Modeling the genomic complexity of human cancer using Cas12a mice

Jess D. Hebert[1], Haiqing Xu[3], Yuning J. Tang[1], Paloma A. Ruiz[1], Colin Detrick[1], Jing Wang[2], Nicholas W. Hughes[1], Oscar Donosa[4], Laura Andrejka[1], Saswati Karmakar[1], Irenosen Aboiralor[1], Le Cong[1,5], Julien Sage[1,4,6], Dmitri A. Petrov[3,4,7], and Monte M. Winslow[1,4,5#]

[1] Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA
[2] Department of Radiology, Stanford University School of Medicine, Stanford, CA, USA
[3] Department of Biology, Stanford University, Stanford, CA, USA
[4] Cancer Biology Program, Stanford University School of Medicine, Stanford, CA, USA
[5] Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA
[6] Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA
[7] Chan Zuckerberg Biohub Investigator

[#] Corresponding author:

Monte M. Winslow, Stanford University School of Medicine | 279 Campus Drive, Beckman Center B256, Stanford, CA 94305 | Phone: 650-725-8696 | Fax: 650-725-1534 | E-mail: mwinslow@stanford.edu

ABSTRACT

Somatic genome editing in mouse models has increased our understanding of the *in vivo* effects of genetic alterations in areas ranging from neuroscience to cancer biology and beyond. However, existing models have been restricted in their ability to create multiple targeted edits, which has limited investigations into complex genetic interactions that underlie development, homeostasis, and disease. To accelerate and expand the generation of complex genotypes in somatic cells, we generated transgenic mice with Cre-regulated and constitutive expression of enhanced *Acidaminococcus sp.* Cas12a (enAsCas12a), an RNA-guided endonuclease with unique attributes that enable simple targeting of multiple genes. In these mice, enAsCas12a-mediated somatic genome editing robustly generated compound genotypes, as exemplified by the initiation of oncogene-negative lung adenocarcinoma, small-cell lung cancer, and a canonical genotype of pancreatic ductal adenocarcinoma, all driven by homozygous inactivation of trios of tumor suppressor genes. We further integrated these modular crRNA arrays with clonal barcoding to quantify the size and number of tumors with each array. These Cas12a alleles will enable the rapid generation of disease models and broadly facilitate the high-throughput investigation of coincident genomic alterations in somatic cells *in vivo*.

INTRODUCTION

Human cancer is defined by its genetic complexity. Depending on the cancer type, a given tumor can have tens to hundreds of non-synonymous mutations, in addition to aberrant epigenetic modifications, gene duplications, chromosomal rearrangements, and gains or losses of entire chromosomes[1–3]. Any or all of these alterations could contribute to tumorigenesis, tumor progression, and resistance to therapy. However, the bulk of these epigenetic, genetic and genomic changes are likely to be passengers, rather than drivers of key aspects of cancer. Deconvoluting functional drivers from human tumor data can also be complicated by the limited availability of samples, differences among patients and which therapies they have received, and technical variability of the approaches used to analyze them.

Functional interrogation of the phenotypes resulting from particular alterations is thus of value. Genetically engineered mouse models have been widely used to study putative oncogenes and tumor suppressor genes, as tumors in these models grow entirely within their natural setting, including having appropriate interactions between cancer cells, stromal cells, and the immune system. However, the creation of new mouse alleles has remained a barrier to the study of more complex tumor genotypes, due to the expense and time required to generate a new allele and cross it with other alleles of interest. As a result, some common but complex genotypes have never been modeled[4]. Somatic genome editing with Cas9 has greatly increased the rate at which single genes can be studied, but the ability of Cas9 to model complex genotypes has been limited by the need for each guide to have its own promoter and tracrRNA, thereby requiring the laborious cloning of guides in sequence.

Cas12a (previously known as Cpf1) is a Class 2 Type V CRISPR-Cas system that has unique value in genome editing due to its ability to easily target multiple genomic loci[5–7]. Unlike

Cas9, Cas12a has both RNase and DNase activity, and can process a single pre-crRNA transcript containing numerous spacers (guides) into its constituent crRNAs to direct Cas12a-mediated cleavage of their target regions[5]. Cas12a is also distinct from Cas9 in recognizing a T-rich 4 bp protospacer adjacent motif (PAM) and generating staggered cuts that are distal to the PAM[6]. Cas12a from multiple species has been employed for genome editing *in vitro*, including Cas12a from *Acidaminococcus sp.* (As) and *Lachnospiraceae bacterium* (Lb)[5,7–12]. Large-scale screens have identified Cas12a variants with increased efficiency, broadened PAM recognition ranges, and reduced off-target cleavage[13–16]. In particular, an enhanced version of AsCas12a (enAsCas12a) with several substitutions (E174R/N282A/S542R/K548R) has superior genome editing relative to wild type AsCas12a, in addition to a substantially expanded targeting range[14].

To accelerate and expand our ability to generate complex genotypes in somatic cells, we generated Cre-regulated and constitutive enAsCas12a transgenic mice. Efficiently multiplexed somatic genome editing using Cas12a transgenic mice enables the rapid generation of complex cancer models and should facilitate the high-throughput investigation of coincident genomic alterations *in vivo*.

RESULTS

Generation and validation of Cre-regulated and constitutive enAsCas12a alleles

To enable Cas12a-mediated somatic genomic editing and thus take advantage of the features that distinguish Cas12a from Cas9 (**Supplementary Figure 1a**), we generated transgenic mice expressing enAsCas12a with an optimized nuclear localization sequence (NLS) configuration that increases cutting efficiency and an HA tag (enAsCas12a; **Figure 1a**)[14,17]. We

generated transgenic mice through CRISPR/Cas9-mediated integration of a CAGGS promoter-driven Lox-Stop-Lox (LSL)-enAsCas12a-PolyA cassette into the *H11* locus (*H11*$^{LSL\text{-}Cas12a}$) (**Figure 1a,c**). To characterize the potential for Cas12a expression from the *H11*$^{LSL\text{-}Ca12a}$ allele *in vivo*, we also crossed *H11*$^{LSL\text{-}Cas12a}$ mice to CMV-Cre "deleter" mice[18] to generate a constitutive *H11*^{Cas12a} allele (**Figure 1b,d**). Expression of Cre in fibroblasts from *H11*$^{LSL\text{-}Cas12a}$ mice induced recombination of the LSL cassette and led to expression of Cas12a protein (**Figure 1e**). *H11*^{Cas12a} mice were viable and fertile and had widespread expression of Cas12a across all organs tested (**Figure 1f**), with prominently nuclear protein localization (**Figure 1g** and **Supplementary Figure 1b**).

Cas12a-induced inactivation of *Nf1, Rasa1* and *Pten* generate lung adenocarcinoma

Oncogene-negative lung adenocarcinoma represents ~30% of all lung adenocarcinomas and has been modeled in mice through combinatorial inactivation of the tumor suppressor genes *Nf1*, *Rasa1* and *Pten* in lung epithelial cells[19]. This was initially accomplished using Cas9-mediated somatic genome editing and multi-step sgRNA cloning[19]. As an initial test of the ability of the *H11*$^{LSL\text{-}Cas12a}$ allele to generate tumors with complex genotypes *in vivo*, we generated lentiviral vectors expressing Cre and a pre-crRNA array targeting *Nf1*, *Rasa1*, and *Pten*. Given the ability to synthesize these three-crRNA arrays and their ease of cloning, we generated a pool of 27 lentiviral Cre vectors with every combination of three guides targeting each tumor suppressor gene (**Figure 2a**). We cloned these pre-crRNA arrays into a vector that contained a diverse 16-nucleotide barcode (BC) at the 3' end of the U6 promoter and Cre recombinase (Lenti-U6^{BC}-cr*Nf1/Rasa1/Pten*-Cre; **Figure 2b**). Cre in these vectors induces Cas12a expression in somatic cells in *H11*$^{LSL\text{-}Cas12a}$ mice, and amplification of the BC-crRNA region from bulk tumor-bearing

lungs followed by tumor barcoding and high-throughput barcode sequencing (Tuba-seq^{Ultra}; U6 barcode Labeling with per-Tumor Resolution Analysis) can quantify the size of each clonal tumor[20–22].

We transduced the lungs of *H11*$^{LSL\text{-}Cas12a}$ mice (some of which also contained an *R26*$^{LSL\text{-}Tomato}$ Cre-reporter allele), wild-type negative control mice, and *Kras*$^{LSL\text{-}G12D}$*;R26*$^{LSL\text{-}Tomato}$ (*KT*) positive control mice with Lenti-U6^{BC}-cr*Nf1/Rasa1/Pten*-Cre (**Figure 2c**). Tumors in *KT* mice form due to Cre-mediated expression of oncogenic KRAS in the absence of any Cas12a-mediated gene targeting. We anticipated that Cas12a-mediated homozygous inactivation of *Nf1, Rasa1,* and *Pten* might be limited; therefore, we used a 5-fold higher titer in *H11*$^{LSL\text{-}Cas12a}$ mice relative to *KT* mice in which every transduced cell will express oncogenic KRAS. Unexpectedly, only 9 weeks after tumor initiation, the *H11*$^{LSL\text{-}Cas12a}$ mice developed extremely high tumor burden as assessed by lung weight (>15-folder higher than *KT* mice after subtracting normal lung weight), direct imaging of the lung lobes, and histology (**Figure 2 d-f**). As expected, histology, positive immunohistochemical staining for NKX2-1/TTF-1, and negative staining for UCHL1 confirmed that these tumors were lung adenomas and adenocarcinomas (**Figure 2f** and **Supplementary Figure 2a**)[19]. Neoplastic cells in these tumors expressed nuclear Cas12a (**Supplementary Figure 2b**).

To assess indels at each of the target sites, we PCR-amplified each of the 9 targeted regions from genomic DNA from bulk tumor-bearing lungs from *H11*$^{LSL\text{-}Cas12a}$ mice and normal lung tissue from wild-type mice. Sanger sequence traces for each targeted genomic region (**Figure 2g**) showed a 10-20% indel rate at each target site, totaling ~30% for each gene (**Figure 2h**). Moreover, overall frequencies of indels at each target site were similar for all three genes (**Figure 2h**), consistent with the large numbers of tumors being driven by inactivation of all three genes.

