A Meta-analysis of Lung Cancer Gene Expression Identifies PTK7 as a Survival Gene in Lung Adenocarcinoma

Ron Chen1, Purvesh Khatri2,3, Pawel K. Mazur1, Melanie Polin1, Yanyan Zheng1, Dedeepya Vaka1, Chuong D. Hoang1, Joseph Shrager1, Yue Xu4, Silvestre Vicent1, Atul J. Butte5, and E. Alejandro Sweet-Cordero1

Abstract

Lung cancer remains the most common cause of cancer-related death worldwide and it continues to lack effective treatment. The increasingly large and diverse public databases of lung cancer gene expression constitute a rich source of candidate oncogenic drivers and therapeutic targets. To define novel targets for lung adenocarcinoma, we conducted a large-scale meta-analysis of genes specifically overexpressed in adenocarcinoma. We identified an 11-gene signature that was overexpressed consistently in adenocarcinoma specimens relative to normal lung tissue. Six genes in this signature were specifically overexpressed in adenocarcinoma relative to other subtypes of non–small cell lung cancer (NSCLC). Among these genes was the little studied protein tyrosine kinase PTK7. Immunohistochemical analysis confirmed that PTK7 is highly expressed in primary adenocarcinoma patient samples. RNA interference–mediated attenuation of PTK7 decreased cell viability and increased apoptosis in a subset of adenocarcinoma cell lines. Further, loss of PTK7 activated the MKK7–JNK stress response pathway and impaired tumor growth in xenotransplantation assays. Our work defines PTK7 as a highly and specifically expressed gene in adenocarcinoma and a potential therapeutic target in this subset of NSCLC. Cancer Res; 74(10); 2892–902. ©2014 AACR.

Introduction

Lung cancer is the leading cause of cancer death in the United States and worldwide (1). Despite intensive basic and clinical research, the overall 5-year survival rate of the major histologic subtype, non–small cell lung cancer (NSCLC) has only improved from 14% to 18% since 1975 (1). Recently, targeted treatment based on patient-specific molecular aberrations has led to significant response rates in subsets of patients with NSCLC (2, 3). However, about half of all patients do not harbor known "driver" mutations and cannot be treated with targeted agents (4). Thus, new approaches for identification of novel regulators and potential targets for identification of treatment for lung cancer are needed.

Gene expression analysis has been used to classify cancers, predict clinical outcomes, and discover disease-associated biomarkers. However, gene expression experiments are usually analyzed in isolation and are limited to a small number of samples. Meta-analysis approaches make it possible to combine multiple gene expression datasets and increase the statistical power for gene discovery. Such meta-analysis approaches have been successfully used for cancers of the breast (5, 6), prostate (7), liver (8), and lung (9), as well as broadly across cancers (10–12). Statistically, individual studies of gene expression in cancer are limited by both biologic (e.g., sampling of a particular patient population) and technical (e.g., only using one expression analysis platform) biases that hinder the broader application of their findings and ultimate translation into clinical practice. Meta-analysis can control for such confounding factors by increasing the statistical power to detect consistent changes across multiple datasets. Several groups have made available large NSCLC gene expression datasets that consist of tumor-to-normal comparisons (9, 13–21). These datasets represent a large and as yet not fully tapped resource for discovering novel genes relevant to the pathogenesis of lung cancer. We reasoned that a careful meta-analysis that combined multiple patient populations across many institutions, platforms, and data procurement methods would uncover genes with functional relevance to lung cancer that may have been otherwise overlooked by the isolated analysis of individual gene expression studies.

We applied a recently proposed meta-analysis approach (22) to 13 gene expression datasets consisting of 2,026 lung samples, which enabled the discovery and validation of commonly overexpressed genes in lung adenocarcinoma, the predominant subtype of NSCLC. Among the most consistently overexpressed
genes was PTK7, a member of the receptor tyrosine kinase family conserved across Hydra, Drosophila, Japanese puffer fish, chicken, and human (23). PTK7 contains seven immunoglobulin (Ig) domains, a transmembrane domain, and a catalytically inactive kinase domain within the cytoplasmic tail (24). It was first discovered in melanocytes (24) and subsequently found to be overexpressed in colon carcinoma (25). In Xenopus, PTK7 acts at the level of Frizzled and Dishevelled to regulate the Wnt/planar cell polarity (PCP) pathway (26). Ptk7 knockout mice die perinatally and display developmental defects of the inner ear and of neural tube closure, consistent with a role in regulating PCP (27). Overexpression of PTK7 has been described in several cancers, including colon (25), gastric (28), esophageal (29), and acute myelogenous leukemia (AML; ref. 30). However, the precise role of PTK7 in regulating oncogenesis remains unclear. In colon cancer, PTK7 may play a role in the Wnt/β-catenin pathway (31), and knockdown leads to caspase-10–mediated apoptosis (32).

