Rapid modelling of cooperating genetic events in cancer through somatic genome editing

Citation: Sánchez-Rivera, Francisco J., Thales Papagiannakopoulos, Rodrigo Romero, Tuomas Tammela, Matthew R. Bauer, Arjun Bhutkar, Nikhil S. Joshi, et al. "Rapid Modelling of Cooperating Genetic Events in Cancer through Somatic Genome Editing." Nature 516, no. 7531 (October 22, 2014): 428–431.

As Published: http://dx.doi.org/10.1038/nature13906

Publisher: Nature Publishing Group

Persistent URL: http://hdl.handle.net/1721.1/97539

Version: Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

Published in final edited form as:
Nature. 2014 December 18; 516(7531): 428–431. doi:10.1038/nature13906.

Rapid modeling of cooperating genetic events in cancer through somatic genome editing

Francisco J. Sánchez-Rivera[1,2,*], **Thales Papagiannakopoulos**[1,*], **Rodrigo Romero**[1,2], **Tuomas Tammela**[1], **Matthew R. Bauer**[1], **Arjun Bhutkar**[1], **Nikhil S. Joshi**[1], **Lakshmipriya Subbaraj**[1], **Roderick T. Bronson**[4], **Wen Xue**[1], and **Tyler Jacks**[1,2,3,#]

[1]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02142

[2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142

[3]Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139

[4]Tufts University and Harvard Medical School, Boston, MA 02115

Abstract

Cancer is a multistep process that involves mutations and other alterations in oncogenes and tumor suppressor genes[1]. Genome sequencing studies have identified a large collection of genetic alterations that occur in human cancers[2–4]. However, the determination of which mutations are causally related to tumorigenesis remains a major challenge. Here we describe a novel CRISPR/Cas9-based approach for rapid functional investigation of candidate genes in well-established autochthonous mouse models of cancer. Using a *Kras*^{G12D}-driven lung cancer model[5], we performed functional characterization of a panel of tumor suppressor genes with known loss-of-function alterations in human lung cancer. Cre-dependent somatic activation of oncogenic *Kras*^{G12D} combined with CRISPR/Cas9-mediated genome editing of tumor suppressor genes resulted in lung adenocarcinomas with distinct histopathological and molecular features. This rapid somatic genome engineering approach enables functional characterization of putative cancer genes in the lung and other tissues using autochthonous mouse models. We anticipate that this approach can be used to systematically dissect the complex catalog of mutations identified in cancer genome sequencing studies.

[#]Corresponding author. Communication can be sent to tjacks@mit.edu.
[*]These authors contributed equally to this work.

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Author Contributions

F.J.S.R, T.P. and T.J. designed the study; F.J.S.R, T.P., R.R., M.R.B. and L.S. performed experiments; T.T. generated *Kras*$^{LSL\text{-}G12D/+}$; *Apc*$^{flox/flox}$ data; A.B. conducted bioinformatic analyses; N.S.J. generated GG cells; R.T.B. provided pathology assistance; W.X. gave conceptual advice; F.J.S.R, T.P. and T.J. wrote the manuscript with comments from all authors.

Illumina MiSeq sequence datasets have been deposited into the NCBI repository under BioProjectID PRJNA256245.

The authors declare no competing financial interests.

Supplementary Information. Reference sequences for *Nkx2.1*, *Pten*, *Apc* and top three predicted off-targets of sgNkx2.1, sgPten and sgApc.

Lung cancer genome sequencing studies have revealed a multitude of recurrent mutations and copy number alterations[2–4]. However, the determination of which mutations are causally related to tumorigenesis remains a major challenge. Genetically-engineered mouse models (GEMMs) of lung cancer have assisted in the functional characterization of putative driver events identified in human lung tumors[6,7],but these require modification of the germline and cannot be performed in a highly parallel manner.

Recent work from our laboratory has demonstrated the feasibility of using the CRISPR (clustered regularly interspaced short palindromic repeats)/Cas9 system to directly mutate cancer genes in the liver following hydrodynamic delivery of plasmids carrying the CRISPR components[8], which relies on the efficient transfection of hepatocytes. To rapidly interrogate cancer genes in the lung and other tissues, we developed pSECC (Fig. 1a), a lentiviral-based system that delivers both the CRISPR system and Cre recombinase. In this setting, CRISPR-induced mutation of genes can be examined in the context of several of the well-studied conditional Cre/loxP mouse models of lung cancer[9] and other cancer types. To test this system, we used GEMMs of lung adenocarcinoma, in which tumors are induced in *Kras*$^{LSL\text{-}G12D/+}$ (K) or *Kras*$^{LSL\text{-}G12D/+}$*; p53*$^{flox/flox}$ (KP) mice upon intratracheal administration of lentiviral vectors expressing Cre-recombinase[10,11].

To validate pSECC, we developed the Green-Go (GG) reporter cell line, which expresses GFP following exposure to Cre (Extended Data Fig. 1a–c). To assess the efficiency of Cas9 in tumors *in vivo*, we targeted a Cre-activatable tdTomato knock-in reporter allele[12] with pSECC lentiviruses expressing an sgRNA against tdTomato (sgTom) or an empty vector control (Extended Data Fig. 1d–e). At 10 weeks post-infection, we assessed knockdown of tdTomato expression by immunohistochemistry (IHC). We observed that 28% of tumors lacked tdTomato expression, suggesting that the system was functional *in vivo* by editing an endogenous allele in the context of a lung tumor (Extended Data Fig. 2a–e). Importantly, animals infected with empty pSECC rarely contained non-tumor Tomato-expressing cells (data not shown), indicating that there is minimal infection of non-epithelial cells when using a low lentiviral titer.

We then proceeded to functionally characterize tumor suppressor genes using this approach. Loss of NK2 homeobox 1(*Nkx2.1*),a master regulator of lung development[13], or Phosphatase and tensin homolog(*Pten*), a negative regulator of oncogenic PI3K/Akt signaling[14] accelerates lung tumorigenesis in K and KP lung tumor models[10,15,16]. We infected K and KP animals with pSECC vectors expressing validated sgPten, sgNkx2.1 and controls (sgTom and empty vector) to induce lung tumors. Ten weeks post-infection, we sacrificed animals to assess the effects of CRISPR/Cas9-mediated gene editing in tumors by histopathology, surveyor assays and deep sequencing of the targeted alleles (Fig. 1a). All animals expressing sgRNAs targeting *Pten* or *Nkx2.1* contained tumors with marked histopathological differences compared to controls (Fig. 1b, d, and Extended Data Fig. 3a–d).

Animals infected with sgNkx2.1-pSECC developed mucinous adenocarcinomas (MA) typified by the presence of elongated cells, mucin production and glandular rearrangements, in agreement with previous Cre/loxP-based (*Nkx2.1*$^{flox/flox}$) data[16](Fig. 1b). The majority of

tumors (61%, 54/88 tumors) from sgNkx2.1-pSECC animals lacked Nkx2.1 expression (compared to 0/33 tumors from controls) (Fig. 1b–c). Importantly, 85% (46/54 tumors) of these Nkx2.1-negative tumors stained positively for mucin (Fig. 1c), a biomarker of mucinous adenocarcinomas[16]. Thus, although a subset of tumors appeared to partially or fully escape CRISPR-mediated deletion of *Nkx2.1*, we were able to observe clear phenotypes by examining the full spectrum of tumors generated by sgNkx2.1-pSECC.

Animals infected with sgPten-pSECC demonstrated complete loss of Pten protein in 74% of tumors (40/54 tumors), which was accompanied by a concomitant increase in pAkt (S473),a downstream biomarker of increased PI3-kinase pathway activity (Fig. 1d–e). These results mimic previously published data using a $Pten^{flox/flox}$ allele in K mice[15]. Collectively, these data indicate that CRISPR/Cas9-based gene editing leads to loss-of-function mutations in this model and closely parallels what is seen with the use of traditional conditional alleles.

We next utilized this system to study adenomatous polyposis coli *(Apc)*, a tumor suppressor whose functional role in lung adenocarcinoma has not been characterized. Of note, *Apc* is found in a region that frequently undergoes copy number loss in human lung cancer[4]. We infected animals with pSECC lentiviruses expressing a validated sgRNA[17] targeting *Apc*. At 10 weeks post-infection, we observed a dramatic difference in the histopathology of sgApc tumors compared to controls (Fig. 1f and Extended Data Fig. 3e). Importantly, tumors from $Kras^{LSL\text{-}G12D/+}$; $Apc^{flox/flox}$ mice, which express a conditional allele of *Apc*[18], exhibited identical histopathology (Fig. 1f). Tumors with Cas9-mediated deletion of *Apc* were highly de-differentiated, invasive and had a significant stromal component (Fig. 1f). The majority of these tumors (78%, 91/117) stained strongly for nuclear β-catenin, a marker of *Apc* mutation in colon cancer and other settings[19] (Fig. 1f–g). Furthermore, 77% (70/91) of tumors with nuclear β-catenin stained positive for the transcription factor Sox9, which might reflect a distal embryonic differentiation state[20,21]. Of note, we observed a statistically significantly higher number of Sox9-positive tumors in KP-sgApc (29/33, or 88%) than in K-sgApc mice (41/58 tumors, or 71%), suggesting a possible role for p53 in regulating this change in differentiation (Extended Data Fig. 6b–c).

To further characterize the differentiation state of sgApc tumors, we stained serial sections for lung differentiation markers, including Sox2, Clara Cell Secretory Protein (CCSP), Surfactant Protein C (SP-C), p63, Nkx2.1 and Sox9 (Extended Data Fig. 6a)[22]. Tumors from KP-sgTom mice stained positively for CCSP, SP-C and Nkx2.1 and negatively for Sox2, p63 and Sox9. In contrast, tumors from sgApc mice frequently stained positively for SP-C, Nkx2.1 and Sox9 and negatively for CCSP, Sox2 and p63. A significant number of tumors from sgApc mice had areas with low levels or complete absence of Nkx2.1, which correlated with the levels of the Nkx2.1 transcriptional target SP-C[22] (16/52 tumors, or 31%) (Extended Data Fig. 6d). These data indicate that these tumors are poorly differentiated and that hyperactivation of the canonical Wnt signaling pathway through loss of *Apc*in Kras-driven lung adenocarcinomas results in tumors with varying degrees of differentiation. These results also mimic what we observed in tumors from *Apc* conditional mice (Fig. 1f and Extended Data Fig. 6e) and recapitulate recent observations in a $Braf^{V600E}$-driven mouse model of lung adenocarcinoma upon Wnt pathway hyperactivation[23].