Quantification of Cas12a-mediated tumorigenesis

PCR amplification and high-throughput sequencing of the BC-crRNA region of the integrated Lenti-U6^{BC}-cr*Nf1/Rasa1/Pten*-Cre vectors from bulk tumor-bearing lungs enabled the quantification of the number of cancer cells in each tumor and the number of clonal tumors in each mouse (**Figure 3a**). Overall tumor burden normalized to viral titer was more than 10-fold greater in *H11*$^{LSL\text{-}Cas12a}$ than *KT* mice (**Figure 3b**). We also assessed the number of clonal barcoded tumors in each mouse and found that *H11*$^{LSL\text{-}Cas12a}$ mice had an average of 2-fold more tumors than *Kras*$^{LSL\text{-}G12D}$ mice when corrected for viral titer (**Figure 3c**). These results are consistent with the high tumorigenic potential of lung epithelial cells with combined inactivation of *Nf1*, *Rasa1*, and *Pten*[19,23], as well as efficient multiplexed genome editing by Cas12a. Tumor sizes in *H11*$^{LSL\text{-}Cas12a}$ mice were dramatically larger than in *KT* mice, with the log-normal mean tumor size around 4-fold greater in *H11*$^{LSL\text{-}Cas12a}$ mice (**Figure 3d**), in line with an increased tumor growth rate for cr*Nf1/Rasa1/Pten* tumors compared to those driven by oncogenic KRAS. Thus, in this experiment we validated the ability of combined *Nf1*, *Rasa1*, and *Pten* inactivation to efficiently generate lung tumors *in vivo* and used three guides targeting each gene in all combinations.

Induction of small-cell lung cancer through the Cas12a-mediated inactivation of *Rb1, Trp53,* and *Rbl2*

Small-cell lung cancer (SCLC) is a neuroendocrine cancer that has been modeled in mice through Cre/lox-mediated, and more recently Cas9-mediated, inactivation of *Rb1*, *Trp53* and *Rbl2*[24,25]. We generated a lentiviral vector expressing Cre and a pre-crRNA array with all combinations of three guides targeting *Rb1*, *Trp53*, and *Rbl2* to create 27 vectors for combinatorial

gene inactivation (Lenti-U6^{BC}-cr*Rb1/Trp53/Rbl2*-Cre; **Figure 4a**). We transduced *H11*$^{LSL\text{-}Cas12a}$ mice (some of which also contained an *R26*$^{LSL\text{-}Tomato}$ allele) and wild-type negative control mice with Lenti-U6^{BC}-cr*Rb1/Trp53/Rbl2*-Cre (**Figure 4b**). After 18 weeks, transduced *H11*$^{LSL\text{-}Cas12a}$ mice began to exhibit declining health and had high lung tumor burden as assessed by lung weight, imaging of lung lobes, and histology (**Figure 4c-e**). Tumors in these mice included UCHL1^{positive} bronchiolar early-stage SCLC, along with more poorly-differentiated regions (**Figure 4e**), and stained positive for the Cas12a HA tag (**Supplementary Figure 2b**). Amplification of the BC-crRNA region from bulk tumor-bearing lungs, high-throughput sequencing, and Tuba-seq^{Ultra} analysis revealed that *H11*$^{LSL\text{-}Cas12a}$ mice transduced with Lenti-U6^{BC}-cr*Rb1/Trp53/Rbl2*-Cre had very large numbers of clonal tumors and high tumor burden, consistent with robust tumor growth (**Figure 4f-g**).

Generation of autochthonous PDAC with inactivation of *Trp53, Cdkn2a,* and *Smad4*

TP53, *CDKN2A*, and *SMAD4* are the three most frequently mutated tumor suppressor genes in human pancreatic ductal adenocarcinoma (PDAC), and many diagrams of the genomic progression of this cancer type show the acquisition of mutations/alterations in all three of these genes[26,27]. However, despite two decades of modeling oncogenic KRAS-driven PDAC in genetically engineered mice and extensive work to investigate these tumor suppressor genes in pancreatic carcinogenesis *in vivo*[28–34], a model in which all three tumor suppressor genes are inactivated has yet to be published. To determine whether Cas12a-mediated somatic genome editing would also be efficient in adult pancreatic epithelial cells, we generated a lentiviral vector expressing Cre and pre-crRNA arrays with three gRNAs targeting *Trp53*, *Cdkn2a*, and *Smad4* (Lenti-U6^{BC}-cr*Trp53/Cdkn2a/Smad4*-Cre; **Figure 4a**).

We delivered this pool to the pancreas of *KT;H11*$^{LSL\text{-}Cas12a}$, Cas12a-negative *KT*, and *H11*$^{LSL\text{-}Cas12a}$ mice through retrograde pancreatic ductal injection (**Figure 4b**)[35]. Several *KT;H11*$^{LSL\text{-}Cas12a}$ mice developed large multifocal pancreatic tumors as early as 6 weeks after transduction, with some also showing metastatic dissemination to the liver (**Figure 4c-e**). These primary tumors were PDAC in which 80-90% of the tumor area was poorly differentiated with sarcomatoid morphology and cytological atypia (features associated with aggressive, malignant PDAC in humans), with the remaining tumor area containing mucinous regions and lobular structures (**Figure 4f**, left versus right *KT;H11*$^{LSL\text{-}Cas12a}$ panels, respectively). These more differentiated tumor areas also expressed Cytokeratin-19 and had extensive desmoplastic stroma (**Supplementary Figure 3a**). *KT* mice developed only early-stage PanIN lesions, and the pancreata of transduced *H11*$^{LSL\text{-}Cas12a}$ mice appeared entirely normal, suggesting that even the combined inactivation of these three tumor suppressors has minimal effect in the absence of oncogenic KRAS (**Figure 4f**).

To quantify tumor number and size, we amplified the BC-crRNA region for bulk tumor-bearing pancreas tissue and used Tuba-seq^{Ultra}. *KT;H11*$^{LSL\text{-}Cas12a}$ mice had a much larger normalized tumor burden compared to *KT* mice (**Figure 4g**), although both genotypes developed roughly the same number of barcoded tumors when normalized to viral titer (**Figure 4h**). This discrepancy was attributable to the fact that *KT;H11*$^{LSL\text{-}Cas12a}$ mice had a much larger average tumor size (**Supplementary Figure 3b**), with the largest individual tumors having $>10^7$ neoplastic cells (**Figure 4i**). Tumors in *KT;H11*$^{LSL\text{-}Cas12a}$ mice retained Cas12a expression as evidenced by immunohistochemical staining of the Cas12a HA tag (**Supplementary Figure 2b**). While indel percentages from bulk tissue were variable, most target sites had a ~10-20% indel rate

(**Supplementary Figure 3c**). Thus, these data indicate that Cas12a-mediated inactivation of *Trp53*, *Cdkn2a*, and *Smad4* within the pancreas rapidly led to robust tumor growth.

DISCUSSION

Human cancer has tremendous genomic complexity, yet the genetically engineered mouse models that have been used to study cancer generally have only a few designed genetic changes. While this has enabled a reductionist approach to uncover key attributes controlled by many genes in cancer, it has also limited our ability to model and understand key aspects of cancer that are driven by combinations of genomic alterations. For instance, engineered mouse models often fail to develop spontaneous metastases and those that do rarely metastasize to all organs commonly observed in humans[36]. Indeed, metastases tend to have a greater degree of genomic complexity than primary tumors[37], so modeling metastatic cancer in mice may be facilitated by the generation of tumors with additional engineered alterations.

While Cre/lox- and CRISPR/Cas9-mediated somatic genome editing have been used to generate cancer models driven by complex combinations of loss of function alleles, the generation of new models remains time-consuming[36]. Furthermore, given the complexity of genotypes within human tumors, the generation of a single model of a given cancer type (*e.g.* our model of oncogene-negative lung adenocarcinoma) should not be seen as a model for all patients with that broad class of tumors. Here, we have sought to close the gap between the complexity of human tumors and the relative simplicity of genetically engineered mouse models through the creation of transgenic mice expressing enAsCas12a. We demonstrated that crRNA arrays containing three guides can generate complex genotypes in somatic cells, although previous *in vitro* studies have employed as many as 25 guides together[8], suggesting that even more loci could be targeted together *in vivo* with

this transgenic model. By synthesizing and cloning arrays in a simple pooled manner, this approach should enable the rapid generation of hundreds to thousands of complex genotypes in parallel, dramatically increasing both the scale and the rate at which new models with complex genotypes can be studied.

While our study focused on the use of Cas12a for multiple gene inactivation in cancer models, this system is versatile and should have numerous future applications within and beyond cancer modeling. Cas12a-generated staggered DNA breaks could make it more suitable for the introduction of exogenous DNA sequences through homology-directed repair[10]. Somatic Cas12a-mediated genome editing may also be well-suited for generating chromosomal rearrangements, such as oncogenic translocations that require the simultaneous targeting of pairs of loci, as well as the generation of panels of larger deletions. The multiplicity of Cas12a targeting could also make it suitable for modeling even more complex mutational states, such as chromothripsis or aneuploidy (through chromosomal destruction). Finally, transgenic Cas12a in our models obviates the need for exogenous delivery of the large (4 kb) Cas12a gene and negates concerns of pre-existing immunity, which has been shown to dramatically affect cells with exogenous Cas9 expression[38].

Compared to Cas9, Cas12a creates more diverse indels and continues cutting its target loci for longer due to the greater distance between its PAM and cleavage sites[39]. As a result, Cas12a has been used to generate an evolving barcode system to map cellular phylogenies[39]. Incorporation of this system into *in vivo* models of development and disease could complement Cas9-based approaches, and the multiplexed ability of Cas12a crRNA expression could allow lineage tracing in combination with gene inactivation. This could dramatically increase our ability to both generate and study complex genotypes[40–43].

