Cell Reports

Supplemental Information

ErbB2 Pathway Activation upon *Smad4* Loss Promotes Lung Tumor Growth and Metastasis

Jian Liu, Sung-Nam Cho, Bindu Akkanti, Nili Jin, Jianqiang Mao, Weiwen Long, Tenghui Chen, Yiqun Zhang, Ximing Tang, Ignacio I. Wistub, Chad J. Creighton, Farrah Kheradmand, and Francesco J. DeMayo

Supplemental Information

SUPPLEMENTAL DATA

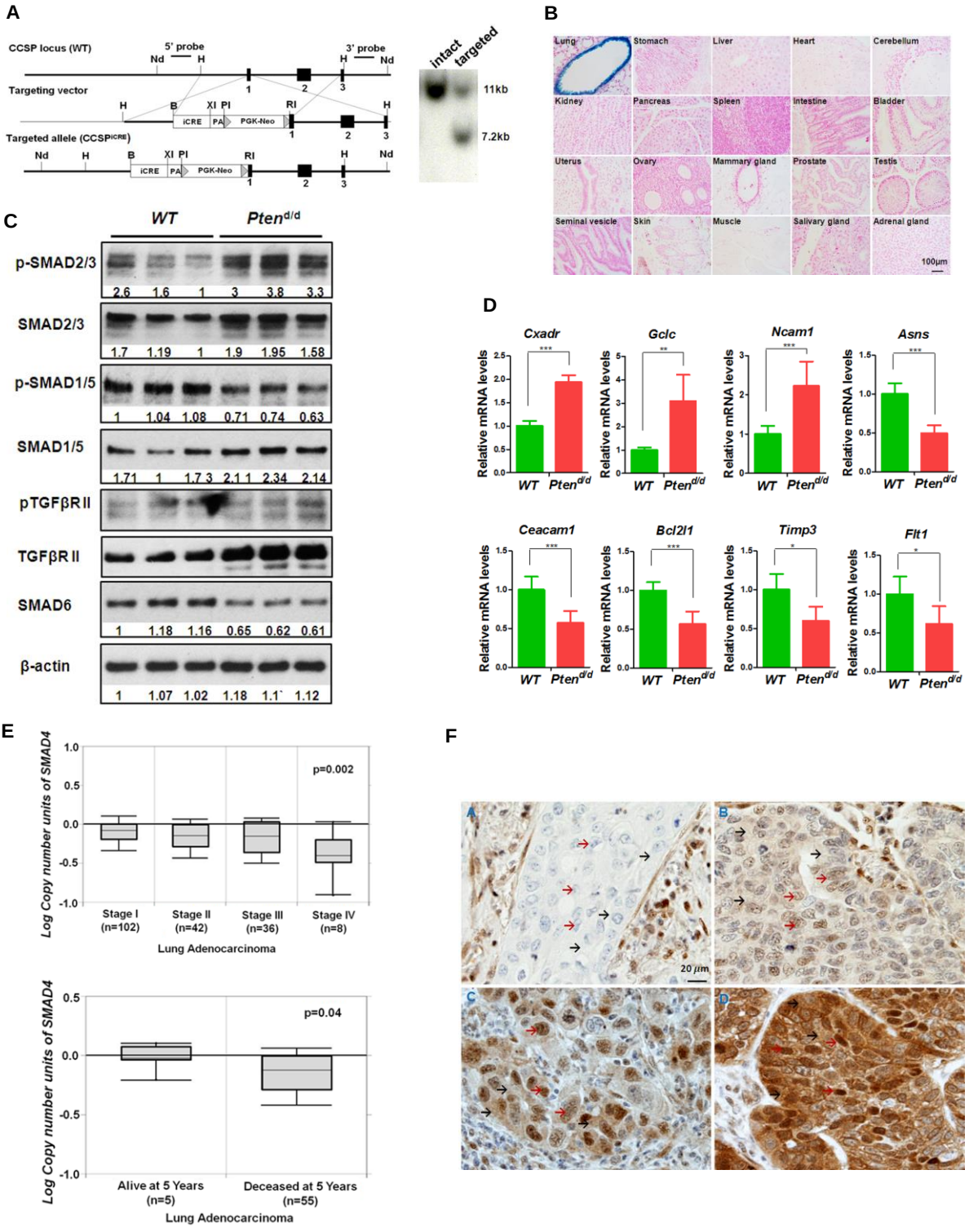


Figure S1, related to Figure 1.

(A) The targeting strategy to generate *CCSP*^{iCre} mice. The *CCSP* WT locus contained the coding sequence exon 1-3 (black boxes) with the putative initiator methionine located at the beginning of exon 1. The targeting vector contained an improved Cre (iCre) recombinase gene sequence followed by SV40 PolyA and an inverted PGK-Neo-cassette flanked by FRT sites (triangle). The left and right arms had 2~3 kb of homologous sequence to *CCSP*. The external 5' probe (rectangle) was used in Southern blot analysis (right panel) of the targeted ES clones. Southern blot analysis showed that the targeted clones generated a novel 7.2 kb fragment different from the 11 kb intact WT fragment.

(B) Tissue distribution of iCre expression by X-gal staining. Tissues from 4-week-old *CCSP*^{iCre}/R26R mice underwent X-gal staining. X-gal staining was not visible in any other tissues other than the lung.

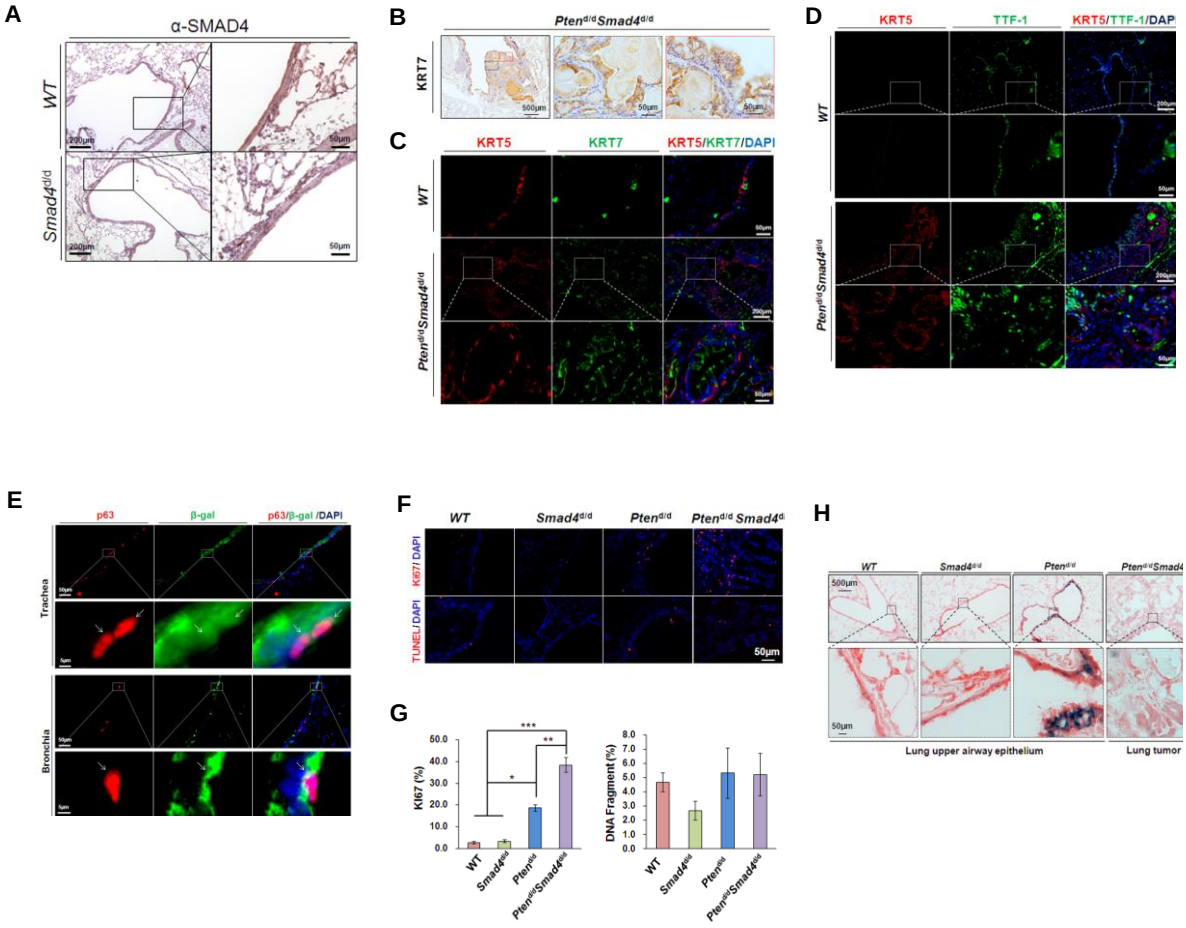
(C) Western blot analysis of the TGFβ/BMP pathway in 7-month-old mouse lungs. The expression of phosphorylation of SMAD2/3 (p–SMAD2/3), SMAD1/5 and TGFβRII increased in the *Pten*$^{d/d}$ mouse lungs when compared to that of wild type mouse lungs while SMAD6, the inhibitor of TGFβ pathway, and phosphorylation of SMAD1/5 (p-SMAD1/5) decreased in *Pten*$^{d/d}$ mouse lungs.

(D) RT-qPCR analysis of some of the significantly altered genes in the TGFβ pathway in the Microarray (*Pten*$^{d/d}$ vs. *WT*). Statistical analysis was performed using Student T-test (p value is *p<0.05, **p<0.01, and ***p<0.001). 6 lungs of wild type (WT) mice and of *Pten*$^{d/d}$ mice were analyzed.

(E) Copy number variations of *SMAD4* in TCGA Lung 2 array. Analysis of copy number of *SMAD4* in different stages of lung adenocarcinoma (upper panel); analysis of copy number of *SMAD4* between alive and dead lung adenocarcinoma patients (lower panel)

and Student T-test was utilized for all statistical analyses. These data indicated *SMAD4* copy number was lost in the human lung cancer development.

(F) Immunohistochemical analysis of SMAD4 in lung cancer specimens. A different score (0-3) is given to each sample based on the intensity of staining signal in the cytoplasm (0-3) and nucleus (0-3). A final expression score (0–300) was obtained by multiplying the intensity and reactivity extension. Black arrows indicate cytoplasmic staining: “0” in A, “1+” in B, “2+” in C and “3+” in D. Red arrows indicate nuclear staining: “0” in A, “1+” in B, “2+” in C and “3+” in D.