Our initial analysis demonstrated significant histological and pathway-specific differences upon deletion of these tumor suppressors in lung tumors. To assess the overall impact of these alterations on tumorigenesis, we measured tumor burden and grade in both K and KP animals. Deletion of *Pten* and *Apc* significantly increased overall tumor burden, which correlated with higher tumor grades (Grade 3 and 4) (Fig. 2a–c and Extended Data Fig. 3f–g). *Nkx2.1* deletion had a significant effect on overall tumor burden only in KP animals; however, we observed a striking transition to highly dedifferentiated MA tumors in both K and KP mice (Fig. 2a–c, Extended Data Fig. 3f–g and Extended Data Fig. 4a–d). Conversely, *Apc* deletion had a significant effect on tumor burden only in K mice (Fig. 2a and Extended Data Fig. 3f–g). Deletion of all three genes led to increased BrdU incorporation, suggesting that the increased tumor burden is partly due to increased proliferation (Extended Data Fig. 3h). These data demonstrate the tumor suppressive role of *Nkx2.1*, *Pten* and *Apc* in the context of oncogenic *Kras.* Furthermore, the unique histopathology observed for each targeted tumor suppressor gene in this *Kras*-driven model illustrates the potential of this approach to rapidly model the effects of cooperative genetic events in lung tumorigenesis and progression.

Using this *in vivo* somatic genome editing approach, we observed inter- and intra-tumoral heterogeneity in terms of CRISPR-based loss-of-function of *Pten* in sgPten animals (Fig. 2d–e and Extended Data Fig. 5). Clones that acquired loss of *Pten* had increased PI3K/Akt signaling and may, therefore, have had a selective advantage over tumors that retained wild-type *Pten* within the same animal. We observed that tumors with complete or sub-clonal loss of *Pten* were significantly larger than tumors that retained *Pten* (Fig. 2d–e).

The histopathological and IHC analyses indicate that the pSECC system is highly efficient *in vivo*, leading to robust target-specific phenotypic differences in lung tumors. To confirm Cas9-mediated editing of the alleles and precisely characterize the events at single-nucleotide resolution, we performed deep sequencing of targetloci from whole lung and tumor DNA. Within a 23bp window (± 10bp flanking the Protospacer Adjacent Motif [PAM] sequence at each locus), the rate of mutations observed in the sgTarget samples was significantly greater than in the control samples (Fig. 3a–c). Using the control samples as a background model to analyze the mutational rate revealed that sgTarget samples were enriched for mutations within 7bp upstream of the PAM sequences in predicted cutting sites, strongly suggesting that they are not secondary consequences of tumor progression (Fig. 3d and Extended Data Fig. 7a–c). The maximum per-base mutation frequency observed in sgTarget samples was 71.7% in *Nkx2.1*, 66.06% in *Pten* and 39.91% in *Apc* (in contrast to control samples: 0.11%, 0.73% and 0.14%, respectively). On average, 27.48% ± 10.3 (*Nkx2.1*), 44.64% ± 5.3 (*Pten*) and 13.54% ± 5.3 (*Apc*) read fragments covering this 7bp locus harbored indels in sgTarget samples. Across all sgTarget samples, > 94% of observed indels constituted nonsynonymous frame-altering events (Extended Data Fig. 7d–e and Supplementary Tables 3–4).

Several studies have reported that Cas9 can bind to sites in the genome other than the intended target site[24–27], which could result in unintended editing at an off-target (OT) site. To assess OT editing, we analyzed the top three predicted[24] loci (Supplementary Tables 2) for each sgRNA by deep sequencing. We observed negligible OT editing (Extended Data

Fig. 8). On average, 0.048% ± 0.031 (for sg*Nkx2.1*), 0.26% ± 0.096 (for sg*Pten*) and 0.051% ± 0.027 (for sg*Apc*) of read fragments harbored indels in the OT sites (Supplementary Tables 5–7). This data suggests that the observations reported for each of the sgRNAs arise from deletion of the intended target and not from editing of another gene.

The goal of cancer genomics is to identify genetic events that underlie cancer initiation and progression. The functional interrogation of putative cancer genes in appropriate experimental models will elucidate which mutations identify bona fide cancer genes. This study presents a novel approach to rapidly evaluate human cancer genome candidates and assess cooperativity between genetic events in the context of well-established mouse models of lung cancer. Moreover, our ability to model different lung adenocarcinoma subtypes allows for the detailed study of subtype-specific molecular mechanisms controlling disease initiation and progression. We anticipate that this approach can be readily adapted to many existing Cre/loxP-based GEMMs of several cancer types to facilitate the rapid functional assessment of new hypotheses generated by cancer genome studies.

Methods

Lentiviral vectors and sgRNA cloning

The U6-sgRNA-EFS-Cas9-2A-Cre (pSECC) lentiviral vector was constructed by assembling four parts with overlapping DNA ends using Gibson assembly. Briefly, a 2.2kb part (corresponding to the U6-Filler fragment from LentiCRISPR[28]), a 0.3kb part (corresponding to the EFS promoter from LentiCRISPR[28]), a 5.3kb part (corresponding to a Cas9-2A-Cre fragment, which was generated by assembly PCR) and a 5.7kb lentiviral backbone were assembled using Gibson assembly following manufacturer guidelines. Detailed cloning strategies and primer sequences are available on request. For sgRNA cloning, the pSECC vector was digested with BsmBI and ligated with BsmBI-compatible annealed oligos (Supplementary Table 1). sgRNAs were designed using CRISPR Design[24] (which was also used to predict potential off-target sites; see Extended Data Fig. 8 and Supplementary Tables 2) or E-CRISP[29], except for sgApc which was previously reported[17]. An extra G (required for U6 transcriptional initiation) was added to the 5′ end of sgRNAs that lacked it.

Lentiviral production

Lentiviruses were produced by co-transfection of 293T cells with lentiviral backbone constructs and packaging vectors (delta8.2 and VSV-G) using TransIT-LT1 (Mirus Bio). Supernatant was collected 48 and 72 hours post-transfection, concentrated by ultracentrifugation at 25,000 RPM for 90 minutes and resuspended in an appropriate volume of OptiMEM (Gibco).

Cell culture and generation of Green-Go cells

Cells were maintained in DMEM supplemented with 10% Fetal Bovine Serum and gentamicin. Green-Go cells were generated by transducing 3TZ cells[30] with a bicistronic retrovirus containing an LTR promoter-driven inverted GFP (flanked by two sets of incompatible loxP sites) and a PGK-driven puromycin resistance cassette. Transduced cells

were selected with puromycin and a single cell clone that expressed high levels of GFP 2–3 days after infection with a lentivirus expressing Cre recombinase was chosen.

Immunobloting

Cells were lysed with ice-cold RIPA buffer (Pierce, #89900) supplemented with 1× Complete Mini inhibitor mixture (Roche, #11 836 153 001) and mixed on a rotator at 4°C for 30 minutes. Protein concentration of the cell lysates was quantified using the Bio-Rad DC Protein Assay (Catalog #500-0114). 50–80 μg of total protein was separated on 4–12% Bis-Tris gradient gels (Life Technologies) by SDS-PAGE and then transferred to nitrocellulose membranes. The following antibodies were used for immunoblotting: anti-FLAG (Sigma, F1804, 1:1,000), anti-Hsp90 (BD, #610418, 1:10,000), anti-Pten (Cell Signaling, 9188, 1:1,000), anti-TTF1/Nkx-2.1 (Epitomics, EP1584Y, 1:1,000).

Mice

All animal studies described in this study were approved by the MIT Institutional Animal Care and Use Committee. All animals were maintained on a mixed C57BL/6J x 129SvJ genetic background. *Kras*$^{LSL\text{-}G12D/+}$and *p53*$^{flox/flox}$ mice have already been described[5,31]. Mice were infected intratracheally with lentiviruses as described[32]. We infected a total of 7 mice with Empty-pSECC and 6 mice with sgTom-pSECC (for a total of 13 control mice), as well as 8 mice with sgNkx2.1-pSECC, 9 mice with Pten-pSECC and 9 mice with Apc-pSECC. No randomization or blinding was used. Total lung area occupied by tumor was measured on hematoxylin and eosin (H&E) stained slides using NIS-elements software.

Immunohistochemistry

Mice were euthanized by carbon dioxide asphyxiation. Lungs were perfused through the trachea with 4% paraformaldehyde (PFA), fixed overnight, transferred to 70% ethanol and subsequently embedded in paraffin. Sections were cut at a thickness of four micrometers and stained with H&E for pathological examination. Immunohistochemistry (IHC) was performed on a Thermo Autostainer 360 machine. Slides were antigen retrieved using Thermo citrate buffer, pH 6.0 in the Pretreatment Module. Sections were treated with Biocare rodent block, primary antibody, and anti-Mouse (Biocare) or anti-Rabbit (Vector Labs) HRP-polymer. The slides were developed with Thermo Ultra DAB and counterstained with hematoxylin in a Thermo Gemini stainer and coverslipped using the Thermo Consul cover slipper. The following antibodies were used for IHC: anti-TTF1/Nkx-2.1 (Epitomics, EP1584Y, 1:1,200), anti-Pten (Cell Signaling, 9559, 1:100), anti-pAkt S473 (Cell Signaling, 4060, 1:100), anti-BrdU (Abcam, 6326, 1:100), anti-β-catenin (BD, 610154, 1:100), anti-Sox9 (Millipore, AB5535, 1:500), anti-RFP (Rockland, 600-401-379, 1:400), anti-Sox2 (Cell Signaling, 3728, 1:250), anti-CCSP (Millipore, 07-623, 1:2,000), anti-SP-C (Chemicon, AB3786, 1:1,000) and anti-p63 (Neomarkers, MS-1081, 1:200). To detect mucin, sections were stained with 1% Alcian Blue pH 2.5 and Periodic Acid-Schiff reagent. All pictures were obtained using a Nikon 80i microscope with a DS-U3 camera and NIS-elements software.