The generation of somatic cells with complex genotypes could be of broad value, both in cancer modeling and beyond. Somatic genome editing with Cas9 transgenic mice has enabled rapid analysis of the *in vivo* effects of genetic alterations in neuroscience[44–47], cancer[21,48–51], immunology[52], and other fields[53,54]. These Cas12a transgenic mice will enable the generation of pairwise and higher-order combinations of genetic alterations to be generated in cell types of interest to map complex phenotypes to complex genotypes.

METHODS

Generation of the *H11*$^{LSL\text{-}Cas12a}$ transgenic allele

A targeting vector containing 5' and 3' homology arms flanking the chicken beta-actin /CMV enhancer/gamma globulin splice acceptor (CAGGS) promoter, *loxP*-STOP(6xSV40 polyA)-*loxP* cassette (LSL), enhanced *Acidaminococcus sp.* Cas12a (E174R/N282A/S542R/K548R; enAsCas12a)[14], a nucleoplasmin nuclear localization signal (NLS), a 3xHA epitope sequence, an SV40 NLS[17], and a rabbit β-globin polyadenylation signal (RBG pA) was used to generate the *H11*$^{LSL\text{-}Cas12a}$ knock-in mice. An sgRNA targeting the mouse *Hipp11* (*H11*) locus (GAACACTAGTGCACTTATCCTGG), the targeting vector, and Cas9 mRNA were co-injected into fertilized C57BL/6J mouse oocytes (Cyagen Biosciences). F0 founder animals were identified by PCR followed by sequence analysis. A founder mouse was bred with C57BL/6J mice to establish the *H11*$^{LSL\text{-}Cas12a}$ line. *H11*$^{LSL\text{-}Cas12a}$ mice were crossed to CMV-Cre "deleter" mice to generate the constitutive *H11*^{Cas12a} mice.

Mice and tumor initiation

The use of mice for this study was approved by the Institutional Animal Care and Use Committee at Stanford University, protocol number 26696. CMV-Cre (RRID:IMSR_JAX:006054)[18], *Kras*$^{LSL\text{-}G12D/+}$ (RRID:IMSR_JAX:008179)[55] and *R26*$^{LSL\text{-}tdTomato}$ (RRID:IMSR_JAX:007914)[56] mice were on a C57BL/6J background and have been described. *H11*$^{LSL\text{-}Cas12a}$ mice (JAX:038388) and *H11*^{Cas12a} mice (JAX:038389) are available through the Jackson Laboratory.

Design, generation, barcoding, and production of lentiviral vectors

Pre-crRNA arrays with three guides were designed with diverse direct repeat sequences flanking each guide[12], with BsmBI cut sites on both ends of the array (all guide and array sequences in **Supplementary Table 1**). Guides for enAsCas12a were designed using CRISPick for the Mouse GRCm38 reference genome (NCBI RefSeq v.108.20200622)[12,57]. Arrays were ordered as single-stranded DNA pools (Twist Biosciences) containing all 27 combinations of 3 guides for each of the 3 genes targeted in an array. Barcoded lentiviral vector backbones were created by cloning a 98 bp oligo with a 16-nucleotide diverse barcode (BC) and two BsmBI restriction sites into the 3' end of the bovine U6 promoter with Gibson Assembly (NEBuilder HiFi, New England Biosciences) in a vector also containing PGK-Cre recombinase[22]. After low-cycle PCR amplification, pre-crRNA arrays were ligated via Golden Gate Assembly with BsmBI into the barcoded vector backbone. The cloning product was then electroporated into competent *E. Coli* (C3020K, New England Biosciences) and plated onto LB-ampicillin plates. To ensure sufficient barcode diversity for Tuba-seq^{Ultra} sequencing and delineation of tumors, ~10^6 colonies were collected for each Lenti-U6^{BC}-crRNA-Cre pool.

Lentiviral vectors were produced using polyethylenimine (PEI)-based transfection of 293T cells with delta8.2 and VSV-G packaging plasmids in 15 cm cell culture plates. Sodium butyrate (B5887, Sigma Aldrich) was added 8 hours after transfection to achieve a final concentration of 20 mM. Media was refreshed 24 hours after transfection. 20 mL of virus-containing supernatant was collected 36, 48, and 60 hours after transfection. The three collections were then pooled and concentrated by ultracentrifugation (112,000 g for 1.5 hours), resuspended overnight in 120 µL PBS, and then frozen at -80°C. Viruses were titered against a lab standard of known titer.

Tumor initiation

Lung tumors were initiated by intratracheal delivery of 60 µl of lentiviral vectors in PBS to isoflurane anesthetized mice as previous described[58]. Pancreatic tumors were initiated by retrograde ductal delivery of 150 µl of lentiviral vectors in PBS to isoflurane anesthetized mice as previous described[35]. For small-cell lung cancer tumors, mice were pre-treated with naphthalene. Briefly, corn oil (C8267, Sigma-Aldrich) was filter-sterilized with a 0.22 µm filter and aliquoted into 5 ml portions. Naphthalene was then dissolved into the corn oil vehicle at a concentration of 50 mg/mL before it was administered into mice via intraperitoneal injections at a dosage of 200 mg/kg. Roughly 46-48 hours later, mice were administered virus intratracheally and supplemented with a bowl of DietGel76A (72-07-5022, ClearH2O) food to promote recovery.

Histology and immunohistochemistry

Tissues were fixed in 4% formalin for 24 hours, stored in 70% ethanol, and paraffin-embedded. and 4 µm thick sections were used for Hematoxylin and Eosin staining and immunohistochemistry. Immunohistochemistry was performed using an Avidin/Biotin Blocking

Kit (SP-2001, Vector Laboratories), Avidin-Biotin Complex kit (PK-4001, Vector Laboratories), and DAB Peroxidase Substrate Kit (SK-4100, Vector Laboratories) following standard protocols. Alcian Blue and Trichrome stains were performed following standard protocols (Histo-Tec Laboratory, Inc.). The following primary antibodies were used: anti-HA (3724, Cell Signaling Technology), anti-TTF1 (NKX2-1; ab76013, Abcam), anti-CK19 (Ab2133570, TROMA-III, Developmental Study Hybridoma Bank), and anti-UCHL1 (HPA005993, Sigma-Aldrich).

Western blot analyses

For immunoblotting, cells and homogenized tissues were lysed in Cell Lysis Buffer (Cell Signaling Technology) containing a complete mini-protease inhibitor cocktail (Roche), then equal quantities of protein lysate (3-10 µg) were separated by SDS-PAGE using 4–12% gradient gels (Invitrogen) and transferred to PVDF membranes. Membranes were blocked with 5% milk in PBS with 0.1% Tween 20 (PBST) for 1 hour at room temperature, followed by incubation with primary antibodies diluted in PBST with 5% milk overnight at 4ºC. After 4x15 minute washes with PBST, membranes were incubated with an HRP-conjugated goat anti-rabbit secondary antibody (12-348, Sigma-Aldrich) diluted 1:5000 in PBST with 5% milk. After 4x15 minute washes with PBST, protein expression was then visualized with enhanced chemiluminescence reagents (PI80196, Fisher Scientific). Primary antibodies were used at the following dilutions: rabbit anti-HA, 1:1000 (3724, Cell Signaling); rabbit anti-Cas12a, 1:1000 (19984, Cell Signaling); and rabbit anti-alpha-Tubulin, 1:2000 (2144, Cell Signaling).

Analysis of indels at the target sites

The target sites for each crRNA were PCR amplified from genomic DNA extracted from tumor-bearing lungs using GoTaq Green® Mastermix (Promega) and the following primer pairs:

Nf1 target site#1 forward 5' ACCCTGAGACTGTGACAACAC 3'

Nf1 target site#1 reverse 5' TCTCTGCTAAGGGCAACCAA 3'

Nf1 target site#2 forward 5' GCCCTGATAATGGGGATGATG 3'

Nf1 target site#2 reverse 5' CCCACACAAAAAGCTCTGCTAC 3'

Nf1 target site#3 forward 5' GCAACCAAAGGACACAATGAGA 3'

Nf1 target site#3 reverse 5' TGGAGATATCTTCTCCAAGTCTGT 3'

Rasa1 target site#1 forward 5' AAACAAATCACAGCGGCAGG 3'

Rasa1 target site#1 reverse 5' GTCCGGGTTCCTGGGGA 3'

Rasa1 target site#2 forward 5' GCCACACTGATACAAACCTTGG 3'

Rasa1 target site#2 reverse 5' CAGGTACCCCTGTTCGTTCC 3'

Rasa1 target site#3 forward 5' AGCGCTTTCCCATTCAATATTAC 3'

Rasa1 target site#3 reverse 5' AGGGAATTTGAGAGGAGGTGA 3'

Pten target site#1 forward 5' ACAGGGTCAGTGCAGTTGAA 3'

Pten target site#1 reverse 5' GCTGTCCCACACCGTCAATA 3'

Pten target site#2 forward 5' GCTTGTTTCCATGTCTAATGAATGC 3'

Pten target site#2 reverse 5' TGGTCCACAAAGAGGGAGGA 3'

Pten target site#3 forward 5' TGGTCCACAAAGAGGGAGGA 3'

Pten target site#3 reverse 5’ TGGTCCACAAAGAGGGAGGA 3’

Cdkn2a target site #1 forward 5’ CGCACGATGTCTTGATGTCC 3’

Cdkn2a target site #1 reverse 5’ ATGGGGTAGAGCCTCCTGAC 3’

Cdkn2a target site #2 forward 5’ GCGGAACGCAAATATCGCA 3’

Cdkn2a target site #2 reverse 5’ TAGAGCCTCCTGACTGTGGA 3’

Cdkn2a target site #3 forward 5’ CCAGCGGAACGCAAATATCG 3’

Cdkn2a target site #3 reverse 5’ AAGAATTCCAAGGCGGGACT 3’

Smad4 target site #1 forward 5’ TGTGGAAGCCACAGGAATGT 3’