A
α-SMAD4
WT
Smad4d/d
B
Ptend/dSmad4d/d
KRT7
C
KRT5
KRT7
KRT5/KRT7/DAPI
WT
Ptend/dSmad4d/d
D
KRT5
TTF-1
KRT5/TTF-1/DAP
WT
Ptend/dSmad4d/d
E
p63
β-gal
p63/β-gal /DAPI
Trachea
Bronchia
F
WT
Smad4d/d
Ptend/d
Ptend/d Smad4d/d
Ki67/ DAPI
TUNEL/ DAPI
50μm
G
Ki67 (%)
DNA Fragment (%)
WT
Smad4d/d
Ptend/d
Ptend/dSmad4d/d
H
WT
Smad4d/d
Ptend/d
Ptend/dSmad4
Lung upper airway epithelium
Lung tumor

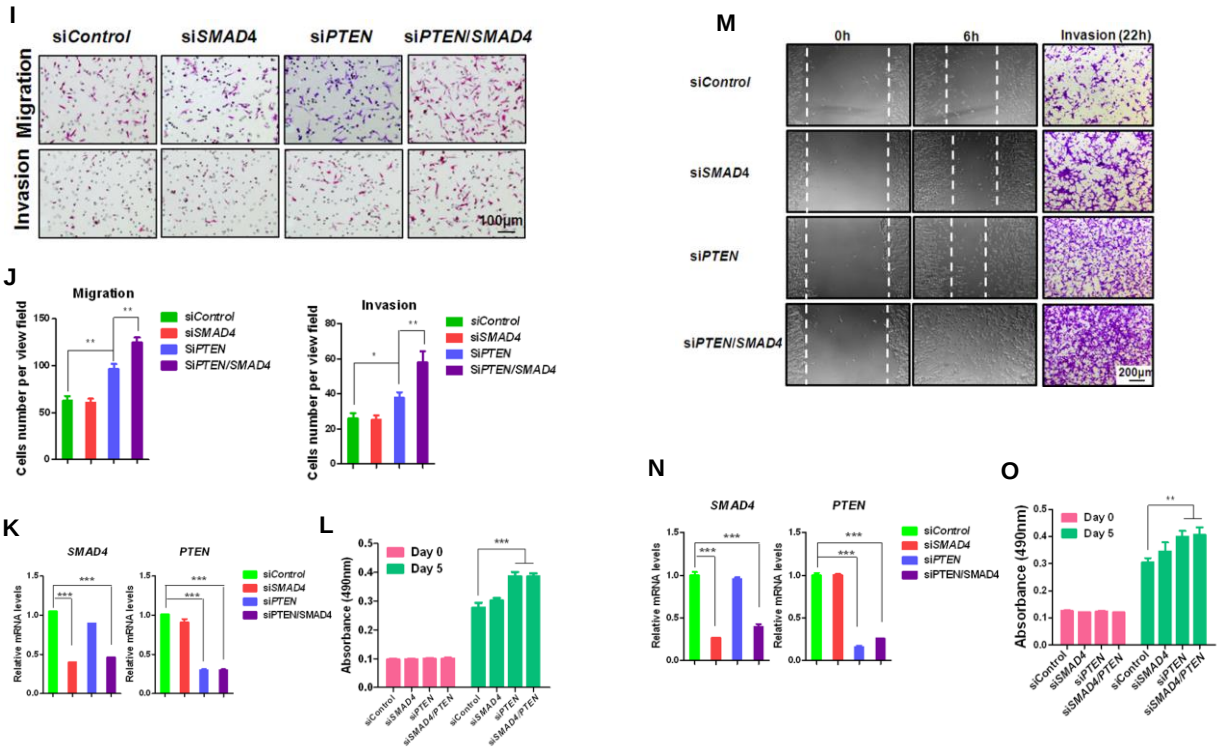

I
siControl
siSMAD4
siPTEN
siPTEN/SMAD4
Migration
Invasion
100μm
M
0h
6h
Invasion (22h)
siControl
siSMAD4
siPTEN
siPTEN/SMAD4
200μm
J
Migration
Invasion
Cells number per view field
siControl
siSMAD4
SiPTEN
SiPTEN/SMAD4
K
SMAD4
PTEN
Relative mRNA levels
siControl
siSMAD4
siPTEN
siPTEN/SMAD4
L
Absorbance (490nm)
Day 0
Day 5
siControl
siSMAD4
siPTEN
siSMAD4/PTEN
N
SMAD4
PTEN
Relative mRNA levels
siControl
siSMAD4
siPTEN
siPTEN/SMAD4
O
Absorbance (490nm)
Day 0
Day 5
siControl
siSMAD4
siPTEN
siSMAD4/PTEN

Figure S2, related to Figure 2.

(A) Immunohistochemical staining of SMAD4 in wild type and *Smad4*$^{d/d}$ mouse lungs. SMAD4 staining is present in the epithelium (brown staining) of wild type mice but lost in the epithelium *Smad4*$^{d/d}$ mice.

(B) Immunohistochemical staining of the lung adenocarcinoma marker KRT7 in *Pten*$^{d/d}$*Smad4*$^{d/d}$ mouse tumors. The regions labeled by black or red dash boxes represent magnified tumor epithelium shown in adjacent panels.

(C) Immunofluorescence of KRT5 and KRT7 in wild type mouse lungs and *Pten*$^{d/d}$*Smad4*$^{d/d}$ mouse tumors. Boxed areas are magnified in the panels directly below.

(D) Immunofluorescence of KRT5 and TTF1 in wild type mouse lungs and *Pten*$^{d/d}$*Smad4*$^{d/d}$ mouse tumors. Boxed areas are magnified in the panels directly below.

(E) Immunofluorescence of p63 and β-Gal (β-galactosidase) in *CCSP*^{iCre}/R26R reporter mouse lungs. The white arrows indicate the cells coexpressing p63 and β-Gal in upper trachea and bronchia. Boxed areas are magnified in the panels directly below.

(F and G) Ki-67 and TUNEL staining were performed on lung sections from wild type, *Smad4*$^{d/d}$, *Pten*$^{d/d}$ and *Pten*$^{d/d}$*Smad4*$^{d/d}$ mice. Each assay was quantified by counting positive nuclei of epithelial cells. All quantitative data (G) were obtained from three to five mice per group. Statistical analysis was performed using one-way ANOVA (p value is *p<0.05, **p<0.01, and ***p<0.001). Cell proliferation significantly increased in the *Pten*$^{d/d}$*Smad4*$^{d/d}$ mice when compared to other three groups.

(H) SA-b-galactosidase (β-Gal) staining of 10-month old mouse lungs and lung tumors. SA-b-galactosidase (β-Gal) (blue staining) is a senescence marker. Boxed areas are magnified in the panels directly below.

(I and J) Two-chamber migration/invasion assays of Beas-2B cells after knock-down of *PTEN* and/or *SMAD4*. (J) Statistical analysis of the results (I) using one-way ANOVA (p value is *p<0.05, **p<0.01).

(K) RT-qPCR analysis of the knock-down of *PTEN* and *SMAD4* in Beas-2B cells. Statistical analysis was performed using one-way ANOVA (p value is *p<0.05, **p<0.01, and ***p<0.001). It indicated there was sufficient knock-down of *PTEN* and *SMAD4* in Beas-2B cells.

(L) MTT assay of cell viability in Beas-2B cells after knock-down of *PTEN* and *SMAD4*. Statistical analysis was performed using one-way ANOVA (p value is *p<0.05, **p<0.01, and ***p<0.001).

(M) Scratch/wound healing assay and invasion analysis of the NL20 cells' migration and invasion, respectively, after knock-down of *PTEN* and *SMAD4*.

(N) RT-qPCR analysis showed the knockdown efficiency of *PTEN* and *SMAD4* in NL20 cells.

(O) MTT assay of cell viability in NL20 cells after knockdown of *PTEN* and *SMAD4*.

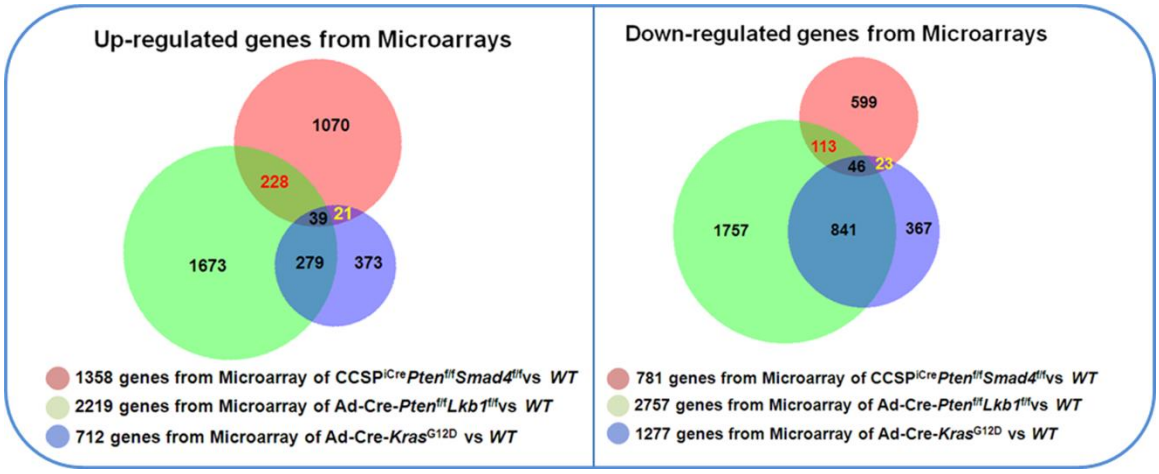
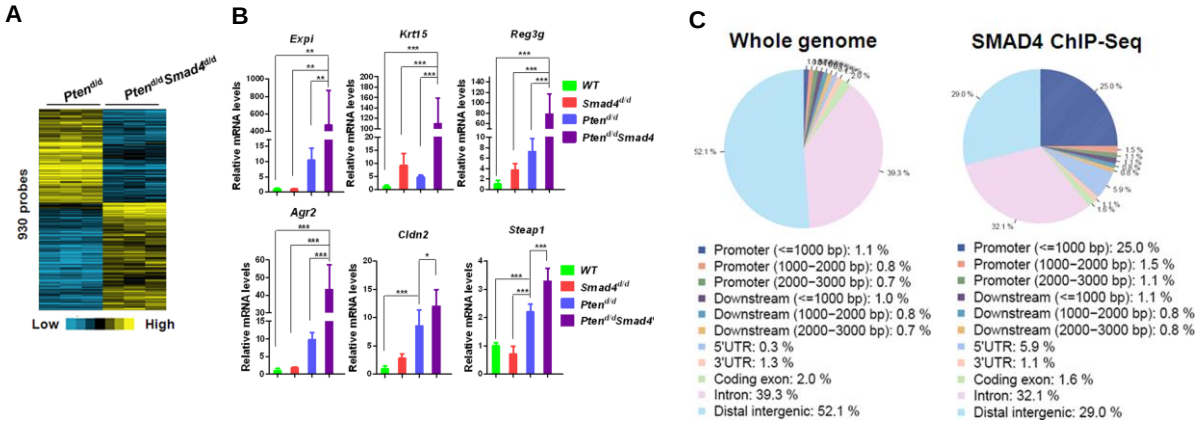
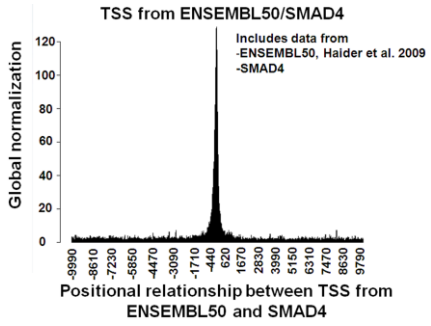
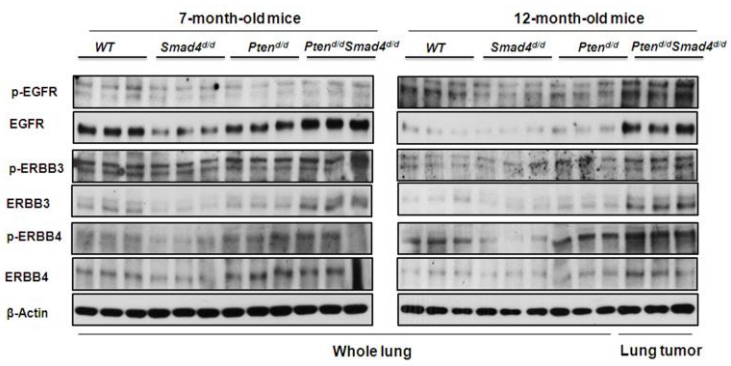


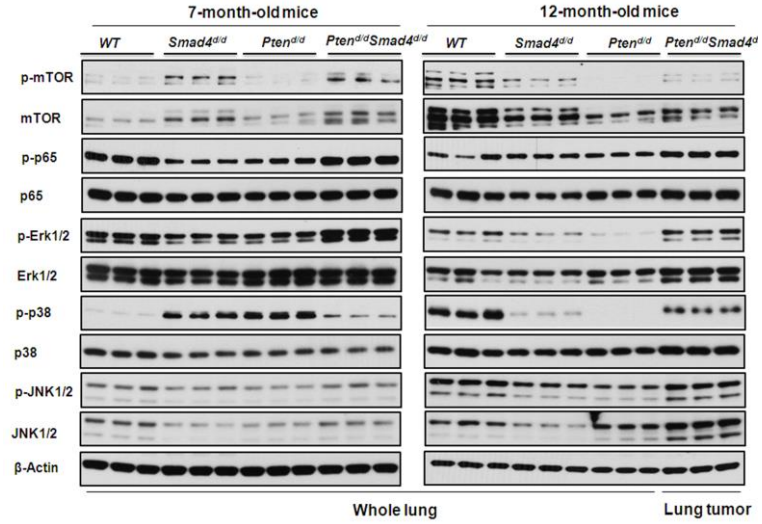
Figure S3, related to Figure 3.