Consistent with the gene expression analysis, we found consistently elevated expression of PTK7 protein in primary adenocarcinoma patient samples. Knockdown of PTK7 demonstrated that it is essential for the viability of a subset of NSCLC cell lines, and PTK7 disruption increased Mapk kinase-mediated apoptosis (32). Xenotransplantation studies revealed the requirement of PTK7 for the viability of a subset of adenocarcinoma patient samples. Knockdown of PTK7 demonstrated that it is essential for the viability of a subset of NSCLC cell lines, and PTK7 disruption increased Mapk kinase-mediated apoptosis (32).

Materials and Methods

Data collection, preprocessing, and normalization

Gene expression data for 13 human lung cancer studies were downloaded from the National Center for Biotechnology Information Gene Expression Omnibus (GEO; accession numbers GSE10072, GSE2514, GSE7670, GSE19188, GSE11969, GSE21933, GSE42127, GSE41271, GSE37745, GSE28571, and GSE20853) and websites http://www.broadinstitution.org/mpr/lung/ and https://array.nci.nih.gov/caarray/project/details.action?project.id=182. The histologic phenotypes were defined as the corresponding original publications. Datasets were curated to include only normal, adenocarcinoma, squamous cell carcinoma (SCC), and large cell carcinoma (LCC) samples. All datasets were normalized individually using GeneSpring (Agilent; ref. 33). For PTK7 gene expression, RNA was extracted and prepared using established protocols and hybridized to Affymetrix Human Gene 1.0 ST Array. Raw data are available in GEO (GSE50138).

Meta-analysis of gene expression data

Two meta-analysis approaches were applied to the normalized data (22). The first approach combines effect sizes from each dataset into a meta-effect size to estimate the amount of change in expression across all datasets. For each gene in each dataset, an effect size was computed using Hedges adjusted g. If multiple probes mapped to a gene, the effect size for each gene was summarized using the fixed effect inverse-variance model. Next, study-specific effect sizes were combined to obtain the pooled effect size and its standard error using the random effects inverse-variance technique. The z-statistic was computed as a ratio of the pooled effect size to its standard error for each gene, and the result was compared with a standard normal distribution to obtain a nominal P value. P values were corrected for multiple hypotheses testing using the Benjamini–Hochberg correction (34).

A second nonparametric meta-analysis that combines P values from individual experiments to identify genes with a large effect size in all datasets was also used. A z-statistic was calculated for each gene in each study. After computing one-tail P values for each gene, these were corrected for multiple hypotheses using the Benjamini–Hochberg correction. Next, Fisher sum of logs method (35) was applied. Briefly, this method sums the logarithm of corrected P values across all datasets for each gene, and compares the sum against a χ2 distribution with 2k degrees of freedom, in which k is the number of datasets used in the analysis.

Leave-one-out validation and classification

To control for the influence of single large experiments on the meta-analysis results, leave-one-out meta-analysis was performed. One dataset at a time was excluded and both meta-analysis methods were applied to the remaining datasets. We hypothesized that the minimal set of genes that are significantly overexpressed, irrespective of the set of datasets analyzed, would constitute a robust gene expression signature of adenocarcinoma across multiple independent cohorts. A very stringent threshold [false discovery rate (FDR) ≤ 1 × 10−5] for selecting differentially overexpressed genes in adenocarcinoma was used. Furthermore, we analyzed heterogeneity of the effect sizes across all studies. Genes with significant heterogeneity (P < 0.05) were removed from the overexpressed genes identified using stringent FDR criteria. The geometric mean of the remaining significant genes was computed and used to create a univariate binomial linear model for classifying a lung sample as a normal or adenocarcinoma sample, or as adenocarcinoma or SCC sample.

