Science Translational Medicine
AAAS

Airway PI3K Pathway Activation Is an Early and Reversible Event in Lung Cancer Development

Adam M. Gustafson, *et al.*

Sci Transl Med **2**, 26ra25 (2010);
DOI: 10.1126/scitranslmed.3000251

CANCER

Airway PI3K Pathway Activation Is an Early and Reversible Event in Lung Cancer Development

Adam M. Gustafson,[1,2]* Raffaella Soldi,[3]* Christina Anderlind,[1] Mary Beth Scholand,[4] Jun Qian,[5] Xiaohui Zhang,[1] Kendal Cooper,[3] Darren Walker,[3] Annette McWilliams,[6] Gang Liu,[1] Eva Szabo,[7] Jerome Brody,[1] Pierre P. Massion,[5] Marc E. Lenburg,[1,2,8] Stephen Lam,[6] Andrea H. Bild,[3]*† Avrum Spira[1,2,8]*†

(Published 7 April 2010; Volume 2 Issue 26 26ra25)

Although only a subset of smokers develop lung cancer, we cannot determine which smokers are at highest risk for cancer development, nor do we know the signaling pathways altered early in the process of tumorigenesis in these individuals. On the basis of the concept that cigarette smoke creates a molecular field of injury throughout the respiratory tract, this study explores oncogenic pathway deregulation in cytologically normal proximal airway epithelial cells of smokers at risk for lung cancer. We observed a significant increase in a genomic signature of phosphatidylinositol 3-kinase (PI3K) pathway activation in the cytologically normal bronchial airway of smokers with lung cancer and smokers with dysplastic lesions, suggesting that PI3K is activated in the proximal airway before tumorigenesis. Further, PI3K activity is decreased in the airway of high-risk smokers who had significant regression of dysplasia after treatment with the chemopreventive agent *myo*-inositol, and *myo*-inositol inhibits the PI3K pathway in vitro. These results suggest that deregulation of the PI3K pathway in the bronchial airway epithelium of smokers is an early, measurable, and reversible event in the development of lung cancer and that genomic profiling of these relatively accessible airway cells may enable personalized approaches to chemoprevention and therapy. Our work further suggests that additional lung cancer chemoprevention trials either targeting the PI3K pathway or measuring airway PI3K activation as an intermediate endpoint are warranted.

INTRODUCTION

Cigarette smoke is the dominant cause of lung cancer in the United States, accounting for an estimated 90% of all cases (*1*). However, the damage caused by cigarette smoke is not limited solely to the lung but rather forms a "field of injury" throughout the entire respiratory tract (*2*–*8*). Tissue from this extended injured area can be used to glean clinically relevant information about smoking-induced damage and disease; cells can be collected from regions of the respiratory tract, such as the bronchial airway, in a less invasive manner than from the lung. Consistent with this idea, our laboratory has developed a gene expression–based biomarker signature in cytologically normal bronchial airway epithelium that can distinguish smokers with and without lung cancer (*9*) and is independent from other clinical factors as a predictor of lung cancer (*9*, *10*). Although this biomarker can predict lung cancer occurrence, it does not identify the signaling pathways underlying these gene expression changes.

One approach to assess signaling pathway activity uses the gene expression pattern resulting from the in vitro activation of the pathway to predict its activity in patient samples. This approach has been successful at predicting pathway status in cell lines as well as tumors with known pathway activation and molecular lesions (*11*). One of the strengths of using gene expression signatures to indicate signaling activity in clinical samples is that they allow for the measurement of multiple pathways in a single microarray experiment. Further, gene expression–based predictions of pathway activity correlate with sensitivity to drugs that target these specific pathways (*11*–*18*). Numerous studies have also found correlation of predicted pathway activity with therapeutic response to the corresponding targeted therapy in clinical trials (*19*–*22*).

Our goal was to understand the signaling pathways that are deregulated at an early stage in lung cancer. In primary tumors from patients with lung cancer, many signaling pathways have previously been found to be deregulated, such as p53, Ras, and phosphatidylinositol 3-kinase (PI3K) (*23*–*25*), although it is not clear which pathways are deregulated before lung cancer development. Molecular changes in signaling pathways such as p53 can occur early in tumorigenesis, as mutant p53 alleles seen within the tumor also occur in histologically normal cells neighboring and more distal to the primary tumor (*5*, *26*). Identification of additional oncogenic pathways that are deregulated in cytologically normal bronchial airway cells from smokers with lung cancer might elucidate mechanisms involved in early tumorigenesis and malignant progression. Furthermore, agents targeting these pathways could provide therapeutic and chemopreventive opportunities at the premalignant stage of lung cancer.

[1]Section of Computational Biomedicine, Department of Medicine and Pulmonary Center, Boston University Medical Center, Boston, MA 02118, USA. [2]Bioinformatics Program, Boston University, Boston, MA 02118, USA. [3]Department of Pharmacology and Toxicology, University of Utah, Salt Lake City, UT 84112, USA. [4]School of Medicine, University of Utah, Salt Lake City, UT 84112, USA. [5]Division of Allergy Pulmonary and Critical Care Medicine, Department of Medicine, Vanderbilt-Ingram Comprehensive Cancer Center and Nashville Veterans Affairs Medical Center, Nashville, TN 37232, USA. [6]Department of Respiratory Medicine, British Columbia Cancer Agency, Vancouver, British Columbia, Canada V5Z 4E6. [7]Lung and Upper Aerodigestive Cancer Research Group, Division of Cancer Prevention, National Cancer Institute, Bethesda, MD 20892, USA. [8]Department of Pathology and Laboratory Medicine, Boston University School of Medicine, Boston, MA 02118, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: aspira@bu.edu (A.S.); andreab@genetics.utah.edu (A.H.B.)

Here, using a gene expression–based approach to define oncogenic pathway signatures from in vitro perturbation experiments, we examined PI3K pathway activity in cytologically normal bronchial airway cells of smokers with lung cancer.

RESULTS

PI3K pathway activation in cytologically normal bronchial airway epithelial cells of smokers with lung cancer

To explore patterns of pathway deregulation in cytologically normal airway epithelial cells from patients with and without lung cancer, we used a whole-genome gene expression data set previously published by our group (GSE4115) consisting of cytologically normal bronchial airway epithelial cell brushings obtained from 129 current and former smokers undergoing flexible bronchoscopy for suspicion of lung cancer (*9*) (see Table 1 for patient demographics). We analyzed these data for oncogenic pathway perturbation using pathway signatures derived from a data set consisting of the gene expression consequences of in vitro pathway activation and a computational approach from our group (*11*, *27*–*29*). Ras, Myc, E2F3, Src, and β-catenin oncogenic pathway signatures had previously been experimentally derived by activating each pathway via expression of a specific oncogene in primary human epithelial cells (*11*). PI3K and ΔNp63 signatures were experimentally derived by expressing the genes in quiescent primary human epithelial cells, and pathway activity was confirmed with Western blotting analysis for the overexpressed gene and downstream pathway activity (fig. S1). A specific gene expression signature was then defined for each oncogenic pathway by identifying genes that are altered after pathway activation (see PI3K pathway heat map in fig. S2). To predict pathway activation in our cytologically normal airway microarray samples, we used oncogenic genomic signatures to train a binary regression model for each of the studied pathways and then applied these signatures to an external gene expression data set to determine respective pathway activity (see Materials and Methods for details). With this methodology, oncogenic pathway activity for the seven signaling pathways (Ras, Myc, E2F3, Src, β-catenin, ΔNp63, and PI3K) was calculated for the samples in the bronchial airway data set (see Fig. 1 for overall study design). It is important to note that, although about half of the samples in this data set were from patients with primary lung cancer, the brushings collected from the proximal mainstem bronchus (that is, not adjacent to the tumor or lung lesion) were cytologically normal and were >90% epithelial.

Of the seven pathways tested, only two were found to have differential pathway activity in the airway of smokers with lung cancer relative to controls with alternate lung pathologies after a random permutation analysis: ΔNp63 and PI3K ($P < 0.001$) (Fig. 2A). Furthermore, genes that have been found to play roles in the phosphatidylinositol signaling system pathway (*30*) were also significantly upregulated in lung cancer patients relative to patients without lung cancer according to gene set enrichment analysis (GSEA) [$P = 0.034$, false discovery rate (FDR) $Q = 0.099$] (*31*). We subsequently validated our gene expression findings on an additional series of bronchial airway brushings collected from smokers undergoing bronchoscopy for suspicion of lung cancer ($n = 35$) and found a significantly higher PI3K pathway activity in the normal airway of smokers with lung cancer ($P = 0.033$) (Fig. 2B). The ΔNp63 pathway, however, was not differentially activated between smokers with and without lung cancer in this cohort ($P = 0.255$) (Fig. 2B).

Table 1. Patient demographics for the cohorts used to explore oncogenic pathway activity in cytologically normal airway epithelium of smokers. Average (± SD) for continuous variables is shown. C, current smoker; F, former smoker.

	Primary data set		Validation data set	
	Lung cancer	No lung cancer	Lung cancer	No lung cancer
Samples	60	69	18	17
Age	64.1 (9)*	49.8 (15.2)*	66.1 (11.4)	62.2 (11.1)
Pack-years	57.4 (25.6)*	29.4 (27.3)*	46.7 (28.8)	60.0 (44.3)
Smoking status	51.7% F, 48.3% C	37.3% F, 62.3% C	66.7% F, 33.3% C	52.9% F, 47.1% C
Months since quitting (formers)	113 (118)	158 (159)	153 (135)	93 (147)

*$P < 0.05$, significantly different between the classes of that data set.

