that have previously been shown to be critical in both embryonal tissue development (JAK/STAT, Wnt, BMP, and Hedgehog) (12–15) and more specifically lung development (MAP kinase, PI3 kinase, and JAK/STAT) (Table S1) (15–17).

TTF-1, NKX2–8, and PAX9 Activation Correlates with Molecular and Pathological Phenotypes of NSCLC. Using robust predictors of activation of TTF-1, NKX2–8, and PAX9, we explored the possibility that signatures representative of deregulated developmental pathways may further dissect lung cancer phenotypes. Fig. 2A depicts an analysis of pathway status of the 3 transcription factors in a cohort of 91 NSCLC tumors described above (GSE3141). Following prediction of transcription factor activation probabilities, the resulting probabilities of a tumor having activation of TTF-1, NKX2–8, or PAX9 pathways in the NSCLC cohort were grouped using hierarchical clustering. Three distinct clusters were observed, with cluster 1 consisting of NKX2–8/TTF-1 coactivation, cluster 2 consisting of NKX2–8 activation, and cluster 3 consisting of PAX9/TTF-1 coactivation (Fig. 2A). Interestingly, PAX9, TTF-1, and NKX2–8 activation status correlates with the tissue subtype of non-small-cell lung cancer; the majority of the adenocarcinoma samples exhibit a high probability of PAX9 and TTF-1 activation (Fig. 2B) relative to the squamous cell carcinoma samples, in line with published reports that TTF-1 protein expression is higher in adenocarcinoma than in squamous cell carcinoma (18–21). In contrast to TTF-1 and PAX9, NKX2–8 deregulation is more prominent in squamous cell carcinomas. (Fig. 2B). Furthermore, an examination of KRAS and epidermal growth factor receptor (EGFR) mutational status in these tumors identified 11 of 91 (12%) samples with KRAS (codon 12) mutations (11/45, 24.4% of adenocarcinomas) and only a small number of samples, 6 of 91 (6.5%), contained EGFR mutations (exons 18, 19, and 21), all restricted to adenocarcinomas (Fig. 2A), consistent with current literature (22, 23). Interestingly, among the 2 largest clusters of patients, only 3 of 11 (27%) of KRAS mutations were confined to tumors with NKX2–8 and TTF-1 coactivation while 5 of 11 (45%) were confined to tumors with coactivated PAX9 and TTF-1, suggesting that the specific patterns of coactivation of developmental transcription factors (TTF-1, NKX2–8, and PAX9) are independent of either KRAS or EGFR mutational events.

We further explored the pathway coactivation of NKX2–8 and TTF-1 coactivation in a series of tumors in which KRAS was spontaneously activated by homologous recombination in adult mice (24). In these animal lung tumors, TTF-1 and NKX2–8 coactivation was observed in a subgroup, similar to the pattern of coactivation seen in human lung tumors. Importantly, not all of the KRAS-dependent murine tumors had evidence of TTF-1/NKX2–8 coactivation (Fig. S3). Taken together, these results further support the hypothesis that coactivation of the developmental pathways TTF-1 and NKX2–8 represents an important, potentially aggressive phenotype of NSCLC that is not defined just by the RAS pathway.

Hsu et al.

PNAS | March 31, 2009 | vol. 106 | no. 13 | 5313
**TTF-1 and NKX2–8 Coactivation and Survival Patterns in NSCLC.** Using survival as a clinically relevant phenotype, we investigated the extent to which analysis of the activation status of the developmental pathways delineates prognosis in lung cancer. Kaplan–Meier survival curves of the 3 independent clusters identified in Fig. 2A were analyzed independent of tumor histology (squamous vs. adenocarcinoma); concerted activation of TTF-1 and NKX2–8 target genes (cluster 1) identified a population of patients (22/91, 24%) with poor survival when compared to cluster 2 (activation of NKX2–8 only) (\( P = 0.002 \)) or cluster 3 (activation of PAX9 and TTF-1) (\( P = 0.02 \)) (Fig. 3). When subgroup analyses involving only activation of any one transcription factor were performed, no statistically significant survival difference was observed (\( P > 0.3 \)) (Fig. S4). Importantly, the poor prognosis cluster of patients represented by coactivation of both TTF-1 and NKX2–8 had a survival pattern similar to that of patients with metastatic NSCLC (Fig. S5, \( n = 37 \)), suggesting that the coactivation of developmental pathways TTF-1 and NKX2–8 in a group of early stage patients portends a poor overall survival that phenotypically mimics metastatic disease.