Genomic DNA isolation and Surveyor assay

Genomic DNA from entire snap-frozen left lung lobes or microdissected tumors was isolated using the High Pure PCR Template Preparation Kit (Roche) following manufacturer guidelines. PCR products for surveyor assay were amplified using Herculase II Fusion DNA polymerase (Agilent) (see Supplementary Table 1 for primers used for surveyor assay), gel purified and subsequently assayed with the Surveyor Mutation Detection Kit (Transgenomic). DNA was separated on 4–20% Novex TBE Gels (Life Technologies) and stained with ethidium bromide.

Deep sequencing and bioinformatic analysis of Cas9 target loci

For each target gene or potential off-target site, a genomic region containing the target sequence was amplified using Herculase II Fusion DNA polymerase and gel purified (primer sequences are shown in Supplementary Table 1). Sequencing libraries were prepared from 50ng of PCR product using the Nextera DNA Sample Preparation Kit (Illumina) and sequenced on Illumina MiSeq machines. In order to retain high quality sequence for mutation analysis, Illumina MiSeq reads (150bp paired-end) were trimmed to 100mer paired end reads to drop lower quality 3′ ends of reads. Traces of Nextera adapters were clipped from PE1 and PE2 100mer reads using the FASTX toolkit (Hannon Lab, CSHL). Reads greater than 15 nucleotides in length were retained. Additionally, reads with 50% or more bases below a base quality threshold of Q30 were dropped from subsequent analysis. Reference sequences with 10bp genomic flanks were indexed using the Burrows-Wheeler Aligner (BWA) IS linear time algorithm[33] and reads were aligned using the BWA aligner. Reads with mapping quality greater than zero were retained. Overlapping alignments of paired end reads due to short inserts were resolved in order to avoid double counting of coverage and/or mutations observed in a single fragment. In order to minimize alignment ambiguity in the presence of mutations (including indels), the GATK Toolkit[34] was used to realign pooled cohorts mapping to a given locus. Mutations (base substitutions, insertions and deletions) were assessed using a combination of Samtools[35] and Annovar[36] (indel quantification and annotation), NGSUtils/BAMutils software suite[37] (total mutations per position), and custom scripts. Mutation frequencies were adjusted for sample purity (see next section) and per base substitution, insertion, and deletion frequencies were determined. Significance of overall mutation rates across 10bp flanking the target locus was assessed using the Wilcoxon rank sum test comparing control and sgTarget sample events. Positional enrichment for mutation frequency compared to control samples was assessed using the conditional maximum likelihood odds ratio estimate (Fisher's exact test) and was mean centered and scaled (z-scores) across a 10bp flank on either side of the PAM sequence in each sample. A number of other utilities/tools were used to enable various parts of the analysis, including: BEDTools[38], the Integrated Genome Viewer (IGV)[39], and Picard (http://picard.sourceforge.net/). Statistical analyses and sequence enrichment plots were implemented in R (http://www.R-project.org). Illumina MiSeq sequence datasets have been deposited into the NCBI repository under BioProjectID PRJNA256245.

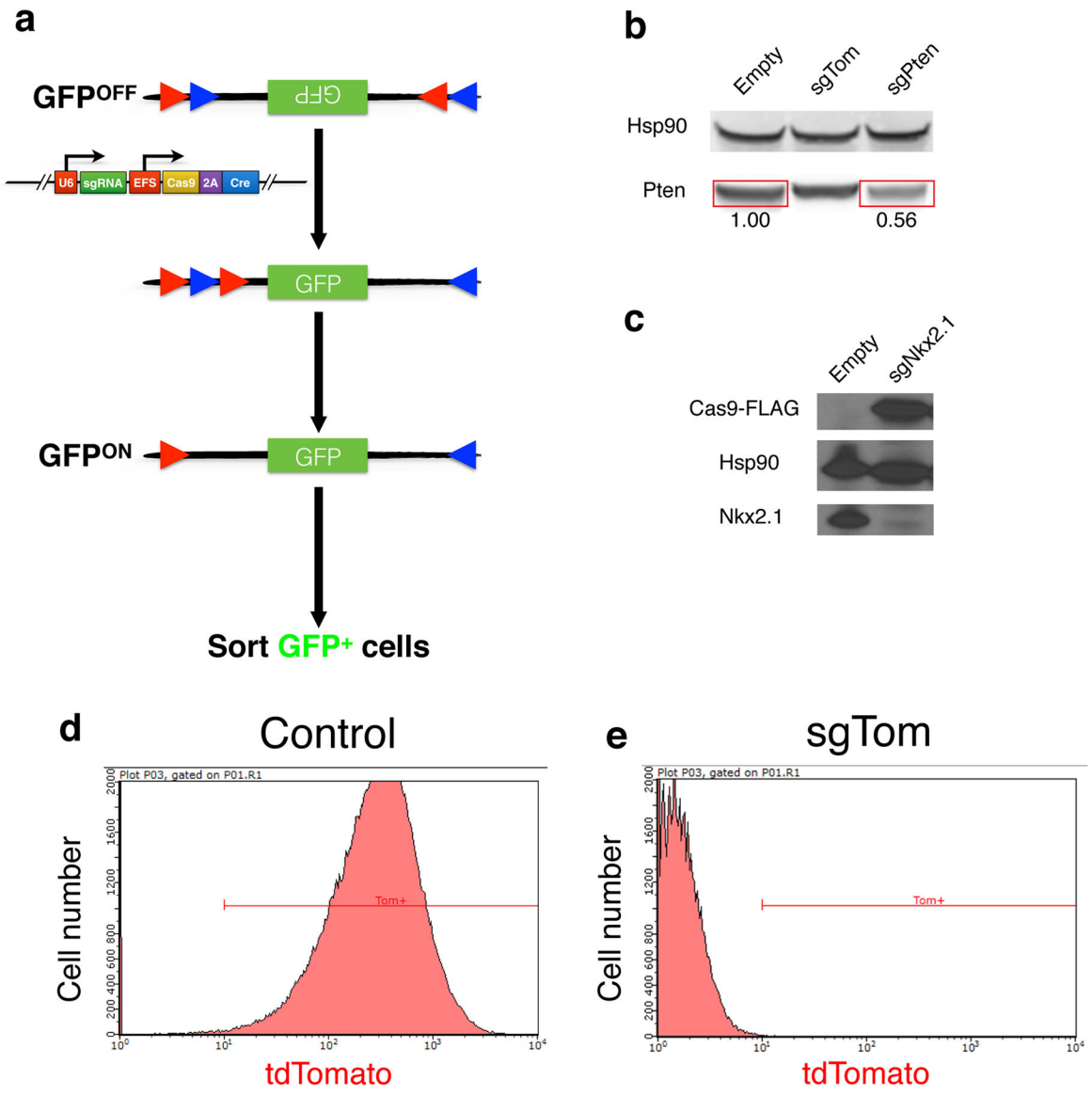
Tumor purity correction

Lung lobe and microdissected tumor genomic DNA was used to perform real-time PCR based analysis to detect the relative levels of the un-recombined *Kras*$^{LSL\text{-}G12D}$ allele (from non-tumor tissue) using forward primer 5′-CTCTTGCCTACGCCACCAGCTC-3′ and reverse primer 5′-AGCTAGCCACCATGGCTTGAGTAAGTCTGCA-3′. To correct for DNA loading of each sample, we amplified the chr5:10054507-10054621 region using forward primer 5′-GAAGAAATTAGAGGGCATGCTTC-3′ and reverse primer 5′ CTTCTCCCAGTGACCTTATGTA-3′. Real-time PCR reactions were performed using KAPA Fast SYBR master mix in a Roche LightCycler Real-Time PCR instrument. To calculate percent purity we performed the following calculations for each sample: $\Delta Cp^{tumorX} = Cp^{Chr5} - Cp^{K\text{-}rasLSL\text{-}G12D/+}$ to normalize for sample loading followed by $1/\Delta\Delta Cp = 1/(\Delta Cp^{tumorX} - \Delta Cp^{LungControl})$ for each sample to compare relative purity to lung tissue from *Kras*$^{LSL\text{-}G12D/+}$ animals that were not infected with Cre. To validate the assay, we generated mouse embryonic fibroblasts from *Kras*$^{LSL\text{-}G12D/+}$ mice treated with Cre recombinase (or control FlpOrecombinase). Purity values are reported in Supplementary Table 3.

Statistics

P-values were determined by Student’s t-test for all measurements of tumor burden and IHC quantifications except for contingency tables, in which Fisher’s exact test or Chi-square test were used. All error bars denote s.e.m.