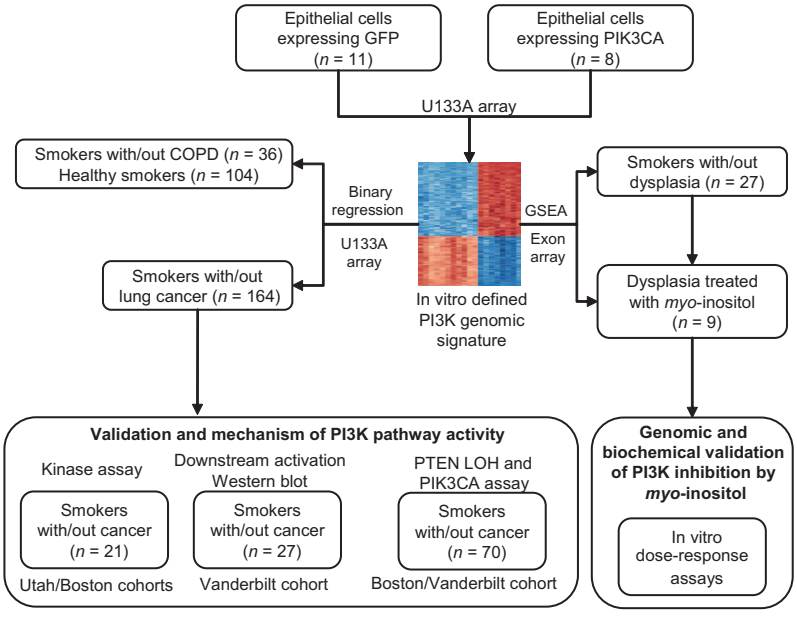

Fig. 1. Study design depicting the various patient populations analyzed and experiments performed. PI3K pathway activity was measured both computationally and biochemically in the cytologically normal bronchial airway of smokers with and without (with/out) lung cancer. Bronchial airway brushings were collected in each cohort, except for the Western blot cohort, which had airway biopsies. LOH analysis was performed on 44 of the 71 samples. Computational methodologies (binary regression and GSEA) and Affymetrix array platforms (U133A and exon array) used in the analysis are listed where applicable. Experimental perturbation and computational analysis of pathway activity were performed for multiple oncogenic pathways (*11*), but the PI3K pathway is the focus of this study, chosen because of its statistical significance.

To validate that our PI3K pathway gene expression signature (derived in vitro; fig. S2) reflects actual PI3K activity in vivo, we explored a gene expression data set from primary breast cancer tumors of the basal-like and HER2-overexpressing subtypes with known PTEN copy number status (*32*). Because PTEN negatively regulates the PI3K pathway, a loss of PTEN should result in higher PI3K pathway activity. When we used the gene expression signature in a regression model to predict the PI3K pathway activity of these samples, we found that samples lacking PTEN had elevated PI3K pathway activity relative to samples with normal PTEN ($P = 2.7 \times 10^{-5}$) (fig. S3). This finding supports the validity of the genomic signature model methodology as well as the PI3K pathway signature in a clinically relevant data set of primary tumor samples.

Additionally, we explored how the in vitro genomic signature for PI3K (gene names listed in table S1) overlaps with genes previously described as playing a role in mediating the effects of this pathway. The genomic signature for the PI3K pathway is made up of genes with the greatest and most consistent messenger RNA (mRNA) expression changes after in vitro activation of this pathway (see fig. S2). This distinction is important, as we are only looking for mRNA expression changes and not for the most immediate downstream activation events or changes in phosphorylation status, which are often the focus in experiments on the PI3K pathway in the literature. However, there is a significant enrichment of genes in the PI3K signature that are involved in DNA-dependent DNA replication [$P \leq 0.0001$, Bayes factor = 26, Gene Ontology (GO) term: 0006261]. It has been well established that PI3K signaling stimulates cellular replication. Additionally, we see significant enrichment of genes involved in the cell cycle, a downstream component of the PI3K pathway ($P \leq 0.0001$, Bayes factor = 14, GO term: 0007049). Further, the PI3K signature includes genes involved in DNA repair (MSH2, MSH6, GADD45B, BARD1, and POLD3), another important effect of PI3K signaling (*33*, *34*). Lastly, there are a number of genes directly involved in the PI3K pathway in the signature, such as JUN family members (JUN, JUNB, and JUND) and cyclins/cyclin-dependent kinase (CDK) pathway genes (CCNE1, CCNE2, and CDK4). Other genes deregulated by PI3K expression known to be a part of the PI3K signaling pathway are shown in fig. S4.

PI3K activation does not correlate with smoke exposure or chronic obstructive pulmonary disease

We next determined whether the increased PI3K pathway activation in cytologically normal airway cells is specifically associated with lung cancer or is the result of other potentially confounding factors, such as differences in cumulative smoke exposure (patients with cancer had higher cumulative exposure in the primary data set but not the additional validation set; see Table 1), or other pulmonary diseases, such as chronic obstructive pulmonary disease (COPD). In linear models that included both cancer status and cumulative smoke exposure, PI3K activity was significantly higher in smokers with cancer ($P < 0.001$), and PI3K activation probabilities were not significantly associated with cumulative tobacco use ($P = 0.57$). Additionally, using a whole-genome gene expression data set of bronchial airway epithelium collected from current ($n = 52$), former ($n = 31$), and never ($n = 21$) healthy smokers (*35*) (Table 2), we found that the activity of neither pathway was associated with smoking history, although the ΔNp63 pathway was slightly activated when comparing never and current smokers ($P = 0.09$) (Fig. 2C). Finally, neither PI3K nor ΔNp63 activity varied in the bronchial airway between smokers with ($n = 17$) and with-

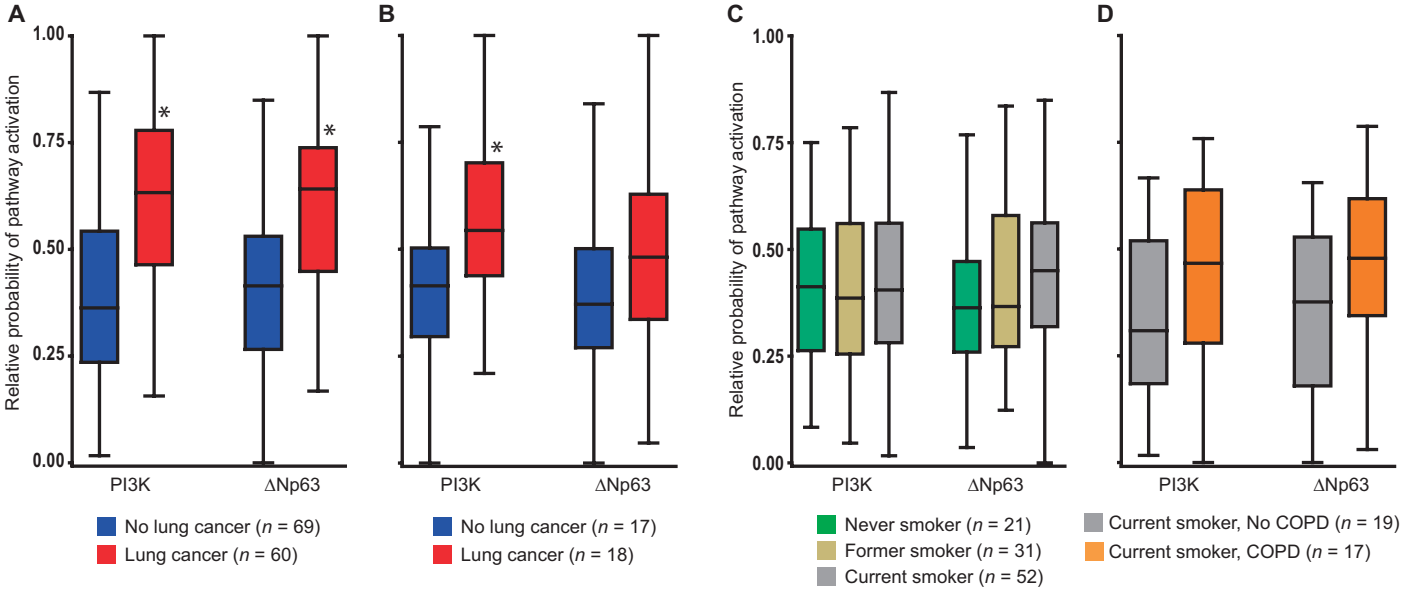

Fig. 2. Differential activation of PI3K in the airway of smokers with lung cancer. By using computational models trained on in vitro gene expression signatures, pathway activation probabilities were calculated in samples obtained from the cytologically normal airway ($n = 129$). Pathway activation is summarized with box plots, where the bar represents the median value, the box denotes the range of the data points from the 25th to the 75th percentile, and the whiskers specify the range of the remaining 1st and 4th quartile. (**A**) PI3K (*$P < 0.001$) and ΔNp63 (*$P < 0.001$) pathway activation status varies in cytologically normal airway epithelium between patients with and without lung cancer ($n = 129$). (**B**) To validate results from (A), we used an additional prospectively collected group of samples ($n = 35$). In this second cohort, PI3K was significantly differentially activated (*$P = 0.033$), whereas ΔNp63 was not. (**C** and **D**) PI3K and ΔNp63 pathway activation status did not vary among healthy never, former, or current smokers (C) or current smokers with or without COPD (D).

out (n = 19) COPD (*9*, *35*) (Fig. 2D; see Table 2 for patient demographics). These data suggest that the differential activation of the oncogenic pathway PI3K in the normal bronchial airway is specific to individuals with lung cancer. Because only the PI3K pathway was differentially activated in both the primary (n = 129) and the secondary (n = 35) data sets, we chose to focus on the potential role of the PI3K pathway in early lung cancer development.