Immunohistochemistry

Immunohistochemistry was performed as previously described (36) with the following antibodies: rabbit antibody to phospho-histone H3 (pH3; 1:500; Upstate), rabbit antibody to cleaved caspase-3 (CC3; 1:400; Cell Signaling Technology; 9664), rabbit antibody to PTK7 (1:1,000; Sigma; SAB3500340). PTK7 staining was performed using pepsin antigen retrieval.

Human primary adenocarcinoma samples

This study complied with federal, state, and local regulations of the Human Research Protection Program and was approved by the Stanford Institutional Research Board. Informed consent was obtained from all patients included in the study.

Establishment of patient-derived xenograft tumors from primary human adenocarcinoma

Surgically removed human NSCLC tumor tissues were kept in ice-cold Hank’s Balanced Salt Solution (Life Technologies) until use. Tumors were cut into 1-mm pieces and implanted in...
the subrenal capsules in NOD-SCID-IL2Rg (NSG) mice (The Jackson Laboratory).

**Tissue microarray**

Lung adenocarcinoma tissue microarray with normal lung tissue, containing 20 cases of lung adenocarcinoma and 10 normal lung tissue (BC04119b; US Biomat) was immunohistochemically stained for PTK7. Staining was scored as negative (0), weak (1), or strong (2).

**Cell culture**


**Short hairpin RNA and virus production**

Human short hairpin RNA (shRNA) constructs against PTK7 were purchased from OpenBiosystems. The human PTK7 shRNA set was cat # RHS4533-NN_008281. Of this set, TRCN0000006434 and TRCN0000006435 were labeled as shPTK7–1 and 2, respectively. Control hairpins against GFP and luciferase in the pLKO.1 backbone were used (see Supplementary Methods for target sequences). Transfection-quality DNA was extracted using Qiagen DNA kits. Lentiviruses were produced by transfection into 293FT cells as previously described (38), filtered, and applied directly to cells. Puromycin selection was started 2 days after lentiviral infection for a duration of 2 days at 2 μg/mL.

**Quantitative real-time PCR analysis**

RNA was isolated 5 days after lentiviral infection and puromycin selection with TRIzol reagent (Invitrogen) following the manufacturer’s protocol. CDNA was synthesized with a Dynabead cDNA Synthesis Kit (F470; New England Biolabs), and quantitative real-time PCR (qRT-PCR) was carried out in triplicate by SYBR Green (Quanta Biosciences) using a C1000 Thermal Cycler (BioRad). See Supplementary Methods for primer sequences.

**Cell proliferation assay**

Cells were trypsinized and plated in triplicate into 96-well plates (day 0). Cell viability was measured by treating cells with MTT. Absorbance measurements were recorded using a SpectraMax 340 (Molecular Devices) at 570 nm on the indicated days (Cell Proliferation Kit I; Roche). All data were normalized to experimental day 0.

**Fluorescence-activated cell sorting analysis**

Cells were scraped from the plate and single cell suspensions were made by passing cells through 28G1/2 insulin syringes (BD) 10 times. Cells were then washed in PBS and resuspended in PF10 (10% FBS in PBS), stained at 4 °C for 30 minutes in the dark with rabbit anti-PTK7 (GTI014510; GeneTex) or IgG2a control (Clone 188B isotype mouse, BD). For Annexin-V staining, cell lines were plated at 5 × 10⁵ cells per 6-cm plates and stained with Annexin-V (FITC-Annexin-V; 556419; BD Biosciences) and propidium iodide (50 pL/mL final concentration) following the manufacturer’s protocol at indicated time points. A C6 Flow Cytometer (Accuri) was used for fluorescence-activated cell sorting (FACS) analysis and the manufacturer’s software and FlowJo v5 for analysis.

