Supplemental Methods

Progression of human bronchioloalveolar carcinoma to invasive adenocarcinoma is modeled in a transgenic mouse model of K-ras-induced lung cancer by loss of the TGF-β type II receptor

Alain C. Borczuk, Marieta Sole, Ping Lu, Jinli Chen, May-Lin Wilgus, Richard A. Friedman, Steven M. Albelda, Charles A. Powell.

Genomics Methods

Description of Gene Expression Profiles

A. Expression Profiles Generated

a. *KrasTgfbr2* $^{-/-}$ mouse model

Tumor Isolation
Mice were sacrificed at 6-7 weeks after administration of Adeno-Cre. Lungs and mediastinum were removed enbloc. One lung was snap frozen and embedded in OCT for storage at -80 C and the other lung and mediastinum were placed in formalin for histological analysis.

RNA isolation and preparation for microarray analysis
To enrich for tumor and stromal cells, we used the PALM MicroBeam LCM system (Carl Zeiss, Jena, Germany) system to acquire cells for mRNA extraction using the RNAqueous-Micro Scale RNA Isolation kit (Ambion, Austin, TX). From each specimen, approximately 600,000 μ^2 of cells were dissected, yielding 20 ng RNA. RNA yield and quality was assessed for each specimen using the RNA 6000 Nano Chip Assay (Agilent Technologies, Santa Clara, CA). Specimens with high quality RNA were then used to prepare cRNA for hybridization to Affymetrix (Santa Clara, CA) oligonucleotide arrays.

PCR analysis of tumor DNA
DNA was isolated using QIAamp DNA Mini Kit (Qiagen, Valencia, CA). We used the following primers to examine recombination of the Tbr2 allele in *KrasTgfbr2* $^{-/-}$ lungs (5' primer TAAACAAGGTCCGGAGCCCA, 3' primer AGAGTGAAGCCGTGGTAGGTGAGCTTG. PCR was performed in the ABI 9700 thermocycler. The detection of a 720 bp band by gel electrophoresis confirmed recombination (Supplemental Figure 1).

qRT-PCR analysis of tumor mRNA:
Samples from the *KrasTgfbr2*^{WT} and *KrasTgfbr2*$^{-/-}$ lungs were homogenized in TRIzol Reagent (Invitrogen, Carlsbad, CA). Three mg of RNA from each group was reverse-transcribed using 0.5 μg oligo (dT) (Promega), 1 mM dNTPs (Clontech, Palo Alto, CA), and 1 unit of SuperScript III Reverse Transcriptase in 1X First-Strand Buffer and 10 mM DTT (Invitrogen), for 60 minutes at 50˚C. Primer sequences are shown in the attached Table. Semi-quantitative analysis of gene expression was performed using a Cepheid Smart Cycler (Sunnyvale, CA) following the manufacturer's protocol for SYBR Green kit supplied by Roche. cDNA concentrations from each pool were normalized using β-actin and GAPDH as control genes. Primers for human Sprr1a were purchased from Origene (Part number HP209018) and for murine Sprr1a from ABI (Part number Mm01962902_s1). Quantitative real-time PCR was performed using SYBR Green Master Mix (Applied Biosystems) on an ABI 7500 cycler, using β-Actin as endogenous control.

Relative level of expression of each of the selected genes comparing the *KrasTgfbr2*^{WT} and *KrasTgfbr2*$^{-/-}$ lungs (fold change) were determined. Each sample was run in triplicate.

Name of primer	Sequence (5' to 3')
Arginase-1-5'	GATTGGCAAGGTGATGGAAG
Arginase-1-3'	TCAGTCCCTGGCTTATGGTT
Ccl5-5'	ACCATGAAGATCTCTGCAGC
Ccl5-3'	TGAACCCACTTCTTCTCTGG
Cxcl-10-5'	CAATGATCTCAACACGTGGG
Cxcl-10-3'	GTCTGAGTGGGACTAAGGG
Cxcl12-5'	GCTCTGCATCAGTGACGGTA
Cxcl12-3'	GCTTTCTCCAGGTACTCTTGGA
Cxcl3-5'	CACCAACCACCAGGCTACAG
Cxcl3-3'	ACCGTTGGGATGGATCGCTT
Cxcl5-5'	CGCTGGCATTTCTGTTGCTG
Cxcl5-3'	ACTGCGAGTGCATTCCGCTT
Fn1-5'	AGCCATTTGCTCCTGCACGT
Fn1-3'	TCACTGGGGTGTGGATTGAC
Tgfrb2-5'	ACGTGGAGTCGTTCAAGCAG
Tgfrb2-3'	CACACACTGTGCTGTGAGAC