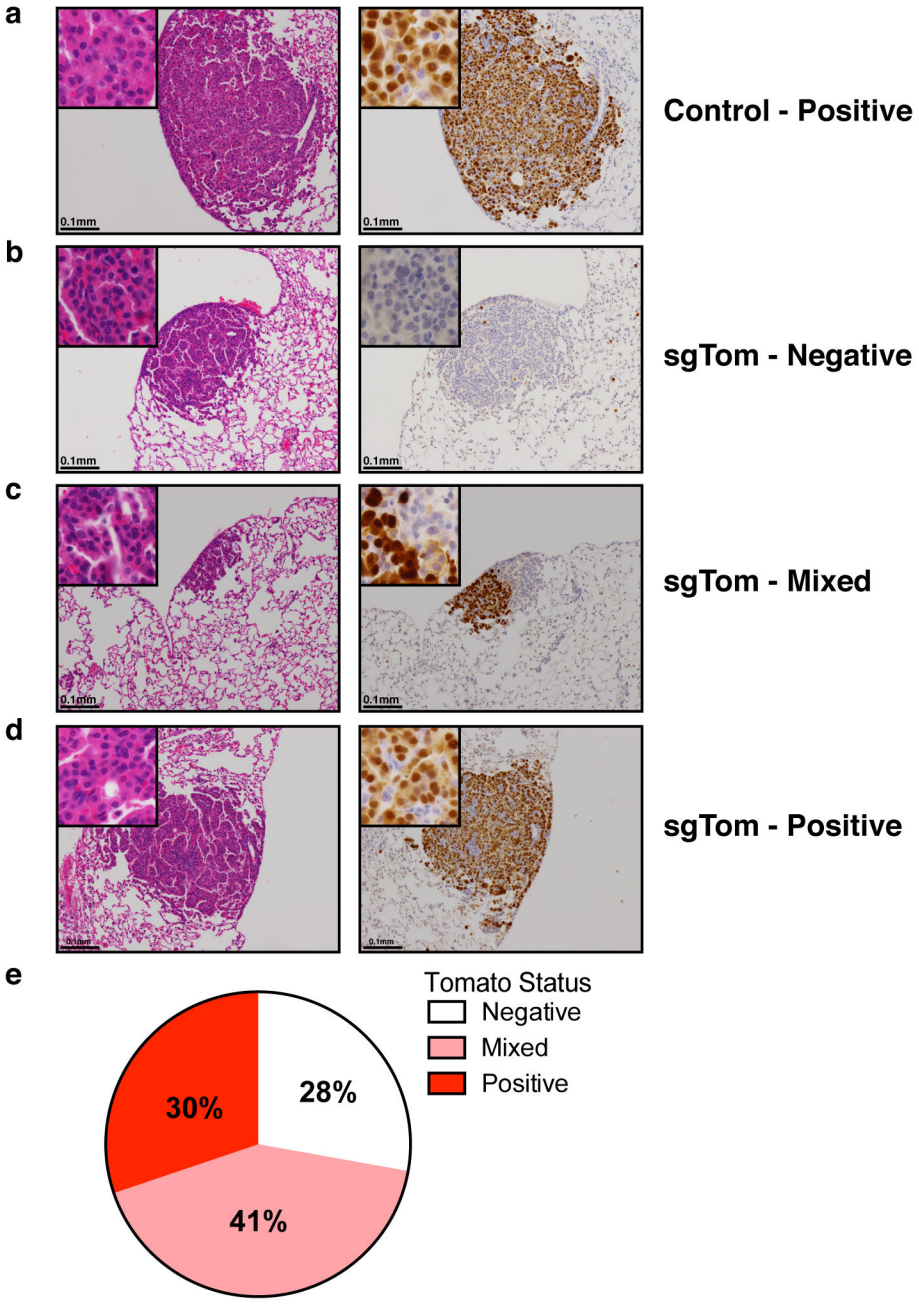
Extended Data



Extended Data Figure 1. *In vitro* validation of pSECC

(a) The Green-Go Cre-reporter cell line used to validate pSECC lentiviruses *in vitro*. Upon infection with a Cre-containing lentivirus, such as pSECC, cells become GFP^{+}, allowing for purification of pSECC-containing cells by FACS. Red and blue triangles denote pairs of loxP sites, with red loxP sites being able to recombine only with other red loxP sites and blue loxP sites being able to recombine only with other blue loxP sites. **(b)** Validation of sgPten-pSECC. Numbers below the bands denote quantitation of protein level relative to empty vector control. **(c)** Validation of sgNkx2.1-pSECC in a cell line that expresses Nkx2.1. **(d–e)** Validation of sgTom-pSECC by Fluorescence Activated Cell Sorting (FACS). Briefly, a cell line obtained from a *Kras*$^{LSL\text{-}G12D/+}$; *p53*$^{flox/flox}$ *Rosa26*$^{LSL\text{-}tdTomato/LSL\text{-}tdTomato}$ (KPT) mouse was infected with either Empty-pSECC (d) or

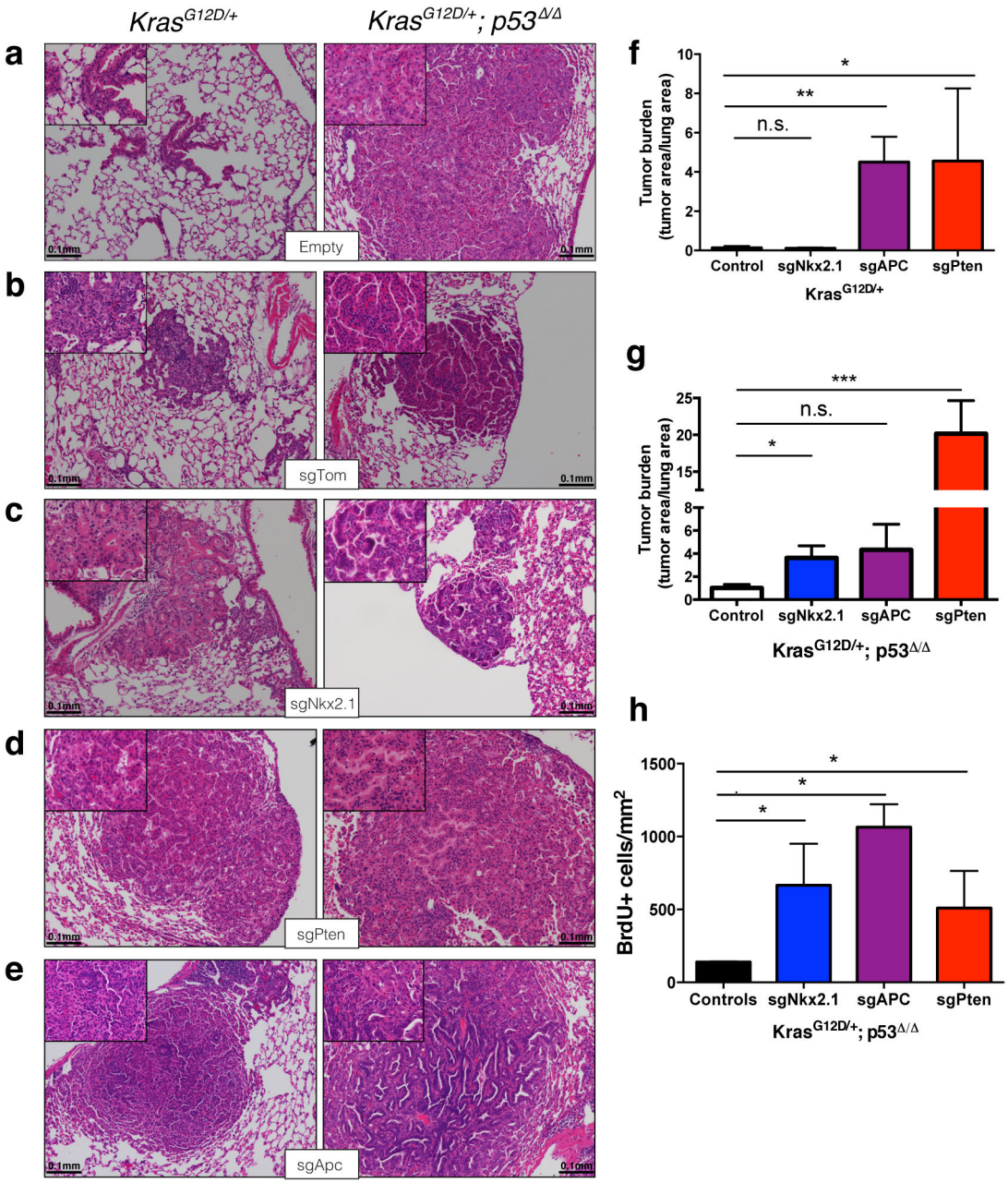
sgTom-pSECC (e) and cultured for 10 days post-infection, after which the cells were harvested and analyzed by FACS.



Extended Data Figure 2. ***In vivo*** **validation of pSECC**

(a) Representative H&E and tdTomato IHC staining of serial sections from lung tumors of KPT mice infected with Empty-pSECC. **(b–d)** Representative H&E and IHC staining of serial sections from **(b)** negative, **(c)** mixed and **(d)** positive lung tumors of KPT mice infected with sgTom-pSECC (n=6). **(e)** Distribution of lung tumors from all mice infected

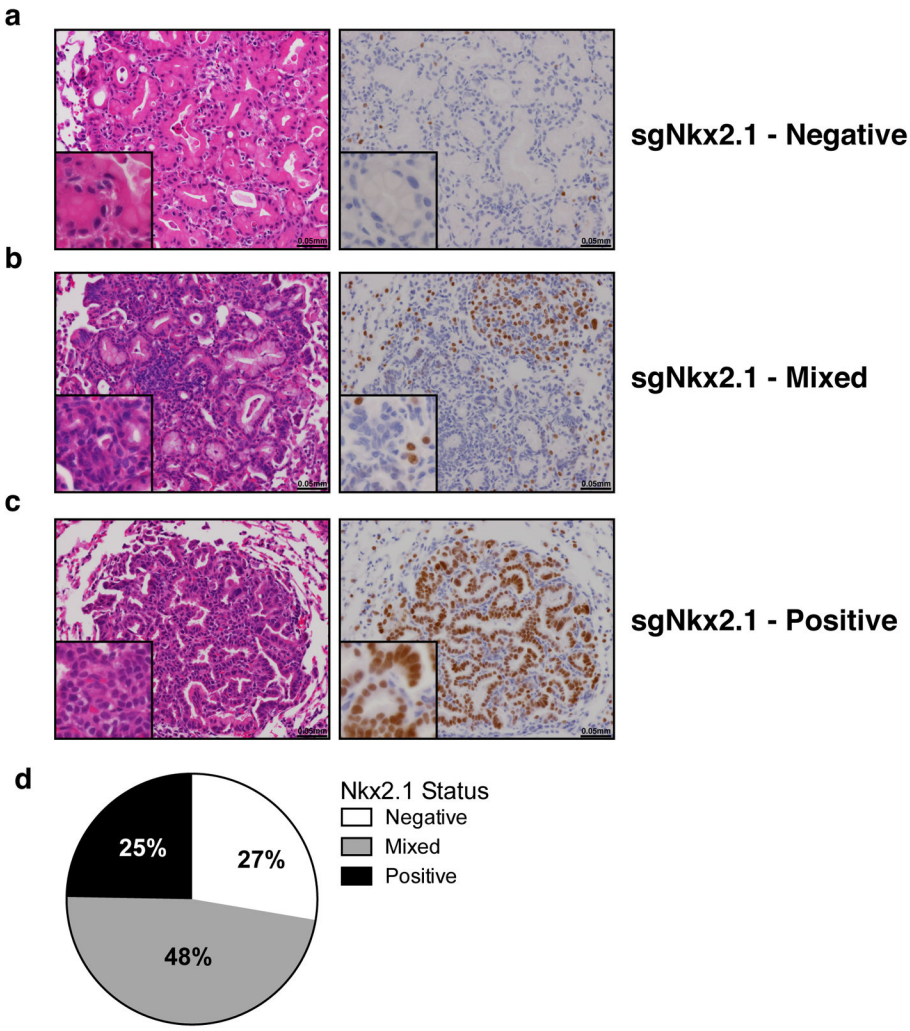
with sgTom-pSECC (n=6) that were scored as negative, mixed or positive based on tdTomato IHC.