**Immunoblotting**

Cells were scraped and lysed in radioimmunoprecipitation assay buffer with protease inhibitor cocktail (Roche), 25 mmol/L sodium fluoride, 1 mmol/L phenylmethylsulfonyl fluoride, and 1 mmol/L sodium orthovanadate (Sigma-Aldrich). Protein samples were resolved by SDS-PAGE, transferred to Amersham Hybond-P membranes (GE Healthcare), and blocking buffer (5% bovine serum albumin in TBS-T) for 1 hour before addition of primary antibody, which was applied overnight at 4 °C. The following antibodies from Cell Signaling Technology were used (1:1,000 dilution unless otherwise noted): rabbit anti-Cleaved PARP (#5625), mouse anti-pERK (# 9106), rabbit anti-ERK (#4695), rabbit anti-pJAK (T180/Y182; #9211), and rabbit anti-P38 (#9272), rabbit anti-p-AKT (S473; #4060), rabbit anti-AKT (#9272), rabbit anti-pJNK (T183/Y185; #9251), rabbit anti-JNK (#9252), rabbit anti-β-actin (#60008), rabbit anti-p-β-actin (S422/423; #9245), rabbit anti-actin (S422/423; #8457), rabbit anti-phospho-PTK7 (T901; # 9255), rabbit anti-PTK7 (R131; #9250). Mouse anti-β-actin (Sigma-Aldrich, clone 1A4; 1:10,000) was used as a loading control.

**Xenograft**

Cell lines infected with specific shRNA were resuspended in serum-free RPMI and injected subcutaneously at 1 million cells per tumor into the two lower flanks of nude mice (The Jackson Laboratory). Between 4 and 8 tumors were injected for each condition. One week after injection, tumor dimensions were measured approximately every 3 days and tumor volume was calculated using the formula π/6 × [(length+width)/2]³ (39). All animal experiments were approved by the Stanford University School of Medicine Committee on Animal Care.

**Statistical analyses**

Unpaired two-tailed t tests were used for comparisons between different groups. Error bars correspond to SEM. Significant P values in the text correspond to <0.05 (*), <0.01 (**), or <0.001 (***)

**Results**

**Meta-analysis of five adenocarcinoma gene expression datasets identifies 11 significantly overexpressed genes in adenocarcinoma**

We identified and manually curated seven published and publicly available adenocarcinoma datasets (9, 13–15, 17–19). Only datasets with accessible unprocessed raw data and containing both normal and adenocarcinoma samples were used for further analysis. We identified five datasets that met these criteria (Table S1; refs. 9, 14, 17–19). We applied a recently proposed meta-analysis approach (22) to these five datasets as outlined...
Because an experiment with a large number of samples can have significant influence on meta-analysis results, we performed a leave-one-out analysis, in which we excluded one dataset at a time and applied our meta-analysis approach on the remaining datasets (see Materials and Methods). We hypothesized that the minimal set of genes identified as significantly overexpressed by both methods, irrespective of the datasets used for analysis, would constitute a robust gene expression signature of adenocarcinoma across multiple independent cohorts.

Figure 1. Meta-analysis of NSCLC gene expression datasets identified 11 significantly overexpressed genes in NSCLC. A, meta-analysis and functional validation pipeline. In silico bioinformatics data analysis workflow consisted of the curation of five publicly available datasets, data preprocessing, discovery meta-analysis, independent validation, and subtype analysis followed by validation in primary human patient samples and in vitro and in vivo functional validation. B, the 11 genes cluster adenocarcinoma (ADC) versus normal across each dataset. Each column is a sample and each row is the expression level of a gene. The color scale represents the raw Z-score ranging from blue (low expression) to yellow (high expression). Dendrograms above each heatmap correspond to the Pearson correlation-based hierarchical clustering of samples by expression of the 11 genes.
identified 14 genes significantly and highly overexpressed in adenocarcinoma across all five datasets. In addition, we estimated between-studies heterogeneity in effect sizes. Three genes (ARGHAP8, CEACAM1, and DDR1) had significant between-studies heterogeneity ($P \leq 0.05$) and were removed from the further analysis because we did not have sufficient information available to further explore heterogeneity in these genes (Supplementary Table S1 and Supplementary Fig. S1). As expected, hierarchical clustering of the five discovery datasets using the remaining 11 genes distinguished normal lung from adenocarcinoma samples (Fig. 1B). We further validated the 11 genes in three additional independent datasets (16, 20, 21; 146 adenocarcinoma, 91 normal; Supplementary Table S2). All genes were significantly overexpressed in the validation datasets at FDR $\leq 5\%$ with at least one meta-analysis method, and all except three (CACNB3, SLC2A5, and SULT1C2) were significantly overexpressed as assessed by both methods (Supplementary Fig. S2 and Supplementary Table S3). Geometric mean of the 11 overexpressed genes was significantly higher for adenocarcinoma compared with normal samples in each of the three validation datasets (Fig. 2A). A univariate linear regression model using the geometric mean had the area under receiver operating characteristic (ROC) curve ranging from 0.809 to 0.993 (Fig. 2B). However, because of a very small number of normal samples in the Takeuchi and colleagues dataset, high area under the curve (AUC) value should be interpreted with caution. Thus, publicly available datasets can consistently identify a small set of genes as overexpressed in human adenocarcinoma. However, as only adenocarcinoma samples were used in this analysis, it is possible that these 11 genes are upregulated in other cancers compared with normal tissue. Therefore, we sought to determine whether the 11-gene signature was specific for adenocarcinoma or whether it was also upregulated in other NSCLC subtypes.

