Article

# A prime editor mouse to model a broad spectrum of somatic mutations in vivo

Zackery A. Ely[1,2,13], Nicolas Mathey-Andrews[1,2,3,13], Santiago Naranjo[1,2], Samuel I. Gould [1,2], Kim L. Mercer [1], Gregory A. Newby [4,5,6], Christina M. Cabana[1,2], William M. Rideout III[1], Grissel Cervantes Jaramillo[1,7], Jennifer M. Khirallah[8], Katie Holland[1,9], Peyton B. Randolph [4,5,6], William A. Freed-Pastor[1,3,10,11], Jessie R. Davis[4,5,6], Zachary Kulstad[1,10], Peter M. K. Westcott[1,12], Lin Lin[1], Andrew V. Anzalone[4,5,6], Brendan L. Horton[1], Nimisha B. Pattada[1], Sean-Luc Shanahan[1,2], Zhongfeng Ye[8], Stefani Spranger[1,2], Qiaobing Xu [8], Francisco J. Sánchez-Rivera [1,2], David R. Liu [4,5,6] & Tyler Jacks [1,2] ✉

Genetically engineered mouse models only capture a small fraction of the genetic lesions that drive human cancer. Current CRISPR–Cas9 models can expand this fraction but are limited by their reliance on error-prone DNA repair. Here we develop a system for in vivo prime editing by encoding a Cre-inducible prime editor in the mouse germline. This model allows rapid, precise engineering of a wide range of mutations in cell lines and organoids derived from primary tissues, including a clinically relevant *Kras* mutation associated with drug resistance and *Trp53* hotspot mutations commonly observed in pancreatic cancer. With this system, we demonstrate somatic prime editing in vivo using lipid nanoparticles, and we model lung and pancreatic cancer through viral delivery of prime editing guide RNAs or orthotopic transplantation of prime-edited organoids. We believe that this approach will accelerate functional studies of cancer-associated mutations and complex genetic combinations that are challenging to construct with traditional models.

Cancer is driven by somatic mutations that accumulate throughout progression and often display extensive intertumoral heterogeneity, occurring in thousands of different combinations across human cancer[1,2]. The precise nature of driver mutations and their combinations can profoundly influence how cancers initiate, progress and respond to therapy, establishing tumor genotype as a critical determinant of disease outcome[3,4]. Emerging precision oncology treatment paradigms aim to match specific therapies with tumor genotypes, and this strategy has shown promise for several driver mutations[5,6]. To expand the promise of precision oncology to more patients, it is critical to develop tools to systematically interrogate the effects of distinct genetic lesions and combinations thereof on the overall tumor phenotype, particularly in vivo.

[1]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA, USA. [2]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, USA. [3]Harvard Medical School, Boston, MA, USA. [4]Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [5]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [6]Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA. [7]Harvard-MIT Division of Health Sciences and Technology, Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [8]Department of Biomedical Engineering, Tufts University, Medford, MA, USA. [9]Department of Biology, Angelo State University, San Angelo, TX, USA. [10]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [11]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [12]Cold Spring Harbor Laboratory, Huntington, NY, USA. [13]These authors contributed equally: Zackery A. Ely, Nicolas Mathey-Andrews. ✉e-mail: tjacks@mit.edu

Genetically engineered mouse models (GEMMs) have proven invaluable for elucidating the mechanisms by which cancer drivers promote tumor development and progression in vivo[7,8]. However, generating new GEMMs using traditional approaches is an expensive, laborious and time-consuming process. Established GEMMs can also take months for investigators to acquire and often require laborious breeding programs to combine multiple alleles of interest and to establish a colony of sufficient size for experimental cohorts. These factors impede studies aimed at developing precision oncology treatments for tumors driven by specific genetic variants, which continue to be identified on a regular basis[9].

Genome editing technologies like CRISPR–Cas9 can be used to rapidly engineer somatic mutations when delivered exogenously or when installed as germline alleles[10–14]. While these models have accelerated the study of putative cancer driver genes, they are most frequently used to induce DNA double-stranded breaks (DSBs), leading to inactivation of tumor suppressor genes via error-prone repair and frameshifting insertion/deletion (indel) formation. Although CRISPR-based homology-directed repair (HDR) has been used to model precise single nucleotide variants (SNVs) in Cas9-knockin mice, this method requires an exogenous DNA donor template and is limited by low efficiency and high rates of indel byproducts[15]. Furthermore, the requirement for DSBs to induce frameshifts or HDR-based precise edits can lead to confounding genotoxic effects, including on-target chromothripsis events and artificial fitness costs incurred through continued disruption of edited oncogenes[16,17].

Precision genome editing technologies like base editing[18] can be used to model cancer in mice by installing specific transition mutations with high efficiency and negligible indel byproducts[11]. Although precise and highly efficient, base editors also have limitations, including the requirement for different base editor enzymes depending on the mutation being studied (for example, cytosine base editor (CBE) or adenine base editor (ABE)), and their propensity for bystander editing, which can prohibit introducing desired amino acid substitutions. While the recent development of C:G and A:Y transversion base editors will expand the scope of cancer modeling[19–22], current base editing technology is not amenable to modeling the full spectrum of small somatic mutations.

In contrast to base editing and standard CRISPR–Cas9, prime editing enables engineering the full spectrum of single nucleotide substitutions and indels with high product purity[23,24]. Prime editors employ a Cas9 nickase coupled with a reverse transcriptase that complexes with prime editing guide RNAs (pegRNAs). pegRNAs encode mutations of interest within a reverse transcriptase template (RTT)[23,24], enabling highly precise and programmable editing. Prime editing thus offers a versatile approach to study the full spectrum of cancer driver mutations, their combinations and the growing catalog of secondary mutations that confer resistance to targeted therapies[25–28].

Beyond editing versatility, prime editing also avoids the formation of indel byproducts associated with DSBs. This is particularly important for studying SNVs with putative neomorphic qualities in tumor suppressor genes, as HDR-directed mutations would be diluted by the higher rate of naturally selected indels. Prime editing also exhibits lower rates of unintended activity at off-target loci, reducing the risk of confounding off-target effects[24,29]. These advantages, combined with broad editing capacity, provide an unprecedented opportunity to generate faithful models of human cancer.