PI3K is activated in lung cancer tissue

The deregulation of PI3K in cytologically normal airway epithelial cells of patients with lung cancer led us to examine PI3K activity in lung cancer tissue. We would expect that processes involved in early lung cancer development might remain activated in lung tumors. For this analysis, we examined PI3K pathway activity in a published gene expression data set composed of lung adenocarcinoma and matched adjacent nontumor tissue (n = 107) (*36*). Malignant tissue showed a pattern of gene expression, indicating a significant increase in PI3K pathway activity when compared to the adjacent nontumor tissue (P < 0.001 by random permutation) (Fig. 3). The ΔNp63 pathway signature was also increased, but to a lesser extent, in tumor cells (P = 0.0098). It is well documented that other oncogenic pathways, including Myc and E2F, are deregulated in rapidly dividing tumor cells, a result that is verified by our results (Fig. 3). The increased expression of the PI3K pathway signature in tumor tissue relative to adjacent normal tissue supports our hypothesis that PI3K activation is important for lung cancer tumorigenesis. Our findings suggest that, PI3K activity is elevated in the normal and premalignant airway epithelial tissue, and this increased activity persists in lung cancer.

Biochemical analysis of PI3K pathway activity in the bronchial airway

To determine whether increased expression of the PI3K pathway signature represents an actual increase in PI3K activity in bronchial airway epithelium, we measured PI3K kinase activity in a prospectively collected cohort of subjects undergoing bronchoscopy for clinical suspicion of lung cancer and from whom cytologically normal airway epithelial samples were obtained. Bronchial airway brushings were independently obtained from patients at the Boston University Medical Center (Boston, MA) and the University of Utah Hospital (Salt Lake City, UT) between October 2007 and June 2008. Subjects were followed after bronchoscopy until a final diagnosis of lung cancer or alternate lung pathology was made. Subjects without lung cancer had a range of other pathologies, including metastatic cancer of nonlung origin, sarcoidosis, septic emboli, and pneumonia (see table S2).

After biochemical measurement by kinase assay, we found significant elevation of PI3K kinase activity measured in the airway of smokers with lung cancer relative to smokers without lung cancer (P = 0.007) (Fig. 4A and table S3). These results indicate that PI3K activity is increased in the normal airway of patients with lung cancer and validate our computational predictions of PI3K activity calculated from gene expression data. Because our control subject group included only patients with a variety of pulmonary pathologies as opposed to healthy controls (with no existing pathologies), these results are particularly noteworthy.

To further strengthen the in vivo biochemical findings above, and to determine the specific downstream pathways activated in this

Table 2. Patient demographics for the healthy smokers and COPD cohorts used to explore oncogenic activity in the cytologically normal airway. Average (± SD) for continuous variables is shown.

	Never/former/current smoker data set			COPD data set	
	Never	Former	Current	No COPD	COPD
Samples	21	31	52	20	17
Age	32.3 (10.7)*	55.9 (14.7)*	48.6 (15.2)*	43.6 (9.3)*	60.4 (12.3)*
Pack-years		34.0 (30.1)	34.5 (34.2)	33.5 (25.9)*	57.4 (27.9)*
Smoking status		100% F	100% C	100% C	100% C
Months since quitting (formers)		145.2 (162.8)			

*P < 0.05, significantly different between the classes of that data set.

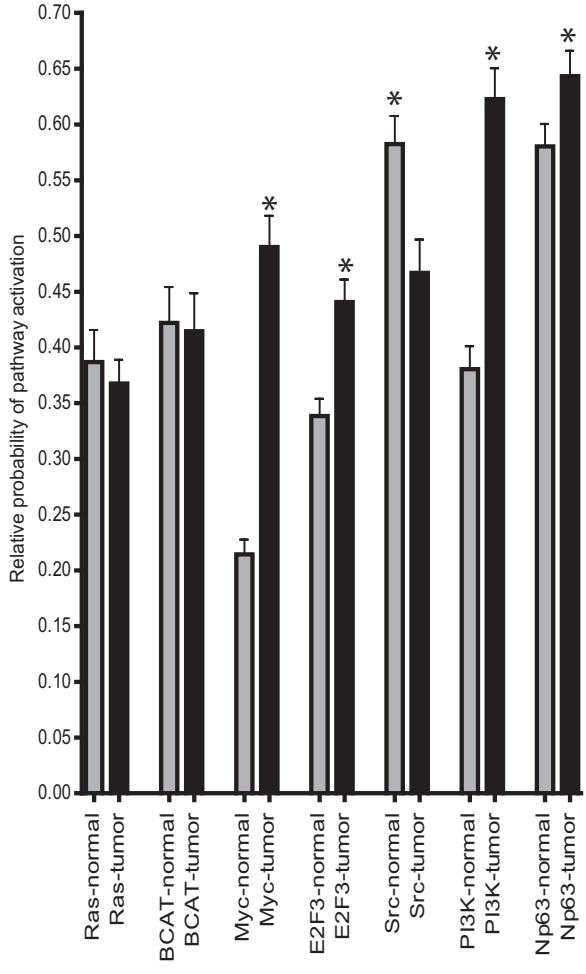

Fig. 3. Oncogenic pathway activity in lung tumor and adjacent normal tissue. Oncogenic pathway activity was calculated for a microarray data set of lung adenocarcinoma and adjacent normal tissue (*36*). An increase in PI3K (P < 0.001) and ΔNp63 (P = 0.0098) was observed when comparing adjacent normal and its paired tumor sample. Increases were also seen in Myc and E2F3, and there was a decrease in Src. Error whiskers are reported as SEM; pathways denoted by asterisks were statistically significantly different (P < 0.01).

setting, we evaluated activation of different arms of the PI3K pathway by Western blot analysis. Figure 4B shows the results for two main downstream pathways, phospho-AKT and phospho–protein kinase C (PKC), on a second cohort of endobronchial biopsies (as opposed to brushings) that had been collected at Vanderbilt University (table S2). Banked bronchial brushings were not available on this cohort, and there was insufficient protein obtained from the endobronchial biopsies to validate PI3K kinase activity on these same samples. After quantification and normalization of the Western blots, we found a significant increase in activated PKC ($P = 0.01$), but not AKT ($P = 0.61$), in the airway of smokers with lung cancer as compared to smokers without lung cancer (Fig. 4B). PKC is downstream of growth factor–induced PI3K activity and drives cellular transformation and bronchiolar stem cell expansion and lung tumorigenesis (*22*, *37*–*41*).

Activation of PI3K is not a result of PTEN loss or PI3K amplification

To determine whether the activation of PI3K pathway was a consequence of a PTEN-defective phenotype or PI3K amplification, we evaluated the PTEN and PI3K genomic status in a panel of genomic DNA isolated from noncancerous bronchial areas of two independent cohorts of patients with or without lung cancer (see table S4 for patient demographics). To evaluate for PTEN deficiency, we examined the genotypes of the Boston cohort ($n = 43$) at exons I and II of the gene on 10q23.3, known to have represent loss of heterozygosity (LOH) frequently in lung cancer (*42*, *43*). TaqMan quantitative polymerase chain reaction (PCR) analysis did not reveal PTEN LOH in noncancerous airway tissue (fig. S5). This result was not totally unexpected; previous studies had reported the absence of PTEN LOH or mutation in precancerous lesions or in early stages of certain cancers even when the actual tumor tissue presents PTEN LOH (*44*). To explore the potential role for PI3K genomic amplification, we performed real-time PCR conditions and copy number calculations for the PI3K locus as described previously (*45*). Only 1 of the 71 patients studied (a former smoker without lung cancer) in both the Boston and the Vanderbilt cohorts showed amplification of the PI3K locus (fig. S5), arguing against this mechanism playing a significant role in the activation of PI3K in the airway of smokers with lung cancer.

PI3K pathway activation in smokers with dysplasia

To evaluate the hypothesis that activation of PI3K is an early event in the development of lung cancer, we compared PI3K pathway activity in cytologically normal airway epithelium from healthy smokers ($n = 11$) and smokers with mild to moderate airway dysplasia found on autofluorescence bronchoscopy ($n = 16$; Table 3). Mild to moderate dysplasia represents an early abnormal change in the cytology of cells that is noncancerous but is often considered a precancerous state. Samples from smokers with airway dysplasia were from a phase 1 clinical study of the chemopreventive agent *myo*-inositol (*46*). In that study, current and former smokers with mild to moderate dysplasia were treated with *myo*-inositol for 3 months, with cell brushings of the uninvolved bronchial airway collected before and after treatment.

Because dysplasia is considered to be a preneoplastic event, if smokers with dysplasia have higher levels of PI3K than healthy smokers, it would suggest that PI3K pathway activation in the airway epithelium precedes the development of lung cancer. Using available gene expression data for both the dysplasia and healthy smoker cohort, that had previously been run on Affymetrix human exon arrays, we sought to computationally explore PI3K activity in these healthy and dysplastic groups. Differences in the microarray platforms used to generate the PI3K pathway signature (U133A 3′ biased arrays) and the dysplasia cohort (exon arrays) prevented us from using the same regression methodology to calculate pathway activity. Instead, a list of genes from the in vitro defined PI3K pathway genomic signature was used with GSEA to explore the differential activity of the PI3K pathway between smokers with and without dysplasia (see table S5 for PI3K pathway gene list that overlaps between two array platforms). First, the in vitro PI3K pathway signature was split into two gene sets grouped by the genes that are either induced or repressed with PI3K activation (PI3K_Up and PI3K_Down). Because there were significant differences in cumulative tobacco exposure between the group of healthy smokers and those with dysplasia, the dysplasia coefficient from a linear model incorporating batch effects, pack-years, and presence of dysplasia was used to rank the genes for the GSEA analysis. Genes that are induced after PI3K activation in vitro were among the genes most induced in the cytologically normal airway of subjects with dysplasia, whereas genes that are repressed after PI3K activation showed the converse pattern (PI3K_Up gene set: $P < 0.001$,

