Supplementary data

Materials and Methods

Mice

[tetO]-*Foxa2* (rat *Foxa2*) mice were obtained from Jeffrey Whitsett and Gang Chen at Cincinnati Children's Hospital Medical Center (CCHMC) and the University of Cincinnati College of Medicine, Cincinnati, OH (1) and crossed with *Scgb1a1-rtTA;*[tetO]-*EGFR*^{L858R} or *Scgb1a1-rtTA;*[tetO]-*Kras4b*^{G12D} (2, 3) to develop *Scgb1a1-rtTA;*[tetO]-*EGFR*^{L858R}*;*[tetO]-*Foxa2* (FVB/N;B6;CBA mixed strain) or *Scgb1a1-rtTA;*[tetO]-*Kras4b*^{G12D}*;*[tetO]-*Foxa2* (FVB/N strain) as described previously (4). Transgenic mice were provided chow containing doxycycline (625 mg/kg chow) beginning at 4–5 weeks of age. Mouse maintenance and procedures were done in accordance with the institutional protocol guidelines of Cincinnati Children's Hospital Medical Center Institutional Animal Care and Use Committee. Mice were housed in a pathogen-free barrier facility in humidity and temperature-controlled rooms on a 12:12-h light/dark cycle and were allowed food and water *ad libitum*. Since all transgenic mice were healthy before doxycycline administration, all of them were enrolled for the study. MicroCT imaging and measurement of tumor number and volume were performed as previously described (4). For the measurement of tumor number and volume, at least 6 mice in each genotype were enrolled. For the histological study, at least 6 mice for each genotype that developed lung tumors were enrolled. Littermates were used for the comparison. See Supplementary Table S1 to S4 for further mouse information.

Histology and immunohistochemistry

Staining (H&E, Alcian blue and immunohistochemistry) was performed using 5 μm paraffin-embedded lung sections as described previously (4). Goat anti-FOXA2 (1:1,000; cat# AF2400; RRID:AB_2294104), mouse anti-HNF4A (1:100; cat# PP-H1415-00; RRID:AB_2263954) from R&D Systems (Minneapolis, MN), rabbit anti-phosphohistone H3 (PHH3; 1:500; cat# sc-8656-R; RRID:AB_653256), rabbit anti-MUC5B (1:100; cat# sc-20119; RRID:AB_2282256), goat anti-FOXA3 (1:200; cat# sc-5361; RRID:AB_647544) from Santa Cruz Biotechnology (Dallas, TX), rabbit anti-caspase-3 (1:200; cat# 9662S; RRID:AB_331439) from Cell Signaling Technology (Danvers, MA), mouse anti-MUC5AC (1:1,000; clone 45M1; cat# ab3649; RRID:AB_2146844) from Abcam, mouse anti-NKX2-1 (1:1,000; cat# WMAB-8G7G31; RRID:AB_451726) from Seven Hills Bioreagents (Cincinnati, OH) and guinea pig anti-AGR2 (1:1,000), guinea pig anti-SPDEF (1:1,000) from Jeffrey Whitsett at Cincinnati Children's Hospital Medical Center were used for immunohistochemistry as primary antibodies. Antigen retrieval was performed by incubating sections in pH 6.0 citrate at 112.5ºC for 15 min using a decloaking chamber from Biocare Medical (Pacheco, CA). The number of different types of tumors per H&E-stained section was counted in at least 3 mice of each group (see Supplementary Table S3 and S4 for details).

Cell culture, lentivirus and/or retrovirus infection, siRNAs, CRISPRi, immunoblotting, co-immunoprecipitation, RNA-seq, TaqMan gene expression analysis

H441 human lung papillary adenocarcinoma cells, A549 human lung carcinoma cells and BEAS-2B human transformed bronchial epithelial cells were obtained from ATCC (Manassas, VA). Lentiviral vector delivering FOXA1, FOXA2 or NKX2-1 was made by inserting mouse *Foxa1*, rat *Foxa2* or rat *Nkx2-1* into the PGK-IRES-EGFP vector as described previously (4). CRISPR interference (CRISPRi; CRISPR/dCas9-KRAB) (5) lentiviral vector (pLV hU6-sgRNA hUbC-

dCas9-KRAB-T2a-Puro; Plasmid #71236), pBabe (Plasmid #1764) and pBabe K-Ras 12V (human *KRAS*^{G12V}; Plasmid #12544) retroviral vectors were obtained from Addgene (Cambridge, MA). Lentivirus and retrovirus were made at the Viral Vector Core at CCHMC (Cincinnati, OH). The siRNA-mediated knockdown analysis was performed as described previously (6) using *Silencer*™ Select Negative Control No. 2 siRNA (cat# 4390846), FOXA2 siRNA #1 (sense: CCAUUAUGAACUCCUCUUATT; antisense: UAAGAGGAGUUCAUAAUGGGC) and FOXA2 siRNA #2 (sense: CCAUGAACAUGUCGUCGUATT; antisense: UACGACGACAUGUUCAUGGAG) from ThermoFisher (Waltham, MA). Forty-eight hours after transfection using Lipofectamine™ RNAiMAX Transfection Reagent (ThermoFisher; cat# 13778075) in A549 cells, RNA was extracted using TRIzol™ Reagent (ThermoFisher; cat# 15596026) according to the manufacturer's instructions.

H441 cells that stably express dCas9-KRAB were developed using the CRISPR/dCas9-KRAB lentiviral vector as we described previously (7-9). H441 cells that stably express dCas9-KRAB were infected with the PGK-IRES-EGFP lentiviral vector carrying rat *Foxa2* or empty control to develop H441 cells that stably express dCas9-KRAB with or without FOXA2 for the CRISPRi experiments. CRISPRi using these H441 cells were performed as described previously (7-9) by transiently transfecting synthetic sgRNA from the Invitrogen custom TrueGuide gRNA (sgRNA) ordering tool (ThermoFisher, Waltham, MA). Non-targeted gRNA (sgRNA) was used as a negative control (cat# A35526, ThermoFisher). The synthetic sgRNA target sequences are described in Fig. 6A (bottom panel). H441 or BEAS-2B cells were infected with the lentiviral vector expressing mouse *Foxa1*, rat *Foxa2* and/or the retroviral vector expressing *KRAS*^{G12V}. A549 cells were infected with the PGK-IRES-EGFP lentiviral vector carrying rat *Foxa2*, rat *Nkx2-1* or

empty control to develop A549 cells with or without ectopic FOXA2 or NKX2-1. The empty vectors were used as controls.

Protein and RNA were extracted as described above and previously (4). Co-immunoprecipitation analysis was performed as described previously (10) by immunoprecipitation using Flag antibody (Sigma-Aldrich; cat# F2426; RRID:AB_2616449) in the presence of ethidium bromide (200 mg/ml; ThermoFisher; cat# 15585011) (11) with cell extracts from A549 cells that were transiently transfected with Flag-SPDEF vector (12). Flag antibody immunoprecipitates were immunoblotted with FOXA2 antibody to detect SPDEF-FOXA2 protein-protein interaction. Immunoblotting assays were performed as described previously (10) using mouse anti-FOXA1 (1:5,000; clone 2F83; cat# WMAB-2F83; RRID:AB_451717) from Seven Hills Bioreagents (Cincinnati, OH), goat anti-FOXA2 (1:5,000; cat# AF2400; RRID:AB_2294104) from R&D Systems, rabbit anti-ACTA1 (actin; 1:5,000; cat# A2066; RRID:AB_476693) from Sigma-Aldrich (St. Louis, MO) and rabbit anti-$KRAS^{G12V}$ (1:5,000; cat# 14412S; RRID:AB_2714031) from Cell Signaling Technology (Danvers, MA).

RNA-seq using biological triplicates were performed as described previously (8) except that RNA was obtained from H441 cells that were infected with lentivirus (Control or rat *Foxa2*). Quality assessment and pre-processing of RNA-seq reads were performed using FASTQC, Trim Galore and SAMtools. Reads were then aligned to hg38 genome using Bowtie2. Low-quality alignments and PCR duplicates were removed using SAMtools and Picard MarkDuplicates tool. Gene expression was counted using htseq-count. Differential expression analysis was performed using Bioconductor DESeq2 package. Differential expression with at least a 2-fold change and false discovery rate < 0.1 was considered significant. Heatmap visualizations of gene expression

were performed using R package pHeatmap using variance stabilizing transformed expression values from the DESeq2 analysis.

TaqMan gene expression analysis was performed as described previously (4) with TaqMan probes (Hs00356521 m1 for human *AGR2*; Hs04187555_m1 for human *FOXA1*; Hs00232764_m1 for human *FOXA2*; Hs00270130_m1 for human *FOXA3*; Hs00230853_m1 for human *HNF4A*; Hs00873651_mH for human *MUC5AC*; Hs00861595_m1 for *MUC5B*; Hs00968940_m1 for human *NKX2-1*; Hs00167036_m1 for human *SFTPB*; Hs00171942_m1 for human *SPDEF*; Rn01415600_m1 for rat *Foxa2*; Rn01512482_m1 for rat *Nkx2-1*; Mm00484713_m1 for mouse *Foxa1* and Hs02758991 g1 for human *GAPDH*). A probe for *dCas9* is described in Stuart et al. (7).