With these considerations in mind, we developed both conditional and tissue-restricted prime editing GEMMs (PE GEMMs) that eliminate the need for exogenous delivery of prime editors, which can be challenging given their significant size[30,31]. Encoding the prime-editing machinery within the mouse germline also minimizes confounding acute or chronic anti-tumor immune responses that could be induced by exogenous delivery of a Cas9-based fusion protein[32–34]. In conjunction with the development of PE GEMMs, we also developed a range

of DNA vectors and engineered pegRNAs (epegRNAs) that promote efficient prime editing in a variety of cell lines and organoids derived from these mice. With this toolset, we established organoid models harboring *Trp53* mutations frequently found in patients with pancreatic cancer but not modeled by current GEMMs of the disease, as well as a clinically relevant *Kras* mutation associated with resistance to KRAS^G12C inhibitors. We further showed that PE GEMMs enable efficient prime editing in vivo via viral or nonviral delivery of pegRNAs to a variety of tissues. Extending these studies, we harnessed PE GEMMs to model cancer in vivo through somatic initiation of autochthonous lung and pancreatic adenocarcinomas, and by orthotopic transplantation of prime-edited pancreatic organoids. We also investigated the oncogenic potential of a variety of primary *Kras* mutations in the lung, including the poorly understood *Kras*^G12A mutation present in more than 10% of patients with lung adenocarcinoma. We expect PE GEMMs to both expand the landscape of achievable cancer-associated mutations and accelerate the techniques required to study their function and associated therapeutic vulnerabilities.

## Results

### Quantification of cancer mutations amenable to prime editing

Recent study has shown that base editing can be used to elucidate the function of specific cancer-associated genetic variants[35] and to systematically probe a large fraction of all possible alleles for genes and proteins of interest[36]. Base editors are primarily capable of engineering transition SNVs[23] (A·T > G·C or G·C > A·T), although the base editor architecture has recently been adapted to produce C·G > G·C transversions with variable efficiency[19,20,37–39]. In contrast, prime editors can engineer all transition and transversion SNVs[24], as well as defined indel alleles[40,41], expanding the potential for rapid modeling of genetic variants even further. To define the expanded editing capacity afforded by prime editing, we quantified the abilities of both base and prime editing to install specific somatic mutations identified from a cohort of 43,035 genetically-profiled patients with cancer from the Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT) dataset (Fig. 1a,b and Supplementary Fig. 1)[9,35]. Of 422,822 mutations identified from the targeted exon sequencing of 594 cancer-associated genes, 82.3% are SNVs, while the remaining 17.7% are deletions (DEL), insertions (INS) and di-nucleotide variant (DNV)/oligo-nucleotide variants (ONV), in descending order of frequency (Fig. 1a).

To estimate what fraction of common cancer-associated mutations are captured in currently available transgenic mouse models, we analyzed a dataset curated from the Mouse Genome Informatics database (Methods)[42,43]. We found that 65 of the 100 most frequent SNVs in MSK-IMPACT, including 50 of 84 missense SNVs, are not represented by published mouse cancer models (Supplementary Tables 1 and 2). Notably, the majority of these SNVs are transitions, which comprise 61.8% of all SNVs in the overall MSK-IMPACT dataset and are theoretically compatible with engineering using base editors (Fig. 1c). In general, 38.4% of all mutations in the dataset are amenable to base editing using a canonical NGG PAM sequence[23,35] (Fig. 1d). The total mutation coverage with base editing increases to 51% when accounting for base editors that use more abundant NG PAM sequences.

With base editors, adjacent identical nucleotides can be collaterally edited and result in undesired editing outcomes. When considering only mutations without identical bases present within one adjacent nucleotide, the total mutation coverage drops to 29.6% (Fig. 1d). This analysis does not account for the location of a desired edit within the protospacer, which can influence base editing efficiency and the total fraction of amenable mutations (Extended Data Fig. 1b).

We used a similar approach to quantify the modeling capabilities of prime editors that use an NGG or NG PAM coupled with variable RTT lengths encoded within pegRNAs (Supplementary Fig. 1). Using an NGG PAM and RTT length of 30 base pairs (bp), excluding the additional
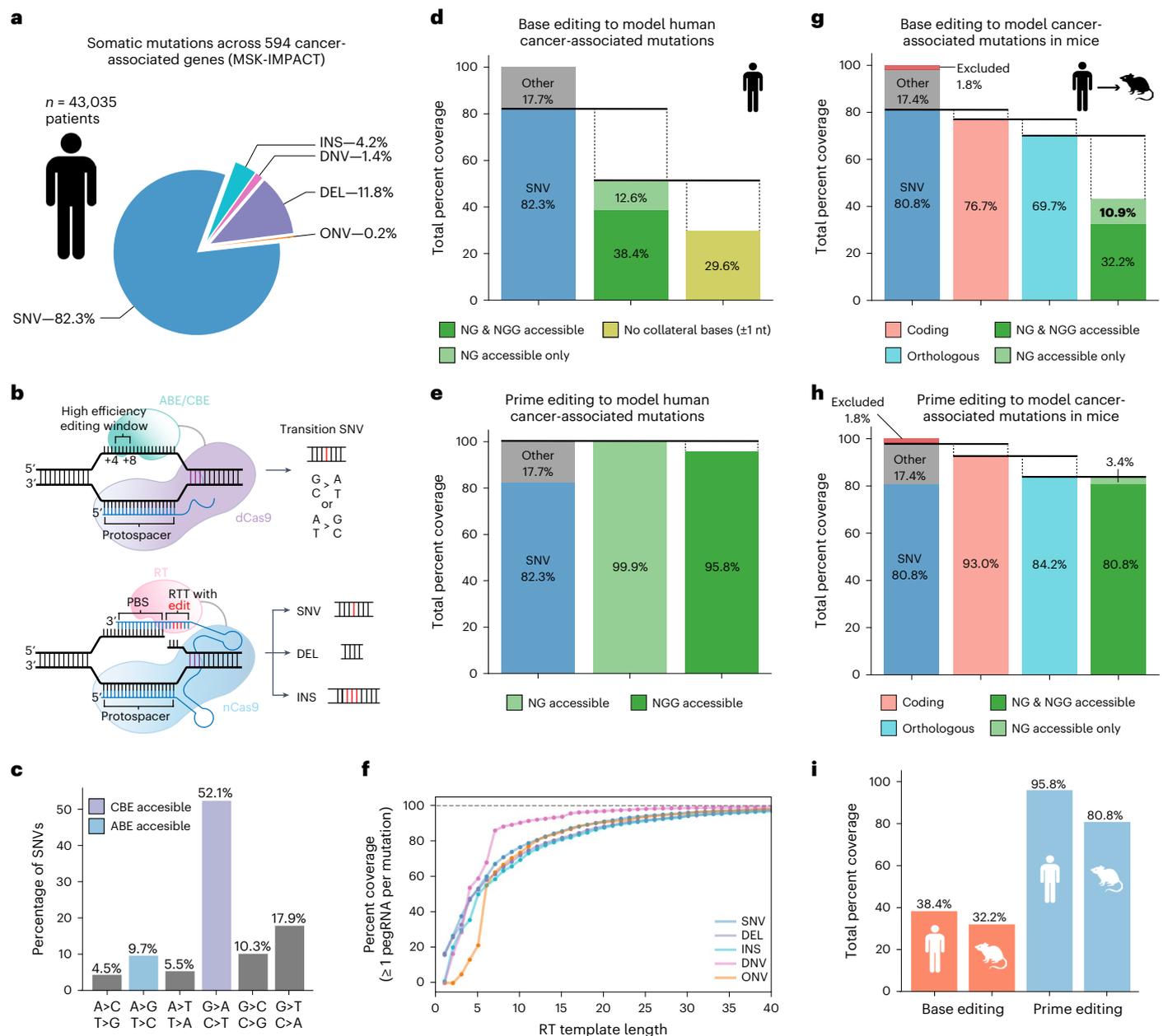
**Fig. 1 | Quantification of cancer-associated mutations amenable to modeling by base editing or prime editing. a**, Distribution of somatic variant types in a cohort of 43,035 patients with 422,822 mutations observed in 594 cancer-associated genes. **b**, Schematic of the modeling capabilities of base editing (top) and prime editing (bottom). **c**, Quantification of somatic SNVs by type. SNVs amenable to modeling by CBEs are shown in purple, while SNVs amenable to ABEs are shown in blue. Transversions are shown in gray. **d**, Quantification of mutations amenable to modeling with CBEs or ABEs that use an NG (light green) or NGG PAM (dark green). All percentages are given as a percentage of all mutations in the dataset. **e**, Quantification of mutations amenable to modeling by a prime editor using an NGG PAM (dark green) coupled with a pegRNA with an RT template length of 30 nucleotides. **f**, Percentage of mutations with at least one suitable pegRNA as a function of the RT template length of the pegRNA, excluding the additional length of a homologous region in the RT template.