**Six of the 11 genes are overexpressed in adenocarcinoma compared with other NSCLC subtypes**

We analyzed an additional eight gene expression datasets consisting of 1,040 human NSCLC samples (16, 20, 40–43; 687 adenocarcinoma, 353 SCC; Supplementary Table S4). Six of the 11 genes were significantly overexpressed in adenocarcinoma compared with SCC, whereas two genes were
PTK7 is overexpressed in adenocarcinoma

PTK7 belongs to the receptor tyrosine kinase family. It is located on chromosomal locus 6p21, which is within the top 4% of all regions of most frequent copy number gain in an independent lung cancer dataset (Supplementary Table S3A). The geometric mean of the 5-gene signature in adenocarcinoma was also significantly higher than that in LCC in the two additional datasets (46 LCC; Supplementary Fig. S3A). A univariate linear regression model using the geometric mean had the area under ROC curve ranging from 0.603 to 0.766 for discriminating between adenocarcinoma and other NSCLC subtypes (Supplementary Fig. S3B).

In summary, a meta-analysis of 13 independent NSCLC datasets consisting of 2,026 human lung samples (1,430 adenocarcinoma, 353 SCC, 46 LCC, and 197 normal) identified five genes (ARHGEF16, CACNB3, MPZL1, PTK7, and RUNXI) that are significantly and consistently overexpressed in adenocarcinoma across all datasets, and are able to distinguish adenocarcinoma samples from normal lung samples or other NSCLC subtypes. Consistent overexpression of these genes despite the potential presence of biologic and technological confounding factors, including different cohorts, treatments, and microarray technologies strongly suggests that at least some of these genes may be functionally important in adenocarcinoma and may play a role in the pathogenesis of this disease.

The receptor tyrosine kinase, PTK7, is overexpressed in adenocarcinoma

PTK7 belongs to the receptor tyrosine kinase family. It is located on chromosomal locus 6p21, which is within the top 4% of all regions of most frequent copy number gain in an independent lung cancer dataset (ref. 44, analyzed using Oncomine; ref. 45, data not shown). In addition, 6p21 has been associated with poor survival in NSCLC (46). PTK7 was significantly underexpressed in adenocarcinoma compared with SCC (Supplementary Table S5). One of the six overexpressed genes, SULT1C2, had significant between-studies heterogeneity (P ≤ 0.05) in the eight datasets. Therefore, it was removed from further analysis. Geometric mean of the remaining five overexpressed genes was significantly higher for adenocarcinoma compared with SCC in seven of the eight datasets (Supplementary Fig. S3A). The geometric mean of the 5-gene signature in adenocarcinoma was also significantly higher than that in LCC in the two additional datasets (46 LCC; Supplementary Fig. S3A). A univariate linear regression model using the geometric mean had the area under ROC curve ranging from 0.603 to 0.766 for discriminating between adenocarcinoma and other NSCLC subtypes (Supplementary Fig. S3B).

In summary, a meta-analysis of 13 independent NSCLC datasets consisting of 2,026 human lung samples (1,430 adenocarcinoma, 353 SCC, 46 LCC, and 197 normal) identified five genes (ARHGEF16, CACNB3, MPZL1, PTK7, and RUNXI) that are significantly and consistently overexpressed in adenocarcinoma across all datasets, and are able to distinguish adenocarcinoma samples from normal lung samples or other NSCLC subtypes. Consistent overexpression of these genes despite the potential presence of biologic and technological confounding factors, including different cohorts, treatments, and microarray technologies strongly suggests that at least some of these genes may be functionally important in adenocarcinoma and may play a role in the pathogenesis of this disease.