Comparison of the significantly changed genes ($p<0.01$, fold change (log-transformed data) > 1.4) in the lung tumors of mice with *CCSP*^{iCre} ablation of *Pten* and *Smad4* (red), Adenoviral deletion of *Pten* and *Lkb1* (green) and Adenoviral Cre activation of *Kras*^{G12D} (blue). The microarrays of Ad-Cre-*Pten*$^{f/f}$*Lkb1*$^{f/f}$ and Ad-Cre-*Kras*^{G12D} are from the GSE54353 dataset.


A
Pten d/d
Pten d/d Smad4 d/d
930 probes
Low
High
B
Expi
Krt15
Reg3g
Agr2
Cldn2
Steap1
Relative mRNA levels
WT
Smad4 d/d
Pten d/d
Pten d/d Smad4
C
Whole genome
SMAD4 ChIP-Seq
Promoter (<=1000 bp): 1.1 %
Promoter (1000−2000 bp): 0.8 %
Promoter (2000−3000 bp): 0.7 %
Downstream (<=1000 bp): 1.0 %
Downstream (1000−2000 bp): 0.8 %
Downstream (2000−3000 bp): 0.7 %
5'UTR: 0.3 %
3'UTR: 1.3 %
Coding exon: 2.0 %
Intron: 39.3 %
Distal intergenic: 52.1 %
Promoter (<=1000 bp): 25.0 %
Promoter (1000−2000 bp): 1.5 %
Promoter (2000−3000 bp): 1.1 %
Downstream (<=1000 bp): 1.1 %
Downstream (1000−2000 bp): 0.8 %
Downstream (2000−3000 bp): 0.8 %
5'UTR: 5.9 %
3'UTR: 1.1 %
Coding exon: 1.6 %
Intron: 32.1 %
Distal intergenic: 29.0 %


D
TSS from ENSEMBL50/SMAD4
Includes data from
-ENSEMBL50, Haider et al. 2009
-SMAD4
Global normalization
Positional relationship between TSS from
ENSEMBL50 and SMAD4


E
7-month-old mice
12-month-old mice
WT
Smad4 d/d
Pten d/d
Pten d/d Smad4 d/d
p-EGFR
EGFR
p-ERBB3
ERBB3
p-ERBB4
ERBB4
β-Actin
Whole lung
Lung tumor


F
7-month-old mice
12-month-old mice
WT
Smad4 d/d
Pten d/d
Pten d/d Smad4 d/d
p-mTOR
mTOR
p-p65
p65
p-Erk1/2
Erk1/2
p-p38
p38
p-JNK1/2
JNK1/2
β-Actin
Whole lung
Lung tumor

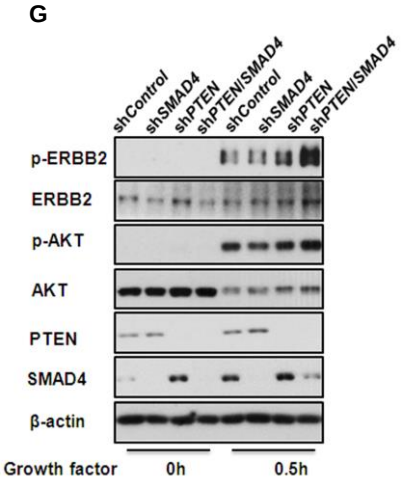

G
shControl
shSMAD4
shPTEN
shPTEN/SMAD4
p-ERBB2
ERBB2
p-AKT
AKT
PTEN
SMAD4
β-actin
Growth factor
0h
0.5h

Figure S4, related to Figure 4.

(A) The heatmap for the 744 differentially regulated genes (930 probes) in the Microarray analysis of lungs from 7-month-old *Pten*$^{d/d}$*Smad4*$^{d/d}$ mice when compared to *Pten*$^{d/d}$ mice. 6 lungs of *Pten*$^{d/d}$ mice and 6 lungs of *Pten*$^{d/d}$*Smad4*$^{d/d}$ mice were analyzed.

(B) RT-qPCR analysis of 6 significantly changed genes in the Microarray (*Pten*$^{d/d}$ *Smad4*$^{d/d}$ vs. *Pten*$^{d/d}$). Statistical analysis was performed using one-way ANOVA (p value is *p<0.05, **p<0.01, and ***p<0.001). 6 lungs of *Pten*$^{d/d}$ mice and 6 lungs of *Pten*$^{d/d}$*Smad4*$^{d/d}$ mice were analyzed.

(C and D) Cistromic (C) and ChIP-Seq online (http://ccg.vital-it.ch/chipseq/) (D) analyses of SMAD4 ChIP-Seq data of 7-month-old *Pten*$^{d/d}$ mouse lungs. Both analyses demonstrate that SMAD4 binding in the whole genome is majorly located in the proximal promoter regions of genes.

(E) Western blot analysis of total and phosphorylated proteins of EGFR, ERBB3 and ERBB4 in mouse lungs and lung tumors.

(F) Western blot analysis of mTOR and MAPK pathways in mouse lungs and lung tumors.

(G) Western blot analysis of phosphorylations of ERBB2 and AKT in BEAS-2B cells after double knock-down of *PTEN* and *SMAD4* with or without the treatment of growth factors.

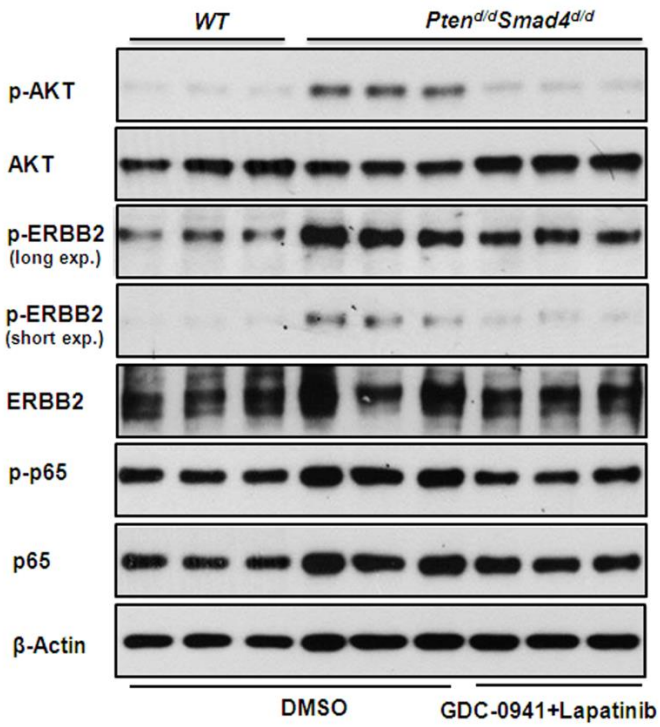


Figure S5, related to Figure 5.

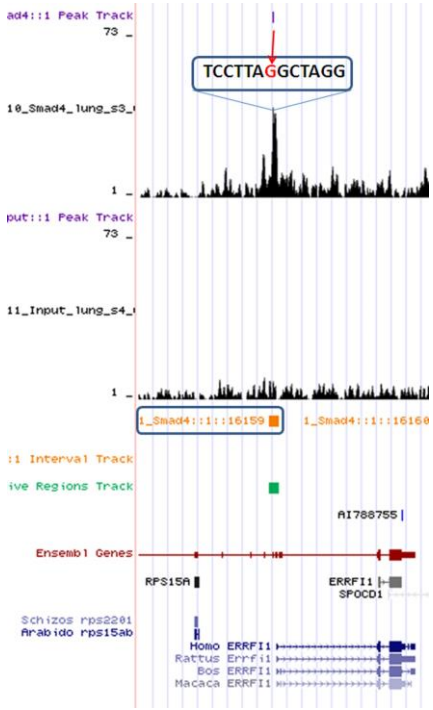
Western blot analysis of the inhibition of GDC-0941 and Lapatinib on the phosphorylations of ERBB2/AKT in mouse lungs. Phosphorylation of p65 (p-p65), which is not effected by the treatment, serves as a negative control.

A

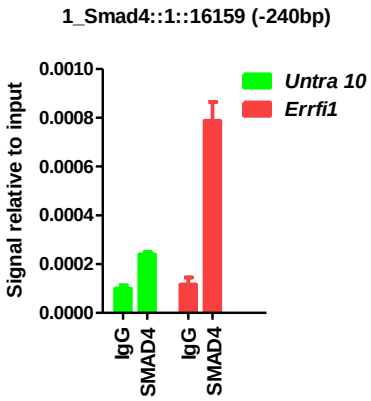
12-month-old mouse tumor Microarray
(*Pten*$^{d/d}$*Smad4*$^{d/d}$ vs *WT*)

Gene symbol	Fold change	P value
Errfi1	0.36	0.0059

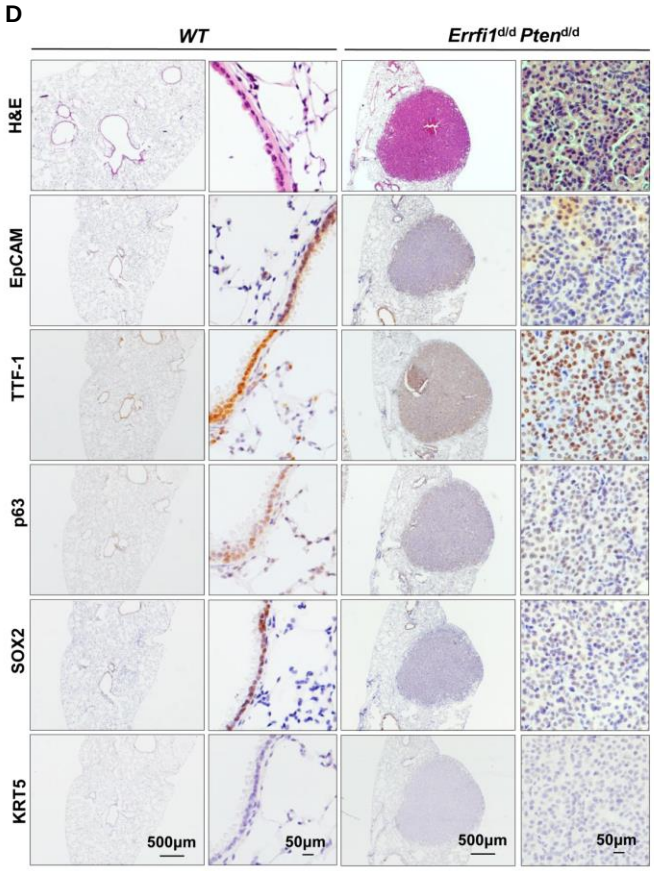
B



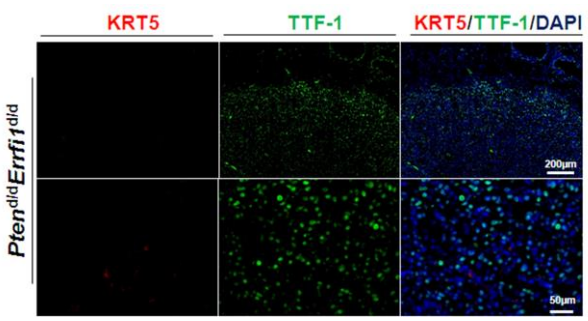
C



D



E



F

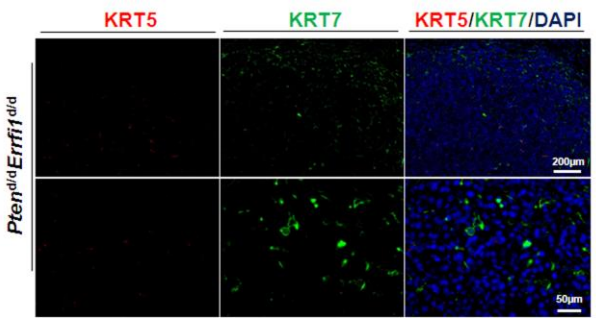


Figure S6, related to Figure 6.