Further validation of the prognostic importance of cluster 1 (TTF-1, NKX2–8 coactivation) was performed in an independent pooled analysis based on tumor samples derived from 2 large independent data sets, 84 adenocarcinoma samples (GSE3593) and 130 squamous carcinoma samples (GSE4573) (25, 26). In this large validation cohort (\( n = 214 \)), coderegulation of TTF-1 and NKX2–8 pathways was also associated with the worst outcome, when compared to other combinations of transcription factor activation (clusters 2 and 3) (Fig. 4A and B). Importantly, a similar trend was seen when we repeated the analysis using either the GSE3593 or the GSE4573 data set alone (Fig. S6 and Fig. S7). As further confirmation, univariate and multivariate analyses (Table 1) revealed a statistically significant (univariate Cox proportional hazards model, \( P = 0.009 \)) prognostic implication of NKX2–8/TTF-1 coactivation (cluster 1), independent of traditional prognostic criteria such as tumor size, vascular invasion, lymph node status, or stage of disease (multivariate Cox proportional hazards model, \( P = 0.03 \)).

As such, it appears that the ability to integrate developmental pathway analysis by identifying patterns of coderegulation does provide a mechanism to better categorize patients with lung cancer. Importantly, this analysis demonstrates the strong association of TTF-1 and NKX2–8 coactivation in non-small-cell lung cancer with an adverse clinical outcome.

**TTF-1 and NKX2–8 Coactivation and Resistance to Cisplatin Therapy.** An ability to identify patients at high risk for an adverse outcome also emphasizes the need for better therapeutic strategies, partic-
usually in NSCLC. Platinum compounds constitute the backbone of combination chemotherapy in NSCLC, although only 30–40% of patients with NSCLC respond to cisplatin-based therapy (27). Using a previously validated gene expression signature of cisplatin resistance, we compared the correlation between the likelihood of TTF-1/NKX2–8 coactivation (the worst prognosis cohort) and cisplatin resistance in a cohort of 56 lung cancer cell lines. As shown in Fig. 5A, clusters similar to those seen in patient samples were also seen in this large cohort of NSCLC cell lines. More importantly, PAX9 activation was associated with cisplatin sensitivity while coactivation of TTF-1 and NKX2–8 had a direct linear relationship with likelihood of cisplatin resistance (P = 0.0001, r = 0.87, Fig. 5B Left). This association was confirmed using drug sensitivity assays, wherein the EC50s of cisplatin in the cluster of NSCLC cell lines that showed coactivation of TTF-1 and NKX2–8 were significantly higher than in the cluster of cell lines that did not demonstrate activation of TTF-1 and NKX2–8 (P = 0.02, r = 0.80, Fig. 5B Right), and is consistent with our previous observation. In our prior work, we used the amplified NCI-H2170 cell system of the squamous cell carcinoma subtype (which lacks transcription of TTF1) to test the tumor maintenance function of NKX2–8 (and PAX9) by stably expressing small hairpin RNA (shRNA) to knock down the protein expression of NKX2–8 (or PAX9) (5). In this study, H2170 was found to have coactivation of NKX2–8 and TTF-1 pathways and the H2170 transfecntant cells of knocked-down expression of NKX2–8 presented an increased sensitivity to cisplatin (data not shown), in line with our statistical prediction and experimental observation that coactivation of TTF-1/NKX2–8 positively correlates with cisplatin resistance.

To determine if the coactivation of TTF-1/NKX2–8 was correlated with other commonly used chemotherapy agents in the treatment of NSCLC, we examined the sensitivity of the lung cancer cell lines with coactivation of TTF-1/NKX2–8 to other cytotoxic agents commonly used as therapeutic options in patients with NSCLC (i.e., pemetrexed, paclitaxel, gemcitabine, and vinorelbine). Fig. S8 shows that coactivation of TTF-1/NKX2–8 is associated with resistance to taxanes, gemcitabine, and vinorelbine (data not shown) but was positively correlated with antifolate therapy with pemetrexed (P = 0.03, r = −0.68), suggesting that pemetrexed should be considered in the initial treatment for this cohort of patients. This finding might explain, at least in part, the increased clinical response of lung adenocarcinoma to pemetrexed.

### Discussion

The capacity to understand biological complexity is often limited by the ability to define relevant phenotypes. There is perhaps no better example of this challenge than that seen in cancer, in particular lung cancer. More than 180,000 new patients are diagnosed with NSCLC each year and 150,000 lung cancer-related deaths occur annually (28), yet only minimal advances in therapeutic strategies that significantly alter patient outcome have been observed over the past 2 decades. This is likely due to the complexity of the oncogenic process that involves the somatic acquisition of large numbers of mutations, coupled with the varied host genetic constitution, which produces a disease of enormous complexity that is extremely difficult to characterize and treat effectively.