TCGA LUAD data and ChIP-seq analysis

Normalized and log2-transformed gene expression data from TCGA LUAD RNA-seq datasets were retrieved from (13). ChIP-seq datasets (called peaks and alignment wiggle files) that identified FOXA1 and FOXA2 binding sites in mouse $Kras^{G12D}$-lung tumor cells in the presence or absence of *Nkx2-1* were retrieved from GSE43252 (14). ChIP-seq data that identified FOXA2 binding sites (irreproducibility discovery rate < 0.05) in human A549 lung carcinoma cells were retrieved from ENCODE using accession number ENCFF686MSH (15 [ENCSR000BRE], 16). ChIP-seq datasets that identified SPDEF binding sites in MCF7 breast adenocarcinoma cells and A549 cells were obtained from GSE48930 (17) and our previous study (12), respectively. To compare with other ChIP-seq data in the present study, genomic locations of SPDEF binding sites were converted to hg38 using UCSC liftover tool. To identify FOXA2 binding sites in H441 cells, we generated H441 cells expressing ectopic FOXA2 using the lentivirus described above and

performed ChIP-seq analysis as described previously (12) except using goat anti-FOXA2 (cat#AF2400) from R&D Systems. We also generated BEAS-2B cells expressing FOXA2 and/or $KRAS^{G12V}$ and performed ChIP-seq as described previously (12) except using rabbit anti-H3K27ac (anti-histone H3 [acetyl K27] antibody; cat# ab4729; RRID:AB_2118291) from Abcam (Cambridge, MA). Illumina DNA sequencing from two biological replicates (IP and input) was performed at the CCHMC DNA Sequencing and Genotyping Core. For each sequenced read, quality assessment and pre-processing were performed using FASTQC, Trim Galore and SAMtools. Reads were aligned to hg38 genome using Bowtie2, using the following setting: --end-to-end –sensitive. Low-quality alignments and duplicate reads were removed using SAMtools and Picard MarkDuplicates tool. To identify FOXA2 binding sites in H441 cells, peaks for each biological replicate were called using MACS2 callpeak v2.1.0 (18; https://github.com/taoliu/MACS/) with the following parameters "-g hs -B --SPMR --keep-dup all -p 0.01 --nomodel --shift 0 --extsize ${FRAGLEN}", where "FRAGLEN" was 241 and 161 for the replicate 1 and 2, respectively. ENCODE blacklist regions (accession number ENCFF356LFX) were filtered. Irreproducibility discovery rate (IDR) framework (v2.0.4.2; 19) was used to compare peaks from the two replicates and identify reproducible peaks with IDR < 0.05. H3K27ac peaks in BEAS-2B cells were called using MACS2 callpeak using the following parameter: "-g hs --SPMR --broad --broad-cutoff=0.1 --keep-dup 1 --bdg --qvalue 0.05". ENCODE blacklist regions (ENCFF356LFX) were filtered. Differentially bound sites (p value < 0.01 and fold value ≥ 2; 20) were identified using DiffBind (v2.14) with edgeR analysis. Visualization of differentially bound sites was performed using R tornadoplot package (21). Visualization of ChIP-seq peaks and signals in genomic regions were performed using UCSC genome browser. ChIP-seq signal tracks were generated using MACS2 bdgcomp to calculate fold

enrichment in treatment over input control. The genome distribution of ChIP-seq peaks was analyzed using CEAS v1.0.0 (22; http://liulab.dfci.harvard.edu/CEAS/) in Cistrome (http://cistrome.org/ap/). Motif enrichment analysis was performed using HOMER v4.10 (23).

References for Supplementary data

1. Chen G, Wan H, Luo F, Zhang L, Xu Y, Lewkowich I, et al. Foxa2 programs Th2 cell-mediated innate immunity in the developing lung. *J Immunol*. 2010;184:6133-41.

2. Fisher GH, Wellen SL, Klimstra D, Lenczowski JM, Tichelaar JW, Lizak MJ, et al. Induction and apoptotic regression of lung adenocarcinomas by regulation of a K-Ras transgene in the presence and absence of tumor suppressor genes. *Genes Dev*. 2001;15:3249-62.

3. Politi K, Zakowski MF, Fan PD, Schonfeld EA, Pao W, Varmus HE. Lung adenocarcinomas induced in mice by mutant EGF receptors found in human lung cancers respond to a tyrosine kinase inhibitor or to down-regulation of the receptors. *Genes Dev*. 2006;20:1496-510.

4. Maeda Y, Tsuchiya T, Hao H, Tompkins DH, Xu Y, Mucenski ML, et al. Kras(G12D) and Nkx2-1 haploinsufficiency induce mucinous adenocarcinoma of the lung. *J Clin Invest*. 2012;122:4388-400.

5. Thakore PI, D'Ippolito AM, Song L, Safi A, Shivakumar NK, Kabadi AM, et al. Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat Methods*. 2015;12:1143-9.

6. Maeda Y, Chen G, Xu Y, Haitchi HM, Du L, Keiser AR, et al. Airway epithelial transcription factor NK2 homeobox 1 inhibits mucous cell metaplasia and Th2 inflammation. *Am J Respir Crit Care Med*. 2011;184:421-9.

7. Stuart WD, Guo M, Fink-Baldauf IM, Coleman AM, Clancy JP, Mall MA, et al. CRISPRi-mediated functional analysis of lung disease-associated loci at non-coding regions. *NAR Genom Bioinform*. 2020;2:lqaa036.

8. Stuart WD, Fink-Baldauf IM, Tomoshige K, Guo M, Maeda Y. CRISPRi-mediated functional analysis of NKX2-1-binding sites in the lung. *Commun Biol*. 2021;4:568.

9. Fink-Baldauf IM, Stuart WD, Brewington JJ, Guo M, Maeda Y. CRISPRi links COVID-19 GWAS loci to LZTFL1 and RAVER1. *EBioMedicine*. 2022;75:103806.

10. Maeda Y, Hunter TC, Loudy DE, Davé V, Schreiber V, Whitsett JA. PARP-2 interacts with TTF-1 and regulates expression of surfactant protein-B. *J Biol Chem*. 2006;281:9600-6.

11. Lai JS, Herr W. Ethidium bromide provides a simple tool for identifying genuine DNA-independent protein associations. *Proc Natl Acad Sci U S A*. 1992;89:6958-62.

12. Guo M, Tomoshige K, Meister M, Muley T, Fukazawa T, Tsuchiya T, et al. Gene signature driving invasive mucinous adenocarcinoma of the lung. *EMBO Mol Med*. 2017;9:462-481.

13. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511:543-50.

14. Snyder EL, Watanabe H, Magendantz M, Hoersch S, Chen TA, Wang DG, et al. Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. *Mol Cell*. 2013;50:185-99.

15. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489:57-74.

16. Gertz J, Savic D, Varley KE, Partridge EC, Safi A, Jain P, et al. Distinct properties of cell-type-specific and shared transcription factor binding sites. *Mol Cell*. 2013;52:25-36.

17. Fletcher MN, Castro MA, Wang X, de Santiago I, O'Reilly M, Chin SF, et al. Master regulators of FGFR2 signalling and breast cancer risk. *Nat Commun*. 2013;4:2464.

18. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 2008;9:R137.

19. Li Q, Brown JB, Huang H, Bickel PJ, Measuring reproducibility of high-throughput experiments. *Ann Appl Stat*. 2011;5:1752-79.