b. Human Lung Adenocarcinoma

Tumor Isolation
Cases of lung adenocarcinoma resected from 2002 to 2006 were examined with Columbia University Medical Center Institutional IRB approval (Supplemental Table 2). Tumors were snap frozen at resection and embedded in OCT for storage at -80 C. Serial sections were examined to confirm the histology.

RNA isolation and preparation for microarray analysis.
Using the PALM LCM system, tumor cells from invasive foci were acquired from invasive adenocarcinomas while visible tumor cells in the field were acquired from *in-situ* bronchioloalveolar tumors. Procedures for RNA extraction and preparation for micorarray analysis were identical to those used for murine tissues.

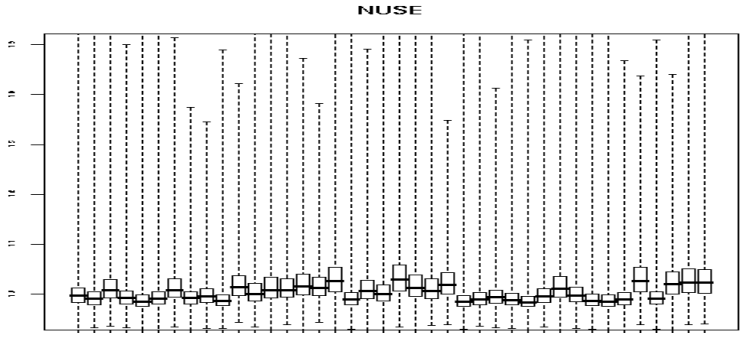
c. RNA preparation and array hybridization
RNA isolated from tumor and from stroma was used to create target for hybridization to DNA microarrays after two rounds of amplification. First strand cDNA synthesis was generated using a T7-linked oligo-dT primer, followed by second strand synthesis. An in vitro transcription reaction using the Megascript T7 kit (Ambion) was used to generate cRNA, which was cleaned up using the GeneChip Sample Cleanup Module (Qiagen).

For the second round of amplification, 400-500 ng of cRNA was used for first strand cDNA synthesis using random primers, followed by second strand synthesis. An *in vitro* transcription reaction using the GeneChip Expression 3'-Amplification kit (Affymetrix) was used to generate cRNA containing biotinylated UTP and CTP, which was chemically fragmented at 94 C for 35 minutes. 12-15 micrograms of fragmented, biotinylated cRNA was hybridized to Affymetrix 430 2.0 arrays (for murine tumors) or HG-133 Plus 2.0 arrays (for human tumors) at 45 C for 16 hours. Arrays were washed and stained with streptavidin-phycoerythrin. Signal amplification was performed using biotinylated anti-streptavidin antibody followed by a second staining reaction before scanning on Affymetrix scanners.

d. Array preprocessing

Quality Assessment- Human Lung Adenocarcinoma (n=40)

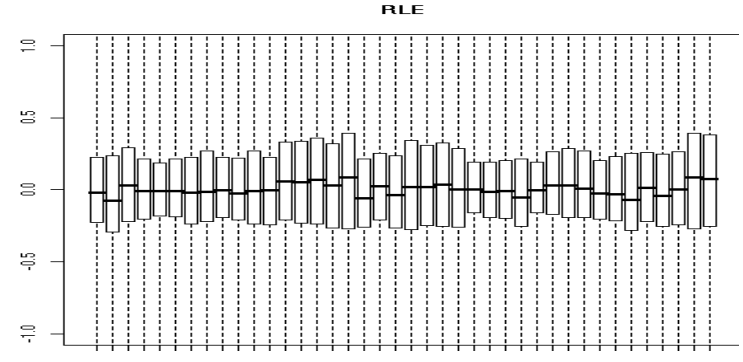
Several quality assessment procedures were performed to determine probe-level qualities (1). First, the NUSE (Normalized Unscaled Standard Error) measures the relative spread of probes within each probeset between arrays The NUSE Plot is:



The NUSE value and spread between arrays is comparable. Quantitatively, the interquartile test shows that all medians of each array are within 1.5 interquatile ranges of the first and third quartiles of the medians of all of the arrays. In addition, all interquartile ranges of each array are within 1.5 interquartile ranges of the first and third quartiles of the interquartiles of all of the arrays. This relationship is summarized in the following table:

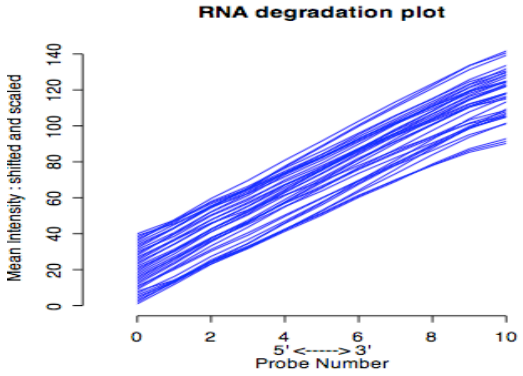
Array		1st quartile		3rd quartile	
Measure:	Parameter	Minimum	-(1.5*IQR)	Maximum	+(1.5*IQR)
NUSE:	Median	0.9838	0.9585	1.0300	1.0441
NUSE:	IQR	0.0211	0.0009	0.0510	0.0647

The RLE (Relative log expression) is a probe-level-model based measure of the relative expression of each probeset on each array. The RLE plot and interquartile range summary indicate that the arrays are similar.



Array		1st quartile		3rd quartile	
Measure:	Parameter	Minimum	-(1.5*IQR)	Maximum	+(1.5*IQR)
RLE:	Median	-0.0740	-0.0879	0.0867	0.0983
RLE:	IQR	0.3484	0.2718	0.6696	0.7027

RNA quality for cRNA generated from the Modified Eberwine double amplification protocol was assessed by the RNA degradation plot which indicates that all arrays are within the interquartile range estimator, with the exception of one array, TBR2_16.cel, which was as an outlier, suggestive of increased RNA degradation and was thus excluded from analysis. .



Normalization and Filtering.
All array data were imported into BRB Array Tools (version 4.1, National Center for Biotechnology Information, Bethesda, MD at http://linus.nci.nih.gov/BRB-ArrayTools.html) and were normalized using the RMA algorithm (2). For the murine tumor arrays, to exclude genes with minimal variation across the specimens, the data were filtered to remove genes with variance of log ratios for each gene < 75 percentile, thus leaving 11,275 genes for analysis. The HG-133 Plus 2.0 human array data were not filtered.

The .cel files, expression datasets and sample labels are deposited in GEO, accession GSE27719.

e. Hierarchical Clustering

We performed unsupervised average linked hierarchical clustering of human adenocarcinoma tumors with Pearson correlation and identified two main clusters (Figure 5). The mixed-subtype adenocarcinomas were grouped in cluster 1 (15/23) while the BACs were grouped in cluster 2 (13/17); Fisher p=0.01.

f. Differential Expression and Ontology Analysis

We identified genes that were differentially expressed between tumors of *KrasTgfbr2* $^{-/-}$ and those of *KrasTgfbr2* ^{WT} mice using a permuted random variance t-test with a two-fold change cut-off and the Significance of Microrarrays (SAM) method (3).

Genes were considered to be statistically significant if their permuted t-test p value was less than 0.001 and the estimated false discovery rate was less than 0.01 (delta value 0.23655), Supplemental Table 1.

We identified gene ontology (GO) groups of genes whose expression was differentially regulated among the classes. We considered all GO categories with between 5 and 100 genes represented on the array. A GO Biological Process group was considered to be significantly differentially regulated between *KrasTgfbr2* $^{-/-}$ and *KrasTgfbr2* ^{WT} tumors if the Fisher (LS) statistic and the Kolmogorov-Smirnov (KS) statistic p values were less than 0.00001 and the Efron-Tibshirani's GSA test p-value was less than 0.005 (4). Supplemental Table 3.

