**Meta-analysis of lung cancer gene expression identifies *PTK7* as a survival gene lung adenocarcinoma**

Ron Chen, Purvesh Khatri, Pawel K. Mazur, Melanie Polin, Yanyan Zheng, Dedeepya Vaka, Chuong D. Huong, Joseph Shrager, Yue Xu, Silvestre Vicent, Atul Butte, E. Alejandro Sweet-Cordero

**SUPPLEMENTARY FIGURE LEGENDS**

**Supplementary Figure S1.** Forest plots summarizing the overexpression of the 11 genes from discovery meta-analysis. Each row represents a study with standardized mean difference between ADC and normal (dark blue box) and the confidence interval of 95% (light blue line). The size of the dark blue box is proportional to the relative effect size of each study. The dotted vertical line at 0.0 represents the null hypothesis (see Forest plot for MPZL1). Thus, positive values represent overexpression in ADC.

**Supplementary Figure S2.** Forest plots summarizing the overexpression of the 11 genes from validation meta-analysis. Each row represents a study with standardized mean difference between ADC and normal (dark blue box) and the confidence interval of 95% (light blue line). The size of the dark blue box is proportional to the relative effect size of each study. The dotted vertical line at 0.0 represents the null hypothesis (see Forest plot for CACNB3). Thus, positive values represent overexpression in ADC.

**Supplementary Figure S3.** Six genes of the original 11 gene ADC signature are also over-expressed in ADC relative to other NSCLC histological subtypes. **A.** Swarm plots of the geometric mean of the seven genes in ADC, SCC or LCC samples across the six subtype-specific data sets. **B.** Performance of a univariate classification model based on the geometric mean of the six genes comparing ADC vs SCC measured by ROC curves. **C.** PTK7 is over-expressed in ADC compared to SCC across four of the six data sets. Data are represented as Forest plots summarizing the overexpression of *PTK7* from the subtype-specific meta-analysis.

Each row represents a study with standardized mean difference between ADC and SCC (dark blue box) and the confidence interval of 95% (light blue line). The size of the dark blue box is proportional to the standard error of the effect size of each study. The dotted vertical line at 0.0 represents the null hypothesis. Thus, positive values represent over-expression of in ADC.

**Supplementary Figure S4.** Expression of PTK7 in an ADC tissue microarray and in a mouse model of lung adenocarcinoma. **A.** Representative images from each specimen on the tissue microarray. Detailed specimen clinical information and scores are listed in Supplemental Table 6. **B.** Ptk7 expression from gene expression microarray analysis of sorted cells from wild-type and KrasLSL [1] lungs using a mouse breeding and FACS method detailed in Zheng et al. [2] **C.** Staining of Ptk7 by IHC in wild-type and KrasLSL lungs through various stages of lung cancer progression. AAH = Atypical adenomatous hyperplasia.

**Supplementary Figure S5.** Correlation of sensitivity to PTK7 knock-down with mutations and expression. **A.** Validation of PTK7 transcript reduction in all NSCLC cell lines after shPTK7 infection. **B.** Scatter plot of PTK7 sensitivity vs expression in the panel of cell lines. PTK7 sensitivity was calculated by averaging the relative viability across both hairpins after PTK7 knock-down. PTK7 expression for each cell line was defined as the $2^{(\Delta\Delta CT)}$ qRT-PCR value relative to A549, which was given a value of 1. **C.** Known mutations of oncogenes and tumor suppressors within the cell line panel as defined by the Sanger COSMIC database. Black squares represent genetic alterations. **D.** QPCR validation of PTK7 knock-down in H1299, H2009, and H23 cell lines.

**Supplementary Figure S6.** MKK7-JNK pathway dysregulation upon PTK7 knock-down. **A.** Validation by qRT-PCR of PTK7 siRNA knock-down in H1299 cells relative to a control non-targeting siRNA. **B.** Immunoblot of pMKK7 and pJUN in H1299. β-actin used as a loading control. **C.** PTK7 knock-down validation and **D.** immunoblots in H2009. **E.** PTK7 knock-down validation and **F.** immunoblots in H23. **G.** Microarray analysis of PTK7 knockdown in H1299

and H2009 cells. Heatmap depicting genes with fold change > 2.0 by SAM [3] and visualized in the “HeatmapImage” module in Genepattern (http://www. broad.mit.edu/cancer/software/genepattern/). RNA was isolated using TRIzol 4 days after infection of cell lines with lentiviral shRNA. RNA was further prepared by passage over an RNeasy column with DNAse1 in-column digestion. cDNA synthesis, biotinylation of cRNA, and hybridization to human gene 1.0 ST array was performed according to the manufacturer’s instructions (Affymetrix) by the Stanford Protein and Nucleic Acid Facility. Microarray data was normalized in R using RMA. Raw data are available in Gene Expression Omnibus (GEO accession GSE50138). **H.** Webgestalt [4] was used to identify enriched transcription factor binding sites in the promoter regions of differentially expressed genes with fold change > 1.5 by SAM. Genes with an FDR below 5% and fold change over 1.5 were included in the respective up/down lists for Webgestalt.

**SUPPLEMENTARY TABLES**

**Supplementary Table S1.** Data sets used in discovery meta-analysis. Institutes and original sources are listed alongside sample numbers.

**Supplementary Table S2.** Data sets used in validation meta-analysis. Institutes and original sources are listed alongside sample numbers.

**Supplementary Table S3.** Gene-level results in validation meta-analysis. Statistical measures of meta-analyses are listed.

**Supplementary Table S4.** Data sets used in NSCLC histological subtype meta-analysis. Institutes and original sources are listed alongside sample numbers.

**Supplementary Table S5.** Gene-level results in NSCLC histological subtype (ADC vs SCC) meta-analysis. Statistical measures of meta-analyses are listed.

**Supplementary Table S6.** Tissue microarray specimen annotations and PTK7 staining results.

**Supplementary Table S7.** Cell line annotations. Original disease states of the cell line panel used in this study are listed.

## REFERENCES


1. Jackson EL, Willis N, Mercer K, et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. Genes & development 2001;15(24): 3243-8.
2. Zheng Y, de la Cruz CC, Sayles LC, et al. A Rare Population of CD24(+)ITGB4(+)Notch(hi) Cells Drives Tumor Propagation in NSCLC and Requires Notch3 for Self-Renewal. Cancer cell 2013;24(1): 59-74.
3. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proceedings of the National Academy of Sciences of the United States of America 2001;98(9): 5116-21.
4. Zhang B, Kirov S, Snoddy J. WebGestalt: an integrated system for exploring gene sets in various biological contexts. Nucleic acids research 2005;33(Web Server issue): W741-8.