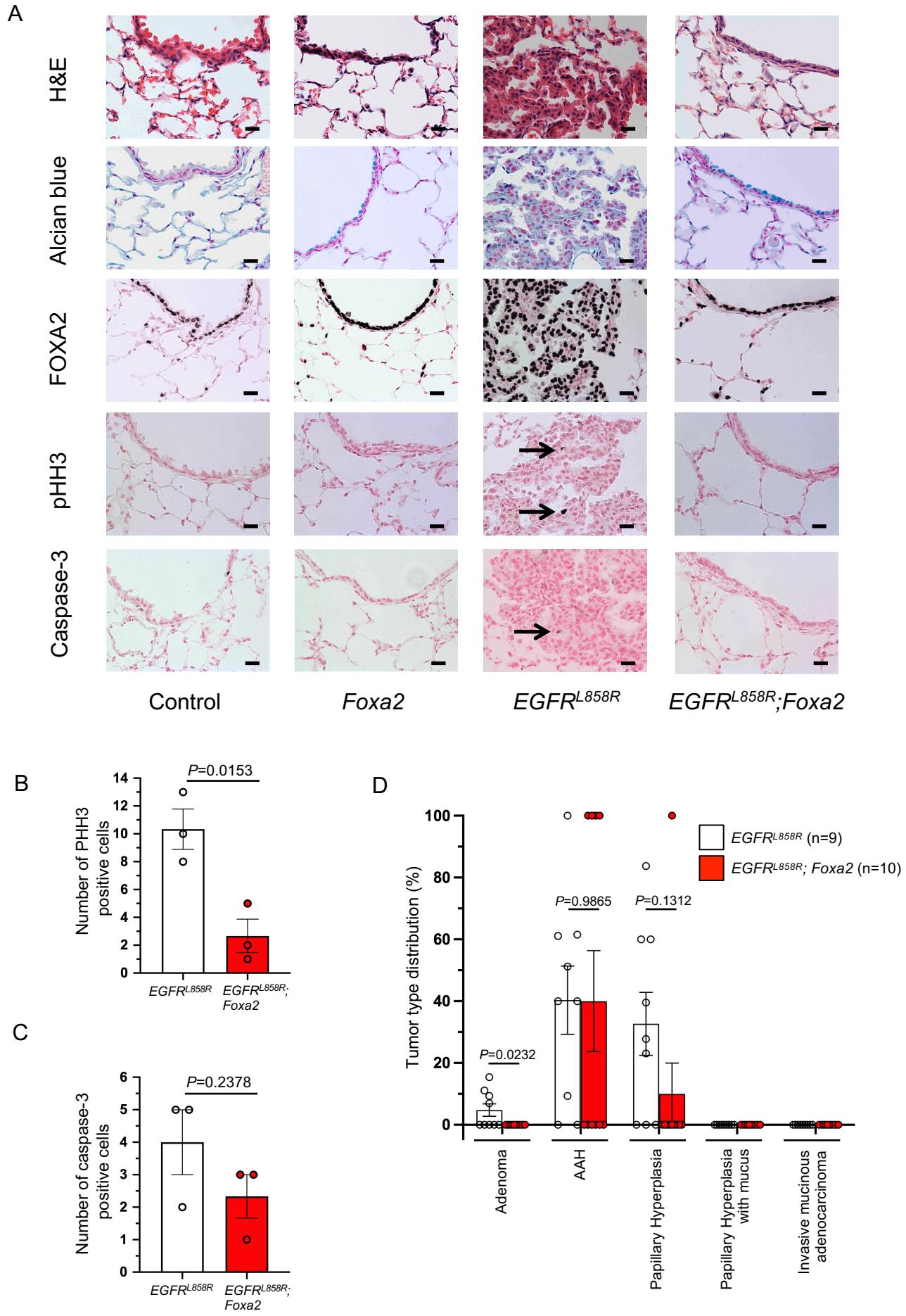
20. Ross-Innes CS, Stark R, Teschendorff AE, Holmes KA, Ali HR, Dunning MJ, et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature*. 2012;481:389-93.

21. Liu Z, Kraus WL. Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci. *Mol Cell*. 2017;65:589-603.e9.

22. Shin H, Liu T, Manrai AK, Liu XS. CEAS: cis-regulatory element annotation system. *Bioinformatics*. 2009;25:2605-6.

23. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38:576-89.

24. Chang JC, Offin M, Falcon C, Brown D, Houck-Loomis BR, Meng F, et al. Comprehensive Molecular and Clinicopathologic Analysis of 200 Pulmonary Invasive Mucinous Adenocarcinomas Identifies Distinct Characteristics of Molecular Subtypes. *Clin Cancer Res.* 2021;27:4066-76.


A
H&E
Alcian blue
FOXA2
pHH3
Caspase-3
Control
Foxa2
EGFR^L858R
EGFR^L858R;Foxa2
B
P=0.0153
Number of PHH3 positive cells
EGFR^L858R
EGFR^L858R; Foxa2
C
P=0.2378
Number of caspase-3 positive cells
EGFR^L858R
EGFR^L858R; Foxa2
D
EGFR^L858R (n=9)
EGFR^L858R; Foxa2 (n=10)
Tumor type distribution (%)
P=0.0232
P=0.9865
P=0.1312
Adenoma
AAH
Papillary Hyperplasia
Papillary Hyperplasia with mucus
Invasive mucinous adenocarcinoma

Supplementary Figure S1. FOXA2 inhibits the proliferation of EGFR^{L858R}-lung tumor cells in mice.

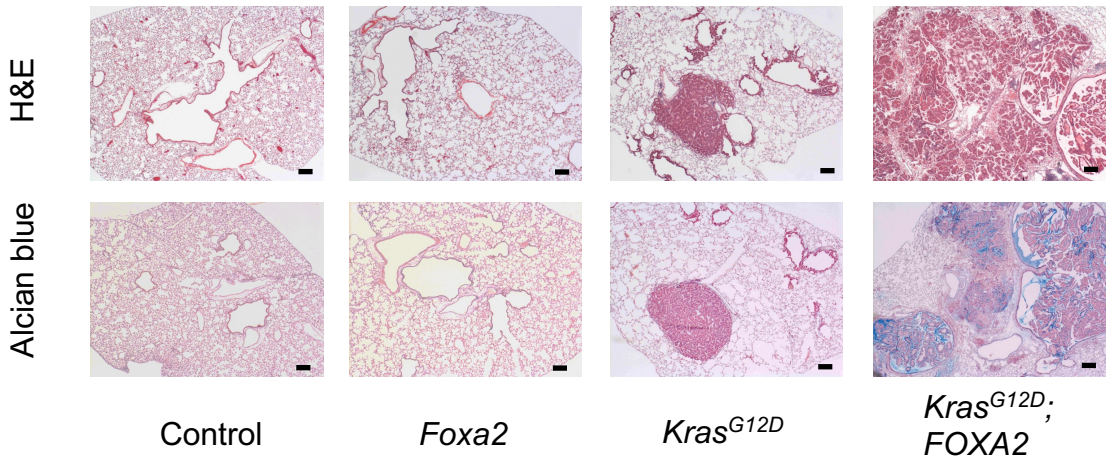
(A) Analysis of histology and immunohistochemistry on mouse lungs that conditionally expressed FOXA2 and/or EGFR^{L858R} was performed. Alcian blue staining was performed to detect mucins. Staining of PHH3 (phosphohistone H3; a marker for proliferation) and caspase-3 (a marker for apoptosis) was performed to assess whether FOXA2 influenced proliferation and apoptosis of EGFR^{L858R}-lung tumors. Black dots in the nuclei indicate the expression of FOXA2. H&E, hematoxylin and eosin. X-axis indicates mouse genotype. Scale bar: 20 μm.

(B) Number of PHH3 positive cells in lungs of different mouse groups was counted in 3 different views per section per mouse. PHH3 was significantly reduced in lungs of *EGFR^{L858R}*;*Foxa2* mice compared to those of *EGFR^{L858R}* mice (n = 3 mice).

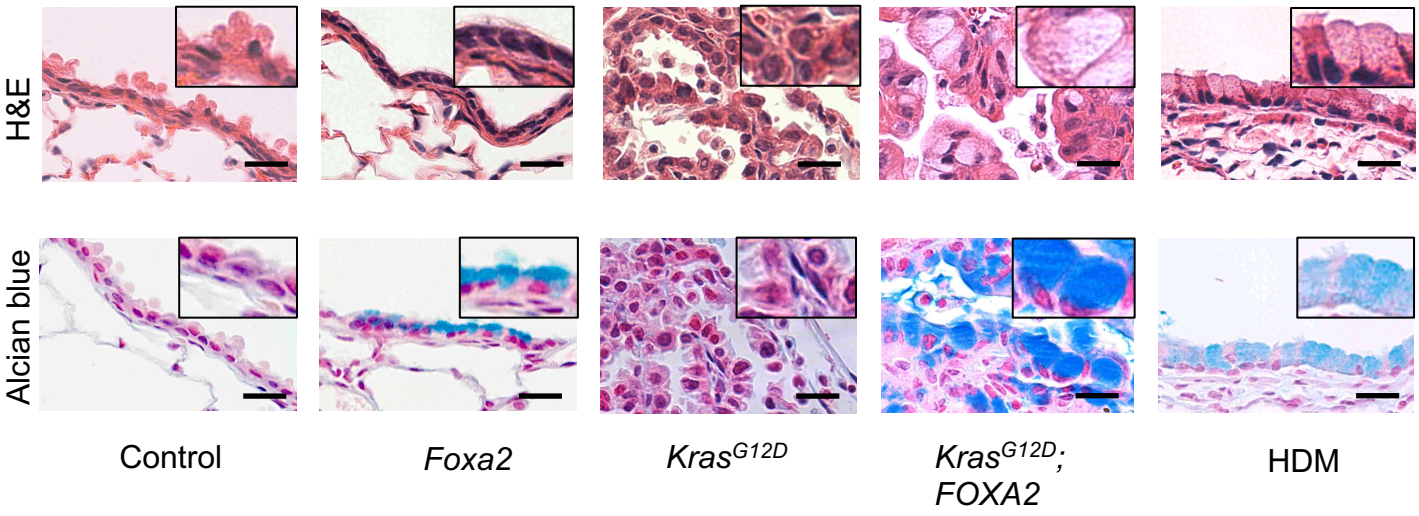
(C) Number of caspase-3 positive cells in lungs of different mouse groups was counted in 3 different views per section per mouse. Caspase-3 (a marker for apoptosis) was not altered in lungs of different mouse groups (n = 3 mice).

(D) Differences in histology of mouse lung tumors developed by EGFR^{L858R} and/or FOXA2 were assessed. Adenoma developed by EGFR^{L858R} was significantly suppressed by FOXA2 ($n \geq 9$). AAH: Atypical adenomatous hyperplasia.