Extended Data Figure 3. Histological analysis of lung tumors obtained from mice infected with pSECC lentiviruses

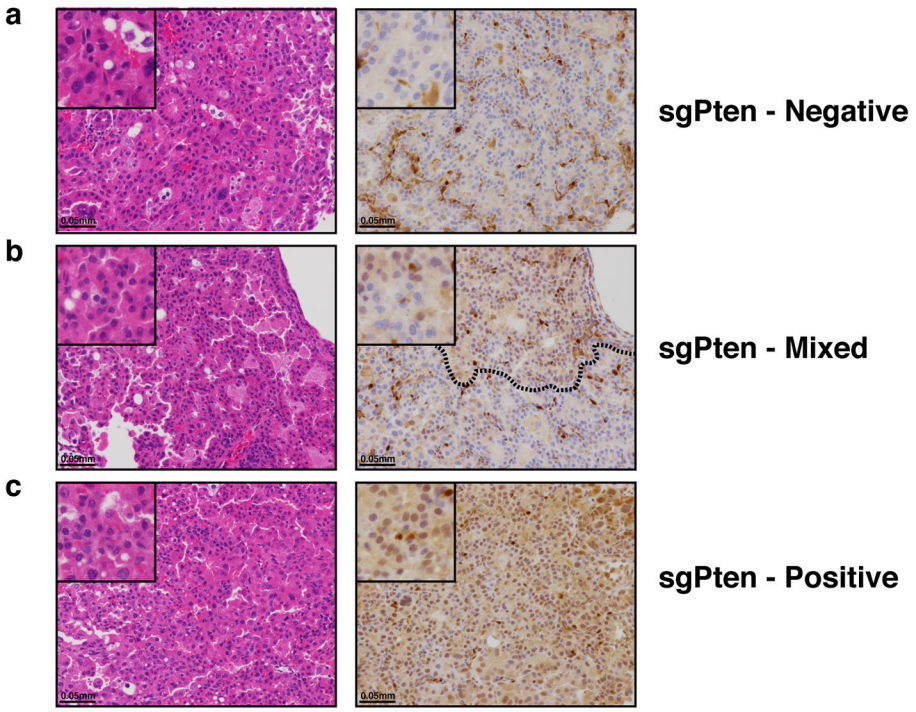
(a) Empty-pSECC, **(b)**sgTom-pSECC, **(c)** sgNkx2.1-pSECC, **(d)** sgPten-pSECC and **(e)** sgApc-pSECC. **(f–g)** Quantitation of tumor burden (total tumor area/total lung area) in K (f) or KP (g) animals 10 weeks after infection with pSECC lentiviruses expressing: control (empty or sgTom, n=4 K and 7 KP), sgNkx2.1 (n=2 K and 6 KP), sgApc (n=3 K and 6 KP) and sgPten (n=4 K and 3 KP). **(h)** Quantitation of BrdU incorporation (BrdU+ cells/mm^2) to

assess proliferation of tumor cells from lung tumors in KP animals 10 weeks after infection with pSECC lentiviruses expressing: control (empty or sgTom, n=4 tumors), sgNkx2.1 (n=11 tumors), sgApc (n=10 tumors) and sgPten (n=15 tumors). Mice were given a pulse of BrdU for 4 hours before being sacrificed. Note: n.s. = not significant, * = $p<0.05$, ** = $p<0.01$, *** = $p< 0.001$ obtained from two-sided Student's t-test. All error bars denote s.e.m.



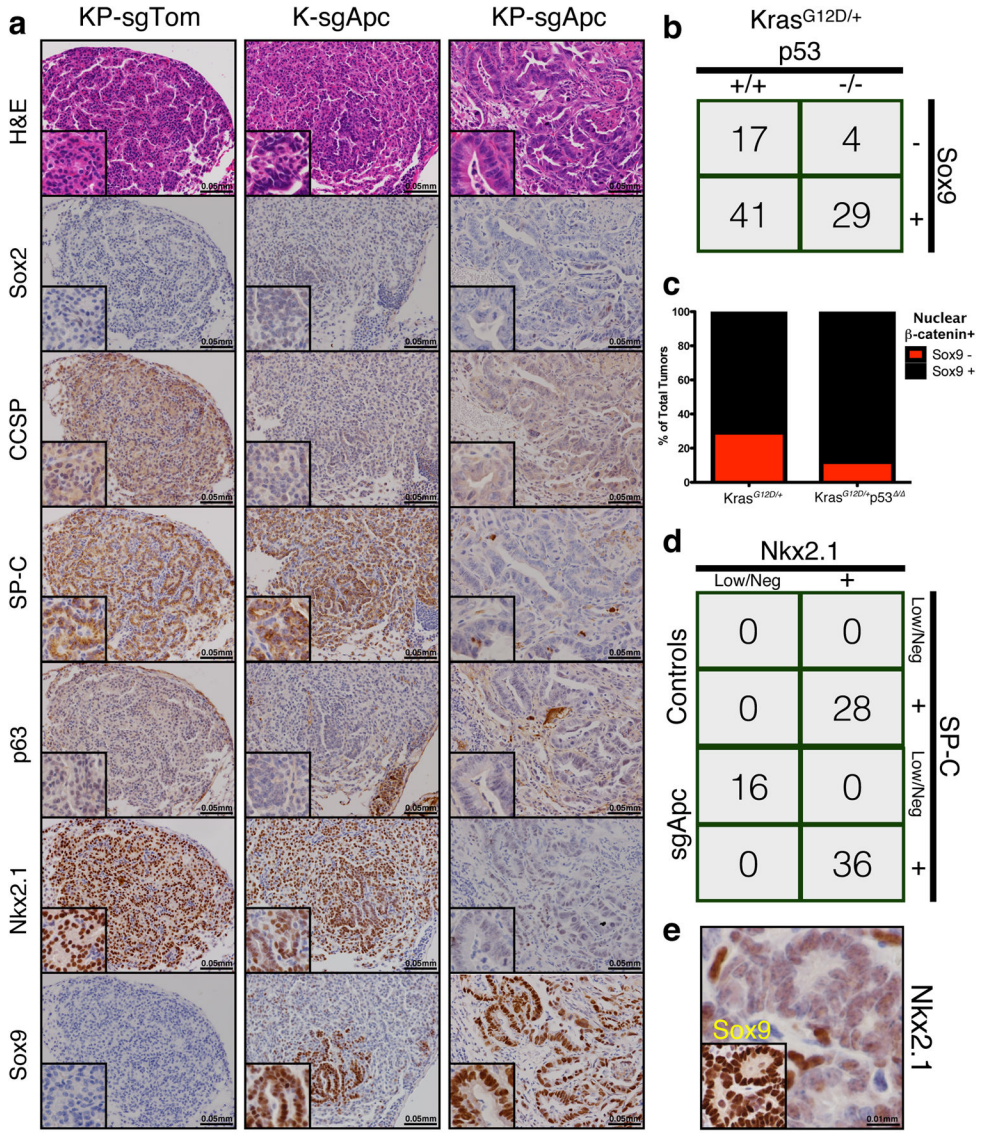
Extended Data Figure 4. IHC-based analysis of mice infected with sgNkx2.1-pSECC

(a) Negative, **(b)** mixed and **(c)** positive lung tumors of mice infected with sgNkx2.1-pSECC. **(d)** Distribution ofNkx2.1 IHC staining status in all sgNkx2.1-pSECC infected animals (n=8) represented as percent of negative, mixed and positive tumors. Positive tumor = ~100% of the tumor cells stained positive for Nkx2.1. Mixed tumor = at least ~30% of tumor cells stained positive for Nkx2.1. Negative tumor = <25% of the tumor cells stained positive for Nkx2.1.