Calculations assume an NGG PAM. **g**, Quantification of orthologous coding mutations potentially amenable to modeling by base editing in mice. Mutations are defined as orthologous if they derive from a wild-type amino acid conserved in the mouse ortholog, as determined by pairwise protein alignment between human and mouse protein sequences. The rightmost bar indicates the fraction of orthologous coding mutations that can be modeled by base editors that recognize NG or NGG PAMs. 'Excluded mutations' refers to mutations that fall in a gene lacking an ortholog. **h**, Quantification of orthologous coding mutations potentially amenable to modeling by prime editing. The rightmost bar indicates the ability of an NG or NGG prime editor to model these mutations, assuming an RT template greater than 30 nt. **i**, Summary of the mutation modeling capabilities of base and prime editing assuming an NGG PAM. SNV, single nucleotide variants; DEL, deletions; INS, insertions; DNV, di-nucleotide variants; ONV, oligo-nucleotide variants.

length of a homologous region in the RTT, prime editing theoretically reaches 95.8% coverage of all mutations in this dataset (Fig. 1e). This value increases to 99.9% for prime editors that could theoretically use an NG PAM (Fig. 1e). Moreover, analysis of the relationship between

RTT length and modeling capabilities reveals that ~85% of mutations in this dataset can be modeled by placing the mutation within the first 15 bp of the RTT (Fig. 1f). These parameters are well within the recommended guidelines for pegRNA RTT length, even with the additional

size required for a region of homology[23]. Collectively, this analysis suggests that both base editing and prime editing can serve as versatile technologies for modeling cancer-associated mutations.

We also sought to determine the fraction of cancer-associated mutations that derive from protein sequences conserved in mouse orthologs. We reasoned that only this subset of conserved sequences, when mutated in mouse systems, could be expected to mimic effects seen in human cancer. To quantify the ability of base and prime editors to model cancer-associated mutations in mice, we performed pairwise alignment on orthologous mouse and human proteins to define whether mutations derive from a conserved wild-type amino acid and reside in a region of homology (Supplementary Fig. 1). Of the SNVs that occur in coding sequences, 90.9% derive from codons that encode conserved amino acids between mouse and human. Of these conserved, cancer-associated SNVs, 61.8% are amenable to base editing (NG or NGG PAM), which translates to 43.1% of all mutations in the dataset (Fig. 1g). In contrast, NG or NGG prime editors are capable of modeling 100% of coding mutations that occur at conserved amino acid residues in mice (84.2% of all mutations in the dataset) (Fig. 1h). In total, 80.8% of human cancer-associated mutations observed in this dataset could be modeled in mice with prime editors using a traditional NGG PAM (Fig. 1f,i). This same pattern holds when filtering the dataset to only mutations that occur in multiple patients, and when considering various stringencies of homology in the regions flanking the mutations of interest (Extended Data Figs. 1f and 2c). In total, these results demonstrate that prime editing could substantially broaden both the diversity and number of human cancer-associated mutations that can be rapidly modeled in mouse orthologs.

## Development of a Cre-inducible prime editor allele

We sought to develop a transgenic system capable of precisely engineering the majority of cancer-associated mutations without requiring exogenous delivery of a prime editor enzyme. To accomplish this, we targeted a transgene expression cassette encoding the PE2 enzyme and the mNeonGreen (mNG)[44] fluorescent reporter, separated by the P2A ribosome skipping sequence, into the *Rosa26* locus[10,45] (Fig. 2a). Like the previous Cre-inducible *Rosa26* alleles[10,46,47], transgene expression is driven by the CAG promoter and is induced by Cre-mediated excision of a loxP-stop-loxP (LSL) cassette. A neomycin resistance gene was included to enable the selection of cells containing the targeted allele. We also incorporated FRT/FRT3 sequences flanking the central construct to enable Flp recombinase-mediated replacement of the *Rosa26^{PE2}* allele with future generations of prime editor enzymes or other desirable editors[29,48]. This vector was targeted to *Trp53^{flox/flox}* *C57BL/6J* ES cells, where *Trp53* can be deleted upon expression of Cre recombinase (Supplementary Fig. 2). Chimeric mice were then crossed to wild-type *C57BL/6J* mice to generate pure strain heterozygous *Trp53^{flox/+}*;*Rosa26^{PE2/+}* mice. These mice were subsequently crossed with *Trp53^{+/+}* and *Trp53^{flox/flox}* mice to generate *Rosa26^{PE2/+}* mice on wild-type and *Trp53^{flox/flox}* backgrounds.