Smad4 target site #1 reverse 5’ AAAGTAGCTAGAGCTGATAAAATGC 3’

Smad4 target site #2 forward 5’ TTACACATCTATGAAGGTGCTGC 3’

Smad4 target site #2 reverse 5’ TTACACATCTATGAAGGTGCTGC 3’

Smad4 target site #3 forward 5’ GCTGCAGTGGCAGTCAACTTA 3’

Smad4 target site #3 reverse 5’ CCCCTCCTGATGATACTTGGG 3’

Trp53 target site #1 forward 5’ GGGTGTCTGTAAATCCTGCG 3’

Trp53 target site #1 reverse 5’ ACTCGGAGGGCTTCACTTG 3’

Trp53 target site #2 forward 5’ GTCACAGCACATGACGGAGG 3’

Trp53 target site #2 reverse 5’ CGGCAGCTTGCACCTCTAAG 3’

Trp53 target site #3 forward 5’ CTGGTAAGGCCCAGAGCAGA 3’

Trp53 target site #3 reverse 5’ CACACGAAAGACAACTCCCC 3’

Amplicons were run on 1% agarose gels, gel-extracted (QIAquick PCR & Gel Cleanup Kit, Qiagen), and Sanger sequenced (Elim Biopharm, Inc). The Sanger sequencing traces were analyzed with TIDE[59] to estimate the percent of the DNA with indels at each site. As these lungs each contain many clonal tumors, the DNA should be a mix of that from non-neoplastic cells (normal lung and stromal cells), tumors with indels at that site, and tumor with indels at the target sites of the other crRNAs targeting that gene.

Tuba-seq^{Ultra} library preparation and analysis

Whole lungs were removed from the freezer and allowed to thaw at room temperature. Benchmark control cell lines containing unique barcodes (50,000 cells each) were added to each lung sample prior to lysis to enable the calculation of the absolute number of neoplastic cells in each tumor[22]. Bulk tumor-bearing tissues were homogenized in 6 ml of cell lysis buffer (Puregene, 158063, Qiagen) using a FastPrep-24 5G tissue homogenizer (116005500, MP Biomedicals) and digested overnight with proteinase K (AM2544, Life Technologies). To remove RNA, samples were incubated with RNase A prior to genomic DNA extraction using a Puregene kit according to the manufacturer's recommended protocol. Q5 High-Fidelity 2x Master Mix (M0494X, New England Biolabs) was used to amplify the U6-BC-crRNA region from 32 μg of genomic DNA in a total reaction volume of 800 μl per sample using unique dual-indexed primers as described[22]. The concentration of the amplified barcode product in each PCR was measured using D1000 screentape reagents (5067-5582, Agilent Technologies) and the Tapestation instrument (Agilent Technologies). Amplicons were pooled at equal molar ratios of barcode product, normalized to the estimated burden of tumors in each mouse lung sample (measured lung mass minus an estimated normal lung weight of 0.15 g). Pooled PCR products were purified using Agencourt AMPure XP

beads (A63881, Beckman Coulter) using a double size selection protocol and sequenced on the Illumina NextSeq 6000 platform (read length 2x150bp).

Paired-end reads were processed with regular expressions to identify the three gRNA sequences (henceforth referred to as the tumor-genotype) and clonal barcodes. When identifying gRNA sequences, we strictly required a perfect match with the designed sequences. The 12-nucleotide random clonal barcode sequence possesses a high theoretical diversity of approximately 4^{12} (> 10^7). Given the virus titer we used ($2x10^5$ ifu and $1x10^6$ ifu per mouse), there are typically fewer than 10,000 clonal tumors for each tumor-genotype per mouse. As such, the probability of two clonal tumors with the same tumor-genotype possessing two barcodes within a 1-hamming distance of each other is extremely low. Consequently, when we encountered low-frequency clonal barcodes within a 1-hamming distance of high-frequency clonal barcodes, we attributed them to sequencing or PCR errors. These low-frequency barcodes were merged with barcodes of higher frequencies. After extracting barcode and crRNA information, we converted the read counts associated with clonal tumors into absolute neoplastic cell numbers. This conversion was accomplished by normalizing the reads of the clonal tumor to the number of reads of the "spike-in" benchmark cell lines added to each sample prior to lung lysis and DNA extraction. We imposed a minimum tumor size cutoff of 300 cells for downstream analysis.

Cell line generation

Tail tip fibroblasts were generated by mechanically dissociating depilated tail tips, digesting them in 0.25% trypsin (Life Technologies) for 30 minutes, collecting tissue fragments with centrifugation at 500 rcf for 5 minutes, then resuspending and plating in high glucose Dulbecco’s Modified Eagle Medium (Life Technologies) containing 20% fetal bovine serum and

2% penicillin-streptomycin (Thermo Fisher Scientific). Tumor-derived cell lines were generated by the same method, except that mechanically dissociated tumors were digested with Hank's Balance Salt Solution (Corning) containing trypsin, collagenase type 4 (Fisher Scientific), and dispase (Corning), then passed through 40 µm cell strainers (Falcon, Fisher Scientific) to break up cell clumps.

ACKNOWLEDGEMENTS

We thank the Stanford Veterinary Animal Care Staff for expert animal care and members of the Winslow laboratory for helpful comments. J.D.H was supported by an American Cancer Society Fellowship (PF-21-112-01-MM) and a TRDRP Postdoctoral fellowship (T31FT1619). Y.J.T was partly supported by the Canadian Institute of Health Research (CIHR) postdoctoral fellowship (MFE-176568). P.A.R. was supported by the National Science Foundation Graduate Research Fellowship (DGE-2146755) and the Lucille P. Markey Stanford Graduate Fellowship. This work was supported by NIH R01-CA230025 (to M.M.W), NIH R01-CA231253 (to M.M.W and D.A.P), NIH R01-CA234349 (to M.M.W and D.A.P.), NIH R35-CA231997 (to J.S.), NIH R35-HG011316 and R01-GM141627 (to L.C.), and in part by the Stanford Cancer institute support grant (NIH P30-CA124435).

References

1. Vogelstein, B. *et al.* Cancer Genome Landscapes. *Science* **339**, 1546–1558 (2013).
2. Aaltonen, L. A. *et al.* Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
3. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
4. Tang, Y. J., Shuldiner, E. G., Karmakar, S. & Winslow, M. M. High-Throughput Identification, Modeling, and Analysis of Cancer Driver Genes In Vivo. *Cold Spring Harb. Perspect. Med.* **13**, a041382 (2023).
5. Paul, B. & Montoya, G. CRISPR-Cas12a: Functional overview and applications. *Biomed. J.* **43**, 8–17 (2020).
6. Khan, S. & Sallard, E. Current and Prospective Applications of CRISPR-Cas12a in Pluricellular Organisms. *Mol. Biotechnol.* **65**, 196–205 (2023).
7. Zetsche, B. *et al.* Multiplex gene editing by CRISPR–Cpf1 using a single crRNA array. *Nat. Biotechnol.* **35**, 31–34 (2017).
8. Campa, C. C., Weisbach, N. R., Santinha, A. J., Incarnato, D. & Platt, R. J. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat. Methods* **16**, 887–893 (2019).
9. Dede, M., McLaughlin, M., Kim, E. & Hart, T. Multiplex enCas12a screens detect functional buffering among paralogs otherwise masked in monogenic Cas9 knockout screens. *Genome Biol.* **21**, 1–19 (2020).
10. Dai, X. *et al.* Massively parallel knock-in engineering of human T cells. *Nat. Biotechnol.* **41**, 1239–1255 (2023).

11. Magnusson, J. P., Rios, A. R., Wu, L. & Qi, L. S. Enhanced Cas12a multi-gene regulation using a CRISPR array separator. *eLife* **10**, e66406 (2021).

12. DeWeirdt, P. C. *et al.* Optimization of AsCas12a for combinatorial genetic screens in human cells. *Nat. Biotechnol.* **39**, 94–104 (2021).

13. Ma, E. *et al.* Improved genome editing by an engineered CRISPR-Cas12a. *Nucleic Acids Res.* **50**, 12689–12701 (2022).

14. Kleinstiver, B. P. *et al.* Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.* **37**, 276–282 (2019).

15. Gier, R. A. *et al.* High-performance CRISPR-Cas12a genome editing for combinatorial genetic screening. *Nat. Commun.* **11**, 3455 (2020).

16. Zhang, L. *et al.* AsCas12a ultra nuclease facilitates the rapid generation of therapeutic cell medicines. *Nat. Commun.* **12**, 3908 (2021).

17. Liu, P. *et al.* Enhanced Cas12a editing in mammalian cells and zebrafish. *Nucleic Acids Res.* **47**, 4169–4180 (2019).

18. Schwenk, F., Baron, U. & Rajewsky, K. A cre-transgenic mouse strain for the ubiquitous deletion of loxP-flanked gene segments including deletion in germ cells. *Nucleic Acids Res.* **23**, 5080–5081 (1995).

19. Yousefi, M. *et al.* Combinatorial Inactivation of Tumor Suppressors Efficiently Initiates Lung Adenocarcinoma with Therapeutic Vulnerabilities. *Cancer Res.* **82**, 1589–1602 (2022).

20. Rogers, Z. N. *et al.* A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. *Nat. Methods* **14**, 737–742 (2017).

21. Cai, H. *et al.* A functional taxonomy of tumor suppression in oncogenic KRAS-driven lung cancer. *Cancer Discov.* (2021) doi:10.1158/2159-8290.CD-20-1325.

22. Tang, Y. J. *et al.* A functional map of epigenetic regulators in lung carcinogenesis. *Manuscr. Prep.*

23. Yousefi, M. *et al.* Fully accessible fitness landscape of oncogene-negative lung adenocarcinoma. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2303224120 (2023).

24. Schaffer, B. E. *et al.* Loss of p130 Accelerates Tumor Development in a Mouse Model for Human Small-Cell Lung Carcinoma. *Cancer Res.* **70**, 3877–3883 (2010).