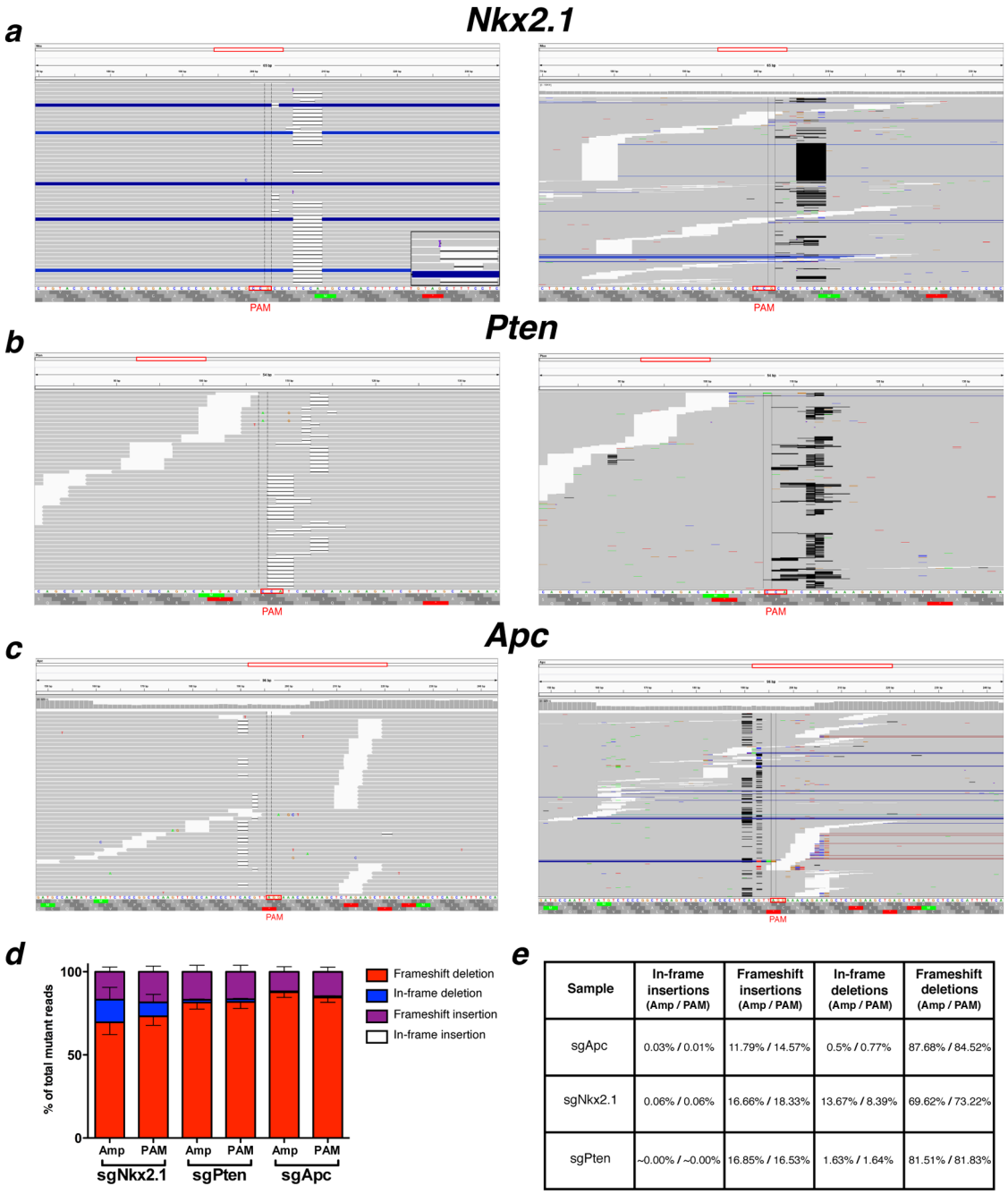
Extended Data Figure 5. IHC-based analysis mice infected with sgPten-pSECC

(a) Negative, **(b)** mixed and **(c)** positive lung tumors of mice infected with sgPten-pSECC (n=9). Positive tumor = ~100% of the tumor cells stained positive for Pten. Mixed tumor = at least ~30% of tumor cells stained positive for Pten. Negative tumor = <25% of the tumor cells stained positive for Pten. Note: dashed line in (b) demarcates positive/negative tumor area.



Extended Data Figure 6. IHC-based analysis of K- and KP-sgApc tumors

(a) Representative H&E and IHC staining of serial sections from KP-sgTom (control), K-sgApc and KP-sgApc lung tumors. CCSP = Clara Cell Secretory Protein, SP-C = Surfactant Protein C. **(b)** Contingency table demonstrating a statistically significantly higher number of β-catenin/Sox9 double-positive tumors in KP-sgApc mice (29/33 tumors, 88%) vs K-sgApc mice (41/58 tumors, 71%) (one-sided Chi-square test, $p<0.05$). **(c)** Percentage of all tumors that stained positive for nuclear β-catenin that stained positive or negative for Sox9 in K- and KP-sgApc mice. **(d)** Contingency table demonstrating a statistically significantly higher number of tumors with Nkx2.1 Low/Negative areas (which are also SP-C Low/Negative) in sgApc-pSECC animals compared to sgTom-pSECC control animals (two-sided Fisher's exact test, $p<0.0001$). **(e)** Representative IHC staining of serial sections from an Nkx2.1 Low/Neg lung tumor obtained from a *Kras*$^{LSL\text{-}G12D/+}$; *Apc*$^{flox/flox}$ mouse 18 weeks after infection with Adeno-Cre. Inset shows Sox9 staining. Low/Neg = tumor that had areas with clear downregulation or complete loss of Nkx2.1 or SP-C as assessed by IHC staining.

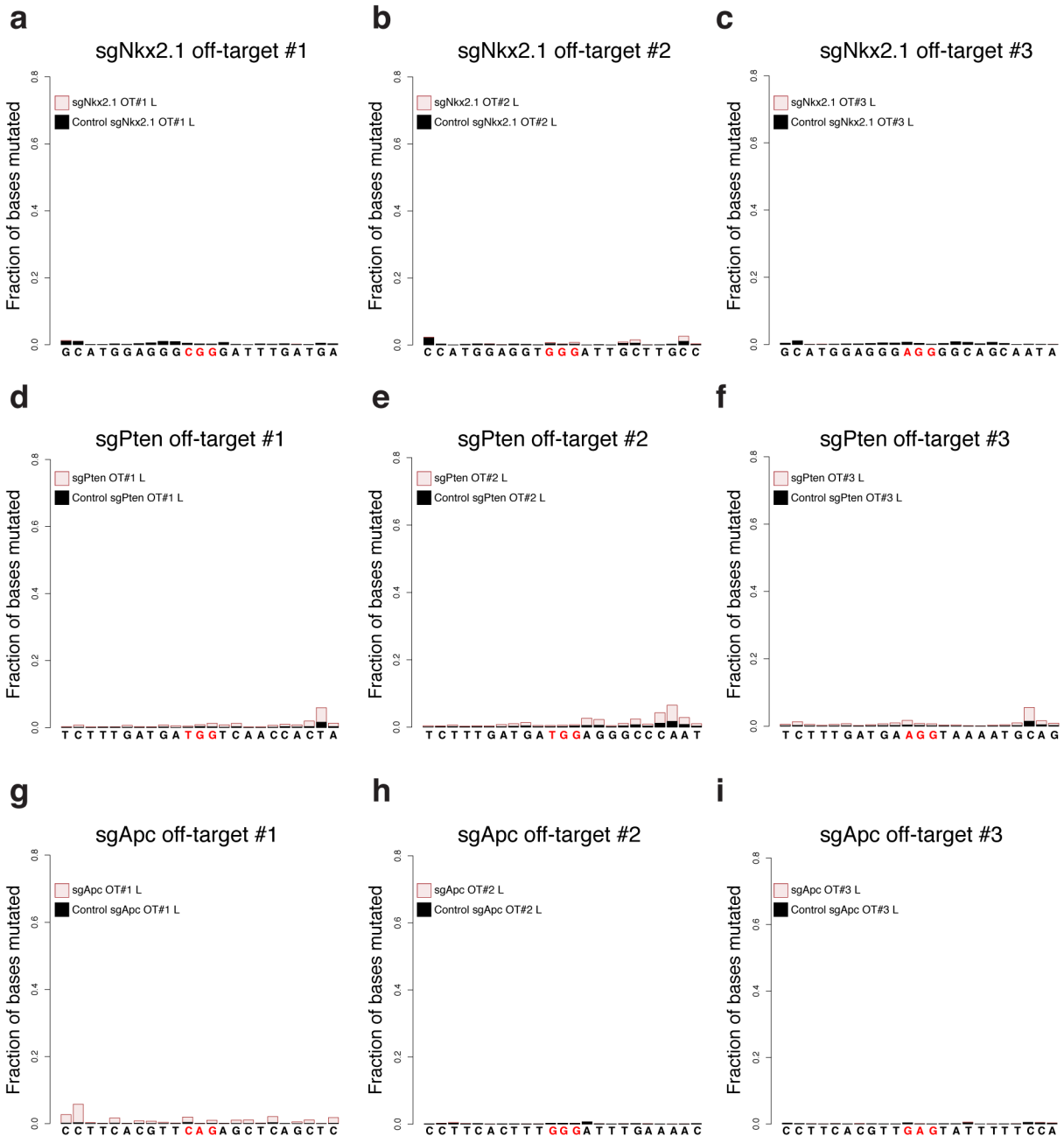


Sample	In-frame insertions (Amp / PAM)	Frameshift insertions (Amp / PAM)	In-frame deletions (Amp / PAM)	Frameshift deletions (Amp / PAM)
sgApc	0.03% / 0.01%	11.79% / 14.57%	0.5% / 0.77%	87.68% / 84.52%
sgNkx2.1	0.06% / 0.06%	16.66% / 18.33%	13.67% / 8.39%	69.62% / 73.22%
sgPten	~0.00% / ~0.00%	16.85% / 16.53%	1.63% / 1.64%	81.51% / 81.83%

Extended Data Figure 7. Representative examples of indels observed in lungs and tumors from mice infected with pSECC lentiviruses

Representative indels observed in the **(a)***Nkx2.1*, **(b)***Pten* and **(c)***Apc* locus from sgNkx2.1T1, sgPtenL1 and sgApcT3 samples, respectively. Left panel: details of sequence alignments around the PAM sequence. Right panel: overview of sequence alignments around the PAM sequence. Deletions and insertions are highlighted in black and purple bars, respectively. Inset in (a) depicts a magnification of an insertion. **(d)** Distribution of indels (in-frame insertions, frameshift insertions, in-frame deletions and frameshift deletions)

observed in samples from mice infected with sgNkx2.1-pSECC, sgPten-pSECC and sgApc-pSECC. Amp: mutations across whole PCR amplicon, PAM: Mutations 7 base pairs upstream of PAM sequence **(e)** Table summarizing percentages of indels from total mutant reads (Left % = Amp/Right % = PAM). All error bars denote s.e.m.