Traditional methods of characterizing tumors rely on gross visual information (size of the tumor, degree of spread, histological characteristics of the tumor) that is coupled to represent the TNM staging system. Although these tools do provide a capacity to define

### Table 1. Univariate and multivariate Cox proportional hazards models of clinical, pathologic, and genomic variables

<table>
<thead>
<tr>
<th></th>
<th>Hazard ratio</th>
<th>Lower 95% CI</th>
<th>Upper 95% CI</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Univariate Cox proportional hazards model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>1.02</td>
<td>0.99</td>
<td>1.05</td>
<td>0.2175</td>
</tr>
<tr>
<td>Tumor size</td>
<td>1.24</td>
<td>1.07</td>
<td>1.44</td>
<td>0.0041</td>
</tr>
<tr>
<td>Gender</td>
<td>1.05</td>
<td>0.57</td>
<td>1.92</td>
<td>0.8797</td>
</tr>
<tr>
<td>Histologic subtype</td>
<td>1.02</td>
<td>0.57</td>
<td>1.83</td>
<td>0.9499</td>
</tr>
<tr>
<td>Vascular invasion</td>
<td>2.89</td>
<td>1.55</td>
<td>5.39</td>
<td>0.005</td>
</tr>
<tr>
<td>Lymphatic invasion</td>
<td>2.49</td>
<td>1.21</td>
<td>5.15</td>
<td>0.0106</td>
</tr>
<tr>
<td>Lymph node status</td>
<td>1.96</td>
<td>1.05</td>
<td>3.67</td>
<td>0.0311</td>
</tr>
<tr>
<td>Pleural invasion</td>
<td>1.69</td>
<td>0.81</td>
<td>3.52</td>
<td>0.1544</td>
</tr>
<tr>
<td>Pathologic stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I/stage II</td>
<td>1.64</td>
<td>0.82</td>
<td>3.3</td>
<td>0.004</td>
</tr>
<tr>
<td>Stage I/stage III</td>
<td>5.62</td>
<td>2.12</td>
<td>14.86</td>
<td></td>
</tr>
<tr>
<td>Genomic cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 2/cluster 1</td>
<td>0.28</td>
<td>0.12</td>
<td>0.66</td>
<td>0.0095</td>
</tr>
<tr>
<td>Cluster 3/cluster 1</td>
<td>0.45</td>
<td>0.22</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td>Cluster 4/cluster 1</td>
<td>0.83</td>
<td>0.34</td>
<td>2.02</td>
<td></td>
</tr>
<tr>
<td><strong>Multivariate Cox proportional hazards model</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tumor size</td>
<td>1.22</td>
<td>1.02</td>
<td>1.46</td>
<td>0.03</td>
</tr>
<tr>
<td>Vascular invasion</td>
<td>2.50</td>
<td>0.97</td>
<td>6.47</td>
<td>0.07</td>
</tr>
<tr>
<td>Lymphatic invasion</td>
<td>0.93</td>
<td>0.3</td>
<td>2.92</td>
<td>0.90</td>
</tr>
<tr>
<td>Lymph node status</td>
<td>0.78</td>
<td>0.27</td>
<td>2.3</td>
<td>0.65</td>
</tr>
<tr>
<td>Pathologic stage</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stage I/stage II</td>
<td>1.42</td>
<td>0.45</td>
<td>4.42</td>
<td>0.11</td>
</tr>
<tr>
<td>Stage I/stage III</td>
<td>4.87</td>
<td>1.09</td>
<td>21.84</td>
<td>0.01</td>
</tr>
<tr>
<td>Genomic cluster</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cluster 2/cluster 1</td>
<td>0.33</td>
<td>0.13</td>
<td>0.86</td>
<td>0.03</td>
</tr>
<tr>
<td>Cluster 3/cluster 1</td>
<td>0.64</td>
<td>0.26</td>
<td>1.54</td>
<td></td>
</tr>
<tr>
<td>Cluster 4/cluster 1</td>
<td>1.28</td>
<td>0.45</td>
<td>3.66</td>
<td></td>
</tr>
</tbody>
</table>