25. Duplaquet, L. *et al.* KDM6A epigenetically regulates subtype plasticity in small cell lung cancer. *Nat. Cell Biol.* **25**, 1346–1358 (2023).

26. Iacobuzio-Donahue, C. A., Velculescu, V. E., Wolfgang, C. L. & Hruban, R. H. Genetic basis of pancreas cancer development and progression: insights from whole-exome and whole-genome sequencing. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **18**, 4257–4265 (2012).

27. Halbrook, C. J., Lyssiotis, C. A., Pasca di Magliano, M. & Maitra, A. Pancreatic cancer: Advances and challenges. *Cell* **186**, 1729–1754 (2023).

28. Izeradjene, K. *et al.* Kras(G12D) and Smad4/Dpc4 haploinsufficiency cooperate to induce mucinous cystic neoplasms and invasive adenocarcinoma of the pancreas. *Cancer Cell* **11**, 229–243 (2007).

29. Hingorani, S. R. *et al.* Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. *Cancer Cell* **7**, 469–483 (2005).

30. Bardeesy, N. *et al.* Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. *Genes Dev.* **20**, 3130–3146 (2006).

31. Aguirre, A. J. *et al.* Activated Kras and Ink4a/Arf deficiency cooperate to produce metastatic pancreatic ductal adenocarcinoma. *Genes Dev.* **17**, 3112–3126 (2003).

32. Hingorani, S. R. *et al.* Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. *Cancer Cell* **4**, 437–450 (2003).

33. Whittle, M. C. *et al.* RUNX3 Controls a Metastatic Switch in Pancreatic Ductal Adenocarcinoma. *Cell* **161**, 1345–1360 (2015).

34. Bardeesy, N. *et al.* Both p16(Ink4a) and the p19(Arf)-p53 pathway constrain progression of pancreatic adenocarcinoma in the mouse. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 5947–5952 (2006).

35. Chiou, S.-H. *et al.* Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. *Genes Dev.* **29**, 1576–1585 (2015).

36. Hebert, J. D., Neal, J. W. & Winslow, M. M. Dissecting metastasis using preclinical models and methods. *Nat. Rev. Cancer* **23**, 391–407 (2023).

37. Nguyen, B. *et al.* Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185**, 563-575.e11 (2022).

38. Li, A. *et al.* AAV-CRISPR Gene Editing Is Negated by Pre-existing Immunity to Cas9. *Mol. Ther. J. Am. Soc. Gene Ther.* **28**, 1432–1441 (2020).

39. Hughes, N. W. *et al.* Machine-learning-optimized Cas12a barcoding enables the recovery of single-cell lineages and transcriptional profiles. *Mol. Cell* **82**, 3103-3118.e8 (2022).

40. Yang, D. *et al.* Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell* **185**, 1905-1923.e25 (2022).

41. Li, L. *et al.* A mouse model with high clonal barcode diversity for joint lineage, transcriptomic, and epigenomic profiling in single cells. *Cell* **186**, 5183-5199.e22 (2023).

42. Bowling, S. *et al.* An Engineered CRISPR-Cas9 Mouse Line for Simultaneous Readout of Lineage Histories and Gene Expression Profiles in Single Cells. *Cell* **181**, 1410-1422.e27 (2020).

43. Chan, M. M. *et al.* Molecular recording of mammalian embryogenesis. *Nature* **570**, 77–82 (2019).

44. Santinha, A. J. *et al.* Transcriptional linkage analysis with in vivo AAV-Perturb-seq. *Nature* **622**, 367–375 (2023).

45. Jin, X. *et al.* In vivo Perturb-Seq reveals neuronal and glial abnormalities associated with autism risk genes. *Science* **370**, eaaz6063 (2020).

46. Zheng, X. *et al.* Massively parallel in vivo Perturb-seq reveals cell type-specific transcriptional networks in cortical development. *BioRxiv Prepr. Serv. Biol.* 2023.09.18.558077 (2023) doi:10.1101/2023.09.18.558077.

47. Lin, R. *et al.* Directed evolution of adeno-associated virus for efficient gene delivery to microglia. *Nat. Methods* **19**, 976–985 (2022).

48. Li, B. E. *et al.* In vivo CRISPR/Cas9 screening identifies Pbrm1 as a regulator of myeloid leukemia development in mice. *Blood Adv.* **7**, 5281–5293 (2023).

49. Romero, R. *et al.* Keap1 loss promotes Kras -driven lung cancer and results in dependence on glutaminolysis. *Nat. Med.* **23**, 1362–1368 (2017).

50. Ng, S. R. *et al.* CRISPR-mediated modeling and functional validation of candidate tumor suppressor genes in small cell lung cancer. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 513–521 (2020).

51. Lee, M. C. *et al.* A multiplexed in vivo approach to identify driver genes in small cell lung cancer. *Cell Rep.* **42**, 111990 (2023).

52. Belk, J. A. *et al.* Genome-wide CRISPR screens of T cell exhaustion identify chromatin remodeling factors that limit T cell persistence. *Cancer Cell* **40**, 768-786.e7 (2022).

53. Song, X. *et al.* Genome Editing with AAV-BR1-CRISPR in Postnatal Mouse Brain Endothelial Cells. *Int. J. Biol. Sci.* **18**, 652–660 (2022).

54. Thürkauf, M. *et al.* Fast, multiplexable and efficient somatic gene deletions in adult mouse skeletal muscle fibers using AAV-CRISPR/Cas9. *Nat. Commun.* **14**, 6116 (2023).

55. Jackson, E. L. *et al.* Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* **15**, 3243–3248 (2001).

56. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).

57. Kim, H. K. *et al.* Deep learning improves prediction of CRISPR-Cpf1 guide RNA activity. *Nat. Biotechnol.* **36**, 239–241 (2018).

58. DuPage, M., Dooley, A. L. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat. Protoc.* **4**, 1064–1072 (2009).

59. Brinkman, E. K., Chen, T., Amendola, M. & van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, e168 (2014).

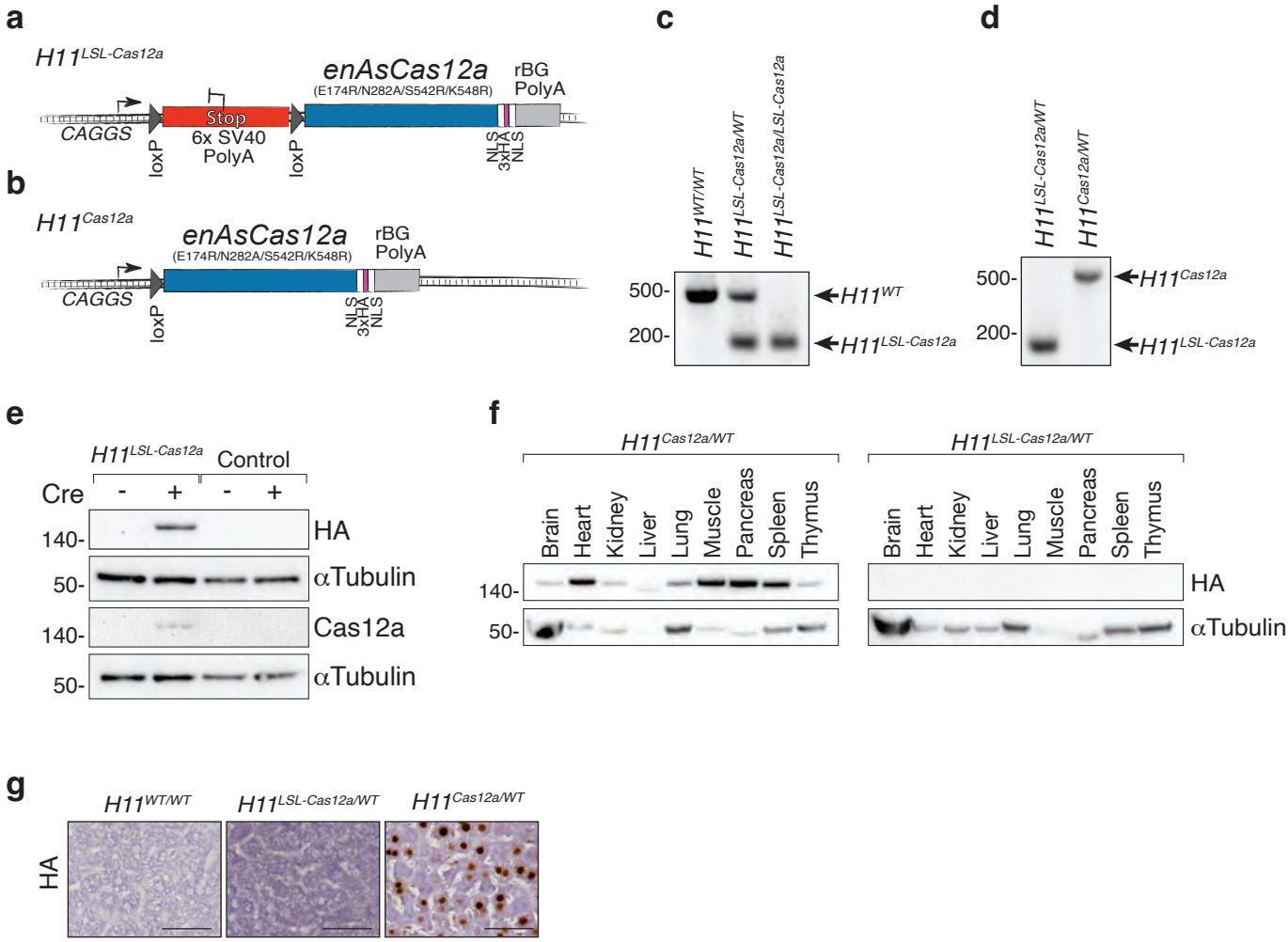


Figure 1. Generation of Cre-regulated and constitutive Cas12a transgenic mice

a,b. Schematic of the *H11*$^{LSL\text{-}Cas12a}$ (**a**) and *H11*^{Cas12a} (**b**) transgenes. Enhanced AsCas12a containg several substitutions to increase on-target efficiency and PAM binding sequence range (Kleinstiver *et al.* (2019). *Nature Biotechnology*), has two nuclear localization sequences (NLS), and three HA tags at the C-terminus (Liu *et al.* (2019), *Nucleic Acids Research*).

c,d. PCR genotyping of mice of the indicated *H11*$^{LSL\text{-}Cas12a}$ and *H11*^{Cas12a} genotypes. DNA ladder sizes in bp are indicated.

e. Western blots on tail tip fibroblasts from *H11*$^{LSL\text{-}Cas12a}$ and wild-type control mice 3 days after Adeno-Cre (Cre) treatment. Cas12a was detected by both anti-HA and anti-Cas12a antibodies. αTubulin shows loading. Protein ladder sizes in kDa are indicated.