Overexpression of PTK7 in primary human and mouse lung adenocarcinoma specimens

Increased expression of PTK7 (Fig. 3A) was assessed by immunohistochemical staining. First, two patient-derived xenografts of human lung adenocarcinoma were analyzed and noted to stain positively for PTK7 to varying intensities (Fig. 3B). Staining patterns were heterogeneous, with both nuclear and cytoplasmic staining present within each specimen. Next, a tissue microarray consisting of human samples across various stages and grades of adenocarcinoma was used to determine PTK7 expression. Fourteen of the 20 adenocarcinoma samples stained positively for PTK7, of which five stained strongly and nine stained weakly to moderately (Fig. 3C; Supplementary Fig. S4A and Supplementary Table S6). In contrast, all normal lung specimens were negative for epithelial staining of PTK7 (Supplementary Fig. S4A). We found no correlation between PTK7 staining and stage or grade, although this may be due to limited sample size. Notably, samples with weak staining tended to have more cytoplasmic rather than nuclear PTK7 (Supplementary Table S6). Thus, PTK7 seems to be overexpressed in most human patient primary adenocarcinoma samples and a significant subset demonstrates high levels of staining.

Consistent with these results, a well-established mouse model of adenocarcinoma (47) also exhibited increased mRNA transcript levels of Ptk7 (Supplementary Fig. S4C) and positive PTK7 protein staining by immunohistochemistry (Supplementary Fig. S4D) relative to normal lung. These results suggest that PTK7 is overexpressed in both human and mouse adenocarcinoma specimens.
Loss of PTK7 in lung cancer cell lines decreases cell viability and induces apoptosis

To determine whether PTK7 could play a functional role in the maintenance of adenocarcinoma, two shRNAs directed against PTK7 were used to decrease expression in a panel of lung cancer cell lines (Supplementary Fig. S5A). Importantly, PTK7 knockdown had no effect on normal lung epithelial cells (Fig. 4A). However, PTK7 depletion using two independent shRNAs led to a significant decrease in viability in a subset of NSCLC cell lines (Fig. 4A, Supplementary Table S7). The dependency on PTK7 was not determined by its basal expression (Supplementary Fig. S5B; \( P = 0.51 \)) or mutations of any known cancer genes (Supplementary Fig. S5C). However, cell lines with chromosomal gains of the locus harboring PTK7 tended to be more sensitive to PTK7 knockdown (\( P = 0.047 \); Fig. 4B).

To further delineate how PTK7 affects cell viability, we chose the three cell lines most sensitive to PTK7 knockdown (H1299, H2009, and H23) for further analysis. Substantial PTK7 knockdown was validated at the RNA level by...
PTK7 Lung Adenocarcinoma Meta-analysis

qRT-PCR (Supplementary Fig. S5D), coinciding with a decrease in protein expression (Fig. 4C). Knockdown of PTK7 in all three cell lines led to an increase in apoptosis as detected by Annexin-V staining (Fig. 4D), as well as by cleaved PARP immunoblotting (Fig. 4E). Thus, knockdown of PTK7 in several adenocarcinoma cell lines in vitro results in decreased cell viability and increased apoptosis whereas normal lung epithelial cells are unaffected. These results suggest that PTK7 may be involved in signaling cascades that regulate survival of a subset of adenocarcinoma.

**Disrupted PTK7 expression activates MKK7–JNK signaling**

During normal development PTK7 is thought to play a role in mediating planar cell polarity (26, 27). However, it remains unclear why disruption of this signaling pathway is important in cancer cells as it is not evident why planar cell polarity mechanisms would be involved in oncogenesis. To understand the signaling pathways that are perturbed upon PTK7 disruption, we surveyed the extracellular signal-regulated kinase (ERK), AKT, JNK, and p38 MAPK pathways by immunoblotting (Fig. 5). We found no significant change in ERK and AKT signaling, and no consistent change in p38 phosphorylation status across the three cell lines. However, phosphorylation of MKK7 at Ser171/Thr275, two residues crucial for MKK7 activity, was clearly increased after PTK7 knockdown. Consistent with the activation of MKK7, phosphorylation of JNK, a downstream target of MKK7, was also increased in response to PTK7 knockdown in the three cell lines tested with two different shRNAs. Activation of MKK7 was also observed when pooled siRNAs were used to deplete PTK7 expression across all three cell lines (Supplementary Fig. S6A–S6F), arguing against an off-target effect.