(A) The change of *Errfi1* expression was identified in the gene Microarray analysis of *Pten*$^{d/d}$/*Smad4*$^{d/d}$ mouse tumors *vs* wild type mouse lungs.

(B and C) Analysis of SMAD4 binding on the *Errfi1* proximal promoter region. (B) The sketch map of peak position of SMAD4 binding on the *Errfi1* proximal promoter region identified in SMAD4 ChIP-Seq analysis of 7-month-old *Pten*$^{d/d}$ mouse lungs. (C) ChIP-qPCR analysis of SMAD4 binding on *Errfi1* promoter region in 7-month-old *Pten*$^{d/d}$ mouse lungs. The primers for *Untra 10* (non-relevant control) target the untranslated region 10.

(D) Immunohistochemical staining of differential markers in wild type mouse lungs and *Errfi1*$^{d/d}$*Pten*$^{d/d}$ mouse lungs. The positive staining for TTF1 (ADC marker) and negative staining for p63/KRT5/SOX2 (SCC marker) indicated the lung tumor in *Errfi1*$^{d/d}$*Pten*$^{d/d}$ belongs to adenocarcinoma.

(E and F) Immunofluorescent staining of *Errfi1*$^{d/d}$*Pten*$^{d/d}$ mouse tumors. Negative staining of KRT5 and positive staining of TTF1 (F) and KRT7 (G) indicated that *Errfi1*$^{d/d}$*Pten*$^{d/d}$ mouse tumors belong to adenocarcinoma.

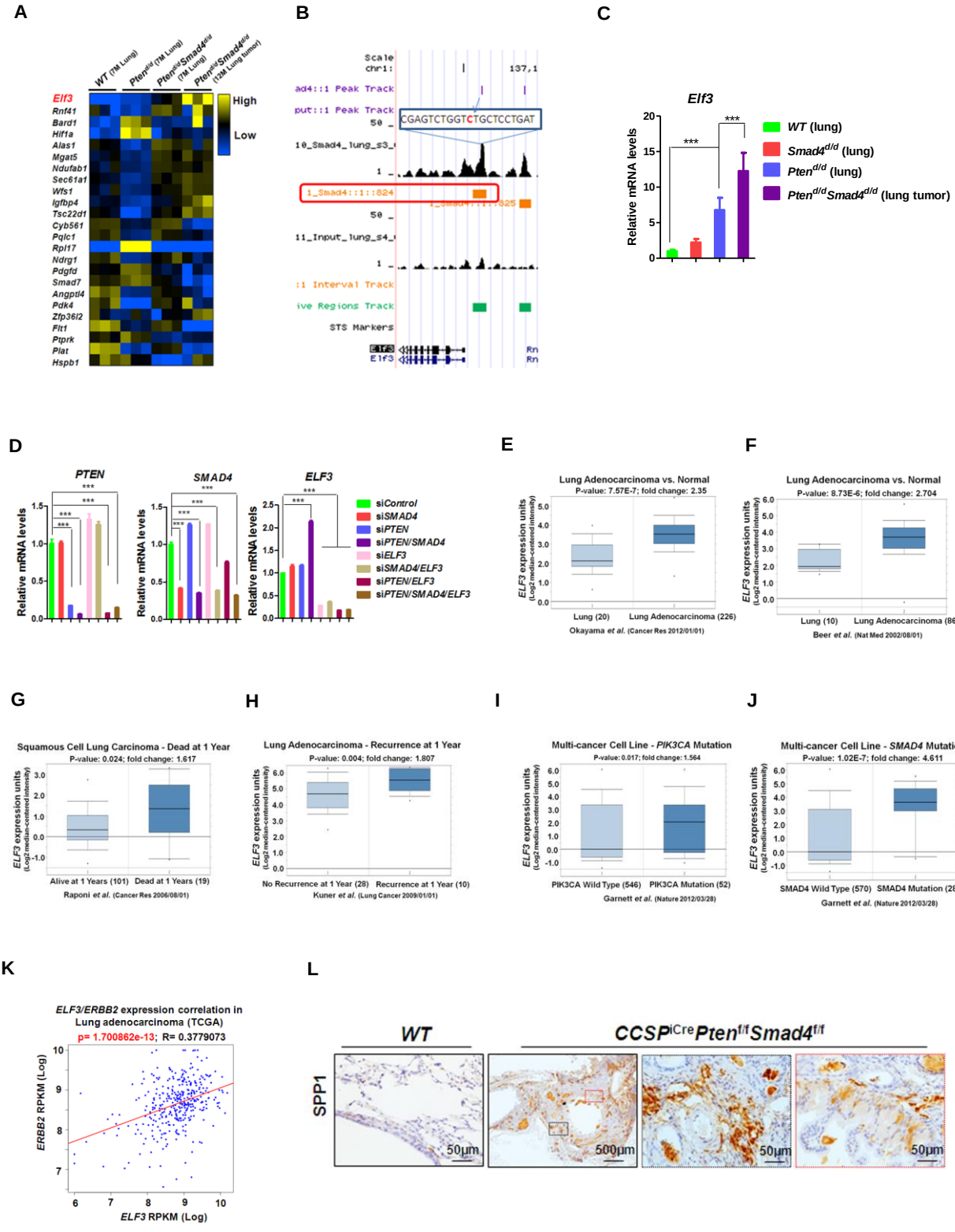

A
WT (7M Lung)
Pten^d/d (7M Lung)
Pten^d/dSmad4^d/d (7M Lung)
Pten^d/dSmad4^d/d (12M Lung tumor)
Elf3
Rnf41
Bard1
Hif1a
Alas1
Mgat5
Ndufab1
Sec61a1
Wfs1
Igfbp4
Tsc22d1
Cyb561
Pqlc1
Rpl17
Ndrg1
Pdgfd
Smad7
Angptl4
Pdk4
Zfp36l2
Flt1
Ptprk
Plat
Hspb1
High
Low
B
Scale
chr1:
137,1
ad4:::1 Peak Track
put:::1 Peak Track
CGAGTCTGGTCTGCTCCTGAT
10_Smad4_lung_s3_
1_Smad4:::1::824
11_Input_lung_s4_
:1 Interval Track
ive Regions Track
STS Markers
ELF3
Elf3
C
Elf3
Relative mRNA levels
***
***
WT (lung)
Smad4^d/d (lung)
Pten^d/d (lung)
Pten^d/dSmad4^d/d (lung tumor)
D
PTEN
SMAD4
ELF3
Relative mRNA levels
siControl
siSMAD4
siPTEN
siPTEN/SMAD4
siELF3
siSMAD4/ELF3
siPTEN/ELF3
siPTEN/SMAD4/ELF3
E
Lung Adenocarcinoma vs. Normal
P-value: 7.57E-7; fold change: 2.35
ELF3 expression units (Log2 median-centered intensity)
Lung (20)
Lung Adenocarcinoma (226)
Okayama et al. (Cancer Res 2012/01/01)
F
Lung Adenocarcinoma vs. Normal
P-value: 8.73E-6; fold change: 2.704
ELF3 expression units (Log2 median-centered intensity)
Lung (10)
Lung Adenocarcinoma (86)
Beer et al. (Nat Med 2002/08/01)
G
Squamous Cell Lung Carcinoma - Dead at 1 Year
P-value: 0.024; fold change: 1.617
ELF3 expression units (Log2 median-centered intensity)
Alive at 1 Years (101)
Dead at 1 Years (19)
Raponi et al. (Cancer Res 2006/08/01)
H
Lung Adenocarcinoma - Recurrence at 1 Year
P-value: 0.004; fold change: 1.807
ELF3 expression units (Log2 median-centered intensity)
No Recurrence at 1 Year (28)
Recurrence at 1 Year (10)
Kuner et al. (Lung Cancer 2009/01/01)
I
Multi-cancer Cell Line - PIK3CA Mutation
P-value: 0.017; fold change: 1.564
ELF3 expression units (Log2 median-centered intensity)
PIK3CA Wild Type (546)
PIK3CA Mutation (52)
Garnett et al. (Nature 2012/03/28)
J
Multi-cancer Cell Line - SMAD4 Mutation
P-value: 1.02E-7; fold change: 4.611
ELF3 expression units (Log2 median-centered intensity)
SMAD4 Wild Type (570)
SMAD4 Mutation (28)
Garnett et al. (Nature 2012/03/28)
K
ELF3/ERBB2 expression correlation in Lung adenocarcinoma (TCGA)
p= 1.700862e-13; R= 0.3779073
ERBB2 RPKM (Log)
ELF3 RPKM (Log)
L
WT
CCSP^iCre Pten^f/f Smad4^f/f
SPP1
50μm
500μm
50μm
50μm

Figure S7, related to Figure 7.

(A) The gene list of ErbB2 pathway revealed by the Ingenuity knowledge-based analysis on 235 intersection genes. In the ErbB2 pathway, *Elf3* was found to be the top changed gene and showed continuously increased expression from wild type group to lung tumor groups.

(B) The sketch map of peak position of SMAD4 binding on the *Elf3* proximal promoter region identified in SMAD4 ChIP-Seq.

(C) RT-qPCR analysis of *Elf3* mRNA expression in 12-month-old wild type, $Smad4^{d/d}$, $Pten^{d/d}$ and $Pten^{d/d}Smad4^{d/d}$ mice. Statistical analysis was performed using one-way ANOVA (p value is *p<0.05, **p<0.01, and ***p<0.001). 3 lungs of wild type (WT) mice and 3 lung tumors of $Pten^{d/d}Smad4^{d/d}$ mice were analyzed.

(D) RT-qPCR analysis of the knockdown efficiency of *PTEN*, *SMAD4* and *ELF3* in BEAS-2B cells. Statistical analysis was performed using one-way ANOVA (p value is *p<0.05, **p<0.01, and ***p<0.001).

(E and F) Oncomine boxed plot of *ELF3* mRNA expression in lung adenocarcinoma.

(G) Oncomine boxed plot of *ELF3* mRNA expression in tumors of squamous cell lung carcinoma patients alive or dead at 1 year.

(H) Oncomine boxed plot of *ELF3* mRNA expression in lung adenocarcinoma patients with or without recurrence.

(I and J) Oncomine boxed plots of *ELF3* mRNA expression in human cancer cell lines. *ELF3* mRNA expression was shown to be upregulated in PIK3CA mutated cancer cell lines (I) and in SMAD4 mutated cancer cell lines (J).

(K) Correlation analysis of *ELF3* mRNA expression and *ERBB2* mRNA expression in lung adenocarcinomas from TCGA dataset. Mean expression values of *ELF3* and *ERBB2* were calculated across all investigated patients. The relative expression was

calculated by normalizing the expression value in each patient to the mean expression value.

(L) Immunohistochemical staining of SPP1 in wild type mouse lungs and *Pten*$^{d/d}$*Smad4*$^{d/d}$ mouse tumors. The regions labeled by black or red dash boxes represent magnified tumor epithelium in adjacent panels.