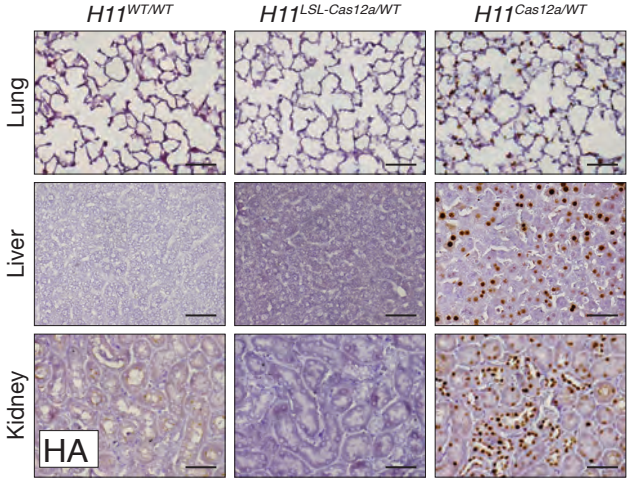
f. Western blots on a panel of tissue lysates from *H11*$^{Cas12a/WT}$ or *H11*$^{LSL\text{-}Cas12a/WT}$ mice. αTubulin shows loading. Protein ladder sizes in kDa are indicated.

g. Immunohistochemical staining for the HA tag on Cas12a on liver sections from the indicated genotypes of mice. Scale bars, 50 μm.

a

	Cas9	Cas12a
Structural features		
RNA scaffold	tracrRNA (~100 nt/guide)	Direct repeats (DR;~20 nt each)
Promoter	one per sgRNA	one per crRNA array
Cloning	one step per sgRNA	one step per crRNA array
Guide uncoupling due to lentiviral template switching	high	low
Ability to sequence across 3 guides	difficult	easy
Cutting		
Cut site relative to PAM	3 bp from PAM	18 bp from the PAM
Cuts	blunt-end	5 nt staggered

b



Supplementary Figure 1. Comparison of salient features of Cas9 and Cas12a, and broad nuclear expression of Cas12a in *H11*^{Cas12a} mice

a. Summary of features of Cas9 and Cas12a relevant to multiplexed genome editing. Note that the potential impact of off-target effects is increased when targeting more genes.

b. Immunohistochemical staining for the Cas12a HA tag in lung, liver and kidney sections from the indicated genotypes of mice. Scale bars, 50 μm. Higher magnification images of these liver sections is shown in Figure 1g.

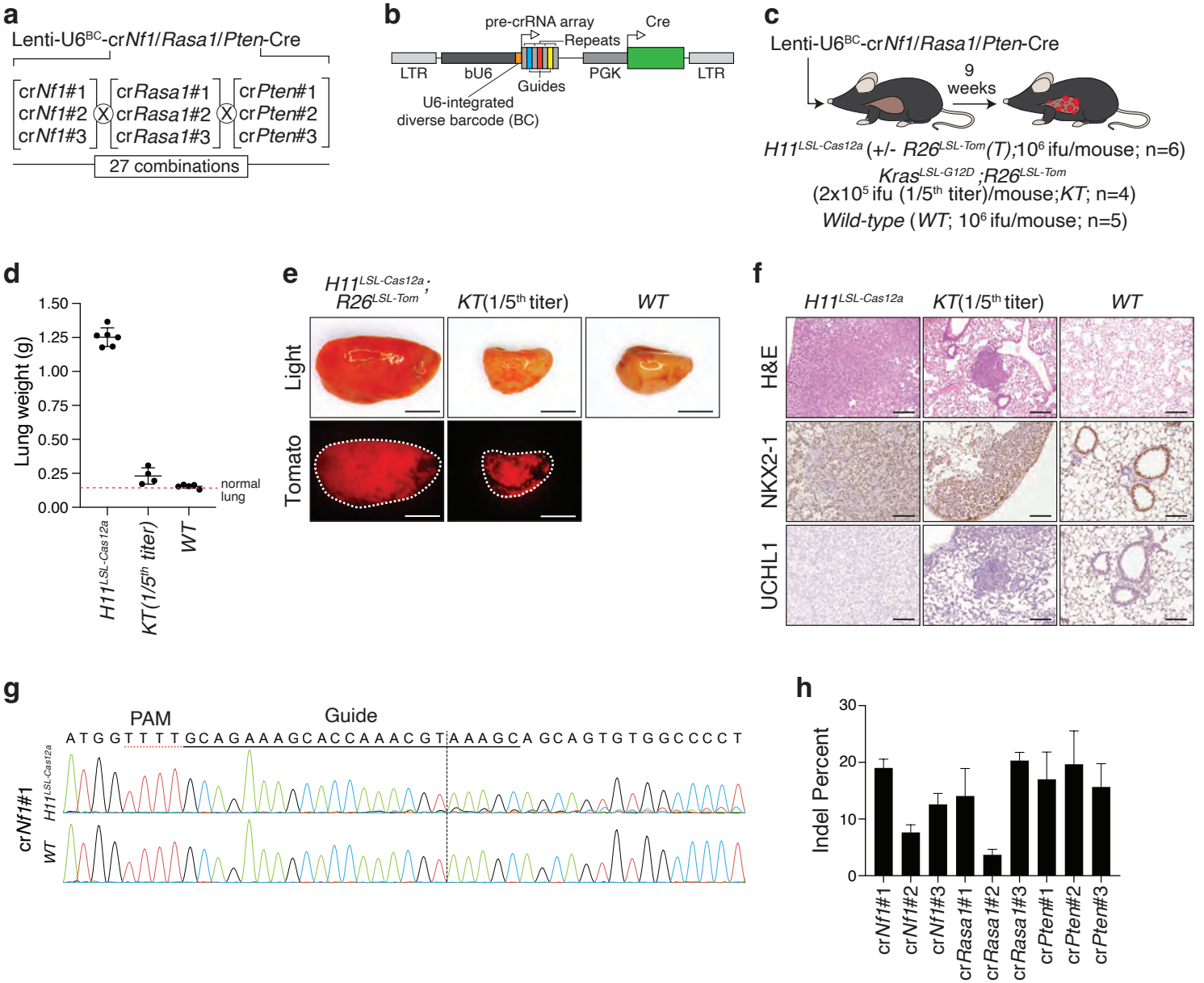


Figure 2. Rapid and efficient generation of oncogene-negative lung tumors through Cas12a-mediated co-incident inactivation of three tumor suppressor genes

a. Lenti-U6^{BC}-cr*Nf1/Rasa1/Pten*-Cre pool has guides targeting *Nf1, Rasa1*, and *Pten*. Each gene is targeted by three crRNAs in all 27 combinations.

b. Design of a lentiviral vector that expresses Cre recombinase and has a pre-crRNA array with three spacers (guides) downstream of a U6 promoter with an integrated diverse barcode region.

c. Intratracheal delivery of Lenti-U6^{BC}-cr*Nf1/Rasa1/Pten*-Cre to *H11*$^{LSL\text{-}Cas12a}$, *Kras*$^{LSL\text{-}G12D}$;*R26*$^{LSL\text{-}Tom}$ (*KT*), and *Wild-type (WT)* mice. Viral titer and mouse numbers are indicated.

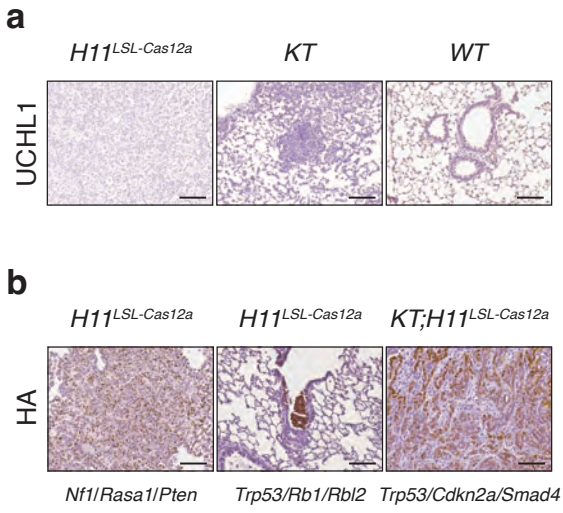
d. Lung weights from mice of the indicated genotypes 9 weeks after transduction with Lenti-U6^{BC}cr*Nf1/Rasa1/Pten*-Cre. Each dot represents a mouse. Mean +/- standard deviation are indicated.

e. Light (upper) and Tomato fluorescent (lower) images of lung lobes from the indicated genotypes of mice 9 weeks after trasnduction with Lenti-U6^{BC}-cr*Nf1/Rasa1/Pten*-Cre. Dashed lines outline tissues. Tumors are Tomato-positive due to recombination of the *R26*$^{LSL\text{-}Tomato}$ allele. Scale bars, 5 mm.

f. Hematoxylin & eosin (upper), NKX2-1 (middle), and UCHL1 (lower) staining of lung sections from the indicated genotypes. Scale bars, 100 µm.

g. Representative Sanger sequencing traces for one genomic region within *Nf1* targeted by cr*Nf1*#1 in *H11*$^{LSL\text{-}Cas12a}$ and *WT* samples, with wild-type sequence shown above. Red dotted line, PAM. Solid black line, guide sequence. Vertical dashed line, estimated cut site on non-template strand (note that Cas12a induced staggered cuts).

h. Indel frequencies for genomic regions targeted by each crRNA within bulk tumor-bearing lung tissue from *H11*$^{LSL\text{-}Cas12a}$ mice. Means +/- standard deviation of three tumor-bearing lungs are shown.