Upon phosphorylation by MKK7, JNK translocates into the nucleus, leading to the heterodimerization of c-Jun with c-Fos to form the AP-1 transcriptional complex. Gene expression profiling following PTK7 knockdown in the H1299 and H2009 cell lines (Supplementary Fig. S6G) revealed an overrepresentation of AP-1 binding sites associated with the differentially expressed genes (Supplementary Fig. S6H). Thus, in some adenocarcinoma cell lines, loss of PTK7 leads to dysregulation of MKK7–JNK and possibly other signaling pathways that ultimately lead to decreased proliferation and increased apoptosis.

**Xenotransplantation of lung cancer cells depleted of PTK7 reduces tumor burden**

The above studies demonstrate that PTK7 is required for cell viability in a subset of adenocarcinoma cell lines in vitro. To test whether PTK7 is important for tumor maintenance in vivo, we examined the effect of PTK7 depletion on xenograft adenocarcinoma tumor growth. Xenotransplantation of H1299 (Fig. 6A and B), H2009 (Fig. 6C and D), and H23 (Fig. 6E and F) cell lines infected with shRNA hairpins against PTK7 into immunocompromised mice resulted in a significant decrease in tumor growth relative to control shRNA-infected cells. PTK7 loss was confirmed by immunohistochemistry for the two PTK7-specific hairpins relative to control (Fig. 6G). The control experiment of infecting normal lung epithelial cell lines was not performed, as these cell lines do not readily grow in vivo. PTK7 depletion led to a significant decrease in pH3, a marker of mitosis (Fig. 6G and H), and a concomitant increase in CC3, a marker of apoptosis (Fig. 6G and I). Taken together, these studies suggest that PTK7 expression may be required for tumor maintenance in vivo for at least a subset of adenocarcinoma.

**Discussion**

The goal of this study was to identify candidate driver genes in lung adenocarcinoma through a meta-analysis of NSCLC gene expression data and to functionally validate their biologic significance. The meta-analysis method described here was designed to account for biases inherent in single studies and nominate commonly overexpressed genes. The explosion of large amounts of gene expression data from many different platforms in multiple independent cohorts has enabled the use of *in silico* tests across numerous patient populations, many more so than a clinical trial could accomplish. With the increasingly widespread availability of gene expression data, a paradigm shift toward inclusion of a “pre-validation” meta-analysis step in biologic studies may accelerate translational research.

PTK7 seems to have a context-specific value as a prognostic marker. In gastric cancers, it is a favorable marker of differentiated cancers (28), but in AML and esophageal cancer, PTK7 is a poor prognostic marker associated with reduced disease-free survival (29, 30). We were unable to find an association between high PTK7 expression and patient survival in NSCLC.
gene expression datasets with clinical annotations (data not shown), although one report suggested a prognostic significance in NSCLC (48).

Functional studies in vitro identified several NSCLC cell lines that were highly sensitive to depletion of PTK7. Reduced expression of PTK7 led to increased cell death both in vitro and in vivo. This increase in cell death was accompanied by increased signaling through the JNK pathway. However, JNK inhibition alone did not rescue cell viability after PTK7 loss (data not shown), suggesting that PTK7 loss alters other signaling pathways and that JNK activation, while present, is not sufficient to induce apoptosis. Further studies will be needed to identify other signaling pathways downstream of PTK7. Importantly, in NSCLC, PTK7 loss did not consistently alter signaling through the ERK, AKT, or p38MAPK pathways.

Approximately 10% of the human kinome consists of pseudokinases on the basis of alterations in residues that are thought to be critical to kinase catalytic activity. The pseudokinase PTK7 is most closely related to the kinase 3-kinase and ERK signaling pathways (51, 52). Pseudokinases can also function as scaffold proteins. For example, KSR protein, which is a member of the EGF receptor family. After ligand binding, HER3 heterodimerizes with HER2, resulting in activation of downstream phosphoinositide 3-kinase and ERK signaling pathways (31, 52). Pseudokinases can also function as scaffold proteins. For example, KSR proteins interact with protein kinases of the mitogen-activated protein kinase pathway to localize the signaling components to the membrane (53). In contrast with HER3, no known ligands are known to bind PTK7 in the human setting, and it is unclear whether PTK7 plays a role as a signal modulator, scaffold protein, or another role.