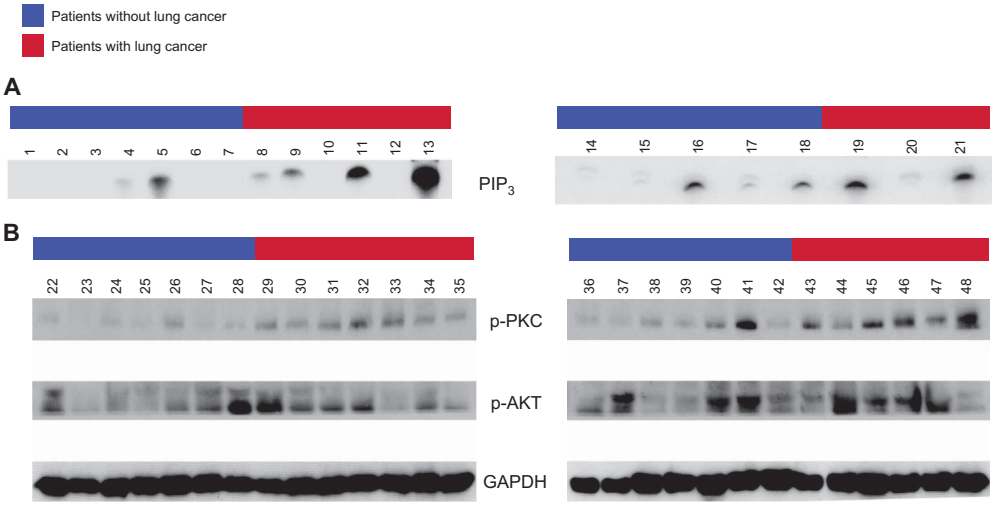

Fig. 4. Biochemical validation of PI3K activity in prospectively collected airway samples. (**A**) Kinase assays measuring PIP_3 were used to identify in vivo levels of PI3K activity in bronchial airway epithelial cell brushings collected via bronchoscopy from patients under suspicion of having lung cancer at the University of Utah (left panel) and Boston University (right panel). Using an ANOVA model that incorporates batch as a variable (that is, Utah or Boston cohort), we found that patients with lung cancer had higher PI3K activity than those without lung cancer ($P = 0.007$). (**B**) In a separate cohort of endobronchial biopsy specimens collected at Vanderbilt University, a Western blot was used to determine the activity of two key downstream members of the PI3K pathway. Patients with lung cancer had increased activated PKC ($P = 0.01$), but not AKT ($P = 0.61$), in their proximal airway by *t* test.

FDR $Q = 0.022$; PI3K_Down gene set: $P < 0.001$, $Q < 0.001$) (Fig. 5 and table S6). These results suggest that PI3K pathway activation is higher in airway epithelium from patients with dysplasia and that activation of PI3K is likely an early event in lung cancer tumorigenesis.

Reversibility of PI3K pathway activation in smokers with dysplasia treated with *myo*-inositol

Because we observed elevated PI3K activity in both high-risk smokers with dysplasia and smokers with lung cancer, we next wanted to determine whether a decrease in PI3K activity correlated with regression of dysplastic lesions. GSEA analysis was again used to measure changes in the expression of genes in the in vitro defined PI3K pathway gene sets (PI3K_Up and PI3K_down), this time comparing gene expression patterns from patients before and after treatment with *myo*-inositol (9 individuals, 18 samples). Subjects with regression of dysplasia (6 individuals, 12 samples) showed increased expression of genes that are repressed on PI3K activation in vitro ($P = 0.04$). Smokers who did not respond to *myo*-inositol treatment (three individuals, six samples) had no change in the in vitro defined PI3K gene sets. The decrease in PI3K activity in patients who respond to *myo*-inositol suggests that regression of dysplasia is associated with a reduction of PI3K pathway activity in the proximal airway.

Myo-inositol as a PI3K inhibitor

Because *myo*-inositol is a precursor of phosphatidylinositol, it is possible that the regression of airway dysplasia after *myo*-inositol treatment is due to an inhibition of PI3K. Derivatives of *myo*-inositol, such as inositol 1,3,4,5,6-pentakisphosphate and inositol hexakisphosphate, have previously been found to act as inhibitors of the PI3K pathway and associated downstream members (*47*, *48*). To further explore whether we are observing a causative or correlative association, we tested the ability of *myo*-inositol to inhibit PI3K in vitro.

The H1299 cancer cell line (derived from lung adenocarcinoma) was treated with varying physiologically relevant doses of *myo*-inositol, and PI3K activity was biochemically quantified with a standard kinase assay protocol (*49*). In this cell line, a dose-dependent decrease in PI3K activity was observed after treatment with *myo*-inositol. The biochemical median effective concentration (EC_{50}) of *myo*-inositol for PI3K inhibition was 7.3×10^{-8} M (Fig. 6A). This dose range is comparable to the doses given for regression of dysplastic lesions to high-risk smokers, which was 9 g twice daily (*46*). Assuming 10% bioavailability and rapid clearance, the peak concentration of *myo*-inositol would exceed 8 mM. To validate the sensitivity of our genomic PI3K signature, we ran whole-genome gene expression arrays on the cell line samples treated with increasing doses of *myo*-inositol and used the regression model to predict PI3K pathway activity. From these results, we calculated the effective concentration of *myo*-inositol that inhibited predicted PI3K activity by 50%. *Myo*-inositol inhibited the genomically defined PI3K activity at EC_{50} values comparable to the biochemical analyses (Fig. 6B). Other proliferative and oncogenic pathways (Ras and E2F3) do not show any loss of activity on *myo*-inositol treatment, highlighting the specificity of *myo*-inositol for the PI3K pathway. Together, these findings suggest that *myo*-inositol is a PI3K inhibitor and that its lung cancer chemoprevention properties may be related to this inhibition.

Table 3. Patient demographics for the cohorts used to explore PI3K activity in patients with preneoplastic airway dysplasias, as well as changes in PI3K activity after treatment with *myo*-inositol. The first column refers to participants in the *myo*-inositol study who had samples collected both before and after treatment with *myo*-inositol. Although 10 patients (20 samples) are listed in the table, 1 subject was removed from our analysis because of an unknown response to treatment. The second column refers to additional subjects with airway dysplasia for whom only a single pretreatment sample was hybridized to a microarray. Clinical variables are shown as means ± SD. All data sets consist of samples collected via brushing of cytologically normal bronchial airway epithelium during bronchoscopy.

	Smokers with dysplasia		Smokers without dysplasia
	Treatment with *myo*-inositol	No treatment	
Patients	10[†]	6	11
Age[*]	61.7 (8.3)	64.5 (7.1)	36.5 (10.0)
Pack-years[*]	42.9 (23.7)	59.8 (17.7)	12.4 (9.2)
Smoking status	70% F, 30% C	100% C	100% C
Months since quitting (formers)	145.7 (89.9)		
Drug response	6 responders, 3 nonresponders, 1 unknown		

[*]Age and pack-years were significantly different between dysplasia and healthy smoker data sets ($P < 0.05$). [†]Two samples were collected per patient: one before and one after treatment with *myo*-inositol.

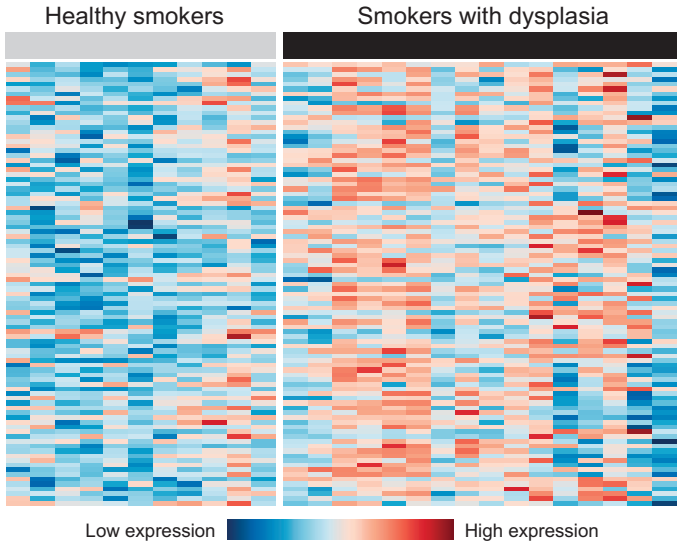

Fig. 5. Increased expression of genes induced by PI3K pathway activation in the airway of smokers with dysplasia. In the cytologically normal airway of smokers with and without dysplasia ($n = 27$), when looking at the relative expression of genes that increase when the PI3K pathway is activated in vitro, a higher activity is observed in smokers with dysplasia. The list of PI3K-induced genes is from the in vitro pathway perturbation experiments (the PI3K_Up gene set). GSEA was used to quantify the enrichment of this gene set ($P < 0.001$, FDR $Q < 0.001$) in this cohort. Dysplasia airway expression data were ranked for GSEA, with a linear model that takes into account dysplasia and cumulative tobacco exposure (measured in pack-years) as well as a random variable accounting for batch effects. The list of genes in this figure is available in table S5.

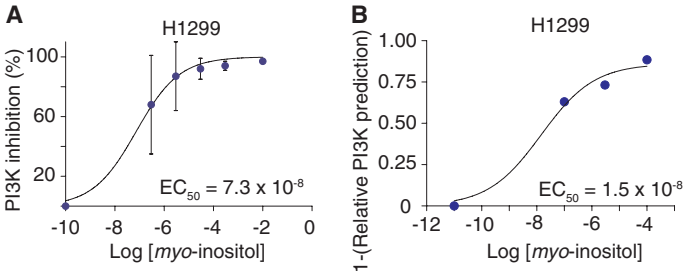

Fig. 6. Inhibition of the PI3K pathway in vitro by *myo*-inositol. (**A** and **B**) Biochemical (A) and genomic (B) analyses of PI3K inhibition by *myo*-inositol treatment. (A) H1299 lung cancer cell lines were treated with varying doses of *myo*-inositol, and PI3K activity was measured by the generation of PIP_3 from PIP_2 in cell lysates. There was a decrease in PI3K activity after treatment with *myo*-inositol, with a calculated EC_{50} at 7.3×10^{-8} M. (B) The H1299 lung cancer cell lines were treated with varying doses of *myo*-inositol, and PI3K pathway activation was predicted, with our PI3K gene expression signature, from whole-genome gene expression data sets generated from each sample. There was a decrease in predicted PI3K activity after treatment with *myo*-inositol, with a calculated EC_{50} at 1.5×10^{-8} M. EC_{50} denotes molarity. Error bars represent the SD for the biochemical measurements (where biological replicates were available). Gene expression signatures of none of the other six oncogenic pathways were affected by increasing doses of *myo*-inositol.