Supplementary Figure 2. Lack of SCLC marker staining in oncogene-negative tumors, and sustained expression of Cas12a in tumors transduced with Lenti-U6^{BC}-crRNA-Cre vectors

a. Immunohistochemical staining for UCHL1 on lung sections from the indicated genotypes of Lenti-U6^{BC}cr*Nf1/Rasa1/Pten*-Cre transduced mice. Scale bars, 100 µm.

b. Immunohistochemical staining for the Cas12a HA tag on tumors from the indicated genotypes of mice (above) transduced with the indicated Lenti-U6^{BC}-crRNA-Cre vectors (below). Scale bars, 100 μm.

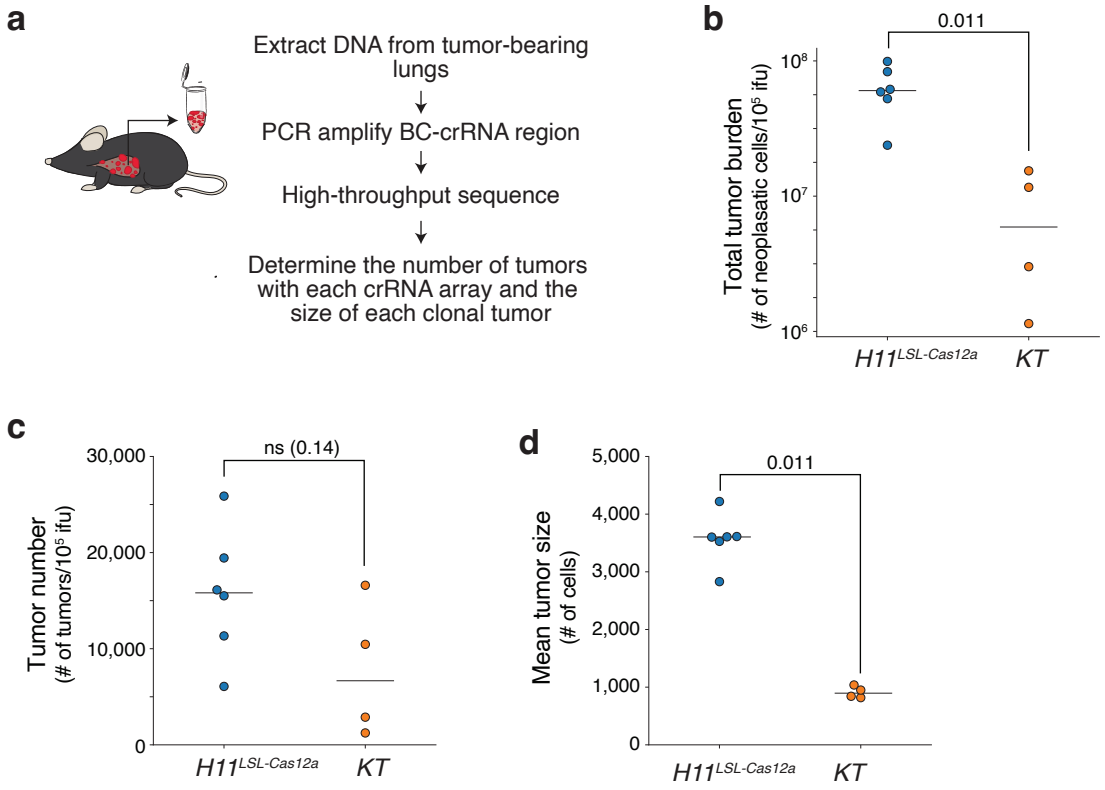


Figure 3. Integration of crRNA arrays with tumor barcoding enables quantification of tumor initiation and tumor size
a. Schematic of tumor barcoding with high-throughput BC-crRNA sequencing to determine the size of each Cas12a-induced clonal tumor.
b. Total number of neoplastic cells (Total tumor burden) in each mouse normalize to viral titer. Each dot represents a mouse and the bar is the mean. P-value calculated with Wilcoxon rank-sum test.
c. The number of tumors in each mouse normalized to viral titer. Each dot represents a mouse and the bar is the mean. P-value calculated with Wilcoxon rank-sum test.
d. The mean tumor size given a log-normal tumor size distribution. Each dot represents a mouse and the bar is the mean. P-value calculated with Wilcoxon rank-sum test.

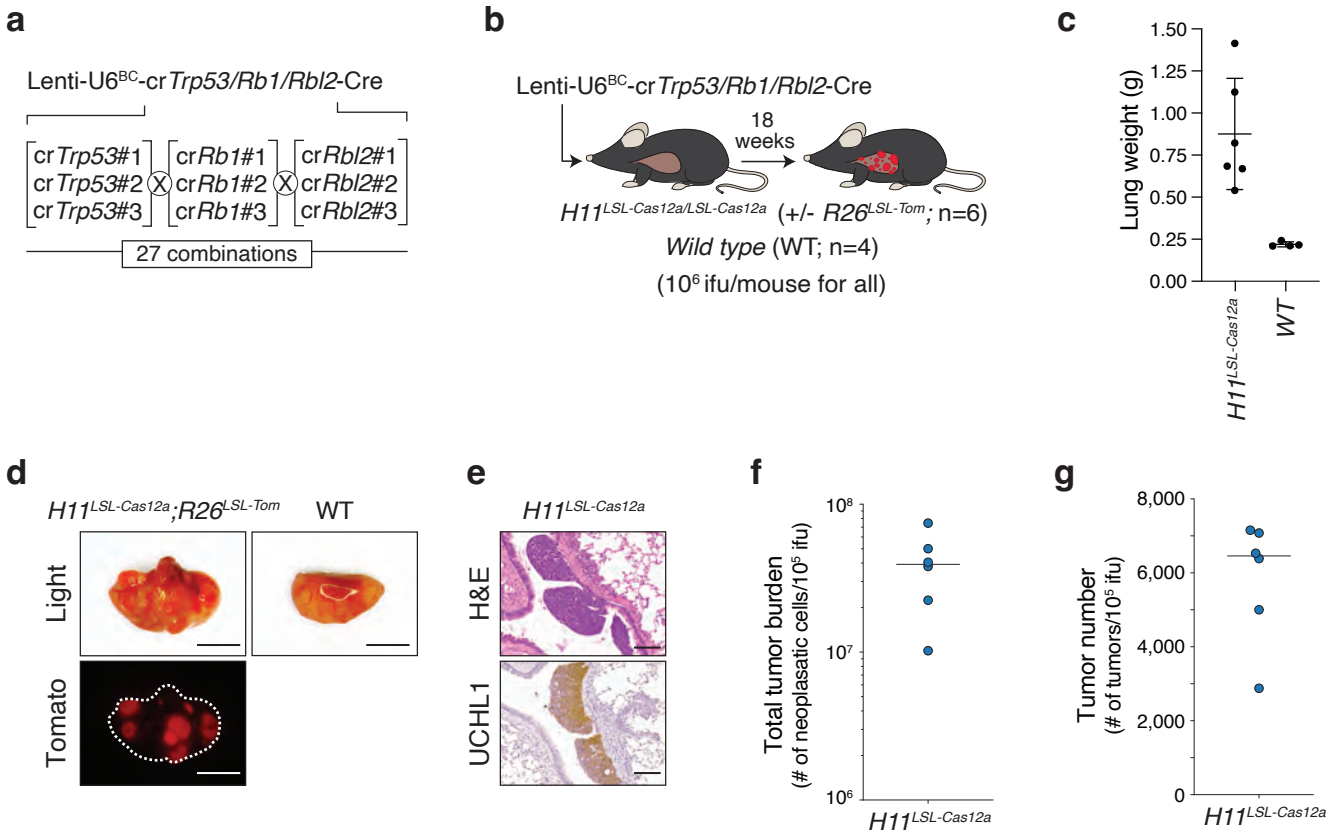


Figure 4. Induction of small-cell lung cancer through the simultaneous Cas12a-mediated inactivation of *Trp53, Rb1* and *Rbl2*

a. Lenti-U6^{BC}-cr*Trp53/Rb1/Rbl2*-Cre pool has guides targeting *Trp53, Rb1*, and *Rbl2*. Each gene is targeted by three crRNAs in all 27 combinations.

b. Intratracheal delivery of Lenti-U6^{BC}-cr*Trp53/Rb1/Rbl2*-Cre to *H11*$^{LSL\text{-}Cas12a}$ and *Wild-type (WT)* mice. Viral titer and mouse numbers are indicated.

c. Lung weights from mice of the indicated genotypes 9 weeks after transduction with Lenti-U6^{BC}-cr*Trp53/Rb1/Rbl2*-Cre. Each dot represents a mouse. Mean +/- standard deviation are indicated.

d. Light (upper) and Tomato fluorescent (lower) images of lung lobes from the indicated genotypes of mice 18 weeks after transduction with Lenti-U6^{BC}-cr*Trp53/Rb1/Rbl2*-Cre. Dashed line outlines tissue. Tumors are Tomato-positive due to recombination of the *R26*$^{LSL\text{-}Tomato}$ allele. Scale bars, 5 mm.

e. Hematoxylin & eosin (upper) and UCHL1 (lower) staining of lung sections from an *H11*$^{LSL\text{-}Cas12a}$ mouse. Scale bars, 100 µm.

f. Total number of neoplastic cells (Total tumor burden) in each mouse normalize to viral titer. Each dot represents a mouse and the bar is the mean.

g. The number of tumors in each mouse normalize to viral titer. Each dot represents a mouse and the bar is the mean.