SUPPLEMENTAL TABLES

Table S1, related to Figure 1

The Gene list in TGFβ1 pathway from the Microarray analysis (*Pten*$^{d/d}$ vs. WT)

Up-regulated genes (67)

Abi2	*Casp1*	*Fzd2*	*Kcnmb1*	*Ncam1*	*Ppp1r14d*	*Tfrc*
Acly	*Casp4*	*Gclc*	*Klk11*	*Nek2*	*Prom1*	*Tmem17*
Adam19	*Cdkn2b*	*Ggt6*	*Krt19*	*Nit1*	*Ptprk*	*Tnfrsf11a*
Adora2b	*Cks2*	*Gja1*	*Ldlr*	*Npas2*	*Rere*	*Unc119*
Afp	*Clu*	*Gnaz*	*Lrba*	*Pfn2*	*Sccpdh*	*Wisp1*
Agtr2	*Ctsc*	*Gsta4*	*Lta4h*	*Pi4k2b*	*Sdc1*	*Wnt4*
B3galt2	*Cxadr*	*Hif1a*	*Mef2c*	*Pla2g4a*	*Slc39a8*	*Wnt5a*
Bsg	*Dsg2*	*Hnmt*	*Mmp11*	*Porcn*	*Sstr2*	
Calml3	*Elf3*	*Hs6st2*	*Muc4*	*Pou4f1*	*Tcf12*	
Calml4	*Fgfr2*	*Idi1*	*Myb*	*Ppfibp2*	*Tert*	

Down-regulated genes (80)

Ace	*Ccbp2*	*Fas*	*Itgb1*	*Mthfd2*	*Rapgef3*	*Solh*
Acvrl1	*Cd163*	*Flt1*	*Itgb3*	*Nfib*	*Rb1*	*Sphk1*
Adm	*Cdc42ep4*	*Fndc3b*	*Klf15*	*Nnmt*	*Rbms3*	*Tgfbr3*
Aldh18a1	*Ceacam1*	*Foxo1*	*Klf9*	*Npr1*	*Rhoc*	*Tgm2*
Aldh2	*Clic4*	*Gfap*	*Kpna3*	*Nup62*	*Rora*	*Timp3*
Angptl4	*Cnn3*	*Gfpt2*	*Lox*	*Pdlim7*	*Sema7a*	*Tsc22d1*
Asns	*Col4a6*	*Ggps1*	*Ltbp1*	*Pecam1*	*Shoc2*	*Ust*
Atf4	*Cyb561*	*Gria1*	*Maf*	*Plaur*	*Sirpa*	*Wdfy3*
Bcl2l1	*Cyp3a13*	*Ifrd1*	*Map2k3*	*Plod2*	*Ski*	
Bin1	*Disp2*	*Igfbp4*	*Mbd1*	*Ppp1r13b*	*Slc39a14*	
Calca	*Ece1*	*Itga5*	*Mbnl2*	*Prkca*	*Slc7a1*	
Cars	*Ercc5*	*Itga6*	*Mfap2*	*Ptk2*	*Smurf2*	

Table S2, related to Figure 1

SMAD4 immunohistochemistry was performed on 479 patients at M.D. Anderson Cancer Center. The average intensity score was compared between different groups utilizing student-t-test ($p<0.05$, statistically significant). Cases without complete survival data were excluded from our analyses (n=21). The average staining score for the nuclear detection group in adenocarcinoma patients was significantly less than that detected in squamous cell carcinoma (121 ± 89.2 SD) versus 156 ±95.1 SD) ($P <0.0004$) although there was no change in recurrence free survival in relation to SMAD4 nuclear or the combined quantification of cytoplasmic and nuclear expression of SMAD4.

		Total (n=479)
Lung Cancer Type	Adenocarcinoma (AD)	318 (66%)
	Squamous Cell Cancer (SCC)	161 (34%)
Gender	Male	239 (50%)
	Female	240 (50%
Smoking	Yes	427 (89%)
	No	52 (11%)

Localization	Types of Cancer	n of 479	Mean Score	SD	P value (AC VS SCC)	Staining Score, n (%)			
						0	1-100	101-200	201-300
Cytoplasm	AD	318	137	90.9	0.25	31(11)	135(42)	99(31)	52(16)
	SCC	161	143	90.6		19(12)	57(35)	55(34)	30(19)
Nuclear	AD	318	121	89.2	0.00005	62(20)	112(35)	105(33)	38(12)
	SCC	161	156	95.1		21(13)	40(25)	65(40)	36(22)
AD: Adenocarcinoma									
SCC: Squamous Cell Carcinoma									

Table S3, related to Figure 3

The Genes related to cell movement identified from the *Pten*$^{d/d}$*Smad4*$^{d/d}$ (tumor) *vs* WT (lung) Microarray

Up-regulated genes (170)

Abcc1	*Crh*	*Fhit*	*Irf6*	*Mmp3*	*Rin1*	*Tacstd2*
Adam8	*Cttn*	*Fhl2*	*Irf7*	*Mst1*	*Ropn1l*	*Tbx1*
Adra2a	*Cxcl14*	*Flnb*	*Itga4*	*Muc4*	*Runx1*	*Tekt4*
Agr2	*Darc*	*Foxa1*	*Itgav*	*Mx1*	*Ruvbl1*	*Tgfbi*
Alox15	*Dbn1*	*Foxc1*	*Itgb4*	*Myo5b*	*S100a14*	*Thbs2*
Aqp4	*Dcn*	*Foxj1*	*Jam3*	*Neo1*	*S100a6*	*Tirap*
Bmx	*Dmp1*	*Frk*	*Klf5*	*Ngfr*	*S100b*	*Tnfaip8*
Brms1	*Dnajb4*	*Fscn1*	*Krt19*	*Nlrp10*	*Scn5a*	*Tnfrsf4*
Bsg	*Dnm2*	*Fst*	*Krt6a*	*Nqo2*	*Selp*	*Tnfsf11*
C5ar1	*Dpysl3*	*Gas8*	*Krt8*	*Ntrk3*	*Serpinb2*	*Tnni3*
Calml3	*Dst*	*Gata6*	*L1cam*	*Pa2g4*	*Serpinb5*	*Tpm1*
Ccl20	*Ebf1*	*Gcnt1*	*Lif*	*Pak1*	*Serpinf1*	*Traf3ip2*
Cd24a	*Ebf3*	*Gna12*	*Llgl1*	*Parva*	*Sfrp1*	*Traf4*
Cdcp1	*Efna5*	*Gnaz*	*Lmcd1*	*Pcsk6*	*Shc4*	*Tubb2b*
Cdh13	*Egln3*	*Gpc1*	*Lrig1*	*Pex7*	*Six1*	*Twist1*
Ceacam1	*Elf3*	*Grid2*	*Lum*	*Pla2g3*	*Skp2*	*Tyro3*
Chl1	*Entpd5*	*Guca2a*	*Ly6d*	*Pln*	*Slc12a2*	*Umod*
Cib1	*Ephb2*	*Has2*	*Madcam1*	*Prdm1*	*Slc9a3r1*	*Vangl1*
Cldn3	*Eps8*	*Hdc*	*Mapk10*	*Prkg1*	*Spa17*	*Vtcn1*
Cldn4	*Erbb2*	*Hoxa1*	*Marcksl1*	*Ptges*	*Spag16*	*Wnt11*
Cldn7	*F2rl1*	*Ift88*	*Mark1*	*Ptn*	*Spag6*	
Clu	*F3*	*Igf2r*	*Matn2*	*Ptp4a1*	*Spdef*	
Cmtm8	*Fbln1*	*Igfbp3*	*Mbp*	*Ptprz1*	*St14*	
Col17a1	*Fbln2*	*Igfbp4*	*Mgat3*	*Ramp1*	*Stk38l*	
Comp	*Fgf13*	*Ikbkg*	*Mmp15*	*Rhobtb2*	*Tac1*	

Down-regulated genes (151)

Ace	*Cd99*	*F10*	*Itgax*	*Nfic*	*Rabep1*	*Tfpi*
Acp5	*Chi3l1*	*Fblim1*	*Itgb2*	*Nod1*	*Rap2a*	*Tgfb1*
Acvrl1	*Cklf*	*Fcer1g*	*Itgb6*	*Nox4*	*Rapgef3*	*Tie1*
Adam17	*Cldn1*	*Fgf7*	*Jun*	*Npr1*	*Rarres1*	*Timp3*
Akt3	*Clec1b*	*Fgfr4*	*Kit*	*Nrp2*	*Reck*	*Tjp1*
Apba1	*Clec5a*	*Fhod1*	*Lcp1*	*Oprm1*	*Robo4*	*Tlr2*
Arf4	*Cnr2*	*Figf*	*Ldlr*	*Pdgfra*	*Rora*	*Tlr7*
Arhgef6	*Csf1*	*Gab1*	*Lims1*	*Pecam1*	*Rps6ka5*	*Tnfrsf9*
Arrb1	*Csf2ra*	*Gab2*	*Lrpap1*	*Plxnc1*	*Rps6ka6*	*Tns1*
B3gnt1	*Ctsd*	*Git2*	*Map3k11*	*Pmp22*	*Saa1*	*Trem2*
Bmpr2	*Ctsk*	*Gna13*	*Map3k8*	*Podxl*	*Sdc3*	*Trpc1*
Braf	*Ctsz*	*Gnaq*	*Mbd2*	*Ppap2a*	*Sema3a*	*Txk*
Calcrl	*Cxcl2*	*Gucy1a3*	*Mertk*	*Ppap2b*	*Sirpa*	*Tyrobp*
Ccl6	*Cybb*	*Gucy1b3*	*Mgat5*	*Pparg*	*Skap1*	*Unc5c*
Ccr5	*Ddx3x*	*Hgf*	*Mgll*	*Ppt2*	*Slc1a2*	*Vav2*
Ccrl1	*Dll4*	*Hipk2*	*Mitf*	*Prkcd*	*Slit2*	*Vegfc*
Ccrl2	*Edn1*	*Id2*	*Mpp1*	*Prkcq*	*Snx27*	*Vim*
Cd274	*Efnb1*	*Il1a*	*Msn*	*Prkd2*	*Sod2*	*Vsnl1*
Cd40	*Efnb2*	*Il1rn*	*Myo1f*	*Prkx*	*Sod3*	*Xcr1*
Cd48	*Egr2*	*Iqgap1*	*Ncf1*	*Ptger2*	*St6gal1*	*Zc3h12a*
Cd93	*Epha4*	*Iqsec1*	*Ndst1*	*Ptgir*	*Tbx3*	
Cd97	*Errfi1*	*Itga1*	*Nfatc1*	*Ptpn12*	*Tbx5*	

Table S4, related to Figure 4

Summary of the SMAD4 ChIP-Seq data

Description	Act Regions	1_Smad4::1	1_Smad4::1 %	2_Input::1	2_Input::1 %
Total Number of Intervals in Build 37_1		23725		7	
Number of Intervals Converted to Build 37_1	23725	23725	100.00%	7	100.00%
Number of Intervals within -10000/+10000 bp of NCBI Genes	18630	18630	78.52%	6	85.71%
Number of Intervals NOT within -10000/+10000 bp of NCBI Genes	5095	5095	21.48%	1	14.29%
Number of NCBI Genes with Intervals within -10000/+10000 bp	13103	13103		6	
Number of Intervals within Promoter Region (-7500/+2500 bp of NCBI Gene Start)	10165	10158	42.82%	2	28.57%
Number of Intervals 500 bp of NCBI Tx Start	7058	6962	29.34%	1	14.29%
Number of NCBI Genes with UCSC CpGs and intervals in Promoters (-7500/2500 bp of start)	9133	9133		5	
Number of NCBI Genes with intervals in Promoters (-7500/2500 bp of start)	10717	10717		5	
Number of Intervals within 200 bp of UCSC CpG Islands	8214	8214	34.62%	5	71.43%
Avg Interval Length	851.4429	852.2869		3405.8333	
Max Interval Length	8542	8543		8396	
Min Interval Length	247	248		897	
Avg Probe Value		2.342326016		2.0975441	
Max Probe Value		312		238	
Total Probe Count		47405430		52937641	

Table S5, related to Figure 3

341 significantly altered genes are only overlapped between the gene microarray of (*CCSP*^{iCre}*Pten*$^{f/f}$*Smad4*$^{f/f}$ vs *WT*) and that of (Ad-Cre-*Pten*$^{f/f}$*Lkb1*$^{f/f}$vs *WT*)

Up-regulated genes (228 genes)