DISCUSSION

On the basis of the concept that genomic changes in the epithelial cells that line the entire respiratory tract reflect host response to and damage from cigarette smoke, we sought to identify early events leading to lung tumorigenesis by identifying pathways that are deregulated in the bronchial airway of smokers with or at risk for lung cancer. To accomplish this, we used pathway activation expression signatures developed through in vitro pathway perturbation (*11*) and trained genomic models to compare oncogenic pathway activity in the cytologically normal airway epithelium of smokers with and without lung cancer. The PI3K pathway is activated in the normal bronchial airway of smokers with lung cancer, and this increased activity is independent of smoking status or other smoking-related disease. Although the PI3K pathway (and the downstream PKC pathway) is known to play a role in some cases of lung cancer (*37*, *50*), our study describes activation of this pathway in distally located (relative to the location of the tumor) and cytologically normal airway epithelium of smokers with lung cancer, suggesting that this pathway is altered early in lung cancer development. Our lung cancer–free controls had other pathologies and were not just healthy volunteers, indicating the specificity of PI3K activation to lung cancer.

Consistent with the hypothesis that PI3K pathway activation is an early event in lung carcinogenesis, we also observed increased PI3K pathway activation in the cytologically normal airway of smokers with dysplastic airway lesions. This finding further supports the hypothesis that PI3K activity is induced before the development of lung neoplasms. Further, we observed higher PI3K activity in lung tumors as compared to adjacent normal tissue, suggesting an additional increase in PI3K activity as cells undergo transformation during lung cancer development.

Of clinical importance is whether a reduction in PI3K pathway activity before the development of lung cancer would offer any therapeutic potential. Current PI3K pathway inhibitors, such as sirolimus, have harmful side effects that prohibit their use as a long-term therapy. We conducted a genomic analysis on a high-risk cohort that had undergone treatment with *myo*-inositol, a candidate lung cancer chemopreventive agent that reduces dysplastic lesions in the airway (*46*). In contrast to sirolimus, *myo*-inositol is very well tolerated and has the potential to be taken orally for long periods of time. In individuals who responded to *myo*-inositol, as evidenced by a regression of dysplasia in the airway, we observed gene expression patterns that reflect a reduction in PI3K activity. Given the relatively limited sample size for this study, we performed in vitro studies that demonstrated that *myo*-inositol inhibits PI3K activity in a dose-dependent manner, further confirming the mechanism of action for this compound in airway epithelial cells. These results are consistent with a recent report of *myo*-inositol inhibiting activation of AKT and extracellular signal–regulated kinase in immortalized human bronchial epithelial cells (*51*). If the chemopreventive effects of *myo*-inositol are confirmed in larger clinical trials, the use of this compound in high-risk smokers with perturbed PI3K activity in the airway could decrease lung cancer occurrence. Airway gene expression profiling on these subjects after treatment may also help identify a subset of patients who would benefit from long-term therapy.

Increased activity of the PI3K pathway has been observed in many different types of malignant tissue, including lung cancer (*52*, *53*). The PI3K pathway can be constitutively activated in tumors as a result of mutations in the tyrosine kinase domain of epidermal growth factor (EGF) receptor; mutations, deletion, or suppression of the tumor suppressor PTEN; increased PI3K gene copy number (*54*); or mutations in p110α, the catalytic subunit of PI3K (*52*). We did not find evidence for loss of PTEN or increase in PI3K gene copy number in the airway of smokers with lung cancer; hence, these two mechanisms do not seem to play a significant role in the activation of PI3K in the proximal airway. We did find activation of PKC, a downstream component of the PI3K pathway, in the normal airway cells of patients with lung cancer compared to those with other pathologies. PKC activity has been previously implicated in lung cancer development (*37*). PKC activation can be downstream of growth factor receptor signaling, including insulin-like growth factor and EGF, and future studies will investigate these pathways as a mechanism for increased PI3K signaling (*40*, *41*).

Our computational approach to identifying the signaling pathways driving lung cancer oncogenesis can identify therapies that may prevent cancer development in high-risk populations. Further, our biochemical measurements correlate with computational gene expression analysis in patient samples and in dose-response studies with cell lines treated with the PI3K inhibitor *myo*-inositol. Our findings suggest that the deregulation of the PI3K pathway in the bronchial airway epithelium is an early, measurable, and reversible step in the development of lung cancer and may serve to guide chemopreventive approaches in high-risk smokers. More broadly, our findings suggest that airway gene expression reflects perturbation of specific oncogenic pathways within a smoker, potentially allowing for personalized approaches to chemoprophylaxis and therapy.

MATERIALS AND METHODS

Patient population

For the primary studies evaluating genomic pathway signatures in the proximal airway of smokers with lung cancer, airway epithelial brushings were collected from current and former smokers under suspicion of lung cancer who were undergoing diagnostic flexible bron-

choscopy from four institutions [Boston University Medical Center, Boston Veterans Administration, Lahey Clinic, and St. James's Hospital (Dublin, Ireland)] for the primary series (n = 129) and a fifth medical center (St. Elizabeth's Hospital, Boston, MA) for the additional validation series (n = 35) [previously described by Spira *et al.* (*9*)] (see demographics in Table 1 for samples used in this study). Additional brushings were collected from volunteer healthy current, former, and never smokers, as well as smokers with COPD, who were undergoing bronchoscopy at Boston University Medical Center [some previously published by Beane *et al.* (*35*); demographics in Table 2].

Brushings from the cytologically normal bronchial airway of current and former smokers with airway dysplasia were collected at the University of British Columbia from volunteers who were between 40 and 74 years of age, had ≥30 pack-years of cumulative smoking history, and had one or more sites of bronchial dysplasia on autofluorescence bronchoscopy (see demographics in Table 3). A subset of these volunteers was treated with *myo*-inositol (18 g daily by mouth) for a period of 2 to 3 months, and additional brushings were collected from cytologically normal airway epithelium at the end of the treatment. Autofluorescence bronchoscopy and endobronchial biopsy were used to measure changes in dysplasia (n = 20 samples, 10 individuals) (*46*). Response to treatment with *myo*-inositol was classified by tracking the progression, regression, or stability in airway dysplasia. Complete response to treatment was defined as a regression in all dysplastic lesions to lesions that were classified as either hyperplastic or normal, as well as the appearance of no new dysplastic lesions that were mild or worse. Progressive disease was defined by the appearance of new lesions classified as mild or worse, or the worsening of existing lesions by two or more grades (for example, mild dysplasia to severe dysplasia). Partial response was defined as the regression of some, but not all, of the dysplastic lesions. Here, nonresponders to *myo*-inositol were defined as participants with either progressive disease or partial response. Biopsies were reviewed by a lung pathologist who was blinded to the treatment.

Prospective samples used for biochemical validation were collected at both Boston University Medical Center and the University of Utah Hospital. Cytologically normal bronchial airway brushings were collected from the mainstem bronchus of patients undergoing bronchoscopy for clinical suspicion of lung cancer (see table S2). Patients were followed after bronchoscopy until a final diagnosis of lung cancer or alternate lung pathology was made. Additionally, for the Western blot studies of phospho-AKT and PKC, endobronchial biopsies were obtained and snap-frozen from the mainstem bronchus of smokers with and without lung cancer at Vanderbilt-Ingram Comprehensive Cancer Center and Nashville Veterans Affairs Medical Center (VAMC) (table S2). Finally, airway DNA for the PTEN LOH and PI3K amplification assays was obtained from brushings of the mainstem bronchus on smokers with and without lung cancer at Boston University Medical Center and Vanderbilt-Ingram Comprehensive Cancer Center–Nashville VAMC (table S3).

The study was approved by the Institutional Review Board of all participating institutions, and all study participants provided written informed consent.

Sample collection and RNA processing

Cytologically normal airway epithelial samples from smokers with and without cancer (n = 129 for primary series and n = 35 for additional validation series, GSE4115), as well as current, former, and never smokers (n = 104, GSE7895) and smokers with (n = 17) and without (n = 19) COPD (subset of GSE4115, GSE7895, and GSE12815), were collected and hybridized onto Affymetrix HG-U133A microarrays as described (*9*, *35*). For lung tumor and adjacent normal studies, GSE10072 was used (*36*). Airway samples collected prospectively for biochemical validation were snap-frozen in liquid nitrogen.

Cytologically normal airway epithelial samples from patients with dysplastic airway lesions at the University of British Columbia were collected before and after 2 to 3 months of treatment with *myo*-inositol (n = 20, 10 patients with two samples each), as well as six additional samples collected before treatment with *myo*-inositol as part of a dose-response study. In that study, bronchial brushing was performed in three separate sixth- to eighth-generation (that is, division of airway branching) bronchial airways using a 1.7-mm-diameter bronchial cytology brush (Hobbs Medical). The brush was retrieved and immediately immersed in RNAlater and kept frozen at −80°C until assayed. Epithelial cell content of representative bronchial brushing samples has been quantified by cytocentrifugation (Thermo Shandon Cytospin) of the cell pellet and staining with a cytokeratin antibody (Signet). The cells in the bronchial brush contained >90% bronchial epithelial cells. Each sample (1 μg) was later hybridized to Affymetrix Human Exon ST microarrays according to the manufacturer's protocol. Data from exon arrays were normalized using RMA-sketch in the Affymetrix Expression Console software.

All new microarray data for this study are available for download at Gene Expression Omnibus (GEO) under accession number GSE12815.