Only significant factors with univariate P < 0.05 were included in the multivariate model. Likelihood-ratio test against full model: P = 0.005. The likelihood-ratio test was done for the multivariate model with and without a genomic cluster. Cl, confidence interval.
tumor subgroups with distinct biology, it is abundantly clear that these classifications are imprecise, creating heterogeneous groupings of tumors and patients. Numerous examples can now be found where gene expression profiles have been able to dissect tumor subtypes (29, 30). The use of gene expression signatures as surrogate phenotypes has been particularly important, linking diverse experimental systems that dissect the complexity of biological systems in a way that was not previously feasible (31, 32). However, none of these studies to date have addressed the potential impact of specific developmental pathways on defining cancer phenotypes. It is very likely that tissue-specific developmental pathways may be critical in dissecting subtypes of cancer and may complement current classification systems. Here we report on the role of critical in dissecting subtypes of cancer and may complement

overlapping with patterns seen in tumor samples (Fig. 3). Predicted probability of cisplatin resistance in the individual lung cancer cell lines is plotted beneath the predicted probability of transcription factor coactivation (red, resistance to cisplatin; blue, sensitive to cisplatin). * KRAS and EGFR mutations. (B) A linear regression analysis of the predicted probability of cisplatin resistance plotted against the predicted probability of transcription factor coactivation of NXK2–8 and TTF-1 (dashed box in A) reveals a positive correlation ($P = 0.0001$). (Right) A linear regression analysis demonstrates a significant relationship between the EC50 of cisplatin and the predicted probability of transcription factor coactivation of NXK2–8 and TTF-1 (dashed box in A) in lung cancer cell lines ($P = 0.02$).

In conclusion, the developmental oncogenes TTF-1, NXK2–8, and PAX9 as coamplified oncogenes with functional synergism suggests that the 14q13.3 amplicon in lung cancer provides selective advantage to tumor cells by way of activation of multiple oncogenes, not just TTF-1 (5). Intriguingly, in this present study, simultaneous activation of TTF-1 and NXK2–8 pathways was associated with poor survival, in contrast to activation of individual genes that did not have prognostic significance. This finding not only echoes our thesis of multiple collaborating driver genes at the 14q13.3 amplicon, but also potentially sheds light on the prognostic value of TTF-1 protein expression, which has thus far been controversial (9, 33, 34). We believe that the prognostic relevance of TTF-1 is best evaluated with consideration of its oncogenic collaborator, NXK2–8. Furthermore, although the prognostic significance of TTF-1 and NXK2–8 coactivation was validated in 2 large patient data sets and the impact on survival in early stage NSCLC was found to be independent of standard clinicopathologic variables such as tumor size and stage (in the univariate and multivariate analyses shown in Table 1), we emphasize, however, that there are several approaches to refining prognosis in early stage NSCLC and the aims of this study were not to develop yet another prognostic model. Instead, in the present context, we use survival merely as a phenotype to demonstrate the biologic relevance of TTF-1 and NXK2–8 coactivation in lung tumors. To further elaborate, 3 large studies in lung cancer have been described, all of which were specifically designed to find a prognostic gene signature in early stage NSCLC using supervised classification on gene expression data from lung tumor samples with known clinical outcomes (25, 35, 36). Although the number of genes that overlap between the 3 independently derived gene sets is minimal, recent data suggest that because all 3 gene signatures could be independently validated in large patient cohorts (25, 35, 36), it is more likely that these gene signatures use different reporter genes to monitor the same biological pathways or processes. When comparing the pathways involved in the TTF-1 and NXK2–8 signatures, coactivation is associated with a higher likelihood of recurrence in our present study. In addition, several biologically relevant pathways are now identified (e.g., MAPKinase signaling, JAK/STAT signaling, Wnt signaling, etc.) that are common among the 3 previously described signatures of recurrence involved in activation of TTF-1 and NXK2–8 (Table S2). Despite this observed overlap in biologic pathways, it is important to note that none of the prognostic models described to date, including our own previous work (25), have shown predictive ability in being able to identify patients resistant to standard cytotoxic therapy or sensitive to specific targeted strategies.

The ability to identify a poor prognosis group of patients with early stage NSCLC characterized by NXK2–8 and TTF-1 coactivation whose survival pattern mimics that of patients with more advanced disease (Fig. S4) suggests that platinum-based chemotherapy, the standard of care for patients with advanced disease, may be beneficial in this “high-risk” group as well. However, in sharp contrast, we found that patients defined by the poor prognosis pattern (of TTF-1 and NXK2–8 coactivation) were in fact more likely to be resistant to cisplatin therapy. This suggests that a rational strategy to targeted therapy in the high-risk phenotype represented by NXK2–8 and TTF-1 coactivation is needed to treat this cohort of patients. We also emphasize that although these findings are encouraging, prospective validation strategies that are currently being initiated are needed before implementation into clinical practice.