4833423E24Rik	*Cdcp1*	*Fscn1*	*Kif21a*	*Oit1*	*Sdcbp2*	*Tmc5*
9930032O22Rik	*Ceacam1*	*Fut9*	*Kirrel3*	*Otx1*	*Sectm1a*	*Tmem158*
Abcc1	*Celsr2*	*Fxyd3*	*Klf5*	*Padi1*	*Sectm1b*	*Tmem40*
Abcc4	*Chac2*	*Fxyd4*	*Klk11*	*Papss1*	*Serinc2*	*Tmprss11a*
Acpp	*Chl1*	*Gabrp*	*Krt14*	*Pax9*	*Serpinb11*	*Tmprss4*
Adam8	*Cib1*	*Gale*	*Krt17*	*Perp*	*Serpinb1a*	*Tnfaip8*
Adh7	*Ckap4*	*Galnt12*	*Krt5*	*Pir*	*Serpinb1b*	*Tnnt2*
Agr2	*Clca2*	*Galnt3*	*Krt6a*	*Pkp1*	*Serpinb2*	*Tpm2*
Aldh18a1	*Clca4*	*Galnt4*	*L1cam*	*Pkp3*	*Sfn*	*Traf4*
Aldh1a3	*Cldn4*	*Gcnt1*	*Leo1*	*Plac8*	*Skp2*	*Trim13*
Alox12e	*Col17a1*	*Glis3*	*Lgals12*	*Plch1*	*Slc16a14*	*Trim15*
Alox15	*Defb1*	*Gpc1*	*Lrig3*	*Plekhg3*	*Slc2a4*	*Trp63*
Anxa8	*Dio2*	*Gpr37l1*	*Lrrc2*	*Plk1*	*Slc35a1*	*Tshz2*
Apobec2	*Dmkn*	*Gpx2*	*Ly6d*	*Plscr1*	*Slc35c1*	*Ttll10*
Aqp3	*Dpysl3*	*Gramd3*	*Ly6g6e*	*Plxna3*	*Slc39a11*	*Ube2f*
Aqp4	*Dsc2*	*Gsta4*	*Mall*	*Pml*	*Slc39a6*	*Upk1b*
Arntl2	*Dsg2*	*Gsto1*	*Marveld3*	*Pof1b*	*Slc44a4*	*Upk3a*
Atp10b	*Dsp*	*H28*	*Matn2*	*Ppap2c*	*Slc45a3*	*Vangl1*
Atp2c2	*Efna5*	*Has2*	*Melk*	*Ppif*	*Slc5a9*	*Vtcn1*
B3galt2	*Ehf*	*Hist1h4j*	*Mfsd4*	*Ppp2r1b*	*Slc9a3r1*	*Zbp1*
Bace2	*Elf3*	*Hmgn3*	*Mgat3*	*Prdm1*	*Slco4a1*	*Zdhhc13*
Bcl11a	*Endod1*	*Hsh2d*	*Mst1*	*Prss22*	*Sorbs2*	*Zdhhc7*
Brms1	*Enpp3*	*Ifi35*	*Myl7*	*Psat1*	*Sostdc1*	*Zfp53*
Bspry	*Entpd7*	*Ifi44*	*Nans*	*Ptges*	*Sox2*	*Zfp750*
Btbd11	*Epsti1*	*Ifit1*	*Net1*	*Ptn*	*Sox21*	
C79407	*Erbb2*	*Igfbp3*	*Nfe2l3*	*Ptpn13*	*Sprr1a*	

Down-regulated genes (113 genes)

Zmynd15	*Slc7a2*	*Ptgs1*	*Man1c1*	*Gtf2b*	*Cys1*	*Arrb1*
Vim	*Slc43a3*	*Prdm2*	*Lonrf3*	*Gprin3*	*Ctsz*	*Arhgef6*
Vav2	*Slc36a2*	*Pqlc3*	*Lhfpl2*	*Gpm6b*	*Ctps2*	*Arhgap6*
Ugcg	*Slc29a1*	*Phf2*	*Lepr*	*Gna13*	*Cnnm3*	*Anxa4*
Trpm6	*Sh3tc1*	*Pecam1*	*Lcp1*	*Frmd4a*	*Cideb*	*AI504432*
Trpc1	*Rora*	*Nr3c1*	*Lass4*	*Fnip1*	*Chst12*	*Adamts8*
Tns1	*Rnf180*	*Npr1*	*Itpkb*	*Fhod1*	*Chi3l1*	*Acp2*
Tbx5	*Rnf146*	*Npnt*	*Itih4*	*Fgf7*	*Cds2*	*Ace*
Tbc1d24	*Rgl1*	*Nox4*	*Itga1*	*Fcer2a*	*Cdc14b*	*Abcc5*
Tacc1	*Rbpms*	*Nfic*	*Iqsec1*	*Errfi1*	*Cd97*	*9030425E11Rik*
Synj1	*Rbms3*	*Nfatc1*	*Iqgap1*	*Epn2*	*Cd302*	*4631416L12Rik*
Stard4	*Rbms2*	*Nebl*	*Hsd17b11*	*Epb4.9*	*Casc4*	
St6gal1	*Rassf8*	*Msn*	*Hoxa5*	*Eif5a2*	*C230081A13Rik*	
Specc1l	*Rassf4*	*Mospd2*	*Hnrpll*	*Egr2*	*C030046E11Rik*	
Spata6	*Rapgef3*	*Mitf*	*Hipk2*	*Dtna*	*Braf*	
Snx27	*Rab12*	*Mapre2*	*H2-DMb2*	*Dst*	*Atxn1*	
Smarca2	*Ptpn12*	*Map4k3*	*Gucy1a3*	*Dhdh*	*Asah1*	

Table S6, The sequence for the gene primers of qPCR

Mouse Primers (5'-3'):

Cxadr	Forward primer	TCTTCTGCTGTCACAGGAAACG
Cxadr	Reverse primer	CGGTCTTGTAAGCGTACTTGAA
Gclc	Forward primer	GGGGTGACGAGGTGGAGTA
Gclc	Reverse primer	GTTGGGGTTTGTCCTCTCCC
Ncam1	Forward primer	ACCACCGTCACCACTAACTCT
Ncam1	Reverse primer	TGGGGCAATACTGGAGGTCA
Asns	Forward primer	GCAGTGTCTGAGTGCGATGAA
Asns	Reverse primer	TCTTATCGGCTGCATTCCAAAC
Ceacam1	Forward primer	ATTTCACGGGGCAAGCATACA
Ceacam1	Reverse primer	GTCACCCTCCACGGGATTG
Bcl2l1	Forward primer	GACAAGGAGATGCAGGTATTGG
Bcl2l1	Reverse primer	TCCCGTAGAGATCCACAAAAGT
Flt1	Forward primer	TGGCTCTACGACCTTAGACTG
Flt1	Reverse primer	CAGGTTTGACTTGTCTGAGGTT
Timp3	Forward primer	CTTCTGCAACTCCGACATCGT
Timp3	Reverse primer	GGGGCATCTTACTGAAGCCTC
Agr2	Forward primer	GGAGCCAAAAAGGACCCAAAG
Agr2	Reverse primer	CTGTTGCTTGTCTTGGATCTGT
Itgb4	Forward primer	GCAGACGAAGTTCCGACAG
Itgb4	Reverse primer	GGCCACCTTCAGTTCATGGA
Cldn4	Forward primer	GTCCTGGGAATCTCCTTGGC
Cldn4	Reverse primer	TCTGTGCCGTGACGATGTTG
Muc4	Forward primer	CCTCCTCTTGCTACCTGATGC
Muc4	Reverse primer	GGAACTTGGAGTATCCCTTGTTG
Erbb2	Forward primer	GAGACAGAGCTAAGGAAGCTGA
Erbb2	Reverse primer	ACGGGGATTTTCACGTTCTCC
Id2	Forward primer	ATGAAAGCCTTCAGTCCGGTG
Id2	Reverse primer	AGCAGACTCATCGGGTCGT
Efnb2	Forward primer	ATTATTTGCCCCAAAGTGGACTC
Efnb2	Reverse primer	GCAGCGGGGTATTCTCCTTC
Ctsd	Forward primer	GCTTCCGGTCTTTGACAACCT
Ctsd	Reverse primer	CACCAAGCATTAGTTCTCCTCC
Expi	Forward primer	CCACAGTCTTTGTTCTGGTAGC
Expi	Reverse primer	CACAAGTTCCAAAACTGCGTG
Krt15	Forward primer	AGCTATTGCAGAGAAAAACCGT
Krt15	Reverse primer	GGTCCGTCTCAGGTCTGTG
Reg3g	Forward primer	ATGCTTCCCCGTATAACCATCA
Reg3g	Reverse primer	GGCCATATCTGCATCATACCAG
Fxyd3	Forward primer	ATTTGTGCAGGGATTCTCTGTG
Fxyd3	Reverse primer	GCCTGGTGTGATGAGAGGTG
Cldn2	Forward primer	CAACTGGTGGGCTACATCCTA
Cldn2	Reverse primer	CCCTTGGAAAAGCCAACCG
Steap1	Forward primer	GGTCGCCATTACCCTCTTGG

Steap1	Reverse primer	GGTATGAGAGACTGTAAACAGCG
Elf3	Forward primer	GCTGCCACCTGTGAGATCAG
Elf3	Reverse primer	GTGCCAAAGGTAGTCGGAGG
Errfi1	Forward primer	TGGCCTACAATCTGAACTCCC
Errfi1	Reverse primer	GACCACACTCTGCAAAGAAGT

Human Primers (5'-3'):

Elf3	Forward primer	CAACTATGGGGCCAAAAGAA
Elf3	Reverse primer	TTCCGACTCTGGAGAACCTC

Table S7, The sequence for the gene primers of Chip-qPCR

Mouse Primers (5'-3'):

Elf3	Forward primer	GGTTGCTGAAGCGGTAGAGA
Elf3	Reverse primer	GGGTAATCAGGAGCAGACCA
Errfi1	Forward primer	ATCCAAGTGCATCTCAGCGT
Errfi1	Reverse primer	TCTCCTGAGGGGACGTTTCT

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Generation of *Pten*$^{d/d}$, *Smad4*$^{d/d}$, *Pten*$^{d/d}$ *Smad4*$^{d/d}$, *Errfi*$^{d/d}$ and *Errfi*$^{d/d}$*Pten*$^{d/d}$ mice

All animal experiments were performed in accordance with an IACUC approved protocol (Baylor College of Medicine, Houston, TX). Generation of *CCSP*^{iCre} mice were carried out as previously described (Li et al., 2008) with some modifications. An *iCre-SV40 PolyA* fragment was amplified from the *Cyp17*iCre plasmid by PCR and subcloned immediately upstream to a *Frt*-flanked phosphoglycerol kinase neomycin-resistance cassette (*PGK-Neo*). The *iCre-SV40 PolyA-PGK-Neo* fragment was again subcloned into a *pGEM-4Z* plasmid harboring a 5.3-kb mouse *CCSP* genomic sequence (Margraf et al., 1993) and the targeting plasmid was linearized and transfected into R1 ES cells by electroporation. Genomic DNA was isolated from Neomycin-resistant clones and digested with *Nde I* and *Xho I* and were screened by Southern blot analysis using a 5' external probe. Selected targeted clones were microinjected into blastocysts from C57BL/6 mice. Chimeras were bred to C57BL/6 mice and germ line transmission of the iCre gene to F1 offspring was validated with PCR genotyping. The CCSP^{icre} mice were crossed with *Pten*$^{f/f}$ (Lesche et al., 2002), *Smad4*$^{f/f}$ (Chu et al., 2004, Biondi et al., 2007) and *Errfi1*$^{f/f}$ (Jin et al., 2007) mice to generate single *Pten*$^{d/d}$, *Smad4*$^{d/d}$, *Errfi*$^{d/d}$, *Smad4*$^{d/d}$*Pten*$^{d/d}$ and *Errfi*$^{d/d}$*Pten*$^{d/d}$ transgenic mice.

Histopathology, Immunohistochemistry, Immunofluorescence staining and Western blotting

Mice lungs were fixed in 4% paraformaldehyde and paraffin-embedded according to standard protocols. H&E and immunohistochemical analysis of 5 μm lung sections were performed. Sections underwent dewaxing followed by antigen retrieval (Vector, H3300)

and peroxide quenching with 3% H_20_2 in methanol for 15 minutes and blocked with 10% normal goat serum or MOM blocking reagents. Subsequently, the slides were incubated with primary antibodies at 4^{o}C overnight followed with secondary antibodies at RT for 30 min. The slides were then developed using the Vectastain ABC kit and DAB reagents (Vector Laboratory). Apoptosis analysis was performed with the *In Situ* Cell Death Detection Kit (Roche, 11684795910) following the manufacture instructions. The TSA kit was used for developing slides where immunofluorescence was utilized. All quantitative data were obtained from three to five mice per group. Statistical analysis was performed using one-way ANOVA or Student T-test (p value is *p<0.05, **p<0.01, and ***p<0.001). Primary antibodies that were used are listed in the table.