Oncogenic pathway activation probability calculation

Ras, Myc, E2F3, Src, and β-catenin oncogenic pathway gene expression data from a previous study were used, and signatures for use in binary regression models were calculated as detailed previously (*11*). We used human primary mammary epithelial cell cultures to develop the PI3K and ΔNp63 pathway signatures. Recombinant adenoviruses were used to express the p110α isoform of PI3K and ΔNp63 in an otherwise quiescent cell, thereby specifically isolating the events subsequent to activation or deregulation of pathway activity. Eighteen hours after infection, cells were collected for both RNA and protein isolation, and expression of the PI3K p110α subunit and ΔNp63 protein and activation of their downstream targets, phospho-AKT and p21, respectively, were determined by standard Western blotting (fig. S1). RNA from multiple independent infections was collected for DNA microarray analysis using the Affymetrix Human Genome U133 Array. Microarray data were normalized by RMA (robust multiarray average) and then corrected for batch effects with distance-weighted discrimination (DWD) (*55*). Specifically, to standardize expression data for the development of regression models, DWD was applied to correct batch effects between the oncogenic pathway signature microarray samples and bronchial airway microarray samples. For the oncogenic signatures, only control samples were used to correct for batch effects.

Before statistical modeling, gene expression data were filtered to exclude probe sets with signals present at background noise levels and for probe sets that did not vary significantly across samples. A multigene signature was developed to represent the activation of a particular pathway based on first identifying the genes that varied in expression between the control cells and the cells with pathway activity. The expression of these genes in any sample was then summarized as a single value or metagene score corresponding to the value from the first principal component as determined by a singular-value decomposition.

Given a training set of metagene scores from samples representing two biological states (for example, pathway-activated and quiescent control), a binary probit regression model was estimated using Bayesian methods. Applied to metagene scores calculated from gene expression data from a new sample, the model returned a probability for that sample being from either of the two states, which is a measure of how strongly the pathway was activated or repressed in that sample on the basis of gene expression pattern (*11*).

Resulting pathway activity predictions from the gene expression observed in the samples of interest were scaled between 0 and 1. To determine whether a pathway exhibits differential pathway activity between conditions (for example, cancer versus normal), we used a Wilcoxon rank-sum test. For P values <0.05, a random permutation analysis was then performed. During random permutation, gene identifiers in the data set of interest (for example, bronchial airway) were randomized, and a P value from a Wilcoxon rank-sum test was calculated to measure differential activation between class variables (for example, lung cancer versus no lung cancer). This was repeated 1000 times, so the most significant random permutation P value from this analysis was <0.001. The reported P values are the least significant of either the random permutation P value or the original P value.

Analysis of the PTEN breast cancer data set [GSE13787 (*32*)] and the in vitro *myo*-inositol dose-response microarray data set (see Materials and Methods) was accomplished using methodologies similar to those used in the analysis of the airway epithelial data set. Specifically, models were trained on the PI3K gene expression data and applied to the PTEN data set, followed by statistical analysis of differential pathway activity between "PTEN loss" samples and "PTEN normal" using a Wilcoxon rank-sum test. The PTEN data set used in this analysis differs from the one available for download from GEO and was provided by the authors along with a more detailed sample annotation.

Predicted pathway activation probabilities used in this analysis and software to calculate them are available online at http://www.pulmonomics.com/papers/pi3k.

Gene set enrichment analysis

GSEA (*31*) was calculated with GSEA v2. The genes identified as being either induced or repressed by PI3K activation in vitro were used to establish two separate gene sets. Ranking of genes for use in GSEA was done in the following ways: First, for the comparison of smokers with and without dysplasia, the genes were ranked according to the value of the dysplasia coefficient in a mixed-linear model that included cumulative smoking exposure and batch differences as covariates. Two of the 16 dysplasia samples were removed before this analysis for being outliers on principal components analysis quality control. Second, for the comparison of smokers with dysplasia before and after treatment, the genes were ranked using a Wilcoxon rank-sum test. Finally, for comparing smokers with lung cancer and those without, the genes were ranked using the default GSEA metric (signal to noise). In GSEA, 100 permutations were used to calculate significance. The PI3K pathway gene set is available in table S1, and gene rankings used in this analysis are available online at http://www.pulmonomics.com/papers/pi3k.

PI3K kinase assay

Endobronchial brushings for Boston and Utah cohorts were obtained from the mainstem bronchus and snap-frozen immediately in liquid nitrogen. Cell extracts from bronchoscopy brushes were prepared by adding 200 μl radioimmunoprecipitation assay (RIPA) buffer. To facilitate the detachment of the cells from the brushes, we vortexed the tubes three times for 5 s. For cell line kinase assays, 80% confluent H1299 cells growing in Dulbecco's modified Eagle's medium with 10% fetal bovine serum (FBS) for 24 hours were treated with increasing concentrations of *myo*-inositol (Sigma) for 16 hours at 37°C. The cells were lysed in RIPA buffer [20 mM tris (pH 7.4), 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA, 0.1% SDS] containing 0.1 mM sodium orthovanadate, 2 mM phenylmethylsulfonyl fluoride, and 100 μM protease inhibitor cocktail (Sigma). Lysates from both bronchoscopies and cell lines were centrifuged at 14,000 rpm for 20 min at 4°C and incubated with monoclonal antibody to p85 PI3K (Santa Cruz Biotechnology) for 1 hour at 4°C. Bound proteins were precipitated with 50 μl of a 50% slurry of protein G–Sepharose (Sigma) and washed three times with lysis buffer; three times with buffer containing 0.1 mM tris (pH 7.4), 5 mM LiCl, and 0.1 mM sodium orthovanadate; and two times with buffer containing 10 mM tris (pH 7.4), 150 mM NaCl, 5 mM EDTA, and 0.1 mM sodium orthovanadate. The beads were washed in kinase buffer [50 mM tris (pH 7.4), 10 mM $MgCl_2$] containing 20 μM cold adenosine 5′-triphosphate (ATP) (Sigma) and resuspended in 45 μl of kinase buffer containing 5 μl of L-α-phosphatidylinositol 4,5-bisphosphate (PIP_2) (Avanti Polar Lipids) (1 mg/ml) and 20 μCi [γ-^{32}P]ATP for 20 min at room temperature. The reactions were stopped by addition of 100 μl of 1N HCl, and the lipids were extracted with 160 μl of $CHCl_3$-MeOH (1:1). The phosphorylated products were separated by thin-layer chromatography on Silica 60 plates pretreated with potassium oxalate in a $CHCl_3$-MeOH-NH_4 solution (45:35:1.5). The production of phosphatidylinositol 3,4,5-trisphosphate (PIP_3) was evaluated by autoradiography and quantified by densitometry analysis. All experiments were repeated at least twice with similar results. Because the kinase assays for the airway brushings from the Boston and Utah cohorts were run on different days with different lots of ^{32}P-labeled ATP, we used an analysis of variance (ANOVA) model that incorporates batch as a variable when comparing cases and controls.

Western blot analysis

Endobronchial biopsies (Vanderbilt cohort) were obtained from the mainstem bronchus and snap-frozen immediately in liquid nitrogen. The biopsies (2 to 3 mm) were resuspended in RIPA buffer and subjected to sonication (three 5-s pulses) in ice and centrifuged at 14,000 rpm for 20 min at 4°C, and the pellets were discarded. The protein yield was quantified by the Bradford assay, and equivalent amounts of protein were electrophoresed on 7% SDS–polyacrylamide gel electrophoresis gels and transferred to nitrocellulose. The nitrocellulose membranes were blocked for 1 hour in tris-buffered saline (TBS) containing 0.1% Tween 20 and 2.5% bovine serum albumin (BSA), or TBS containing 0.1% Tween 20 and 5% low-fat milk, and placed in primary antibody diluted into TBS containing 0.1% Tween 20, 2.5% BSA, and 0.02% sodium azide overnight at 4°C. The primary antibodies used in this study are the following: rabbit phospho-PKC (pan) (βII Ser^{660}) (1:100; Cell Signaling Technology), rabbit phospho-AKT (Ser^{473}) (1:100; Cell Signaling Technology), and rabbit glyceraldehyde-3-phosphate dehydrogenase (1:1000; Abcam). Membranes were then washed three times in TBS containing 0.1% Tween 20 and/or 0.1% NP-40. Primary antibody was detected using horseradish peroxidase–linked secondary antibody and visualized with the ECL Plus Western Blot Detection System (GE Health-

care). To measure statistical differences in the Western blot experiments, we first quantified the bands (table S3) and then used a *t* test to calculate a *P* value.

Measuring PIK3CA amplification by quantitative real-time PCR

DNA was isolated from airway brushings in smokers with and without lung cancer (table S3). Copy number was determined by quantitative real-time PCR on an iCycler apparatus (Bio-Rad). Tumor DNA content was normalized to that of Line-1, a repetitive element whose copy number per haploid genome is similar among normal and neoplastic tissues. All PCR amplifications were carried out in triplicate, and the threshold cycle numbers were averaged. The experiments were repeated at least twice. The error bars correspond to the SD of the two independent experiment days in triplicates. The following primers were used: PIK3CA, 5′-TCAGATTTACGGCAAGATATGC-3′ (forward) and 5′-TGCCTTACTGGTTACCTACCG-3′ (reverse); Line-1, 5′-AAAGCCGCTCAACTACATGG-3′ (forward) and 5′-TGCTTTGAATGCGTCCCAGAG-3′ (reverse). Real-time PCR conditions and copy number calculations were done as described previously (*45*).