In conclusion, the developmental oncogenes TTF-1, NXK2–8, and PAX9 are important biologically and clinically in lung cancer initiation and progression. Gene signatures reflecting the status of these transcription factors and patterns have prognostic and predictive value and may provide a rational approach to individualizing therapeutic options for a given patient with NSCLC.

**Methods**

**Overexpression of NXK2–8, PAX9, and TTF-1.** Oncogene-immortalized, premalignant normal human lung epithelium (BEAS-2B) was purchased from American Type Cell Collection. Retrovirus-mediated gene transfer was performed as described previously (5) with expression vectors bearing each individual transcription factor and 1 of 3 types of selection markers: puromycin, hygromycin, or blasticidin. In each instance, cells were harvested for preparation of total RNA. Infections of each individual transcription factor were performed 10 times to generate sufficient consistent data for pattern analysis. Quantitative assays including Western analysis and RT-PCR were performed to evaluate the extent of...
overexpression and activation of TTF-1, NKX2–8, and PAX9 (Fig. 1A). Anti-TTF-1, anti-NKX2–8, and anti-PAX9 antibodies were purchased from abCam, Invitrogen, and GenWay, respectively.

Clinical Data. The “discovery” data set consists of 91 early stage NSCLC tumors (Ia/Ib, IIa/Ib, and IIIa) (GSE3141) identified from the Duke Lung Cancer Prognostic Laboratory (10). This data set represents equal numbers of squamous and adenocarcinoma samples. The validation sample set included 214 samples: 84 lung adenocarcinomas (GSE3593) (25) and 130 lung squamous cell carcinomas (GSE4573) (26). Details of the relevant demographic and clinical information by data set are shown in Table S3.

Gene Expression Arrays. Total RNA from stable transfectant cells overexpressing each individual transcription factor was extracted using the Qiagen RNeasy Mini kit. RNAs were quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies) and quality was assessed with an Agilent 2100 Bioanalyzer. Total RNAs were amplified by a modified Eberwine technique (37), using a Message Amp II kit (Ambion) for 2 rounds of amplification. The resulting biotin-labeled RNA samples were then hybridized with Affymetrix U133Plus2.0 GeneChip arrays according to the manufacturer’s instructions. All analyses were performed in a minimal information about a microarray experiment (MIAME)-compliant fashion, as defined in the guidelines established by MGED (www.mged.org). The gene expression data used in the analyses are available at GEO (GSE9212, GSE141, GSE3593, and GSE4573).

Gene Expression Analysis. Gene expression data obtained from cell lines transfected with each individual transcription factor were used to populate a matrix with MATLAB software. Statistical modeling representing activation of the individual transcription factors (NKX2–8, PAX9, and TTF-1) was performed using metagene construction and binary prediction analysis, as described previously (10, 25, 38). Gene expression signatures that reflect the activity of a given transcription factor were applied to a clinically annotated data set of 91 tumor samples (GSE3141) to predict patterns of transcription factor activation. Predictions of transcription factor pathway activation within the tumor samples were evaluated to produce estimated relative probabilities and associated measures of uncertainty of transcription factor activation across the validation set. An estimated probability of 0.5 was classified as high probability of transcription factor activation and a probability of <0.5 was classified as low probability of transcription factor activation. Applications of the transcription factor predictors to other data sets (GSE4573, GSE3593) and the gene expression-based predictor of cisplatin resistance (38) to lung cancer cell lines and lung tumor data sets (GSE3141) were performed as described above. See SI Methods for complete details.

Hierarchical clustering using the gene pattern software (http://www.broad.mit.edu/cancer/software/genepatter) based on the probability of transcription activation of the tumor samples was used to generate clusters with coactivation of specific transcription factors. Specifically, the clustering was performed using a complete linkage with a Pearson correlation metric method on the preprocessed data to organize all of the data elements into a single tree with the highest levels of the tree representing the discovered classes (29).

Standard Kaplan–Meier mortality curves and their significance level were generated to evaluate the prognostic role of individual transcription factors and in clusters of patients with coactivated transcription factors, using the graph pad software. The log-rank test was used to assess the differences between the survival curves and to calculate the nominal P-values between groups.

ACKNOWLEDGMENTS. The authors thank the Duke University microarray core facility and the Cold Spring Harbor Laboratory microarray facility for collecting the microarray data. A.P. was supported by the Emily Brown Cancer Research Fund and by grants from the Jimmy V Foundation, the American Cancer Society, and the Burroughs Wellcome Fund. D.M. was supported by the Joan’s Legacy Foundation, a National Institutes of Health grant (1RO1CA127547-01), and a Penn State University Pathology Internal Fund.