Extended Data Figure 8. Off-target analysis

Analysis of off-target editing for **(a–c)** sgNkx2.1, **(d–f)** sgPten and **(g–i)** sgApc. Briefly, potential off-target cutting at the top three predicted off-target sites (obtained from http://crispr.mit.edu/; see Supplementary Tables 2) for each sgRNA was assayed by Illumina

MiSeq. Each plot corresponds to the fraction of bases mutated per position in 10bp flanks on either side of the PAM sequence (highlighted in red). Samples were obtained from entire lobes (L) from mice 10 weeks after infection with pSECC lentiviruses expressing sgNkx2.1, sgPten, sgApc or sgTom (control).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank D. McFadden and Y. Soto-Feliciano for critical reading of the manuscript, H. Yin, S. Levine and T. Mason for MiSeq sequencing support, R. Stott, J. Bartlebaugh and C. Shivalila for technical assistance and K. Cormier and C. Condon from the Hope Babette Tang (1983) Histology Facility for technical support. This work was supported by the Howard Hughes Medical Institute, the Ludwig Center for Molecular Oncology at MIT and in part by Cancer Center Support (core) grant P30-CA14051 from the National Cancer Institute. T.P. is supported by the Hope Funds for Cancer Research. T.J. is a Howard Hughes Medical Institute Investigator, the David H. Koch Professor of Biology, and a Daniel K. Ludwig Scholar. The authors acknowledge the service to the MIT community of the late Sean Collier. F.J.S.R. dedicates this work to the memory of Blanca Rivera.

References

1. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 2000; 100:57–70. [PubMed: 10647931]
2. Imielinski M, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012; 150:1107–1120. [PubMed: 22980975]
3. Govindan R, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012; 150:1121–1134. [PubMed: 22980976]
4. Network, T. C. G. A. R. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014
5. Jackson EL, et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. Genes & development. 2001; 15:3243–3248. [PubMed: 11751630]
6. McFadden DG, et al. Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. Cell. 2014; 156:1298–1311. [PubMed: 24630729]
7. Frese KK, Tuveson DA. Maximizing mouse cancer models. Nat Rev Cancer. 2007; 7:645–658. [PubMed: 17687385]
8. Xue W, et al. CRISPR-mediated direct mutation of cancer genes in the mouse liver. Nature. 2014
9. Farago AF, Snyder EL, Jacks T. SnapShot: Lung cancer models. Cell. 2012; 149:246–246 e241. [PubMed: 22464334]
10. Winslow MM, et al. Suppression of lung adenocarcinoma progression by Nkx2-1. Nature. 2011; 473:101–104. [PubMed: 21471965]
11. DuPage M, et al. Endogenous T cell responses to antigens expressed in lung adenocarcinomas delay malignant tumor progression. Cancer Cell. 2011; 19:72–85. [PubMed: 21251614]
12. Madisen L, et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat Neurosci. 2010; 13:133–140. [PubMed: 20023653]
13. Rock JR, Hogan BL. Epithelial progenitor cells in lung development, maintenance, repair, and disease. Annu Rev Cell Dev Biol. 2011; 27:493–512. [PubMed: 21639799]
14. Song MS, Salmena L, Pandolfi PP. The functions and regulation of the PTEN tumour suppressor. Nat Rev Mol Cell Biol. 2013; 13:283–296. [PubMed: 22473468]
15. Curry NL, et al. Pten-null tumors cohabiting the same lung display differential AKT activation and sensitivity to dietary restriction. Cancer Discov. 2013; 3:908–921. [PubMed: 23719831]
16. Snyder EL, et al. Nkx2-1 represses a latent gastric differentiation program in lung adenocarcinoma. Mol Cell. 2013; 50:185–199. [PubMed: 23523371]
17. Schwank G, et al. Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients. Cell stem cell. 2013; 13:653–658. [PubMed: 24315439]

18. Cheung AF, et al. Complete deletion of Apc results in severe polyposis in mice. Oncogene. 2010; 29:1857–1864. [PubMed: 20010873]
19. Moon RT, Kohn AD, De Ferrari GV, Kaykas A. WNT and beta-catenin signalling: diseases and therapies. Nat Rev Genet. 2004; 5:691–701. [PubMed: 15372092]
20. Pacheco-Pinedo EC, et al. Wnt/beta-catenin signaling accelerates mouse lung tumorigenesis by imposing an embryonic distal progenitor phenotype on lung epithelium. J Clin Invest. 2011; 121:1935–1945. [PubMed: 21490395]
21. Kormish JD, Sinner D, Zorn AM. Interactions between SOX factors and Wnt/beta-catenin signaling in development and disease. Dev Dyn. 2010; 239:56–68. [PubMed: 19655378]
22. Hogan BL, et al. Repair and Regeneration of the Respiratory System: Complexity, Plasticity, and Mechanisms of Lung Stem Cell Function. Cell stem cell. 2014; 15:123–138. [PubMed: 25105578]
23. Juan J, Muraguchi T, Iezza G, Sears RC, McMahon M. Diminished WNT -> beta-catenin -> c-MYC signaling is a barrier for malignant progression of BRAFV600E-induced lung tumors. Genes & development. 2014; 28:561–575. [PubMed: 24589553]
24. Hsu PD, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology. 2013; 31:827–832.
25. Fu Y, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology. 2013; 31:822–826.
26. Wu X, et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Nature biotechnology. 2014; 32:670–676.
27. Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. Nature biotechnology. 2014; 32:677–683.
28. Shalem O, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. Science. 2014; 343:84–87. [PubMed: 24336571]
29. Heigwer F, Kerr G, Boutros M. E-CRISP: fast CRISPR target site identification. Nature methods. 2014; 11:122–123. [PubMed: 24481216]
30. Psarras S, et al. Gene transfer and genetic modification of embryonic stem cells by Cre- and Cre-PR-expressing MESV-based retroviral vectors. The journal of gene medicine. 2004; 6:32–42. [PubMed: 14716675]
31. Jackson EL, et al. The differential effects of mutant p53 alleles on advanced murine lung cancer. Cancer Res. 2005; 65:10280–10288. [PubMed: 16288016]
32. DuPage M, Dooley AL, Jacks T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. Nat Protoc. 2009; 4:1064–1072. [PubMed: 19561589]
33. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010; 26:589–595. [PubMed: 20080505]
34. McKenna A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010; 20:1297–1303. [PubMed: 20644199]
35. Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009; 25:2078–2079. [PubMed: 19505943]
36. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010; 38:e164. [PubMed: 20601685]
37. Breese MR, Liu Y. NGSUtils: a software suite for analyzing and manipulating next-generation sequencing datasets. Bioinformatics. 2013; 29:494–496. [PubMed: 23314324]
38. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010; 26:841–842. [PubMed: 20110278]
39. Thorvaldsdottir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform. 2013; 14:178–192. [PubMed: 22517427]

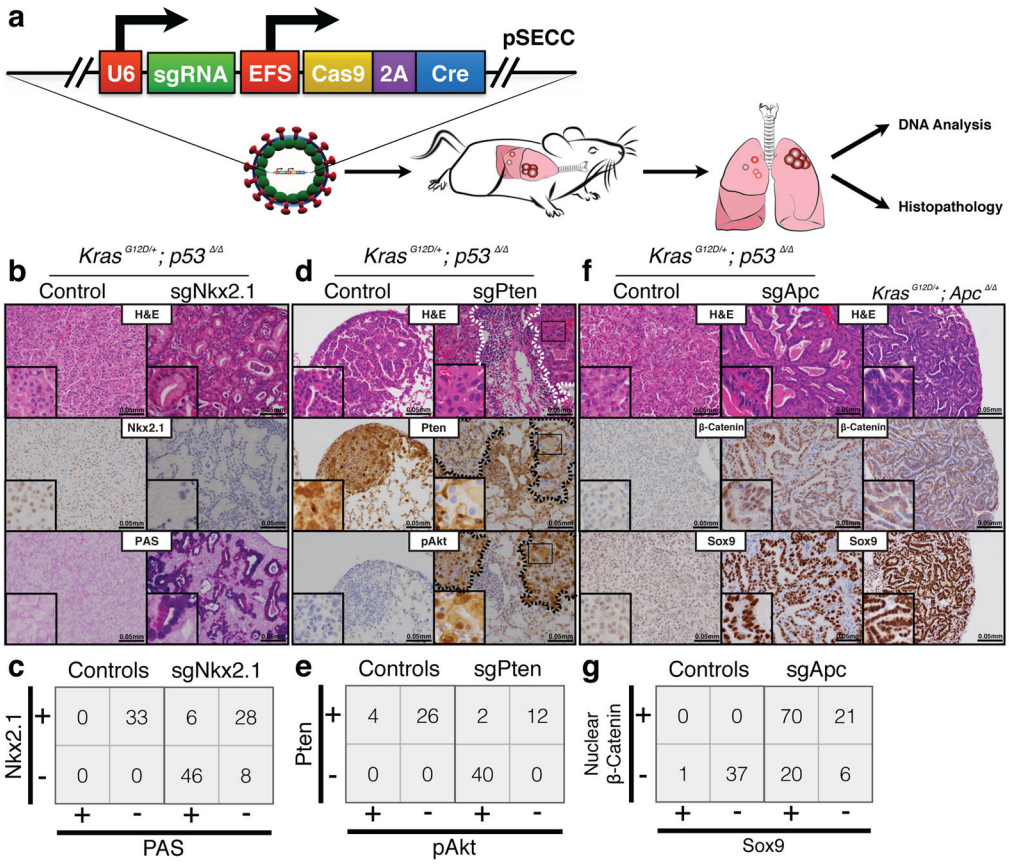


Figure 1. CRISPR/Cas9-mediated somatic gene editing in an autochthonous mouse model of lung cancer

(a)pSECC lentiviruses are intratracheally delivered into the lungs of mice to delete genes of interest. DNA extracted from tumor-bearing lungs is analyzed by high-throughput sequencing and surveyor assays to identify gene-editing events. The remaining tissue is analyzed by histopathology. **(b)** Representative H&E and IHC stainings of serial sections from lung tumors of mice 10 weeks after infection with sgTom-pSECC (left panel) or sgNkx2.1-pSECC (right panel). Alcian Blue/PAS (Periodic Acid-Schiff) stain for mucin. Note the accumulation of mucin only in tumors from sgNkx2.1-pSECC mice. **(c)** Contingency tables demonstrating anti-correlation between Nkx2.1 expression and mucin production (PAS stain) (two-sided Fisher's exact test, $p<0.0001$). **(d)** Representative H&E and IHC stainings of serial sections from lung tumors of mice 10 weeks after infection with sgTom-pSECC (left panel) or sgPten-pSECC (right panel). Note: dashed lines demarcate tumor boundaries on each consecutive histological section. **(e)** Contingency tables demonstrating anti-correlation between Pten expression and Akt phosphorylation (two-sided Fisher's exact test, $p<0.0001$). **(f)** Representative H&E and IHC stainings of serial sections from lung tumors of mice 10 weeks after infection with sgTom-pSECC (left panel) or sgApc-pSECC (middle panel). The far right panel corresponds to serial sections from lung tumors of *Kras*$^{LSL\text{-}G12D/+}$; *Apc*$^{flox/flox}$ mice 18 weeks after infection with Adeno-Cre.**(g)** Contingency tables demonstrating positive correlation between β-Catenin expression and

Sox9 expression (two-sided Fisher's exact test, $p<0.0001$). These data are representative of at least 3 independent K or KP mice infected with each pSECC sgRNA.