LOH analysis of PTEN

Bronchial genomic DNA from a subset of patients with or without lung cancer was assessed for LOH of *PTEN*. We used TaqMan copy number assays (Applied Biosystems) to analyze two regions of the *PTEN* gene. The location of the first set of primers was Chr10:89632002 [National Center for Biotechnology Information (NCBI) build 36] within intron 1. The second set was located at Chr10:89669070 (NCBI build 36) within intron 2. Real-time PCR was performed using the LightCycler 480 System (Roche). In both assays, ribonuclease P (RNase P) was coamplified as an internal reference gene. For each assay, the following components were added to a total of 10 μl per reaction: 10 ng of template DNA, 1× TaqMan Genotyping Master Mix, 1× PTEN assay mix, and 1× RNase reference assay mix. Thermal cycling was initiated with a polymerase activation for 10 min at 95°C. After this, 40 cycles of 15-s denaturation at 95°C and 60-s elongation at 60°C followed. FAM signals of RNase P amplification were detected by 483 to 533 nm, and HEX signals of PTEN assays were detected by 523 to 568 nm. Analysis was done using advanced relative quantification with LightCycler 480 software.

In vitro genomic drug response assays

Following treatment with *myo*-inositol, cells were subjected to RNA extraction with RNeasy Mini kit (Qiagen). RNA quality was determined by using the Experion RNA StdSens Analysis kit (Bio-Rad). RNA was processed and hybridized onto the Affymetrix HG-U133A 2.0 array, and the microarray data were MAS5.0-normalized and logged before analysis.

SUPPLEMENTARY MATERIAL

www.sciencetranslationalmedicine.org/cgi/content/full/2/26/26ra25/DC1

Fig. S1. Generation of PI3K and ΔNp63 pathway signatures.

Fig. S2. Heat map of in vitro defined PI3K pathway signature.

Fig. S3. Computationally predicted PI3K activity in breast cancer samples with known PTEN copy number status.

Fig. S4. Gene expression changes after PI3K activation in vitro that are known PI3K pathway family members.

Fig. S5. PTEN loss and PIK3CA amplification in the airway of smokers with and without lung cancer.

Table S1. Pathway gene lists for the in vitro derived PI3K pathway.

Table S2. Patient demographics for prospective series of smokers undergoing bronchoscopy for suspicion of lung cancer.

Table S3. Quantified Western blot analysis for PI3K pathway components.

Table S4. Demographics for DNA airway cohorts used for PTEN LOH and PI3K amplification studies.

Table S5. List of PI3K signature genes that had counterparts on the Affymetrix exon array.

Table S6. Ordering of genes that make up the heat map in Fig. 5.

REFERENCES AND NOTES

1. U.S. Department of Health and Human Services, The health consequences of smoking: A report of the Surgeon General—Executive summary (U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Rockville, 2004).
2. C. A. Powell, S. Klares, G. O'Connor, J. S. Brody, Loss of heterozygosity in epithelial cells obtained by bronchial brushing: Clinical utility in lung cancer. *Clin. Cancer Res.* **5**, 2025–2034 (1999).
3. I. I. Wistuba, S. Lam, C. Behrens, A. K. Virmani, K. M. Fong, J. LeRiche, J. M. Samet, S. Srivastava, J. D. Minna, A. F. Gazdar, Molecular damage in the bronchial epithelium of current and former smokers. *J. Natl. Cancer Inst.* **89**, 1366–1373 (1997).
4. M. Guo, M. G. House, C. Hooker, Y. Han, E. Heath, E. Gabrielson, S. C. Yang, S. B. Baylin, J. G. Herman, M. V. Brock, Promoter hypermethylation of resected bronchial margins: A field defect of changes? *Clin. Cancer Res.* **10**, 5131–5136 (2004).
5. W. A. Franklin, A. F. Gazdar, J. Haney, I. I. Wistuba, F. G. La Rosa, T. Kennedy, D. M. Ritchey, Y. E. Miller, Widely dispersed p53 mutation in respiratory epithelium. A novel mechanism for field carcinogenesis. *J. Clin. Invest.* **100**, 2133–2137 (1997).
6. Y. M. Miyazu, T. Miyazawa, K. Hiyama, N. Kurimoto, Y. Iwamoto, H. Matsuura, K. Kanoh, N. Kohno, M. Nishiyama, E. Hiyama, Telomerase expression in noncancerous bronchial epithelia is a possible marker of early development of lung cancer. *Cancer Res.* **65**, 9623–9627 (2005).
7. O. Auerbach, E. C. Hammond, L. Garfinkel, Histologic changes in the larynx in relation to smoking habits. *Cancer* **25**, 92–104 (1970).
8. A. Spira, J. Beane, V. Shah, G. Liu, F. Schembri, X. Yang, J. Palma, J. S. Brody, Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 10143–10148 (2004).
9. A. Spira, J. E. Beane, V. Shah, K. Steiling, G. Liu, F. Schembri, S. Gilman, Y. M. Dumas, P. Calner, P. Sebastiani, S. Sridhar, J. Beamis, C. Lamb, T. Anderson, N. Gerry, J. Keane, M. E. Lenburg, J. S. Brody, Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat. Med.* **13**, 361–366 (2007).
10. J. Beane, P. Sebastiani, T. H. Whitfield, K. Steiling, Y. M. Dumas, M. E. Lenburg, A. Spira, A prediction model for lung cancer diagnosis that integrates genomic and clinical features. *Cancer Prev. Res.* **1**, 56–64 (2008).
11. A. H. Bild, G. Yao, J. T. Chang, Q. Wang, A. Potti, D. Chasse, M. B. Joshi, D. Harpole, J. M. Lancaster, A. Berchuck, J. A. Olson Jr., J. R. Marks, H. K. Dressman, M. West, J. R. Nevins, Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* **439**, 353–357 (2006).
12. J. Massagué, Sorting out breast-cancer gene signatures. *N. Engl. J. Med.* **356**, 294–297 (2007).
13. F. Huang, K. Reeves, X. Han, C. Fairchild, S. Platero, T. W. Wong, F. Lee, P. Shaw, E. Clark, Identification of candidate molecular markers predicting sensitivity in solid tumors to dasatinib: Rationale for patient selection. *Cancer Res.* **67**, 2226–2238 (2007).
14. M. Sato, D. S. Shames, A. F. Gazdar, J. D. Minna, A translational view of the molecular pathogenesis of lung cancer. *J. Thorac. Oncol.* **2**, 327–343 (2007).
15. D. R. Rhodes, S. Kalyana-Sundaram, S. A. Tomlins, V. Mahavisno, N. Kasper, R. Varambally, T. R. Barrette, D. Ghosh, S. Varambally, A. M. Chinnaiyan, Molecular concepts analysis links tumors, pathways, mechanisms, and drugs. *Neoplasia* **9**, 443–454 (2007).
16. R. S. Finn, J. Dering, C. Ginther, C. A. Wilson, P. Glaspy, N. Tchekmedyian, D. J. Slamon, Dasatinib, an orally active small molecule inhibitor of both the *src* and *abl* kinases, selectively inhibits growth of basal-type/"triple-negative" breast cancer cell lines growing in vitro. *Breast Cancer Res. Treat.* **105**, 319–326 (2007).
17. S. Horvath, B. Zhang, M. Carlson, K. V. Lu, S. Zhu, R. M. Felciano, M. F. Laurance, W. Zhao, S. Qi, Z. Chen, Y. Lee, A. C. Scheck, L. M. Liau, H. Wu, D. H. Geschwind, P. G. Febbo, H. I. Kornblum, T. F. Cloughesy, S. F. Nelson, P. S. Mischel, Analysis of oncogenic signaling networks in glioblastoma identifies *ASPM* as a molecular target. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17402–17407 (2006).