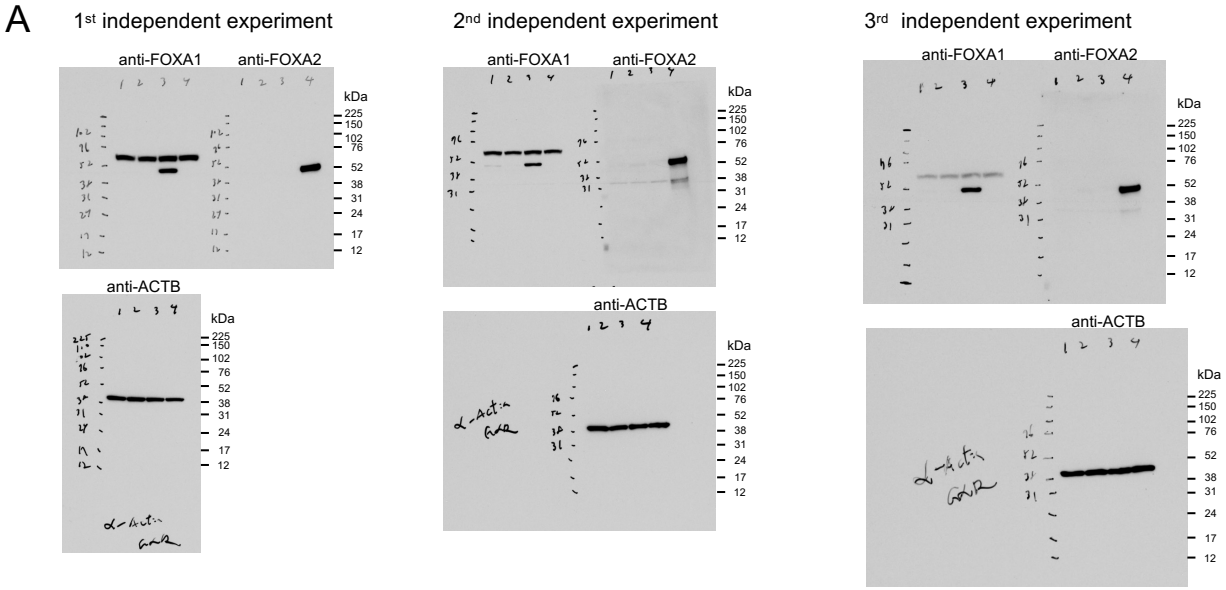
All tests are unpaired, two-tailed Student's t-tests. Data are presented as mean +/- SEM.

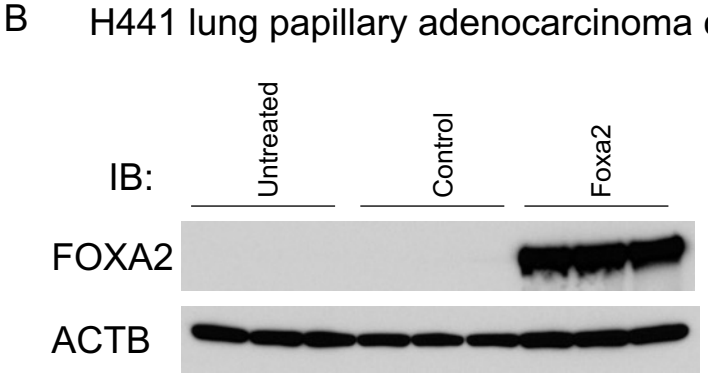


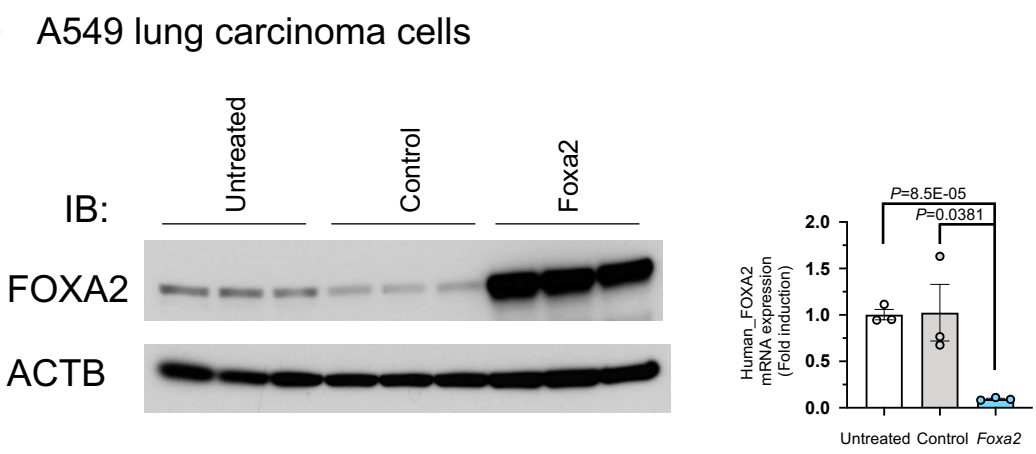
Supplementary Figure S2. FOXA2 promotes the growth of KRAS-mutant lung tumors containing mucinous lung tumors. Histology of mouse lung tumors influenced by the ectopic expression of rat *Foxa2* and/or *Kras*^{G12D} was examined by H&E and Alcian blue staining. H&E, hematoxylin and eosin. Alcian blue staining detects mucins. X-axis indicates mouse genotype. See Fig. 1A for the transgenic mouse model expressing rat *Foxa2* and/or *Kras*^{G12D}. Scale bar: 200 μm.



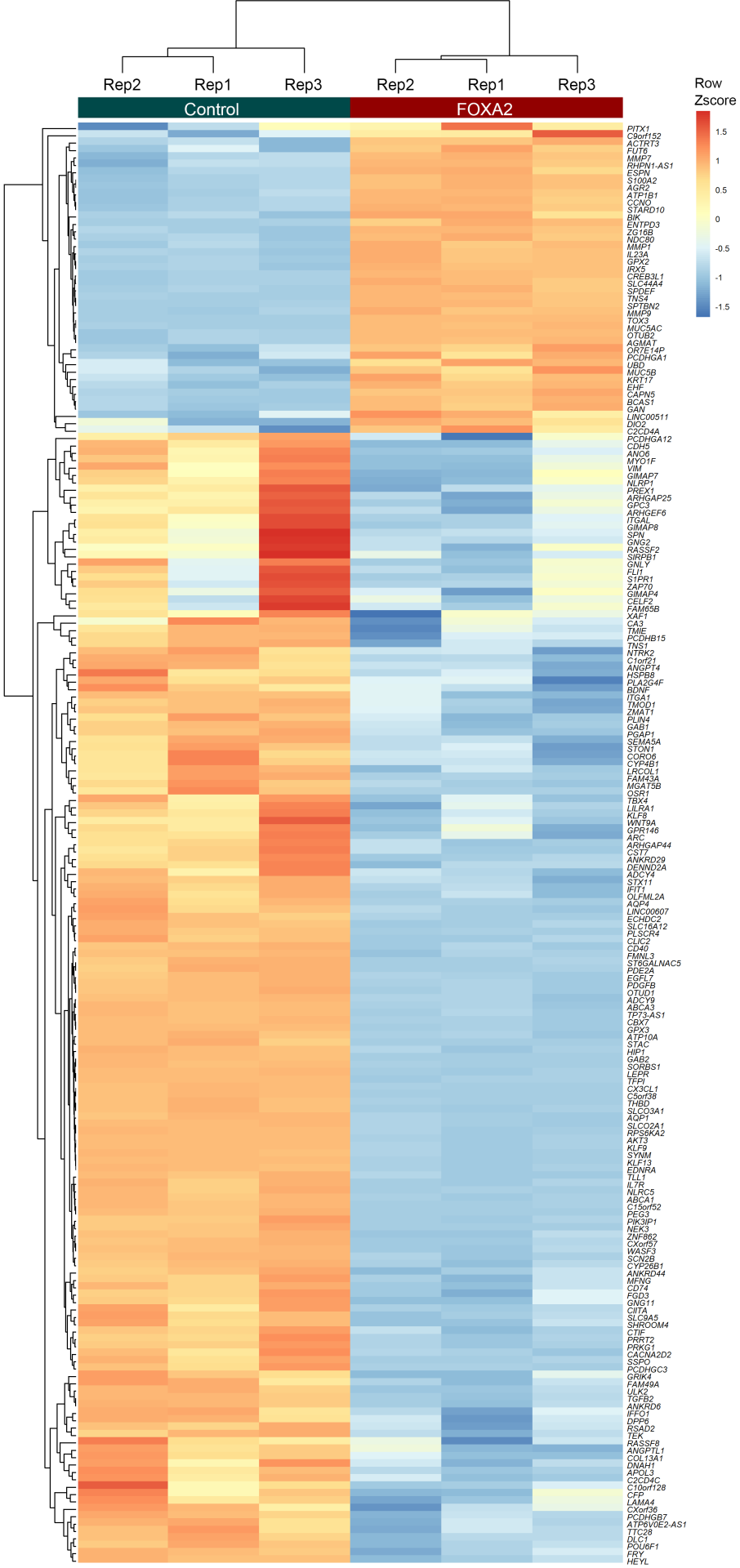
Supplementary Figure S3. Magnified views of immature and mature goblet (mucous) cells in mouse lungs.
Immature goblet (mucous) cells were induced by the ectopic expression of FOXA2 alone (*Foxa2*) whereas mature (or fully differentiated) goblet (mucous) cells were induced by HDM (house dust mite allergen; asthma model) or the co-expression of mutant KRAS and FOXA2 (*Kras*^{G12D};*Foxa2*). H&E, hematoxylin and eosin. Alcian blue staining detects mucins. X-axis indicates mouse genotype. Scale bar: 20 μm.