## Functional validation of the prime editor allele in organoids

To confirm the functionality of the *Rosa26^{PE2}* allele, we developed two lentiviral vectors that coexpress a pegRNA and either Cre recombinase (hU6-pegRNA-EF-1α-Cre (UPEC)) or the red fluorescent protein (RFP), mScarlet[49] (hU6-pegRNA-EFS-mScarlet (UPEmS)) (Fig. 2b). We derived pancreatic organoids from chimeric *Trp53^{flox/flox}*;*Rosa26^{PE2/+}* mice and developed a pure culture of transgene-containing cells via selection with neomycin (Fig. 2c and Supplementary Fig. 3). As expected, these pancreatic organoids displayed Cre-dependent mNG expression upon transduction with UPEC vectors (Fig. 2d and Supplementary Fig. 3). To test the prime-editing functionality of this allele, we designed a *Dnmt1*-targeting pegRNA encoding a +1 CCC INS, which templates a trinucleotide insertion of a GGG codon encoding glycine at residue 4 of *Dnmt1*. UPEC-transduced organoids were selected using nutlin-3a, a mouse double minute 2 homolog (MDM2) inhibitor that induces cell cycle arrest in *Trp53*-proficient (but not *Trp53*-deficient) cells[50], enriching for those *Trp53^{flox/flox}* cells that underwent Cre-mediated recombination following UPEC transduction. After selection, we detected up to 33.8% editing efficiency and minimal indel byproducts at *Dnmt1* (Supplementary Fig. 3). These results validate the functionality of the *Rosa26^{PE2}* allele, including its ability to mediate prime editing of endogenous loci when using optimized pegRNAs.

## Prime editing in organoids derived from the *Rosa26^{PE2}* model

We next sought to evaluate prime editing across multiple tissues. To accomplish this, we derived lung organoids, pancreatic organoids and tail-tip-derived fibroblasts (TTFs) from multiple *Rosa26^{PE2/+}* mice (Fig. 2e). Consistent with results using chimera-derived organoids, we observed highly efficient *Dnmt1* editing across all investigated tissues (Fig. 2f). Corroborating the well-established on-target fidelity of prime editing[24,29,51], we did not detect off-target prime editing across multiple loci prioritized based on protospacer homology[52] (Extended Data Fig. 3c,d).

A previous study established a subset of DNA damage repair (DDR) genes as key factors influencing prime-editing efficiency[29]. Given p53's fundamental role in DDR, we examined whether *Dnmt1^{+GGG}* editing levels differed substantially across *Trp53^{+/+}* and *Trp53^{flox/flox}* conditions. In both TTFs and pancreatic organoids, we noticed a consistent twofold to threefold decrease in *Dnmt1^{+GGG}* editing in *Trp53^{+/+}* relative to *Trp53^{flox/flox}* tissues (Extended Data Fig. 3a,b). This result suggests that *Trp53* status may affect prime-editing efficiency, although we still observe highly efficient editing across loci in *Trp53*-proficient tissues.

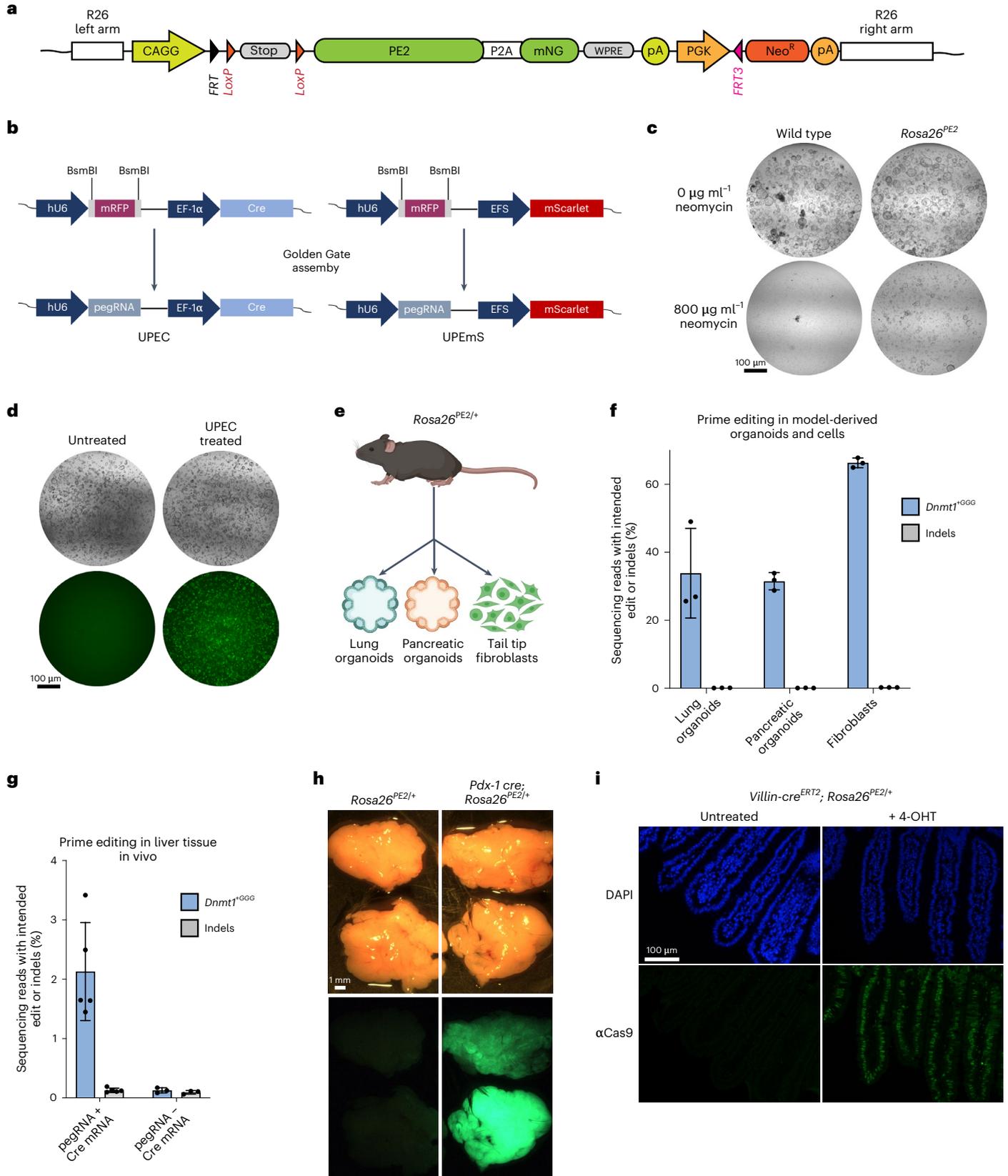## Prime editing in vivo with lipid nanoparticles