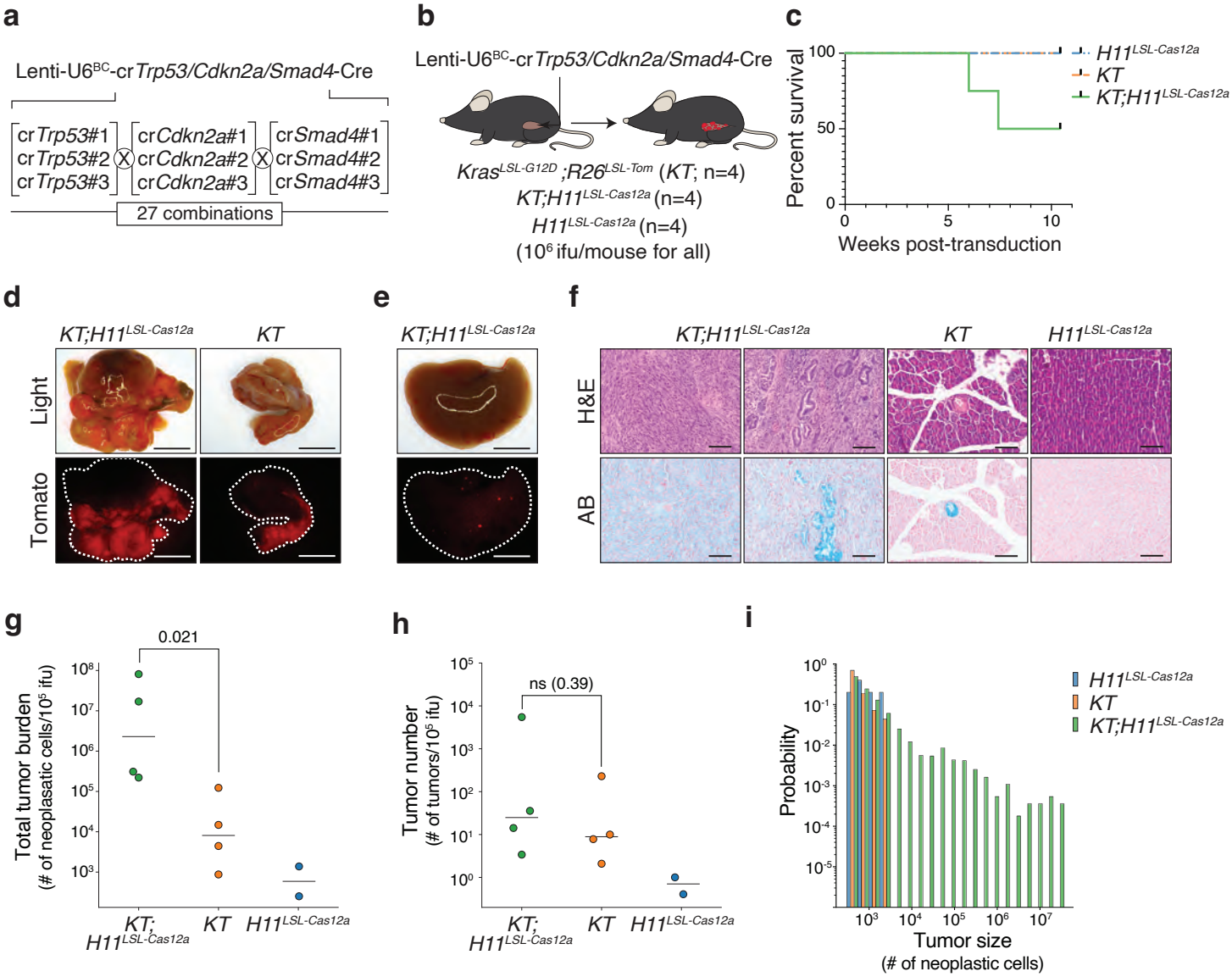


Figure 5. Rapid generation of aggressive PDAC through inactivation of the three most commonly mutated tumor suppressor genes

a. Lenti-U6^{BC}-cr*Trp53/Cdkn2a/Smad4*-Cre pool has guides targeting *Trp53, Cdkn2a*, and *Smad4*. Each gene is targeted by three crRNAs in all 27 combinations.

b. Intrapancreatic delivery of Lenti-U6^{BC}-cr*Trp53/Cdkn2a/Smad4*-Cre to *Kras$^{LSL-G12D}$;R26$^{LSL-Tom}$* (*KT*), *KT;H11$^{LSL-Cas12a}$*, and *H11$^{LSL-Cas12a}$* mice. Viral titer and mouse numbers are indicated.

c. Survival curve of mice of the indicated genotypes transduced with Lenti-U6^{BC}-cr*Trp53/Cdkn2a/Smad4*-Cre.

d. Light (upper) and fluorescent Tomato (lower) pancreas images from the indicated genotypes of mice 6 weeks (*KT;H11$^{LSL-Cas12a}$*) or 11 weeks (*KT*) after transduction with Lenti-U6^{BC}-cr*Trp53/Cdkn2a/Smad4*-Cre. Dashed lines outline tissues. Scale bars, 5 mm.

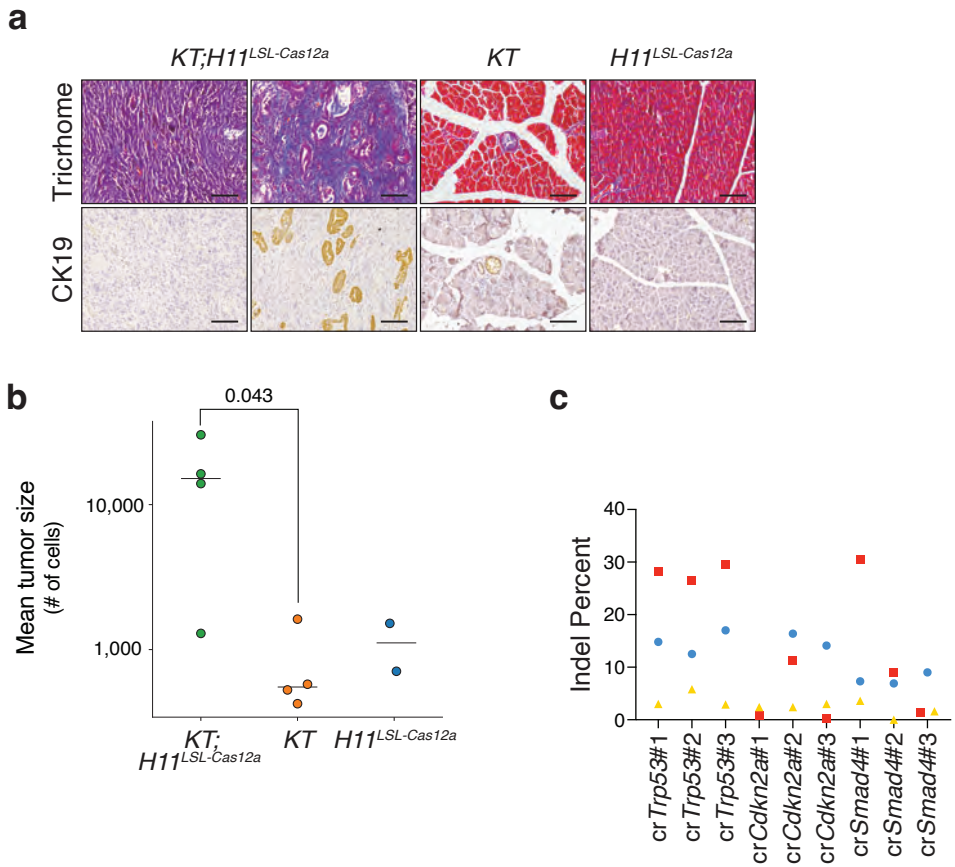
e. Light (upper) and fluorescent Tomato (lower) images of a liver lobe from a *KT;H11$^{LSL-Cas12a}$* mouse. Dashed line outlines tissue. Scale bars, 5mm.

f. Immunohistochemistry images of pancreas sections from the indicated genotypes, showing hematoxylin & eosin (H&E; upper) and Alcian Blue (AB; lower) staining. *KT;H11$^{LSL-Cas12a}$* panels show representative regions with poorly differentiated sarcomatoid carcinoma (left; approximately 80-90% of tumor area) and more differentiated areas with mucinous cells and retaining some lobular structure (right). Scale bars, 100 µm.

g. Total number of neoplastic cells (Total tumor burden) in each mouse normalized to viral titer. Each dot represents a mouse and the bar is the mean. Note that two *H11$^{LSL-Cas12a}$* mice did not have detectable tumor burden and are not plotted. P-value calculated with Wilcoxon rank-sum test.

h.The number of tumors in each mouse normalized to viral titer. Each dot represents a mouse and the bar is the mean. Note that two *H11$^{LSL-Cas12a}$* mice did not have detectable tumor burden and are not plotted. P-value calculated with Wilcoxon rank-sum test.

i. Probability distribution of the size of tumors in the indicated genotypes of mice.



Supplementary Figure 3. Inactivation of *Trp53*, *Cdkn2a*, and *Smad4* greatly increases tumor size and leads to the development of PDAC with differentiated areas with CK19+ cancer cells and desmoplastic stroma as well as more poorly differentiated areas.

a. Immunohistochemistry images of pancreas sections from the indicated genotypes, showing Trichrome (upper) and Cytokeratin-19 (CK19; lower) staining. *KT;H11*$^{LSL\text{-}Cas12a}$ panels show representative poorly differentiated (left) and more differentiated (right) regions. Scale bars, 100 µm.

b. Mean tumor sizes in mice of the indicated genotypes transduced with Lenti-U6^{BC}-cr*Trp53/Cdkn2a/Smad4*-Cre, given a log-normal tumor size distribution. Each dot represents a mouse and the bar is the mean. P-value calculated with Wilcoxon rank-sum test.

c. Indel frequencies for genomic regions targeted by each crRNA within bulk tumor-bearing pancreas tissue from KT;*H11*$^{LSL\text{-}Cas12a}$ mice. Symbol shapes and colors are identical for values from the same mouse.