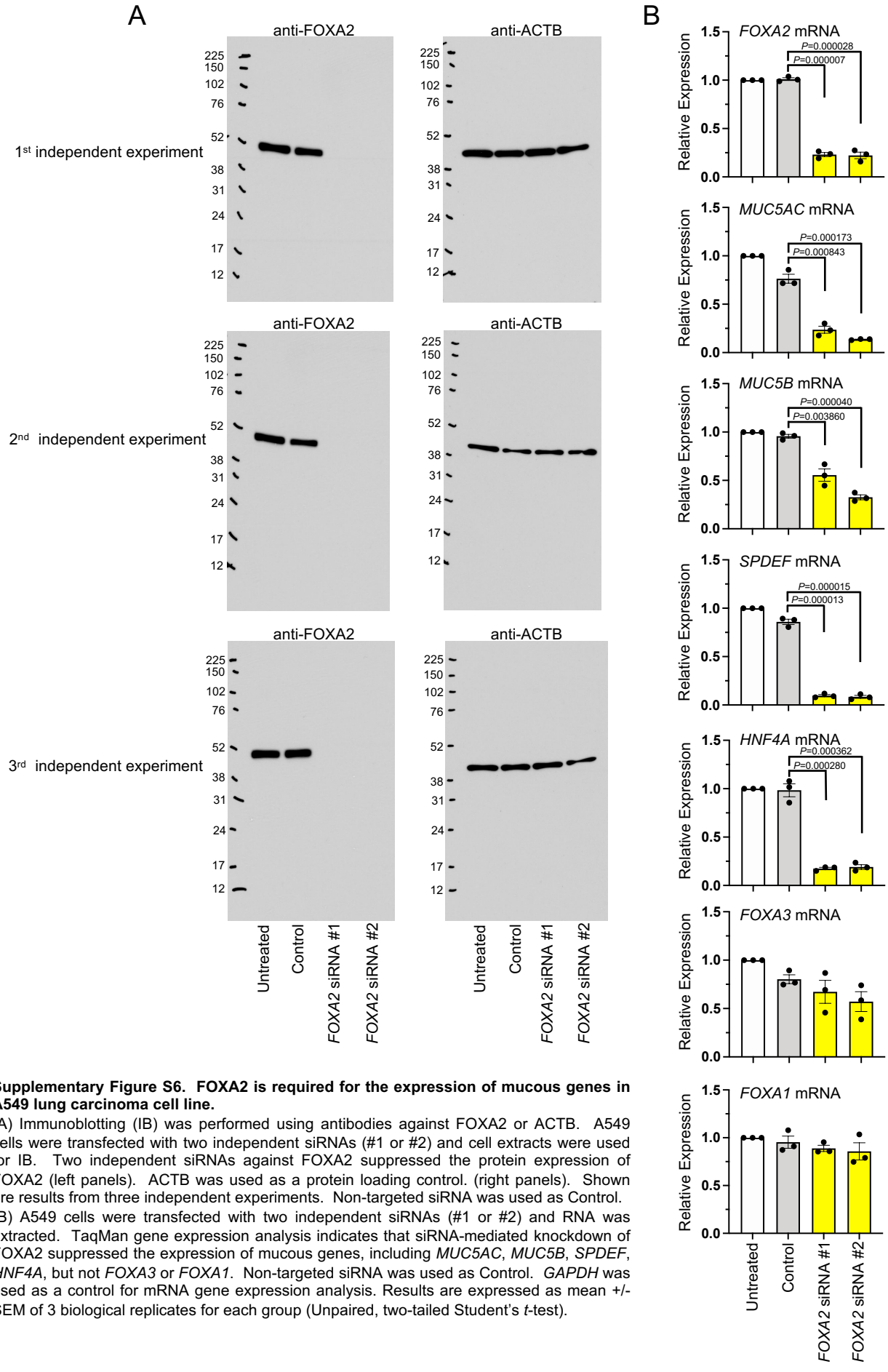


Supplementary Figure S4. FOXA2 suppresses the mRNA expression of FOXA2 in human lung adenocarcinoma cell lines.
(A) Immunoblotting (IB) was performed using antibodies against FOXA1 or FOXA2. H441 human lung papillary adenocarcinoma cells were infected with a lentiviral vector carrying mouse *Foxa1* or rat *Foxa2* and cell extracts were used for IB. Shown are results from three independent experiments. See Fig. 5A for details.
(B) H441 cells were infected with lentivirus carrying rat *Foxa2* or empty control from which protein and RNA were extracted. TaqMan gene expression analysis indicates that the ectopic expression of FOXA2 detected by an immunoblot (left panel) repressed the mRNA expression of endogenous *FOXA2* (right panel). ACTB was used as a protein loading control (left panel). GAPDH was used as a control for mRNA gene expression analysis (right panel).
(C) A549 human lung carcinoma cells (B) were infected with lentivirus carrying rat *Foxa2* or empty control and protein and RNA were extracted. TaqMan gene expression analysis indicates that the ectopic expression of FOXA2 detected by an immunoblot (left panel) repressed the mRNA expression of endogenous *FOXA2* (right panel). ACTB was used as a protein loading control (left panel). GAPDH was used as a control for mRNA gene expression analysis (right panel). Results are expressed as mean +/- SEM of 3 biological replicates for each group (Unpaired, two-tailed Student's *t*-test).



Supplementary Figure S5. Differentially expressed genes affected by ectopic expression of FOXA2 (rat *Foxa2*) in the H441 human lung papillary adenocarcinoma cell line.

RNA-seq was performed using RNAs from H441 cells expressing ectopic FOXA2 as described in Fig. 5B (n = 3). Control, an empty lentiviral vector. Shown are FOXA2-induced genes that are significantly increased in human invasive mucinous adenocarcinoma of the lung (IMA) compared to normal lung and FOXA2-reduced genes that are significantly decreased in human IMA compared to normal lung (DESeq2: fold change ≥ 2, FDR < 0.1).

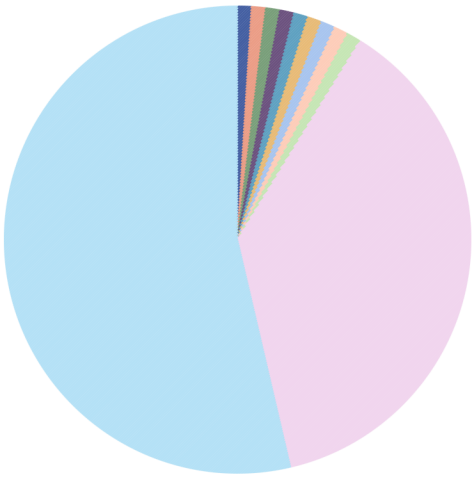


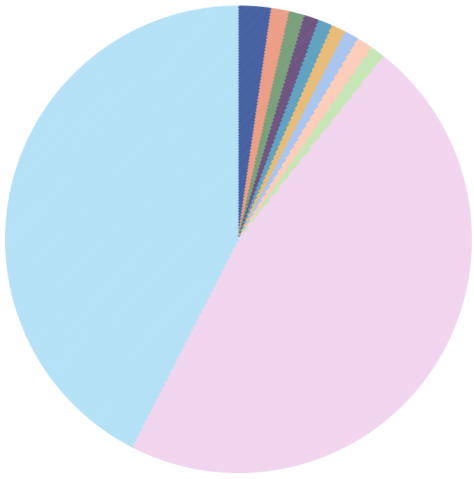
Supplementary Figure S6. FOXA2 is required for the expression of mucous genes in A549 lung carcinoma cell line.

(A) Immunoblotting (IB) was performed using antibodies against FOXA2 or ACTB. A549 cells were transfected with two independent siRNAs (#1 or #2) and cell extracts were used for IB. Two independent siRNAs against FOXA2 suppressed the protein expression of FOXA2 (left panels). ACTB was used as a protein loading control. (right panels). Shown are results from three independent experiments. Non-targeted siRNA was used as Control.

(B) A549 cells were transfected with two independent siRNAs (#1 or #2) and RNA was extracted. TaqMan gene expression analysis indicates that siRNA-mediated knockdown of FOXA2 suppressed the expression of mucous genes, including *MUC5AC*, *MUC5B*, *SPDEF*, *HNF4A*, but not *FOXA3* or *FOXA1*. Non-targeted siRNA was used as Control. *GAPDH* was used as a control for mRNA gene expression analysis. Results are expressed as mean +/- SEM of 3 biological replicates for each group (Unpaired, two-tailed Student's *t*-test).