To determine whether PE GEMMs enable prime editing in vivo, we co-formulated Cre mRNA and a synthetic pegRNA encoding the *Dnmt1^{+GGG}* insertion within lipid nanoparticles (LNPs). We then treated *Rosa26^{PE2/+}* and *Rosa26^{PE2/PE2}* mice with one of two LNP formulations[53] (Methods) via tail vein injection. After 1 week, we observed Cre-induced fluorescence in the livers of mice that received pegRNA-bearing LNPs, but not in a control mouse that received PBS (Supplementary Fig. 4).

**Fig. 2 | Design and functional validation of the *Rosa26^{PE2}* prime editor allele. a**, Schematic depicting the design of the Cre-inducible *Rosa26^{PE2}* allele. **b**, Schematic depicting the formation of UPEC and UPEmS vectors from templates encoding an RFP by Golden Gate assembly. **c**, Bright-field images of pancreatic organoids derived from chimeric prime editor mice and wild-type mice with and without treatment with neomycin. This experiment was completed once. **d**, Bright-field and fluorescent images showing PE2-P2A-mNG expression only after exposure to Cre encoded by a UPEC vector. This experiment was completed more than five times with consistent results. **e**, Schematic depicting the derivation of multiple organoids and a fibroblast cell line from *Rosa26^{PE2/+}* prime editor mice. **f**, Editing efficiency of a trinucleotide (+GGG) insertion located 8 bp downstream of the start codon in *Dnmt1* in pancreatic organoids, lung organoids and TTFs. Unintended indel byproducts in all conditions were present in <1% of sequencing reads. Data and error bars indicate the mean and standard deviation of three independent transductions. **g**, Editing efficiency and indel byproduct frequency of *Dnmt1^{+GGG}* in liver tissue 1 week after tail vein injection with LNPs harboring either Cre mRNA and pegRNA (*n* = 5 mice) (left) or pegRNA alone (*n* = 3 mice) (right). Data and error bars indicate the mean and standard deviation of independent animals. **h**, Bright-field and fluorescent images of pancreases derived from *Rosa26^{PE2/+}* (left) or *Pdx-1 cre;Rosa26^{PE2/+}* mice (right). This experiment was completed twice with consistent results. **i**, Immunofluorescence imaging of intestinal tissue derived from *Villin-creER^{T2}*;*Rosa26^{PE2/+}* mice that were either untreated (left) or exposed to tamoxifen (right; 4-OHT). Tissue slides were stained with the DNA stain DAPI (4′,6-diamidino-2-phenylindole; top) or with an antibody specific to Cas9 (bottom). Scale bar indicates 100 μm. This experiment was completed once.

We also detected moderately efficient prime editing (up to 3.4%) at *Dnmt1* as assessed by bulk liver analysis, and we did not detect significant editing in mice that received LNPs harboring only the *Dnmt1* pegRNA (that is, lacking Cre mRNA) (Fig. 2g). These results confirm that PE GEMMs are amenable to precision edits in vivo.

## Generation of constitutive and inducible PE GEMMs
Prime editing in vivo could be more convenient if the need for Cre co-delivery was eliminated. To demonstrate the compatibility of the conditional PE2 allele with tissue-restricted Cre drivers, we generated additional PE GEMMs through genetic crosses with mice harboring



**a**

**b**

**c**
Wild type / *Rosa26*^PE2

0 µg ml⁻¹ neomycin

800 µg ml⁻¹ neomycin

100 µm

**d**
Untreated / UPEC treated

100 µm

**e**
*Rosa26*^PE2/+

Lung organoids / Pancreatic organoids / Tail tip fibroblasts

**f** Prime editing in model-derived organoids and cells

Sequencing reads with intended edit or indels (%)

Lung organoids / Pancreatic organoids / Fibroblasts

*Dnmt1*^+GGG / Indels

**g** Prime editing in liver tissue in vivo

Sequencing reads with intended edit or indels (%)

pegRNA + Cre mRNA / pegRNA − Cre mRNA

*Dnmt1*^+GGG / Indels

**h** *Rosa26*^PE2/+ / *Pdx-1 cre; Rosa26*^PE2/+

1 mm

**i** *Villin-cre*^ERT2; *Rosa26*^PE2/+
Untreated / + 4-OHT

DAPI

100 µm

αCas9

alleles that express Cre recombinase from endogenous loci. First, we crossed $Rosa26^{PE2/PE2}$ mice to $Pdx\text{-}1\,cre$[54], a pancreas-specific Cre driver allele, and $Villin\text{-}creER^{T2}$, an inducible, intestinal epithelial Cre driver allele[55]. As expected, $Pdx\text{-}1\,cre;Rosa26^{PE2/+}$ mice showed bright and robust evidence of mNG expression in the pancreas (Fig. 2h), and $Villin\text{-}creER^{T2};Rosa26^{PE2/+}$ mice demonstrated PE2 expression in intestinal epithelial cells upon treatment with tamoxifen (Fig. 2i). Notably, histologic analysis of the pancreas and intestinal epithelia, respectively, revealed no gross or pathologic abnormalities, suggesting that constitutive or inducible expression of the PE2 enzyme does not lead to toxicity in these tissues (Supplementary Fig. 4).

## Optimization of *Kras*-targeted pegRNAs

We next sought to empirically identify highly efficient pegRNAs that introduce the $Kras^{G12D}$ transition as an SNV (GGT > GAT). Based on previous study[56], we hypothesized that spacer sequences capable of producing the highest Cas9 indel efficiency in mouse N2A cells would serve as ideal scaffolds for high-efficiency pegRNA designs (Supplementary Fig. 5 and Supplementary Table 3). Using TTFs, we observed up to ~5% editing efficiency of $Kras^{G12D}$ with spacer-optimized pegRNAs (Fig. 3a and Supplementary Fig. 5). To further increase editing efficiency, we modified our best-performing pegRNA with a structured RNA pseudo-knot motif, prequeosine₁-1 riboswitch aptamer (tevopreQ1), recently shown to enhance prime-editing efficiency by more than threefold in cell lines[51]. This resulted in up to ~18.4% editing efficiency of $Kras^{G12D}$ in pancreatic organoids and TTFs (Fig. 3b). We then modified this epegRNA to template the $Kras^{G12C}$ transversion and observed ~0.5% editing efficiency in pancreatic organoids and ~5% in TTFs. We also generated $Kras^{G12A}$ and $Kras^{G12R}$ epegRNAs and observed up to ~30% editing efficiency with both epegRNAs in TTFs (Fig. 3b).