Antibodies for immunostaining

Antigen	**Supplier**	**Cat.#**
Pten	Cell Signaling	CST9559
Smad4	Santa Cruz	sc-7966
Smad4	Cell Signaling	CST9515
CCSP	DeMayo Lab	None
SP-C	Seven Hills	WRAP-SP-C
SPP1	Santa Cruz	sc-21742
EpCAM	Proteintech	21050-1-AP
TTF-1	Dako	M3575
TTF-1	Santa Cruz	sc-13040
p63	Santa Cruz	sc-8431
SOX2	Millipore	AB5603
KRT5	Abcam	ab52635
KRT7	Santa Cruz	sc-23876
KRT14	Covance	Sig-3476
Ki67	Abcam	ab15580
ERBB2	Santa Cruz	sc-284
p-ERBB2	Santa Cruz	sc-293110
AKT	Cell Signaling	CST4691
p-AKT	Cell Signaling	CST4060
ELF3	Sigma	HPA003319

Antibodies for Western Blot

Antigen	Supplier	Cat.#
Pten	Cell Signaling	CST9559
p-AKT	Cell Signaling	CST4060
AKT	Cell Signaling	CST4691
SMAD4	Cell Signaling	CST9515
p-ERBB2	Santa Cruz	SC-293110
ERBB2	Santa Cruz	SC-284
ERRFI1	DeMayo Lab	None
ELF3	R&D	AF5787
p-SMAD2	Cell Signaling	CST3108
SMAD2	Cell Signaling	CST5339
p-SMAD3	Cell Signaling	CST9520
SMAD3	Cell Signaling	CST9523
p-SMAD1/5	Cell Signaling	CST9516
SMAD1	Cell Signaling	CST6944
SMAD5	Cell Signaling	CST9517
p-TGFβRII	Santa Cruz	SC-17005
TGFβRII	Santa Cruz	SC-400
SMAD6	Cell Signaling	CST9519
p-EGFR	Cell Signaling	CST3777
EGFR	Santa Cruz	SC-03
p-ERBB3	Cell Signaling	CST4791
ERBB3	Santa Cruz	SC-285
p-ERBB4	Santa Cruz	SC-33040
ERBB4	Santa Cruz	SC-283
p-mTOR	Cell Signaling	CST2971
mTOR	Cell Signaling	CST2983
p-p65	Cell Signaling	CST3033
p65	Cell Signaling	CST8242
p-ERK1/2	Cell Signaling	CST4370
ERK1/2	Cell Signaling	CST4695
p-p38	Cell Signaling	CST4511
p38	Cell Signaling	CST8690
p-JNK1/2	Cell Signaling	CST4668
JNK1/2	Santa Cruz	SC-572

Cell derivation and culture

Beas-2B (ATCC® CRL-9609™) and NL20 (ATCC® CRL-2503™) cells were purchased from ATCC and cultured by following the ATCC culture condition. The culture medium for Beas-2B cells contains the free medium (CC-3171, LONZA), the growth factor (CC-4175, LONZA), Penicillin (100I.U./mL) and Streptomycin (100µg/mL) (ATCC® 30-2300™). The culture medium for NL20 cells was consist of F-12K medium (21127-022, Life Technologies), 4% FBS (16000-044, Life Technologies) and Pen./Strep.. 293T cells were purchased from Tissue culture core in Baylor College of Medicine and was cultured with DMEM medium (11965-092, Life Technologies), 10% FBS (16000-044, Life Technologies) and Pen./Strep..

Senescence experiment

Frozen lung sections were prepared from 10-month-old mice and treated using the standard protocol of the Senescence Detection Kit (Abcam, Cat. Ab65351).

Real Time Quantitative PCR

Total RNAs were isolated from mouse lungs with TRIzol reagent (Invitrogen) and reversely transcribed into cDNA with the M-MLV kit (Invitrogen). The mRNA expression levels of human and mouse were determined by quantitative Real-time PCR using SYBR Green and Taqman probes (Applied Biosystems). Statistical analysis was performed using one-way ANOVA and Student T- test (p value is *p<0.05, **p<0.01, and ***p<0.001). Mouse lung cancer Microarray: Lung tumors macro-dissected from 12 months old *Pten*$^{d/d}$*Smad4*$^{d/d}$ mice, and control lung tissues harvested from 12 month-old

wild type mice were used for the microarray analysis of late stage (3 mice per group). Lungs from 7 month-old wild type, *Pten*$^{d/d}$ and *Ptend*$^{/d}$*Smad4*$^{d/d}$ mice were utilized for microarray analysis of early stage (6 mice per group). Briefly lung tissues were stored at -80°C and total RNA was extracted using the TRIzol method (TRIzol reagent; Invitrogen, Carlsbad, CA) followed by purification using an RNeasy column (Qiagen, Valencia, CA). RNA quality was assessed by using the Agilent Model 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). Ten micrograms of total RNA was processed for the microarray by using the Affymetrix GeneChip one-cycle target labeling kit (Affymetrix, Santa Clara, CA) according to the manufacturer's recommended protocols. After reverse transcription followed by *in vitro* transcription and biotin labeling, the labeled cRNA was hybridized to Affymetrix mouse genome 430 2.0 arrays (Affymetrix, Santa Clara, CA, USA). Data analysis was performed according to standard protocols including scanning of hybridization images and quantification was done using Microarray suite. A DNA-Chip Analyzer dChip (version 1.3) was utilized on log-transformed data for quantile normalization and a PM-MM model was used to determine expression. A fold change of 1.4 and a p-value of <0.01 (two-sided t-test) were used as cut-offs for differential expression. The murine gene signature that was obtained from the microarray expression analysis was then compared to profiles from the human lung cancer expression datasets, utilizing previously described methods (Gibbons et al., 2009). Taqman probes & primers from Life Technologies were used to detect human *PTEN* and *SMAD4* (Life Technologies). The list of primers for SYBR® Green is shown in **(Table S5)**. All Q-PCR experiments for the validation of the Microarray were done with the same number of mice. All Q-PCR experiments for the human cells were repeated three times in the biological way. Ingenuity Systems Pathway Analysis (IPA) software (http://www.ingenuity.com) was used for gene functional annotations.

Human lung cancer samples and Tissue Microarray

All human lung tissues were obtained from the Lung Cancer Specialized Program of Research Excellence (SPORE) Tissue Bank at the MD Anderson Cancer Center and an IRB was approved for expression studies from this Tissue Bank. For each tissue sample, the percentage of malignant tissue was calculated and the cellular composition of specimens was determined by histological examination (I.I.W.) following Hematoxylin-Eosin (H&E) staining. Specimens resected from NSCLC patients with stages I to IV disease according to the revised International System for Staging Lung Cancer were used for making tissue microarray (TMA) (Mountain, 1997). The TMAs were prepared with a manual tissue arrayer (Advanced Tissue Arrayer ATA100, Chemicon International) using 1-mm-diameter cores in triplicates. Patients who had follow-up length more than 3 months were valid to access into analysis of prognostic overall and recurrence-free survival prediction assay. Patients who had smoked at least 100 cigarettes in their lifetime were defined as smokers, and smokers who quit smoking at least 12 months before lung cancer diagnosis were defined as former smokers. Tissue Micro Array: Immunohistochemistry (IHC) staining was performed as previously described (Tang et al., 2011). Tissue section slides were baked at 56^{o}C overnight, then deparaffinized in xylene and rehydrated through a degraded series of ethanol concentrations. Antigen retrieval was carried out using a decloaker (pH9, 121^{o}C for 30 second and then 90^{o}C for 10 minutes). Intrinsic peroxidase activity was blocked by 3% hydrogen peroxide for 10 minutes and intrinsic avidin and biotin was blocked by incubating sections with avidin and biotin for 10 minutes, respectively then washed in TBS, nonspecific antibody binding was blocked by incubating with 5% goat serum (Sigma) solution at room temperature for 60 minutes. Slides were then incubated with primary antibodies against SMAD4 (Mouse

monoclonal, clone 138G6, Santa Cruz, CA, dilution 1:300) at RT for 60 min. After 3 washes in Tris-buffered saline with Tween-20 (TBST) for 5 min each, the slides were incubated with Vector Elite ABC complex at RT for 30 min. Following additional 3 washes in TBST, the slides were incubated with Dako chromogen substrate for 5 min and counterstained with hemotoxylin for 1 min. Two pathologists (X.T. and I.W.) examined both the intensity and extent of immunostaining by light microscopy using a ×20 magnification objective. Quantification schema was developed as SMAD4 is expressed both in the nucleus and cytoplasm. Cytoplasmic and nuclear expressions were quantified using a 4-value intensity score (0, none; 1+, weak; 2+, moderate; and 3+, strong) and the reactivity extension (0–100) **(Figure S1F)**. A final expression score (0–300) was obtained by multiplying the intensity and reactivity extension. This expression score was utilized to stratify patients into low SMAD4 (<100) and high SMAD4 (>100) groups and Kaplan-Meier estimates were performed to evaluate differences in overall survival and recurrence-free survival between the two groups. Quantification schema was developed as SMAD4 is expressed both in the nucleus and cytoplasm. Cytoplasmic and nuclear expressions were quantified using a 4-value intensity score (0, none; 1+, weak; 2+, moderate; and 3+, strong) and the reactivity extension (0–100) **(Figure S1F)**. A final expression score (0–300) was obtained by multiplying the intensity and reactivity extension. This expression score was utilized to stratify patients into low SMAD4 (<100) and high SMAD4 (>100) groups and Kaplan-Meier estimates were performed to evaluate differences in overall survival and recurrence-free survival between the two groups.

siRNA and shRNA experiments

siRNA experiments: ON-TARGET plus siRNAs SMARTpools (Thermo) were used, consisting of four mixed siRNAs targeting each gene (*PTEN, SMAD4, ELF3*). As a

control, a nontargeting siRNA pool was used. Final concentration for each siRNA was 40nM. The transfection reagent Lipofectamine® RNAiMAX from Life Technologies (Cat. 13778150) was utilized and all processes followed the standard protocol of Lipofectamine® RNAiMAX. shRNA experiments: All the pGIPZ lentivirus shRNA plamids for *PTEN* (Cat. RHS4430-99140716 and RHS4430-98902220), *SMAD4* (Cat. RHS4430-98485365 and RHS4430-98893393) and nontargeting control (Cat. RHS4346_00000000) were purchased from the C-BASS Core at Baylor College of Medicine. Fifteen micrograms of the pGIPZ vector together with 7.5 μg of each packaging vector (pMD2.G and psPAX2) were cotransfected into 293T cells. Supernatant containing lentivirus particles was harvested 48h after transfection, passed through a 0.45-μm membrane filter, and directly used to infect cells in the presence of 6 μg/mL polybrene. After 24 h, the infected cells were maintained in the medium with 2 μg/mL puromycin for 2 weeks before knockdown assessment.