A





- Promoter (<=1000 bp): 0.9 %
- Promoter (1000−2000 bp): 0.8 %
- Promoter (2000−3000 bp): 0.7 %
- Downstream (<=1000 bp): 0.8 %
- Downstream (1000−2000 bp): 0.7 %
- Downstream (2000−3000 bp): 0.7 %
- 5'UTR: 0.4 %
- 3'UTR: 1.0 %
- Coding exon: 0.9 %
- Intron: 38.3 %
- Distal intergenic: 54.8 %

- Promoter (<=1000 bp): 1.8 %
- Promoter (1000−2000 bp): 1.1 %
- Promoter (2000−3000 bp): 1.1 %
- Downstream (<=1000 bp): 1.0 %
- Downstream (1000−2000 bp): 1.1 %
- Downstream (2000−3000 bp): 0.8 %
- 5'UTR: 0.5 %
- 3'UTR: 1.1 %
- Coding exon: 0.6 %
- Intron: 48.7 %
- Distal intergenic: 42.3 %

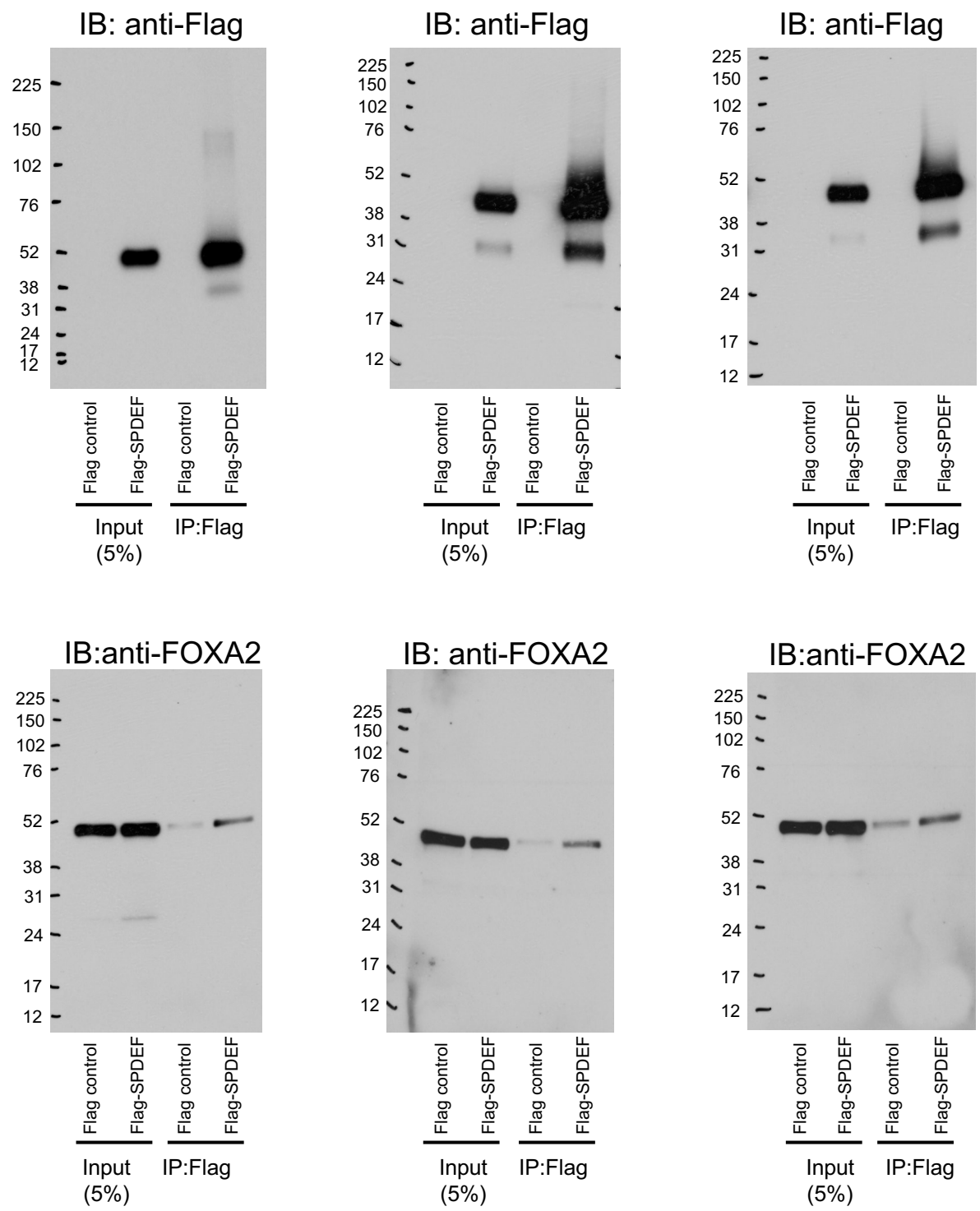
B

Motif	Name	pvalue
	FOXA1(Forkhead)	1e-1770
	FOXA1(Forkhead)	1e-1761
	FOXM1(Forkhead)	1e-1619
	Foxa2(Forkhead)	1e-1600
	Foxa3(Forkhead)	1e-1430
	Fox:Ebox(Forkhead,bHLH)	1e-1243
	Fra1(bZIP)	1e-1017
	Atf3(bZIP)	1e-1012
	JunB(bZIP)	1e-1001
	BATF(bZIP)	1e-994

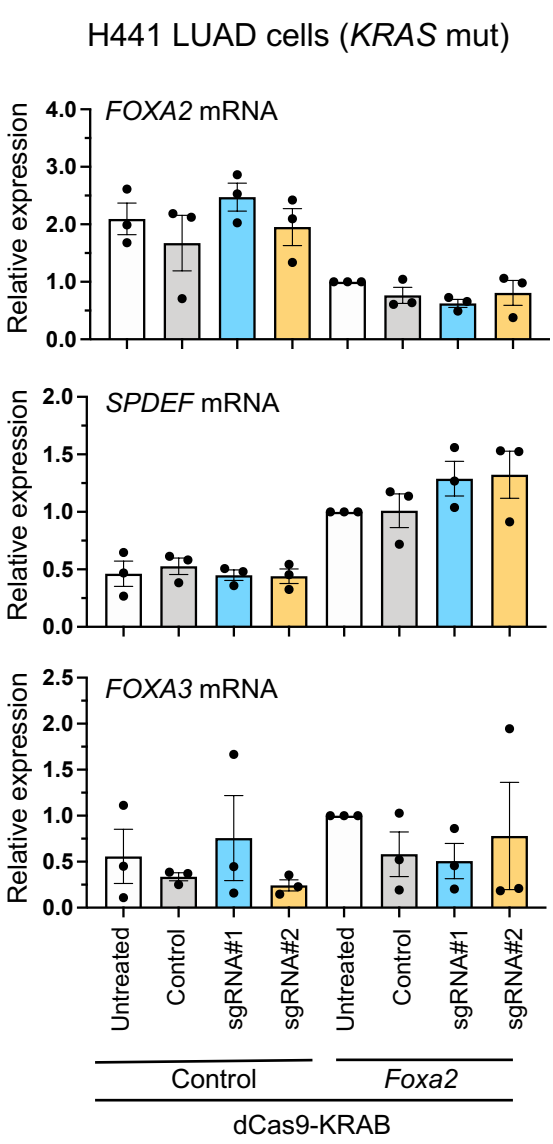
Supplementary Figure S7. Genome-wide location and motif of FOXA2 binding sites in the H441 human lung papillary adenocarcinoma cell line.

(A) ChIP-seq was performed to identify FOXA2 binding sites in H441 cells expressing ectopic FOXA2. Genome-wide locations of FOXA2 binding sites are shown.

(B) Most highly enriched motifs of FOXA2 binding sites obtained from the ChIP-seq in A were identified. Motif enrichment analysis was performed using HOMER on ChIP-seq peaks with an irreproducible discovery rate < 0.05. Top 10 motifs with most significant p-values are shown.