Both $Kras^{G12A}$ and $Kras^{G12R}$ epegRNAs template G·C-to-C·G substitutions, which proceed from C·C mismatch intermediates. These mismatches are not efficiently repaired by mismatch repair (MMR) and are thought to have higher basal prime-editing rates as a consequence[29]. A study by Chen et al. has indicated that co-installation of silent or benign MMR-evasive edits can promote higher prime-editing efficiency, consistent with the increased editing efficiency in producing $Kras^{G12A}$ and $Kras^{G12R}$ over $Kras^{G12C}$ (ref. 29). To further probe this phenomenon, we compared a variety of epegRNAs templating cancer-associated mutations across *Kras*, *Trp53* and *Egfr* to counterparts modified with silent or inconsequential edits. In nearly every case, we found that installing MMR-evasive edits amplified prime-editing efficiencies by more than threefold, often resulting in efficiencies greater than 20% (Extended Data Fig. 4b–d). Collectively, these data demonstrate that the $Rosa26^{PE2}$ allele enables efficient installation of SNVs, multinucleotide alterations and insertions and deletions across a diverse array of cell lines and organoids.

To confirm the functional effects of these mutations, we installed either $Kras^{G12D}$ or $Kras^{G12C}$ mutations in $Trp53^{flox/flox};Rosa26^{PE2/+}$ pancreatic organoids and selected transduced cells with nutlin-3a. We then treated prime-edited organoids with the epidermal growth factor receptor (EGFR) inhibitor, gefitinib, to select for the oncogenic *Kras* mutation[57] and evaluated the fraction of cells containing the intended edits before and after treatment. Consistent with receptor-independent signaling downstream of EGFR, only $Kras^{G12D}$ and $Kras^{G12C}$ prime-edited cells survived treatment with gefitinib, while control cells infected with the template UPEC lacking a pegRNA did not (Fig. 3c and Extended Data Fig. 5a,b). We then tested whether cells transduced with $Kras^{G12C}$ epegRNAs were sensitive to sotorasib, a $KRAS^{G12C}$-specific inhibitor, alone or in combination with gefitinib. Consistent with the previous study[58], we found that $Kras^{G12C}$ pancreatic organoids were uniquely sensitive to the combination of sotorasib and gefitinib, while $Kras^{G12D}$ organoids were unaffected by these treatments (Fig. 3d and Extended Data Fig. 5b). While $KRAS^{G12C}$ inhibition has shown promising signs of clinical efficacy in pancreatic cancer[5,59], current preclinical efforts focused on $KRAS^{G12D}$ inhibition have the potential to benefit a broader fraction of patients with this disease (>38%)[60,61]. Therefore, we treated prime-edited $Kras^{G12D}$ pancreatic organoids with MRTX1133 (ref. 62), a $KRAS^{G12D}$ inhibitor, alone or in combination with gefitinib. Consistent with results using sotorasib, we found that $Kras^{G12D}$ organoids were substantially more sensitive to the combination treatment compared with MRTX1133 alone (Fig. 3d,e), suggesting that concomitant EGFR inhibition may be a broadly effective strategy to augment the overall efficacy of KRAS mutant inhibitors in pancreatic cancer cells.

## Rapid interrogation of resistance mutations

While targeted therapies have revolutionized modern cancer treatment, therapy resistance is common and frequently arises through the acquisition of secondary missense mutations affecting the drugged driver[28,63,64]. A recent study in ref. 63 revealed a class of secondary *KRAS* mutations occurring in over 10% of patients with non-small cell lung cancer and colorectal cancer with acquired resistance to adagrasib, a $KRAS^{G12C}$ inhibitor. Intriguingly, several mutations occur in codons 95–96, which occupy the switch II pocket targeted by adagrasib and sotorasib.