B. Expression profiles previously published

a. Japan Lung Cancer dataset

This dataset was generated at the Aichi Cancer Center by Takeuchi et al (5). It was used to generate rank-ordered gene lists to assess for enrichment of the *KrasTgfbr2* $^{-/-}$ tumor signature. The dataset consists of specimens from 5 nonmalignant lung and 158 lung cancers, of which 90 were lung adenocarcinomas that were hybridized to a customized Agilent oligonucleotide array containing 21,619 probes (GEO Platform ID GPL7015). Unsupervised clustering indicated the presence of three major classes of adenocarcinoma distinguished by morphology and differentiation: terminal respiratory unit (TRU) type B that was similar to bronchioloalveolar carcinoma (n=19); TRU type A that was similar to invasive adenocarcinoma with BAC type morphology (n=34); and non-TRU type with poorly differentiated morphology without TTF-1 expression (n=37).

The expression dataset and sample labels are accessible on GEO (GSE11969). http://www.ncbi.nlm.nih.gov/projects/geo/query/acc.cgi?acc=GSE11969.

C. Application of Gene Set Enrichment Analysis (GSEA) to validate the TBRII model

a. GSEA is a computational method http://www.broadinstitute.org/gsea/index.jsp that determines whether an *a priori* defined set of genes shows statistically significant differences between two phenotypes (6). Using expression datasets of lung adenocarcinoma from New York and Japan, we tested whether several gene sets generated from murine lung cancer models were enriched in invasive vs. non-invasive lung adenocarcinoma (e.g. TRU-A vs. TRU-B in Japan dataset). The weighted t-test metric and permutation of gene sets were used to rank the genes and calculate the normalized enrichment score and significance. To facilitate inter-platform and inter-species analysis, analysis was performed by collapsing probe sets to unique HUGO gene symbols.

b. Gene sets

Supplemental Table 6. Mouse Model Gene Sets Used in GSEA

i. ***KrasTgfbr2^{WT} ; KrasTgfbr2$^{-/-}$*** **model**
192 unique gene symbols upregulated in six-week *KrasTgfbr2* $^{-/-}$ vs. six-week *KrasTgfbr2*^{WT} mouse tumors.

ii. **KrasLA model**
In this model, a latent mutated Kras2 allele is sporadically activated through spontaeous homologous recombination (7). Gene expression profiling was carried out on 31 mouse lung tumors and 19 normal mouse lung samples (8). Published genesets of upregulated genes and down-regulated genes in the KrasLA model vs. normal lung were accessed at http://web.mit.edu/ccr/labs/jacks/sweetcordero_et_al/.

iii. **Kras-Hif2α model**

iv. In this model, LSL-Kras mice were crossed with LSL-HIF2α mice that conditionally overexpressed a nondegradable variant of HIF2α (9). The published geneset of 200 genes upregulated in LSL-Kras;LSL-HIF2α vs. LSL-Kras was accessed at http://www.jci.org/articles/view/38443/sd/1.

v. **Kras-Lkb1 model and Kras-p53 models**
In these models, Kras mice were crossed with animals with homozygous or heteorzygous Lkb1 inactivation and with animals with homozygous p53 inactivation (10). The gene expression dataset was generated from tumors dissected from Kras mice (n=5) and mice with conditional inactivation of Lkb1 (n=13), p53 (n=5), and p16 (n=2) that were hybridized to an Affymetrix Mouse Genome 430A 2.0 array.

.cel files accessed from GEO (GSE6135) were imported into BRB Array tools and normalized with RMA.

Using the approach outlined for generation of the *KrasTgfbr2* $^{-/-}$ gene sets, we identified genes that were differentially expressed between tumors of Kras mice and those with homozygous inactivation of Lkb1 (n=11) and p53 (n=5) mice, using a permuted random variance t-test and a Significance of Microrarrays (SAM) method. Genes were considered to be statistically significant if their permuted t-test p value was less than 0.001 and the estimated false discovery rate was less than 0.01. This yielded 400 genes upregulated and 306 genes down-regulated in *Kras-Lkb1* $^{-/-}$ tumors. Because the number of genes in the p53 group was disproportionately small, we used only the permuted random variance t-test with a cutoff p value < .01 to identify 571 upregulated and 706 down-regulated genes.

D. Stromal Expression Profile in KrasTBRII$^{-/-}$ mouse model

The stromal and advanced tumor gene expression dataset was filtered to remove genes with variance less than the 75th percentile, redundant probe sets (retaining the probe with highest expression), and genes located on chromosome X and Y to reduce bias introduced by gender differences in the two genotypes examined.