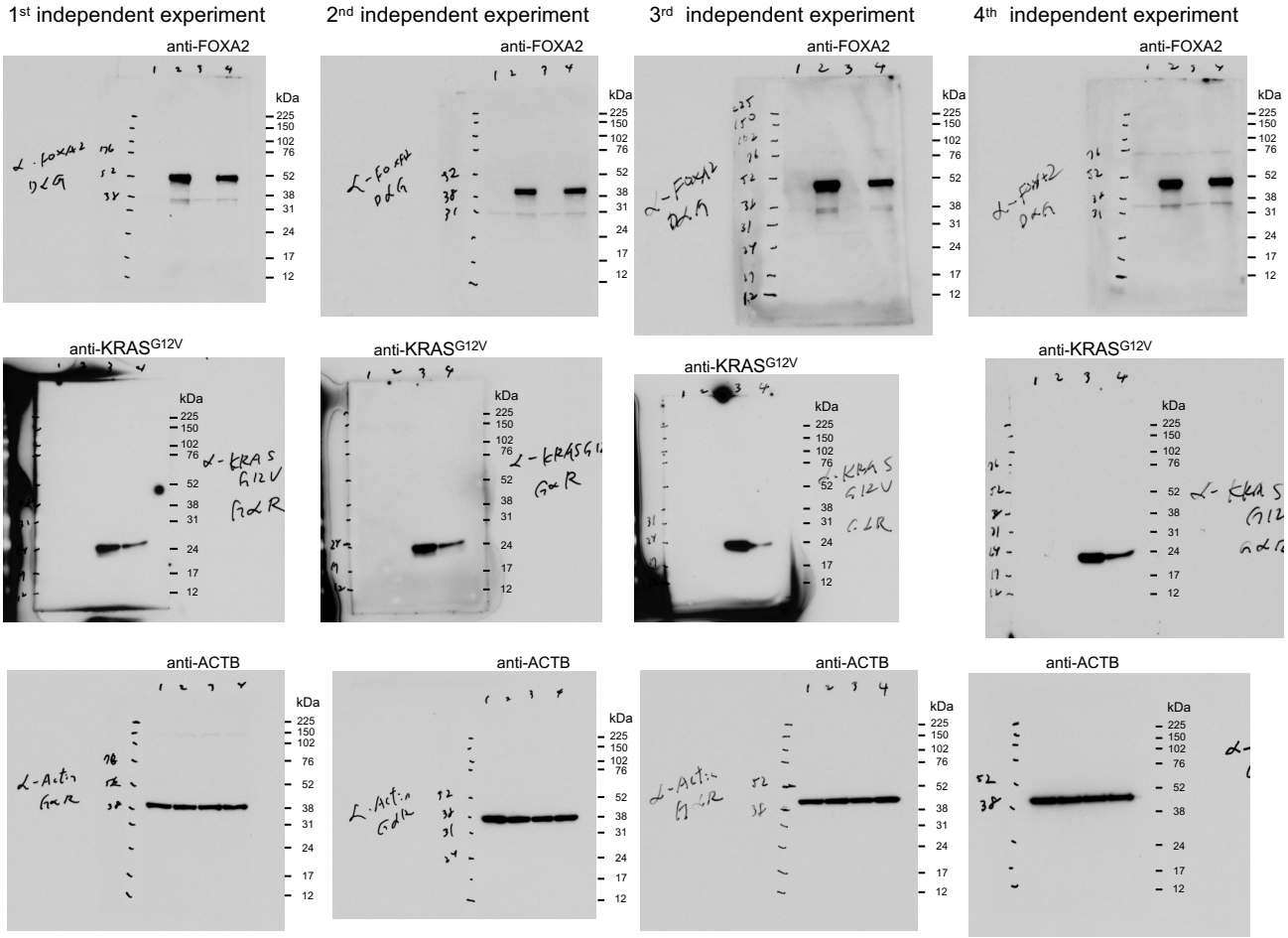


Supplementary Figure S8. SPDEF interacts with FOXA2.
Co-immunoprecipitation experiments were performed using A549 lung carcinoma cells that are transfected with Flag empty vector control or Flag-tagged SPDEF expression vector. Flag antibody immunoprecipitates (IP) were immunoblotted (IB) with antibodies against Flag (top panels) or FOXA2 (bottom panels). Shown are images from three independent experiments.



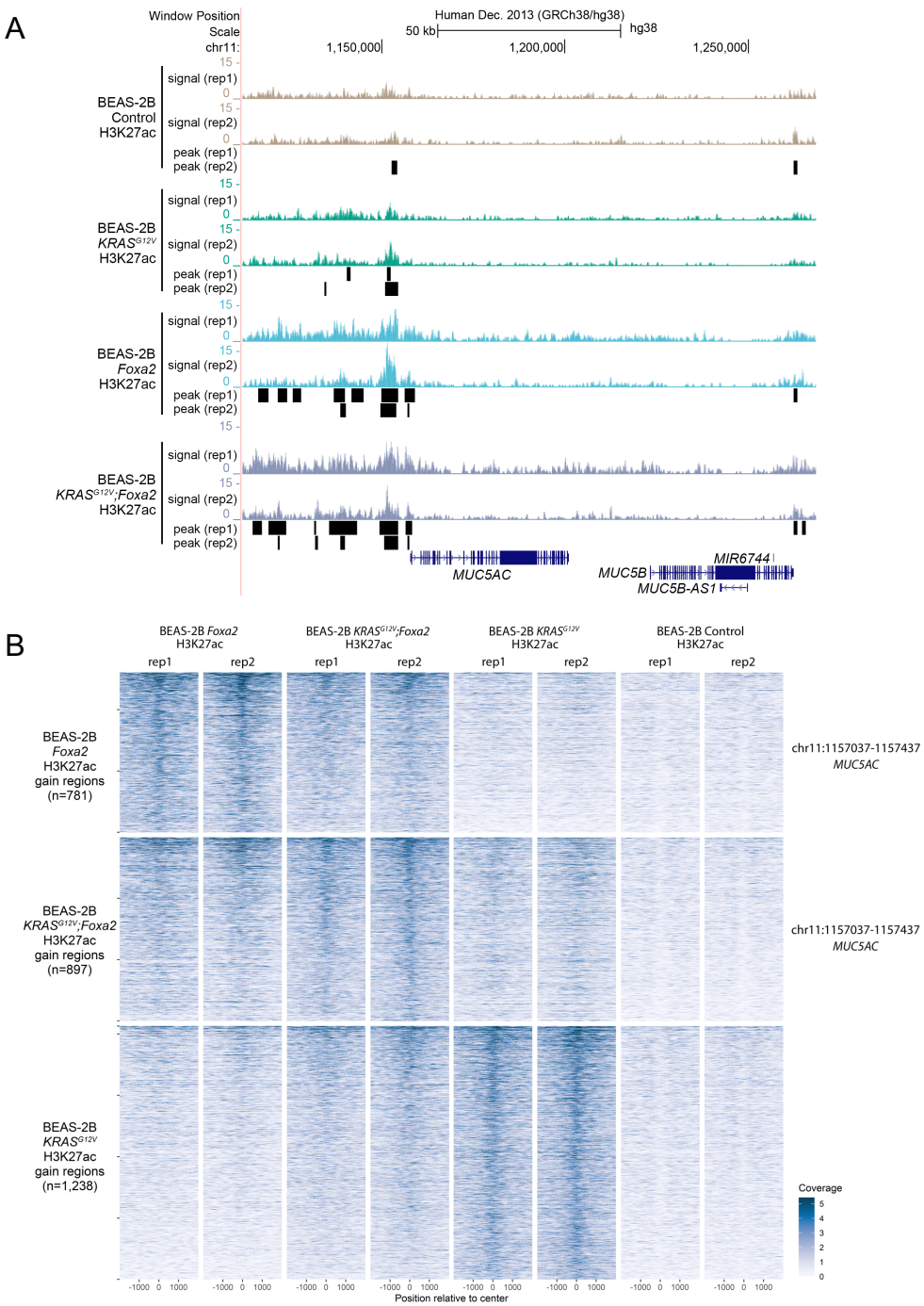
Supplementary Figure S9. Gene expression analysis from H441 *KRAS*-mutant lung adenocarcinoma cells infected with lentivirus carrying *dCas9-KRAB* with or without rat *Foxa2*.

TaqMan gene expression analysis was performed as described in Materials and Methods. H441 cells that stably express dCas9-KRAB along with ectopic FOXA2 were transiently transfected with synthetic sgRNA that target the loci of *MUC5AC* and *MUC5B* as described in Fig. 6 and Materials and Methods. Results are expressed as mean +/- SEM of 3 biological replicates for each group.



Supplementary Figure S10. Validation of the ectopic protein expression of FOXA2 and KRAS^{G12V} in BEAS-2B human transformed bronchial epithelial cells.

Immunoblotting (IB) was performed using antibodies against FOXA2 or KRAS^{G12V}. BEAS-2B transformed human bronchial epithelial cells were infected with a lentiviral vector carrying rat *Foxa2* and/or a retroviral vector carrying human *KRAS^{G12V}* and cell extracts were used for IB. Shown are images from four independent experiments. See Fig. 7A for details.