To test the utility of the $Rosa26^{PE2}$ model to functionally interrogate mutations associated with resistance, we developed an epegRNA designed to introduce the $Kras^{Y96C}$ transversion and tested its capacity to promote resistance in prime-edited $Kras^{G12C}$ pancreatic organoids treated with gefitinib and sotorasib (Extended Data Fig. 5d). All organoids were initially treated with both inhibitors for two passages, followed by continued treatment for three additional passages in one group (continuous treatment) and treatment removal in the second group (transient treatment). Consistent with patient data[63], organoids transduced with the $Kras^{Y96C}$ epegRNA were resistant to combined treatment with gefitinib and sotorasib and exhibited increased allele frequency of the $Kras^{Y96C}$ mutation over time (Fig. 3f). Positive selection for composite $Kras^{G12C;Y96C}$ mutant organoids was not observed in
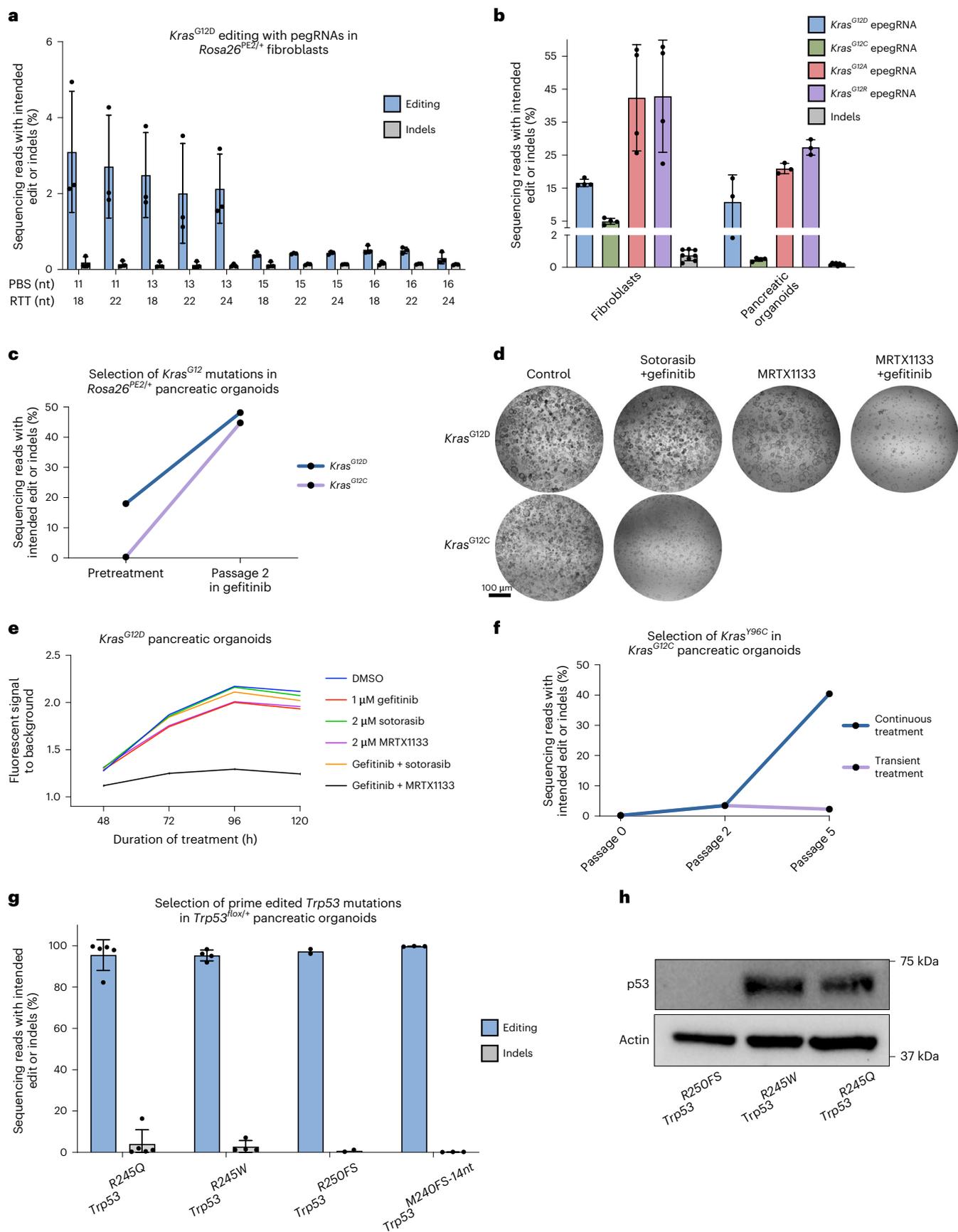
**Fig. 3 | Ex vivo prime editing and functional testing of *Kras* and *Trp53* mutations. a**, Editing efficiency and indel byproduct frequency of the $Kras^{G12D}$ transition mutation templated by pegRNAs based on a single Cas9 spacer ($n = 3$ for each pegRNA). pegRNAs are delineated by differences in the lengths of the primer binding site (PBS) and RTT. Data and error bars indicate the mean and standard deviation of three independent transductions. **b**, Editing activity of four epegRNAs templating $Kras^{G12}$ mutations in TTFs or pancreatic organoids. Data and error bars indicate the mean and standard deviation of three independent transductions. Indel byproduct calculations were pooled from all conditions within each tissue. **c**, Allele frequencies of $Kras^{G12D}$ or $Kras^{G12C}$ mutations in pancreatic organoids before and after two passages of treatment with gefitinib (1 μM; $n = 1$). **d**, Bright-field images of prime-edited $Kras^{G12C}$ or $Kras^{G12D}$ organoids treated for 4 d with either control DMSO, sotorasib (2 μM) and gefitinib (1 μM), MRTX1133 (5 μM) or MRTX1133 and gefitinib. This experiment was repeated three times with consistent results. **e**, Viability of $Kras^{G12D}$ pancreatic organoids under various treatment conditions. **f**, Allele frequency of $Kras^{Y96C}$ in $Kras^{G12C}$ organoids during and after treatment with sotorasib (2 μM) and gefitinib ($n = 1$). After two passages, organoids were split into two groups, which included continued treatment (continuous treatment) in one group and removal of treatment in a second group (transient treatment). **g**, Allele and indel byproduct frequencies of $Trp53^{R245Q}$ ($n = 5$), $Trp53^{R245W}$ ($n = 4$), $Trp53^{R250FS}$ ($n = 2$) and $Trp53^{M240FS\text{-}14nt}$ ($n = 3$) in $Trp53^{flox/+}$ pancreatic organoids treated with nutlin-3a for three to five passages after transduction with UPEC vectors. Note that the highest indel frequency depicted for $Trp53^{R245W}$ derives primarily from a scaffold insertion in a single replicate. $Trp53^{R250FS}$ denotes a dinucleotide deletion. $Trp53^{M240FS\text{-}14nt}$ denotes a fourteen-nucleotide deletion. Data and error bars indicate the mean and standard deviation across three or more independent transductions. **h**, Immunoblot indicating detectable levels of p53 protein in prime-edited $Trp53^{flox/R245Q}$ and $Trp53^{flox/R245W}$ organoids and an absence of detectable protein in $Trp53^{flox/R250FS}$ organoids.

organoids following the removal of gefitinib and sotorasib, confirming the requirement of the selective pressure exerted by the treatment. Although initially discovered in patients with lung cancer treated with sotorasib monotherapy, these data indicate that secondary *KRAS* mutations can also confer therapy resistance in other tissues and combination treatment contexts. The above results demonstrate



**a** *Kras*^G12D^ editing with pegRNAs in *Rosa26*^PE2/+^ fibroblasts

**b** *Kras*^G12D^ epegRNA / *Kras*^G12C^ epegRNA / *Kras*^G12A^ epegRNA / *Kras*^G12R^ epegRNA / Indels (Fibroblasts, Pancreatic organoids)

**c** Selection of *Kras*^G12^ mutations in *Rosa26*^PE2/+^ pancreatic organoids

**d** Control / Sotorasib +gefitinib / MRTX1133 / MRTX1133 +gefitinib (*Kras*^G12D^, *Kras*^G12C^) — 100 µm

**e** *Kras*^G12D^ pancreatic organoids — DMSO / 1 µM gefitinib / 2 µM sotorasib / 2 µM MRTX1133 / Gefitinib + sotorasib / Gefitinib + MRTX1133

**f** Selection of *Kras*^Y96C^ in *Kras*^G12C^ pancreatic organoids — Continuous treatment / Transient treatment

**g** Selection of prime edited *Trp53* mutations in *Trp53*^flox/+^ pancreatic organoids — Editing / Indels

**h** p53 / Actin — 75 kDa / 37 kDa — *Trp53*^R250FS^, *Trp53*^R245W^, *Trp53*^R245Q^

that the $Rosa26^{PE2}$ allele can be harnessed for rapid preclinical evaluation of emerging mechanisms of resistance to targeted therapies in tissues of interest and, ultimately, for testing second-generation therapies designed to overcome resistance.