CHIP-Seq and Chip-qPCR

CHIP-Seq: a. Chromatin Immunoprecipitation: Mouse lung tissues were submersed in PBS + 1% formaldehyde, cut into small (~1 mm^3) pieces with a razor blade and incubated at RT for 15 min. Fixation was stopped by the addition of 0.125 M glycine (final). The tissue pieces were then treated with a Tissue Tearer and finally spun down and washed 2x in PBS. Chromatins were isolated from the sample by the addition of lysis buffer followed by disruption with a Dounce homogenizer. Lysates were sonicated with a microtip in order to shear the DNA to an average length of 300-500 bp. Lysates were cleared by centrifugation and stored at -80 C. Genomic DNA (Input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by phenol/chloroform extraction and ethanol precipitation. Purified DNA was

quantified on a NanoDrop spectrophotometer. Extrapolation to the original chromatin volume allowed quantitation of the total chromatin yield. For each ChIP reaction, 30 ug of chromatin was precleared with protein G agarose beads (Invitrogen). ChIP reactions were set up using precleared chromatin and antibody SMAD4 (R&D Systems, cat. AF2097, Lot# KUH0311081) and incubated overnight at 4°C. Protein G agarose beads were added and incubation at 4°C was continued for another 3 h. Immune complexes were washed, eluted from the beads with SDS buffer, and subjected to RNase treatment and proteinase K treatment. Crosslinks were reversed by incubation overnight at 65°C, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation. Quality of ChIP enrichment was assayed by qPCR using primers against candidate SMAD4 binding sites. qPCR reactions were carried out in triplicate using SYBR Green Supermix (Applied Biosystems). The resulting signals were normalized for primer efficiency by carrying out qPCR for each primer pair using input DNA. b. ChIP Sequencing (Illumina): ChIP and Input DNAs were prepared for amplification by converting overhangs into phosphorylated blunt ends and adding an adenine to the 3'-ends. Illumina genomic adapters were ligated and the samples were size-fractionated (200-250 bp) on a 2% agarose gel. After a final PCR amplification step (18 cycles), the resulting DNA libraries were quantified and sequenced on HiSeq 2000. Sequences (36 nt reads, single end) were aligned to the mouse genome (mm9) using the BWA algorithm. Aligns were extended in silico at their 3'-ends to a length of 150 bp, which is the average genomic fragment length in the size-selected library, and assigned to 32-nt bins along the genome. The resulting histograms (genomic "signal maps") were stored in BAR and bigWig files. Peak locations were determined using the MACS algorithm (v1.3.7.1) with a cutoff of pvalue = 1e-10. Signal maps and peak locations were used as input data to Active Motifs proprietary analysis program, which creates excel tables containing detailed information on sample comparison, peak metrics, peak locations and

gene annotations. Chip-qPCR: ChIP was performed with the SimpleChIP Enzymatic Chromatin IP Kit (Cell Signaling Technology, 9005). The lysate was prepared from 7-month old *Pten*$^{d/d}$ mouse lung (N=3) following the standard protocol. The lysates were sheared by sonication to a size range of 150–900 bp using VirTis Virsonic 100 Ultrasonic Homogenizer. An aliquot of sheared chromatin to be used as input DNA was treated with RNAse A and proteinase K, and cross-linking was reversed by heat. The purified input DNA was quantified on a Biotek Microplate Spectrophotometer. Chromatin/protein complexes (10 µg) were precleared with normal rabbit IgG (sc-2017, Santa Cruz Biotechnology, Inc.). ChIP was performed using an antibody against SMAD4 (R&D, AF2097). Following immunoprecipitation, crosslinking was reversed as described above. Purified DNA was used for real-time qPCR. The primers were designed to amplify SMAD4 binding regions identified by ChIP-Seq (**Table S6**). Real-time qPCR was carried out in duplicate using SYBR Green Supermix (Applied Biosystems). Immunoprecipitation with normal rabbit IgG was performed as a negative control. Untra Primers were designed for a region of the mouse genome devoid of any known genes to serve as a negative qPCR primer set (Active Motif). The resulting signals were normalized to input DNA. Sequence conservation, cis-regulatory Element Annotation System (CEAS), and motif enrichment were performed using the Cistrome Analysis Pipeline software (http://cistrome.org/ap/) under default settings.

Transwell migration assay, invasion assay and MTS assay

Standard 24-well Boyden control and invasion chambers (BD Biosciences) were used to assess cell migration and invasion following the manufacturer's suggestions. Briefly, cells were trypsinized, rinsed twice with PBS, resuspended in serum-free media, and seeded at 1×10^5 cells per well for BEAS-2B and NL20 cells per well. Chambers in

triplicate were placed in growth media as a chemo-attractant and an equal number of cells were seeded in cell culture plates in triplicate as input controls. Following 18h incubation for the migration assay and 22 h incubation for the invasion assay, chambers were fixed in 10% formalin, stained with crystal violet for manual counting. Data were normalized to input cells to control for differences in cell number (loading control). Meanwhile, 5000 cells were seeded into each well in flat bottomed 96-well plates and triplicate wells are placed for each sample. After 6 hours (Day 0) and 5 days (Day 5), CellTiter 96® Aq_{ueous} One Solution Reagent (Promega, G3580) was added to each well according to the manufacturer's instructions. After 1 h in culture the cell viability was determined by measuring the absorbance at 490nm.

Inhibitors treatment

Inhibitors: GDC-0941(LC Laboratories, G-9252); Lapatinib (LC Laboratories, L-4899); Herceptin (BOC Sciences, 180288-69-1); DMSO was used to dissolve these inhibitors and as a control in the experiment. Treatment on BEAS-2B cells: These inhibitors would be added into the growth medium outside the invasion chamber during the whole experiment. The other processes were the same as the invasion chamber experiment as described above. The final concentration of GDC-0941 and Lapatinib was 1μM and that of Herceptin was 10μg/ml. Treatment on mice: 9 9-month-old wild type mice were used for the control. 14 9-month-old *Pten*$^{d/d}$*Smad4*$^{d/d}$ mice were randomly separated into two groups. Then one group (7 mice) was used for DMSO treatment and the other group (7 mice) was used for inhibitors treatment. These inhibitors were fed to the mice by oral gavage twice daily. The final concentration of GDC-0941 was 17 mg/kg each time and that of Lapatinib was 34 mg/kg each time.

Human lung tumor expression datasets

The gene expression array dataset GSE11969 (Takeuchi et al., 2006) was probed using the mouse lung cancer gene signature, using previously described approaches (Minami et al., 2007).

To further understand the expression of SMAD4 in human lung cancer patients, the TCGA 2 lung cancer data was analyzed utilizing *Oncomine*™. In Oncomine™, circular binary segmentation is used to find regions or segments of a chromosome with equal copy number. All genes within that region of the chromosome are then assigned the average of the probe measurements in that segment (Rhodes et al., 2007). These gene measurements can then be reviewed for amplifications (values > 0 on a log2 scale) and deletions (values < 0 on a log2 scale). *Oncomine*™ was used for analysis and visualization of the log median copy number units of *SMAD4* (reporter 18-046837993, NM_005359) in normal samples and samples of lung adenocarcinoma and squamous cell lung cancer. *Oncomine*™ was also utilized to compare the log median copy number units of *SMAD4* in early and advanced stages of lung adenocarcinoma and squamous cell lung cancer. (www.oncomine.com, May 2012, Compendia Bioscience, Ann Arbor, MI).

RNA sequencing data of Lung Squamous Cell Carcinoma (LUSC) and Lung Adenocarcinoma (LUAD) were obtained through TCGA pan-cancer synapse (https://www.synapse.org/#!Synapse:syn300013). The normalized RNA sequencing level 3 data were available in the dataset. In our analysis, the RPKM value of individual genes was subjected to logarithm (loge). When analyzing the RNA expression difference between two groups, log (RPKM) values were used, and two-tailed t test was utilized to

assess the statistical significance. When identifying the RNA expression correlation between two individual genes, log (RPKM) values were used, and Pearson's correlation analysis was employed to determine the correlation coefficient (R) and statistical significance. When identifying the RNA expression correlation between two groups of genes, relative values were used. The relative value of one gene in each sample would be dividing the log (RPKM) value of this gene over mean of the log (RPKM) values of this gene across all samples. If multiple genes were in the group, the relative value of the group would be the multiplication of the relative values of all genes in the group. Then, Pearson's correlation analysis was further used to determine the correlation coefficient (R) and statistical significance. In addition, to classify the patients into two groups based on the expression of specific gene or gene set, the relative expression values were used, a patient was classified into the high expression group if the relative expression value of the target gene or gene set was higher than the mean value, and *vice versa*. The clinical data of Lung Squamous Cell Carcinoma (LUSC) patients were also obtained through TCGA pan-cancer synapse. To investigate the clinical relevance of a specific gene, patients were divided into two groups based on the RNA expression of the gene, and then the vital status and survival length recorded in clinics were incorporated to perform the survival analysis, log-rank test was utilized in determining the difference of survival between two groups of patients.

REFERENCES

BIONDI, C. A., DAS, D., HOWELL, M., ISLAM, A., BIKOFF, E. K., HILL, C. S. & ROBERTSON, E. J. 2007. Mice develop normally in the absence of Smad4 nucleocytoplasmic shuttling. *Biochem J,* 404**,** 235-45.

CHU, G. C., DUNN, N. R., ANDERSON, D. C., OXBURGH, L. & ROBERTSON, E. J. 2004. Differential requirements for Smad4 in TGFbeta-dependent patterning of the early mouse embryo. *Development,* 131**,** 3501-12.

GIBBONS, D. L., LIN, W., CREIGHTON, C. J., ZHENG, S., BEREL, D., YANG, Y., RASO, M. G., LIU, D. D., WISTUBA, II, LOZANO, G. & KURIE, J. M. 2009. Expression signatures of metastatic capacity in a genetic mouse model of lung adenocarcinoma. *PLoS One,* 4**,** e5401.

JIN, N., GILBERT, J. L., BROADDUS, R. R., DEMAYO, F. J. & JEONG, J. W. 2007. Generation of a Mig-6 conditional null allele. *Genesis,* 45**,** 716-21.

LESCHE, R., GROSZER, M., GAO, J., WANG, Y., MESSING, A., SUN, H., LIU, X. & WU, H. 2002. Cre/loxP-mediated inactivation of the murine Pten tumor suppressor gene. *Genesis,* 32**,** 148-9.

LI, H., CHO, S. N., EVANS, C. M., DICKEY, B. F., JEONG, J. W. & DEMAYO, F. J. 2008. Cre-mediated recombination in mouse Clara cells. *Genesis,* 46**,** 300-7.

MARGRAF, L. R., FINEGOLD, M. J., STANLEY, L. A., MAJOR, A., HAWKINS, H. K. & DEMAYO, F. J. 1993. Cloning and tissue-specific expression of the cDNA for the mouse Clara cell 10 kD protein: comparison of endogenous expression to rabbit uteroglobin promoter-driven transgene expression. *Am J Respir Cell Mol Biol,* 9**,** 231-8.

MINAMI, Y., SHIMAMURA, T., SHAH, K., LAFRAMBOISE, T., GLATT, K. A., LINIKER, E., BORGMAN, C. L., HARINGSMA, H. J., FENG, W., WEIR, B. A., LOWELL, A. M., LEE, J. C., WOLF, J., SHAPIRO, G. I., WONG, K. K., MEYERSON, M. & THOMAS, R. K. 2007. The major lung cancer-derived mutants of ERBB2 are oncogenic and are associated with sensitivity to the irreversible EGFR/ERBB2 inhibitor HKI-272. *Oncogene,* 26**,** 5023-7.

MOUNTAIN, C. F. 1997. Revisions in the International System for Staging Lung Cancer. *Chest,* 111**,** 1710-7.

RHODES, D. R., KALYANA-SUNDARAM, S., MAHAVISNO, V., VARAMBALLY, R., YU, J., BRIGGS, B. B., BARRETTE, T. R., ANSTET, M. J., KINCEAD-BEAL, C., KULKARNI, P., VARAMBALLY, S., GHOSH, D. & CHINNAIYAN, A. M. 2007. Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia,* 9**,** 166-80.

TAKEUCHI, T., TOMIDA, S., YATABE, Y., KOSAKA, T., OSADA, H., YANAGISAWA, K., MITSUDOMI, T. & TAKAHASHI, T. 2006. Expression profile-defined classification of lung adenocarcinoma shows close relationship with underlying major genetic changes and clinicopathologic behaviors. *J Clin Oncol,* 24**,** 1679-88.

TANG, X., KADARA, H., BEHRENS, C., LIU, D. D., XIAO, Y., RICE, D., GAZDAR, A. F., FUJIMOTO, J., MORAN, C., VARELLA-GARCIA, M., LEE, J. J., HONG, W. K. & WISTUBA, II 2011. Abnormalities of the TITF-1 lineage-specific oncogene in NSCLC: implications in lung cancer pathogenesis and prognosis. *Clin Cancer Res,* 17**,** 2434-43.