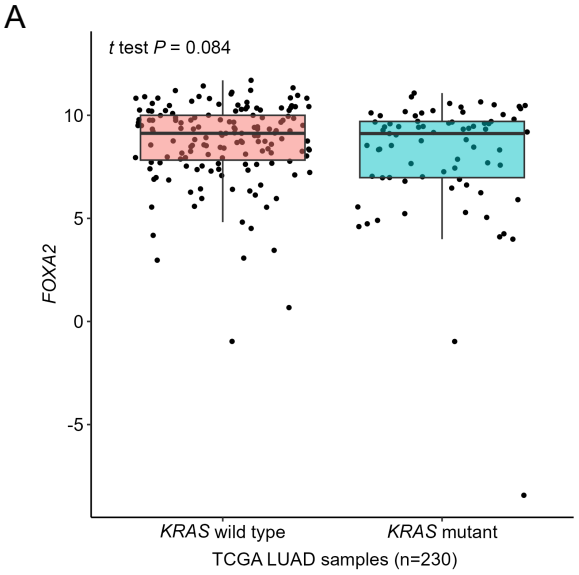


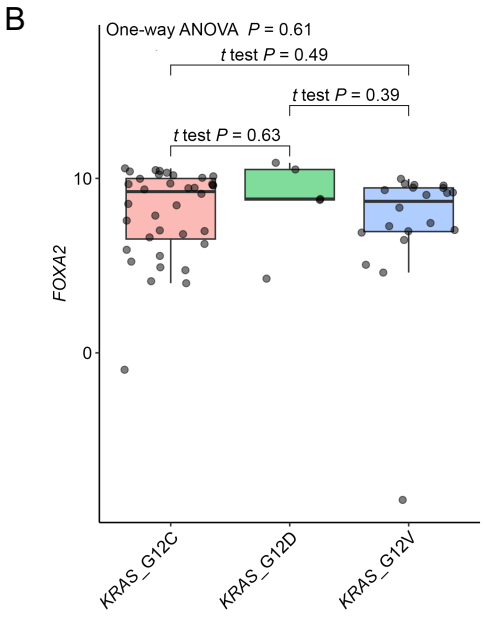
Supplementary Figure S11. FOXA2 but not mutant KRAS initiates H3K27ac at the upstream enhancer region of the human *MUC5AC* gene in BEAS-2B transformed bronchial epithelial cells.

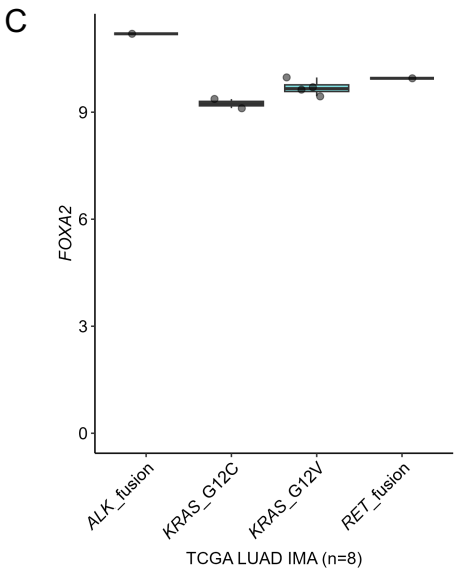
(A) ChIP-seq using an antibody against H3K27ac (an enhancer mark of histone) was performed using BEAS-2B cells that ectopically expressed FOXA2 and/or KRAS^{G12V} as described in Fig. 7. Shown is the genome browser demonstrating two independent datasets showing the gain of H3K27ac marks at the loci of *MUC5AC* in the BEAS-2B cells expressing FOXA2 or KRAS^{G12V};FOXA2 but not KRAS^{G12V} compared to Control.

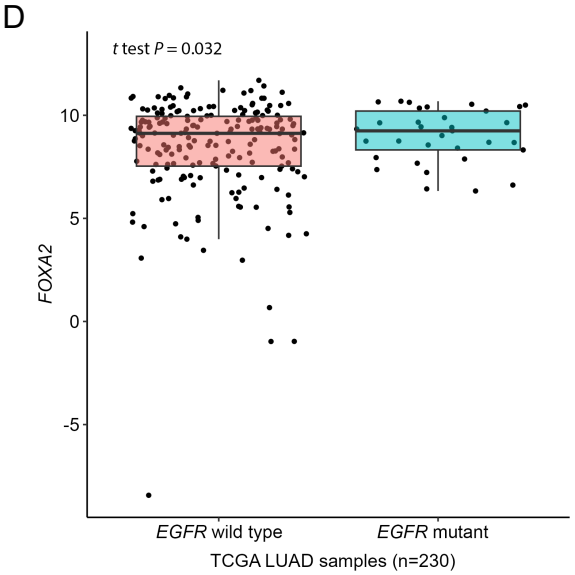
(B) Shown is a heatmap of ChIP-seq signals in genomic regions (n = 2,916) that had enriched H3K27ac in FOXA2, KRAS^{G12V};FOXA2, or KRAS^{G12V} when compared to Control. H3K27ac is located at the *MUC5AC* locus in BEAS-2B cells expressing FOXA2 or KRAS^{G12V};FOXA2 but not KRAS^{G12V}. The heatmap was generated using the R tornadoplot package (https://github.com/teunbrand/tornadoplot).

(C) Principal component analysis from two independent datasets indicates that four different conditions (Control, FOXA2, KRAS^{G12V} and KRAS^{G12V};FOXA2) formed four groups.



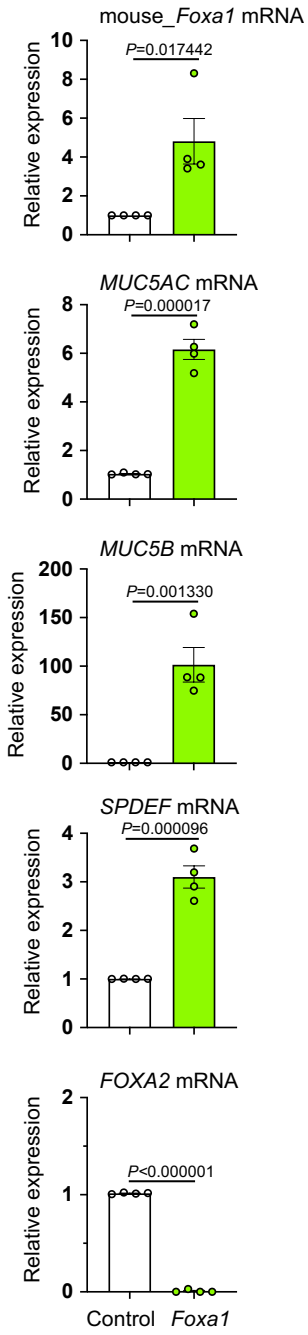






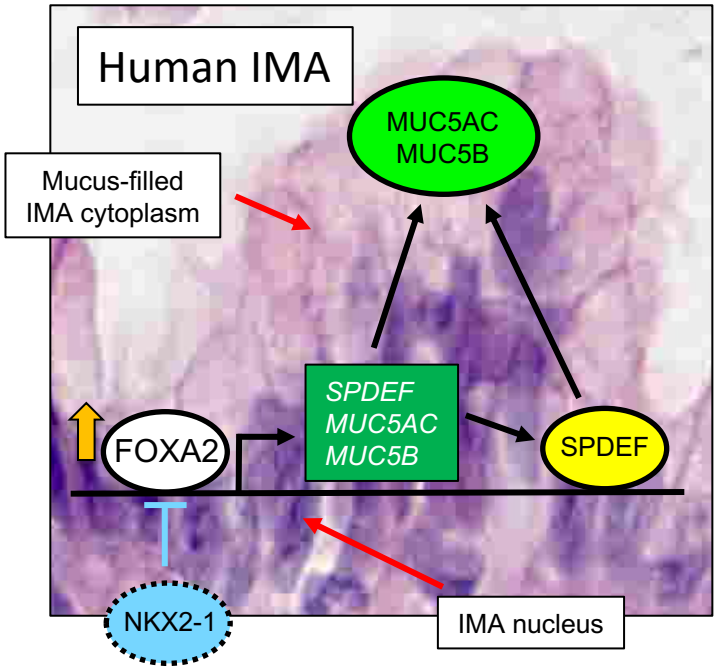
Supplementary Figure S12. *EGFR* mutations but not *KRAS* mutations affect the expression of *FOXA2* mRNA in TCGA LUAD datasets.

Shown is normalized and log2-transformed mRNA expression of *FOXA2* in TCGA LUAD *KRAS* wild type and *KRAS* mutant (A), $KRAS^{G12C}$, $KRAS^{G12D}$ and $KRAS^{G12V}$ in TCGA LUAD (B), $KRAS^{G12C}$, $KRAS^{G12V}$, *ALK* and *RET* fusions in TCGA LUAD IMA (C) and TCGA LUAD *EGFR* wild type and *EGFR* mutant (D). $KRAS^{G12C}$ is the most common *KRAS* mutation in LUAD (13) while $KRAS^{G12D}$ and $KRAS^{G12V}$ are the most common *KRAS* mutations in LUAD IMA (24). $P < 0.05$ was considered significant (Unpaired, two-tailed Welch's *t*-test).



Supplementary Figure S13. FOXA1 induces the expression of mucous genes in H441 human lung papillary adenocarcinoma cell line.

H441 cells were infected with lentivirus carrying mouse *Foxa1* or empty control (Control) and RNA was extracted. TaqMan gene expression analysis indicates that the ectopic expression of *Foxa1* (also confirmed by immunoblots in Fig. 5A; Supplementary Fig. S4A) induced the mRNA expression of mucous genes, including *MUC5AC*, *MUC5B* and *SPDEF*. *GAPDH* was used as a control for mRNA gene expression analysis. Results are expressed as mean +/- SEM of 4 biological replicates for each group (Unpaired, two-tailed Student's *t*-test).



Supplementary Figure S14. Transcriptional regulation of invasive mucinous adenocarcinoma of the lung (IMA).
Schematic indicates a proposed model of the mechanism by which FOXA2 induces human IMA in the presence of a *KRAS* mutation by activating a gene regulatory network critical for IMA.