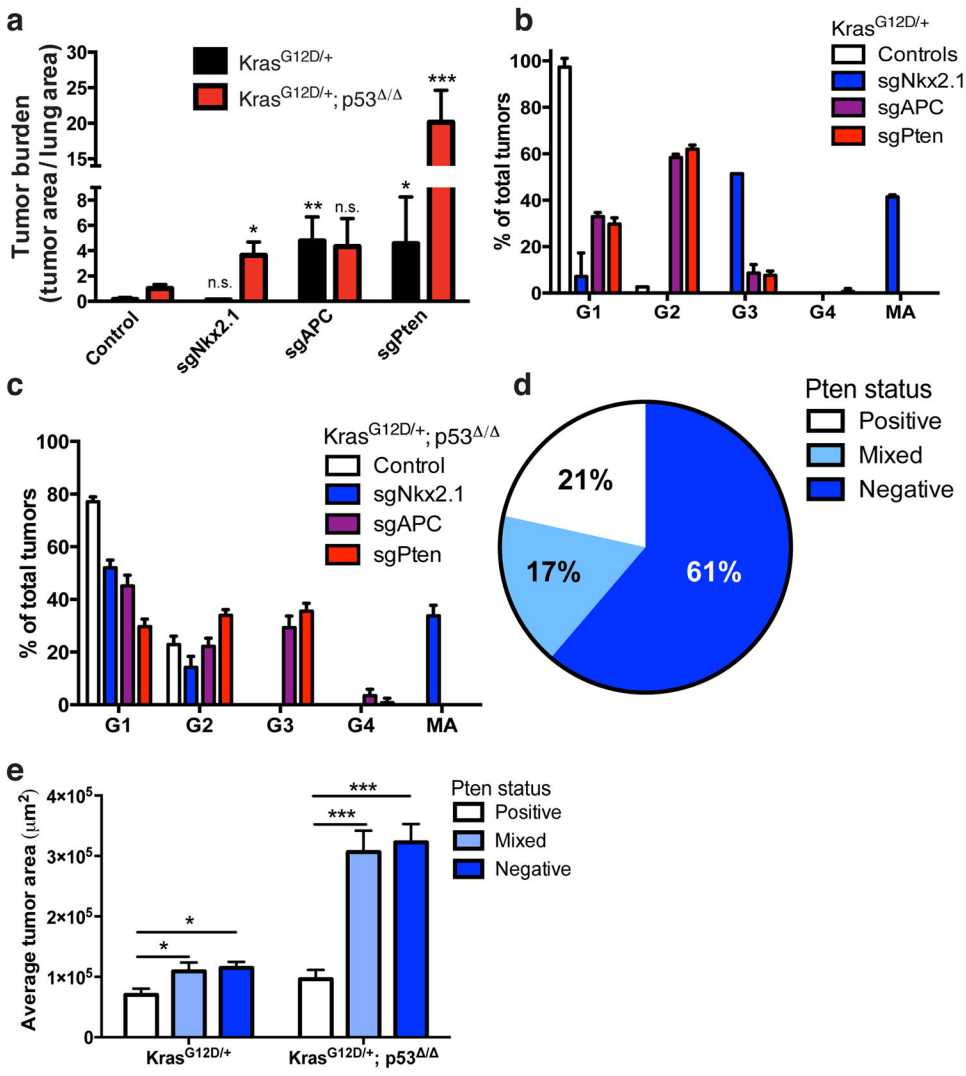


Figure 2. Histopathological characterization of tumors from pSECC infected animals

(a) Combined quantitation of tumor burden (total tumor area/total lung area) in both K and KP animals 10 weeks after infection with pSECC lentiviruses expressing: control (empty or sgTom, n=4 K and 7 KP), sgNkx2.1 (n=2 K and 6 KP), sgApc (n=3 K and 6 KP) and sgPten (n=4 K and 3 KP). The asterisks indicate statistical significance obtained from comparing K-sgTarget samples to K-Control samples or KP-sgTarget samples to KP-Control samples using Student's t-test (two-sided).**(b-c)** Distribution of tumor grades in K (b) or KP (c) animals 10 weeks after infection with pSECC lentiviruses expressing: control (empty or sgTom, n=4 K and 7 KP), sgNkx2.1 (n=2 K and 6 KP), sgApc (n=3 K and 6 KP) and sgPten (n=4 K and 3 KP). G1: grade 1, G2: grade 2, G3: grade 3, G4: grade 4, MA: mucinous adenocarcinoma. **(d)** Distribution of Pten IHC staining status in all sgPten-pSECC infected animals (n=9) represented as percent of negative, mixed and positive tumors. **(e)** Quantitation of average tumor area (μm^2) of tumors staining negative, mixed or positive in all sgPten-pSECC infected animals (n=9). Positive tumor = ~100% of the tumor cells stained positive for Pten. Mixed tumor = at least ~30% of tumor cells stained positive for

Pten. Negative tumor = <25% of the tumor cells stained positive for Pten. Note: n.s. = not significant, * = $p<0.05$, ** = $p<0.01$, *** = $p< 0.001$ obtained from two-sided Student's t-test. All error bars denote s.e.m.

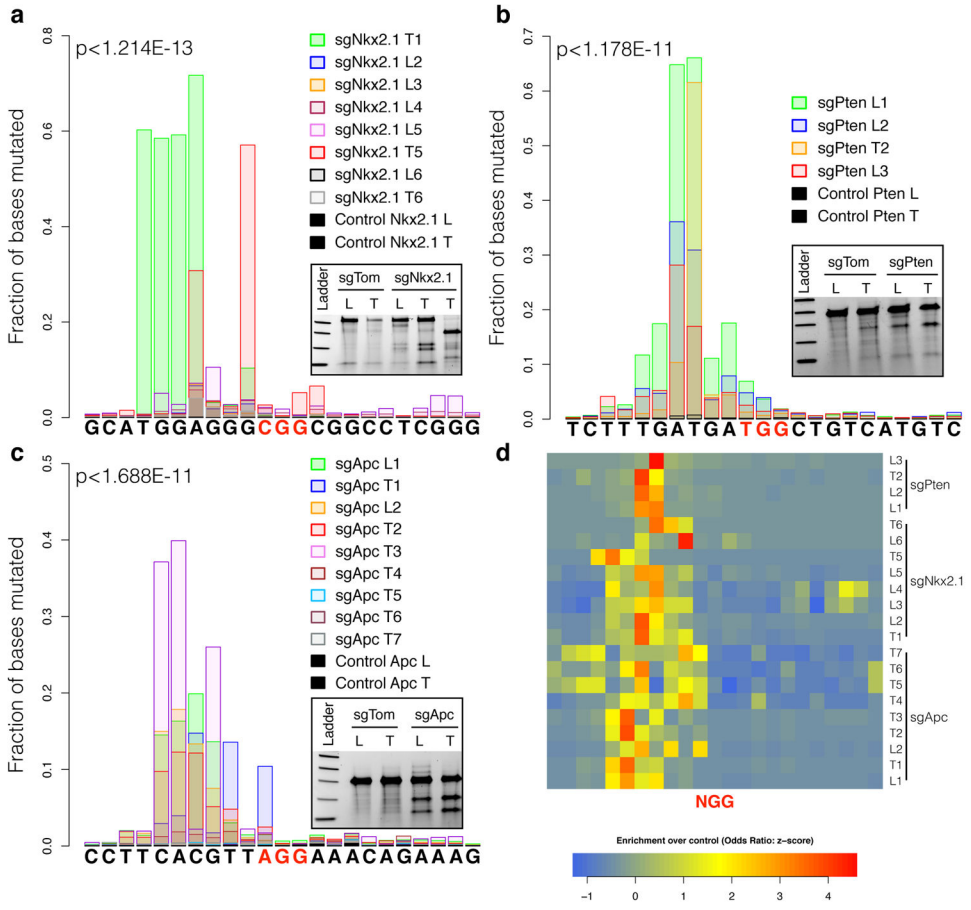


Figure 3. CRISPR/Cas9 efficiently generates indels in autochthonous tumors

(a–c) Fraction of bases mutated per position in 10bp flanks on either side of the Protospacer Adjacent Motif (PAM) sequence (highlighted in red). Samples were obtained from entire lobes (L) or microdissected tumors (T) from mice 10 weeks after infection with pSECC lentiviruses targeting **(a)***Nkx2.1*, **(b)***Pten* or **(c)***Apc*. P-values denote enrichment of mutation rate in sgTarget-pSECC samples compared to sgTom-pSECC control samples (Wilcoxon rank sum test). Insets depict surveyor assays for each of the targets from either entire lobes (L) or microdissected tumors (T) from mice. Samples obtained from mice infected with sgTom-pSECC were used as controls. **(d)** Positional enrichment of mutations in sgTarget-pSECC samples compared to sgTom-pSECC control samples based on all mutations considered at a given position (SNPs, indels). Each row represents a different sgRNA Lung (L) or Tumor (T) sample. Each cell represents the row-normalized (z-score) odds ratio estimate of mutational enrichment over an associated control sample (Fisher's exact test) upstream (+) or downstream (−) of the PAM sequence.