We identified genes that were differentially expressed between stroma of five-week *KrasTgfbr2* $^{-/-}$ and nine-week *KrasTgfbr2* ^{WT} mouse tumors using a permuted t test with p value less than 0.01 (Supplemental Table 4). We identified Immune Response and Developmental genes in this profile by examining the GO ontology for each classifier and selecting genes with a heading or subheading in these categories (Supplemental Table 5).

E. Lung development pathway TGF-β response gene ontology (GO0071560).

a. Lung development pathways

Microarray data was obtained from GEO GSE1423 (11) from branching and non-branching regions of developing mouse lung and imported into BRB Arrays Tools, using RMA normalization. Genes demonstrating differential expression in branching vs. non-branching lung with P <.01 and fold change > 1.75 (476 probe sets) were converted to corresponding probe sets on the Affymetrix Mouse 430 v2 array. This list was used in GSA analysis using Efron-Tibshirani GSA test with 200 and 2000 permutations to determine statistically significant differences between *KrasTgfbr2* $^{-/-}$ and *KrasTgfbr2* ^{WT} mouse tumor and stroma gene expression, using a gene list of 43856 probe sets (filtered for removal of X and Y chromosome and of redundant probes to yield probes for 19750 genes).

b. TGF-β response gene ontology (GO0071560)

To test whether the mouse tumoral stroma showed evidence of TGF-β response, the list of genes under this ontology across all species was downloaded and using gene symbols converted to Affymetrix Mouse 430_2 probes (NetAffxhttp://www.affymetrix.com/analysis/index.affx).

This list was used in GSA analysis using Efron-Tibshirani method with 200 and 2000 permutations to determine statistically significant differences between the *KrasTgfbr2* $^{-/-}$ and *KrasTgfbr2* ^{WT} mouse stroma gene expression, using a gene list of 43856 probe sets (filtered to 19750 genes, as noted above).

References:

1. Bolstad, B.M., Collin, F., Brettscneider, J., Simpson, K., Cope, L.M., Irizarry, R., and Speed, T.P. 2005. Quality Assessment of Affymetrix GeneChip Data. In *Bioinformatics and Computational Biology Solutions using R and Bioconductor*. R. Gentelman, V. Carey, W. Huber, and S. Dudoit, editors. New York: Springer.
2. Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. 2003. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res* 31:e15.
3. Tusher, V.G., Tibshirani, R., and Chu, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 98:5116-5121.
4. Efron, B., and Tibshirani, R. 2007. On Testing the Significance of Genes. *Annals of Applied Statistics* 1:107-129.
5. Takeuchi, T., Tomida, S., Yatabe, Y., Kosaka, T., Osada, H., Yanagisawa, K., Mitsudomi, T., and Takahashi, T. 2006. Expression Profile-Defined Classification of Lung Adenocarcinoma Shows Close Relationship With Underlying Major Genetic Changes and Clinicopathologic Behaviors. *Journal of Clinical Oncology* 24:1679-1688.
6. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 102:15545-15550.
7. Johnson, L., Mercer, K., Greenbaum, D., Bronson, R.T., Crowley, D., Tuveson, D.A., and Jacks, T. 2001. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature* 410:1111-1116.
8. Sweet-Cordero, A., Mukherjee, S., Subramanian, A., You, H., Roix, J.J., Ladd-Acosta, C., Mesirov, J., Golub, T.R., and Jacks, T. 2005. An oncogenic KRAS2 expression signature identified by cross-species gene-expression analysis. *Nat Genet* 37:48-55.
9. Kim, W.Y., Perera, S., Zhou, B., Carretero, J., Yeh, J.J., Heathcote, S.A., Jackson, A.L., Nikolinakos, P., Ospina, B., Naumov, G., et al. 2009. HIF2alpha cooperates with RAS to promote lung tumorigenesis in mice. *J Clin Invest* 119:2160-2170.
10. Ji, H., Ramsey, M.R., Hayes, D.N., Fan, C., McNamara, K., Kozlowski, P., Torrice, C., Wu, M.C., Shimamura, T., Perera, S.A., et al. 2007. LKB1 modulates lung cancer differentiation and metastasis. *Nature* 448:807-810.
11. Lu, J., Qian, J., Izvolsky, K.I., and Cardoso, W.V. 2004. Global analysis of genes differentially expressed in branching and non-branching regions of the mouse embryonic lung. *Dev Biol* 273:418-435.