18. L. A. Garraway, H. R. Widlund, M. A. Rubin, G. Getz, A. J. Berger, S. Ramaswamy, R. Beroukhim, D. A. Milner, S. R. Granter, J. Du, C. Lee, S. N. Wagner, C. Li, T. R. Golub, D. L. Rimm, M. L. Meyerson, D. E. Fisher, W. R. Sellers, Integrative genomic analyses identify *MITF* as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
19. J. Baselga, Targeting tyrosine kinases in cancer: The second wave. *Science* **312**, 1175–1178 (2006).
20. J. G. Paez, P. A. Jänne, J. C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F. J. Kaye, N. Lindeman, T. J. Boggon, K. Naoki, H. Sasaki, Y. Fujii, M. J. Eck, W. R. Sellers, B. E. Johnson, M. Meyerson, *EGFR* mutations in lung cancer: Correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
21. L. J. van't Veer, R. Bernards, Enabling personalized cancer medicine through analysis of gene-expression patterns. *Nature* **452**, 564–570 (2008).
22. B. T. Hennessy, D. L. Smith, P. T. Ram, Y. Lu, G. B. Mills, Exploiting the PI3K/AKT pathway for cancer drug discovery. *Nat. Rev. Drug Discov.* **4**, 988–1004 (2005).
23. M. Marinov, B. Fischer, A. Arcaro, Targeting mTOR signaling in lung cancer. *Crit. Rev. Oncol. Hematol.* **63**, 172–182 (2007).
24. T. Takahashi, M. M. Nau, I. Chiba, M. J. Birrer, R. K. Rosenberg, M. Vinocour, M. Levitt, H. Pass, A. F. Gazdar, J. D. Minna, p53: A frequent target for genetic abnormalities in lung cancer. *Science* **246**, 491–494 (1989).
25. J. Downward, Targeting RAS signalling pathways in cancer therapy. *Nat. Rev. Cancer* **3**, 11–22 (2003).
26. U. Grepmeier, W. Dietmaier, J. Merk, P. J. Wild, E. C. Obermann, M. Pfeifer, F. Hofstaedter, A. Hartmann, M. Woenckhaus, Deletions at chromosome 2q and 12p are early and frequent molecular alterations in bronchial epithelium and NSCLC of long-term smokers. *Int. J. Oncol.* **27**, 481–488 (2005).
27. E. P. Black, T. Hallstrom, H. K. Dressman, M. West, J. R. Nevins, Distinctions in the specificity of E2F function revealed by gene expression signatures. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15948–15953 (2005).
28. E. P. Black, E. Huang, H. Dressman, R. Rempel, N. Laakso, S. L. Asa, S. Ishida, M. West, J. R. Nevins, Distinct gene expression phenotypes of cells lacking Rb and Rb family members. *Cancer Res.* **63**, 3716–3723 (2003).
29. E. Huang, S. Ishida, J. Pittman, H. Dressman, A. Bild, M. Kloos, M. D'Amico, R. G. Pestell, M. West, J. R. Nevins, Gene expression phenotypic models that predict the activity of oncogenic pathways. *Nat. Genet.* **34**, 226–230 (2003).
30. M. Kanehisa, M. Araki, S. Goto, M. Hattori, M. Hirakawa, M. Itoh, T. Katayama, S. Kawashima, S. Okuda, T. Tokimatsu, Y. Yamanishi, KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* **36**, D480–D484 (2008).
31. A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, J. P. Mesirov, Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).
32. B. Marty, V. Maire, E. Gravier, G. Rigaill, A. Vincent-Salomon, M. Kappler, I. Lebigot, F. Djelti, A. Tourdès, P. Gestraud, P. Hupé, E. Barillot, F. Cruzalegui, G. C. Tucker, M. H. Stern, J. P. Thiery, J. A. Hickman, T. Dubois, Frequent PTEN genomic alterations and activated phosphatidylinositol 3-kinase pathway in basal-like breast cancer cells. *Breast Cancer Res.* **10**, R101 (2008).
33. P. Michl, J. Downward, Mechanisms of disease: PI3K/AKT signaling in gastrointestinal cancers. *Z. Gastroenterol.* **43**, 1133–1139 (2005).
34. S. M. Planchon, K. A. Waite, C. Eng, The nuclear affairs of PTEN. *J. Cell Sci.* **121**, 249–253 (2008).
35. J. Beane, P. Sebastiani, G. Liu, J. S. Brody, M. E. Lenburg, A. Spira, Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol.* **8**, R201 (2007).
36. M. T. Landi, T. Dracheva, M. Rotunno, J. D. Figueroa, H. Liu, A. Dasgupta, F. E. Mann, J. Fukuoka, M. Hames, A. W. Bergen, S. E. Murphy, P. Yang, A. C. Pesatori, D. Consonni, P. A. Bertazzi, S. Wacholder, J. H. Shih, N. E. Caporaso, J. Jen, Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One* **3**, e1651 (2008).
37. R. P. Regala, R. K. Davis, A. Kunz, A. Khoor, M. Leitges, A. P. Fields, Atypical protein kinase Cι is required for bronchioalveolar stem cell expansion and lung tumorigenesis. *Cancer Res.* **69**, 7603–7611 (2009).
38. M. A. Gorin, Q. Pan, Protein kinase Cε: An oncogene and emerging tumor biomarker. *Mol. Cancer* **8**, 9 (2009).
39. L. Zhao, P. K. Vogt, Class I PI3K in oncogenic cellular transformation. *Oncogene* **27**, 5486–5496 (2008).
40. R. V. Farese, M. P. Sajan, M. L. Standaert, Atypical protein kinase C in insulin action and insulin resistance. *Biochem. Soc. Trans.* **33**, 350–353 (2005).
41. Q. W. Fan, C. Cheng, Z. A. Knight, D. Haas-Kogan, D. Stokoe, C. D. James, F. McCormick, K. M. Shokat, W. A. Weiss, EGFR signals to mTOR through PKC and independently of Akt in glioma. *Sci. Signal.* **2**, ra4 (2009).
42. S. Petersen, J. Rudolf, U. Bockmühl, K. Gellert, G. Wolf, M. Dietel, I. Petersen, Distinct regions of allelic imbalance on chromosome 10q22-q26 in squamous cell carcinomas of the lung. *Oncogene* **17**, 449–454 (1998).
43. D. H. Teng, R. Hu, H. Lin, T. Davis, D. Iliev, C. Frye, B. Swedlund, K. L. Hansen, V. L. Vinson, K. L. Gumpper, L. Ellis, A. El-Naggar, M. Frazier, S. Jasser, L. A. Langford, J. Lee, G. B. Mills, M. A. Pershouse, R. E. Pollack, C. Tornos, P. Troncoso, W. K. Yung, G. Fujii, A. Berson, R. Bookstein, J. B. Bolen, S. V. Tavtigian, P. A. Steck, *MMAC1/PTEN* mutations in primary tumor specimens and tumor cell lines. *Cancer Res.* **57**, 5221–5225 (1997).
44. Y. L. Li, Z. Tian, D. Y. Wu, B. Y. Fu, Y. Xin, Loss of heterozygosity on 10q23.3 and mutation of tumor suppressor gene PTEN in gastric cancer and precancerous lesions. *World J. Gastroenterol.* **11**, 285–288 (2005).
45. T. L. Wang, C. Maierhofer, M. R. Speicher, C. Lengauer, B. Vogelstein, K. W. Kinzler, V. E. Velculescu, Digital karyotyping. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 16156–16161 (2002).
46. S. Lam, A. McWilliams, J. LeRiche, C. MacAulay, L. Wattenberg, E. Szabo, A phase I study of *myo*-inositol for lung cancer chemoprevention. *Cancer Epidemiol. Biomarkers Prev.* **15**, 1526–1531 (2006).
47. C. Huang, W. Y. Ma, S. S. Hecht, Z. Dong, Inositol hexaphosphate inhibits cell transformation and activator protein 1 activation by targeting phosphatidylinositol-3′ kinase. *Cancer Res.* **57**, 2873–2878 (1997).
48. E. Piccolo, S. Vignati, T. Maffucci, P. F. Innominato, A. M. Riley, B. V. Potter, P. P. Pandolfi, M. Broggini, S. Iacobelli, P. Innocenti, M. Falasca, Inositol pentakisphosphate promotes apoptosis through the PI 3-K/Akt pathway. *Oncogene* **23**, 1754–1765 (2004).
49. T. Kamalati, H. E. Jolin, M. J. Fry, M. R. Crompton, Expression of the BRK tyrosine kinase in mammary epithelial cells enhances the coupling of EGF signalling to PI 3-kinase and Akt, via *erb*B3 phosphorylation. *Oncogene* **19**, 5471–5476 (2000).
50. J. A. Engelman, P. A. Jänne, C. Mermel, J. Pearlberg, T. Mukohara, C. Fleet, K. Cichowski, B. E. Johnson, L. C. Cantley, ErbB-3 mediates phosphoinositide 3-kinase activity in gefitinib-sensitive non-small cell lung cancer cell lines. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 3788–3793 (2005).
51. W. Han, J. J. Gills, R. M. Memmott, S. Lam, P. A. Dennis, The chemopreventive agent myoinositol inhibits Akt and extracellular signal-regulated kinase in bronchial lesions from heavy smokers. *Cancer Prev. Res.* **2**, 370–376 (2009).
52. A. G. Bader, S. Kang, L. Zhao, P. K. Vogt, Oncogenic PI3K deregulates transcription and translation. *Nat. Rev. Cancer* **5**, 921–929 (2005).
53. A. Arcaro, A. S. Guerreiro, The phosphoinositide 3-kinase pathway in human cancer: Genetic alterations and therapeutic implications. *Curr. Genomics* **8**, 271–306 (2007).
54. H. Yamamoto, H. Shigematsu, M. Nomura, W. W. Lockwood, M. Sato, N. Okumura, J. Soh, M. Suzuki, I. I. Wistuba, K. M. Fong, H. Lee, S. Toyooka, H. Date, W. L. Lam, J. D. Minna, A. F. Gazdar, *PIK3CA* mutations and copy number gains in human lung cancers. *Cancer Res.* **68**, 6913–6921 (2008).
55. M. Benito, J. Parker, Q. Du, J. Wu, D. Xiang, C. M. Perou, J. S. Marron, Adjustment of systematic microarray data biases. *Bioinformatics* **20**, 105–114 (2004).
56. M. Whitman, D. R. Kaplan, B. Schaffhausen, L. Cantley, T. M. Roberts, Association of phosphatidylinositol kinase activity with polyoma middle-T competent for transformation. *Nature* **315**, 239–242 (1985).

57. **Acknowledgments:** We thank W. Akerley, B. Walsh, F. Schembri, and K. Steiling for their helpful contributions to this work. Tissue was collected and disbursed with assistance from the Tissue Resource and Applications Core facility at the University of Utah. **Funding:** NIH–National Cancer Institute grant R01CA124640, Doris Duke Charitable Foundation, Vanderbilt SPORE in Lung Cancer grant CA90949, and NIH–National Institute of Environmental Health Sciences grant U01ES016035. **Author contributions:** A.M.G., A.H.B., and A.S. conceived the study; A.M.G. and A.H.B. performed the computational analysis; R.S., A.H.B., X.Z., J.Q., K.C., and G.L. performed the experiments; A.M.G., C.A., M.B.S., D.W., A.M., P.P.M., S.L., and A.S. collected patient samples; A.M.G., R.S., A.H.B., and A.S. wrote the manuscript; J.B., E.S., S.L., and M.E.L. contributed to the study design and editing of manuscript. **Competing interests:** A.S. is a founder of, is on the Board of Directors of, and is a consultant for Allegro Diagnostics, a company specializing in gene expression molecular diagnostic testing for lung cancer and other diseases. J.B. has served as Chief Scientific Officer for and owns equity in Allegro Diagnostics. M.E.L. also owns equity in and is a scientific advisor to the company. A provisional patent related to PI3K gene signature generation and its use, as reported in this article, has been filed by the Boston University and the University of Utah. **Accession numbers:** All new microarray data for this study are available for download at GEO under accession number GSE12815.

Submitted 6 July 2009
Accepted 19 March 2010
Published 7 April 2010
10.1126/scitranslmed.3000251

Citation: A. M. Gustafson, R. Soldi, C. Anderlind, M. B. Scholand, J. Qian, X. Zhang, K. Cooper, D. Walker, A. McWilliams, G. Liu, E. Szabo, J. Brody, P. P. Massion, M. E. Lenburg, S. Lam, A. H. Bild, A. Spira, Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Sci. Transl. Med.* **2**, 26ra25 (2010).