In vertebrates, the PCP pathway regulates convergent extension movements and neural tube closure. PCP also is involved in regulating orientation of stereociliary bundles of sensory hair cells in the inner ear. Studies of Ptk7 C-terminal knockout mice firmly establish this pseudokinase as a regulator of PCP in mammals (27). The precise signaling role of PTK7 in development is not clear, although the drosophila homologue off-track (otk) may act as a co-receptor for plexins to mediate semaphoring signaling (54). Most recently, PTK7/OTK has been suggested to act by inhibiting canonical Wnt signaling (55).

Although loss of PTK7 has been associated with activation of WNT in Xenopus, we did not detect consistent activation of WNT signaling in the absence of PTK7 (data not shown). Thus, the effect of PTK7 loss on lung cancer pathogenesis is likely to be independent of the regulation of WNT signaling. In summary, the data presented here describe PTK7 as a functionally relevant protein in NSCLC. We demonstrate that its disruption in vitro and in vivo decreases tumor cell growth. These observations provide a framework for further studies to characterize PTK7 as a potentially therapeutic target in adenocarcinoma.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: R. Chen, P. Khatr, A.J. Butte, E.A. Sweet-Cordero

Figure 6. PTK7 depletion in three NSCLC cell lines decreases tumor burden in a xenograft transplantation assay. A, tumor volume over time of xenografted H1299 with control shRNA against luciferase (shLUC), or two independent hairpins against PTK7. Error bars represent SEM and significant P values correspond to <0.01 (‘’’) or <0.001 (***) when compared with the shLUC control. Values are mean ± SEM (n = 4–9). B, representative ex vivo images of H1299 tumors. C and D, same figures for H2009 cell line. E and F, same figures for H23 cell line. G, staining of PTK7, pHH3, and CC3 by immunohistochemistry in the H1299 xenografts. H, quantification of pHH3-positive cells per field of view (n = 6). I, quantification of CC3-positive cells per field of view (n = 6).
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc): R. Chen, P. Khatri, P.K. Mazur, M. Polin, Y. Zheng, C.D. Hoang, J. Shrager, Y. Xu, E.A. Sweet-Cordero

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): R. Chen, P. Khatri, M. Polin, D. Vaka, S. Vicent, E.A. Sweet-Cordero

Writing, review, and/or revision of the manuscript: R. Chen, P. Khatri, Y. Zheng, J. Shrager, S. Vicent, A.J. Butte, E.A. Sweet-Cordero

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): P. Khatri, J. Shrager, S. Vicent

Study supervision: R. Chen, P. Khatri, A.J. Butte, E.A. Sweet-Cordero

Acknowledgments

The authors thank Julien Sage, Laura Attard, Steve Artandi, Monte Winslow, and all members of the Sweet-Cordero laboratory for helpful discussions.

References


A Meta-analysis of Lung Cancer Gene Expression Identifies *PTK7* as a Survival Gene in Lung Adenocarcinoma

Ron Chen, Purvesh Khatri, Pawel K. Mazur, et al.

*Cancer Res* 2014;74:2892-2902. Published OnlineFirst March 20, 2014.

**Updated version**
Access the most recent version of this article at:
doi:10.1158/0008-5472.CAN-13-2775

**Supplementary Material**
Access the most recent supplemental material at:
http://cancerres.aacrjournals.org/content/suppl/2014/03/20/0008-5472.CAN-13-2775.DC1

**Cited articles**
This article cites 53 articles, 22 of which you can access for free at:
http://cancerres.aacrjournals.org/content/74/10/2892.full#ref-list-1

**Citing articles**
This article has been cited by 5 HighWire-hosted articles. Access the articles at:
http://cancerres.aacrjournals.org/content/74/10/2892.full#related-urls

**E-mail alerts**
Sign up to receive free email-alerts related to this article or journal.

**Reprints and Subscriptions**
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

**Permissions**
To request permission to re-use all or part of this article, use this link
http://cancerres.aacrjournals.org/content/74/10/2892.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.