## Engineering of common p53 mutations with prime editing

A key advantage of PE GEMMs is the ability to mediate nearly any codon substitution in accessible tissues, enabling tissue-specific functional studies of genetic variants with putative effects on tumor progression. *TP53* is the most frequently mutated gene in human cancer and is often altered via missense mutations that can confer gain-of-function properties in certain contexts[65]. In an analysis of data from cBioPortal[66,67], we found that some of the most frequent p53 amino acid substitutions observed in lung ($TP53^{R158L}$ and $TP53^{R270L}$) and pancreatic adenocarcinomas ($TP53^{R248W}$ and $TP53^{R248Q}$) have not been targeted to the endogenous *Trp53* locus in mouse models (Supplementary Fig. 6), despite having putative gain-of-function effects[68–70]. Notably, three of these mutations are transversions that cannot be modeled using base editing, and the human amino acid (p53$^{R248}$), but not the human codon (CGG versus CGC), is conserved in mouse *Trp53*. Therefore, engineering the $Trp53^{R245W}$ mutation in mice requires a dinucleotide substitution uniquely suitable to prime editing (Supplementary Fig. 6). We developed a suite of epegRNAs designed to introduce both $Trp53^{R245W}$ and $Trp53^{R245Q}$ and two truncating deletions, $Trp53^{R250FS}$ and $Trp53^{M240FS-14nt}$, using a $Trp53^{+/+}$ cell line derived from mouse 3TZ cells (Supplementary Fig. 7). After selection with nutlin-3a, most $Trp53^{flox/+}$;$Rosa26^{PE2/+}$ pancreatic organoids transduced with each of these epegRNAs exhibited a prime-edited allele frequency near 100% (Fig. 3g).

We also observed an average of >90% editing purity in these organoids (Fig. 3g and Extended Data Fig. 6c). Western blots confirmed that $Trp53^{R245Q}$ and $Trp53^{R245W}$ cells retained p53 protein expression, while $Trp53^{R250FS}$ cells did not (Fig. 3h). While the ratio of prime-edited reads to random indel-bearing reads was typically high, we did observe a variable unintended single nucleotide substitution (0.24%–11.34% of reads) attributable to partial RT of the scaffold sequence when prime editing $Trp53^{R245Q}$ (Supplementary Fig. 6). In one instance, we also observed an insertion of the scaffold sequence when prime editing $Trp53^{R245W}$ (~7% of reads). Notably, we did not observe any of these unintended events with an epegRNA templating $Trp53^{M240FS-14nt}$, which was designed to evade MMR and exhibited a high basal editing efficiency (Extended Data Fig. 6a,c).

In all cases, we observed negligible off-target activity at computationally predicted loci, even after more than 4 weeks of culturing organoids with sustained pegRNA expression (Extended Data Fig. 6d–f). This result is most striking for $Trp53^{R245W}$, which is templated by a pegRNA bearing a protospacer that shares 100% sequence homology with an off-target locus on chromosome 17 (Supplementary Table 4). We detected an average of 0.002% editing at this locus, which

was substantially greater than the sequencing error rate in control samples (Extended Data Fig. 6f). No other loci displayed editing levels higher than those observed in controls. Collectively, these results establish the utility of our approach for high-fidelity installation of mutations using systems that can be rationally engineered and easily translated to an in vivo setting.

## Modeling lung and pancreatic adenocarcinomas in vivo

To benchmark the utility of PE GEMMs to model cancer in vivo, we initiated lung and pancreatic adenocarcinomas using autochthonous and orthotopic transplantation strategies (Fig. 4a). To model lung cancer, we intratracheally transduced the lungs of $Trp53^{flox/flox}$;$Rosa26^{PE2/+}$ and $Trp53^{flox/flox}$;$Rosa26^{PE2/PE2}$ mice with UPEC lentiviruses encoding the template vector ($n = 4$) or pegRNAs for $Kras^{G12D}$ ($n = 20$), $Kras^{G12R}$ ($n = 9$), $Kras^{G12A}$ ($n = 10$), $Kras^{G12C}$ ($n = 13$) or the neutral $Dnmt1^{+GGG}$ ($n = 6$). We also infected $Trp53^{+/+}$;$Rosa26^{PE2/+}$ mice with UPEC-$Kras^{G12D}$ to model low-grade lesions and assess in vivo prime editing in a *Trp53*-proficient setting.

In $Trp53^{flox/flox}$ recipients, tumors initiated by UPEC-$Kras^{G12D}$ were readily visible by μCT at 14 weeks postinjection (Supplementary Fig. 8). At 16 weeks, we observed multifocal fluorescent lesions in 16 of 20 (80%) UPEC-$Kras^{G12D}$ recipients and in none of the controls (Fig. 4b). Histopathological analysis confirmed that lesions induced by prime editing recapitulated the full spectrum of lung cancer progression, from grade 1 atypical adenomatous hyperplasia through grade 4 adenocarcinoma. By immunohistochemistry, prime-edited tumors recapitulated the cellular and molecular evolution seen in the classical $Kras^{LSL-G12D/+}$;$Trp53^{flox/flox}$ (KP) GEMM model, demonstrating the downregulation of lung lineage transcription factor *Nkx2-1* and the expression of chromatin regulator *Hmga2* in poorly differentiated, advanced lesions[71–73] (Fig. 4c,d and Extended Data Fig. 7b,c).

We confirmed that tumors were initiated through on-target prime editing by sequencing genomic DNA derived from several bulk tumors (Fig. 4e). Prime editing in vivo did not require a loss of p53, as 2 of 3 $Trp53^{+/+}$;$Rosa26^{PE2/+}$ mice developed fluorescent tumors upon infection with UPEC-$Kras^{G12D}$, consistent with previous studies demonstrating that oncogenic *Kras* is sufficient to drive lung adenoma formation in vivo[74] (Extended Data Fig. 8e,g). These adenomas also harbored the intended $Kras^{G12D}$ mutation.
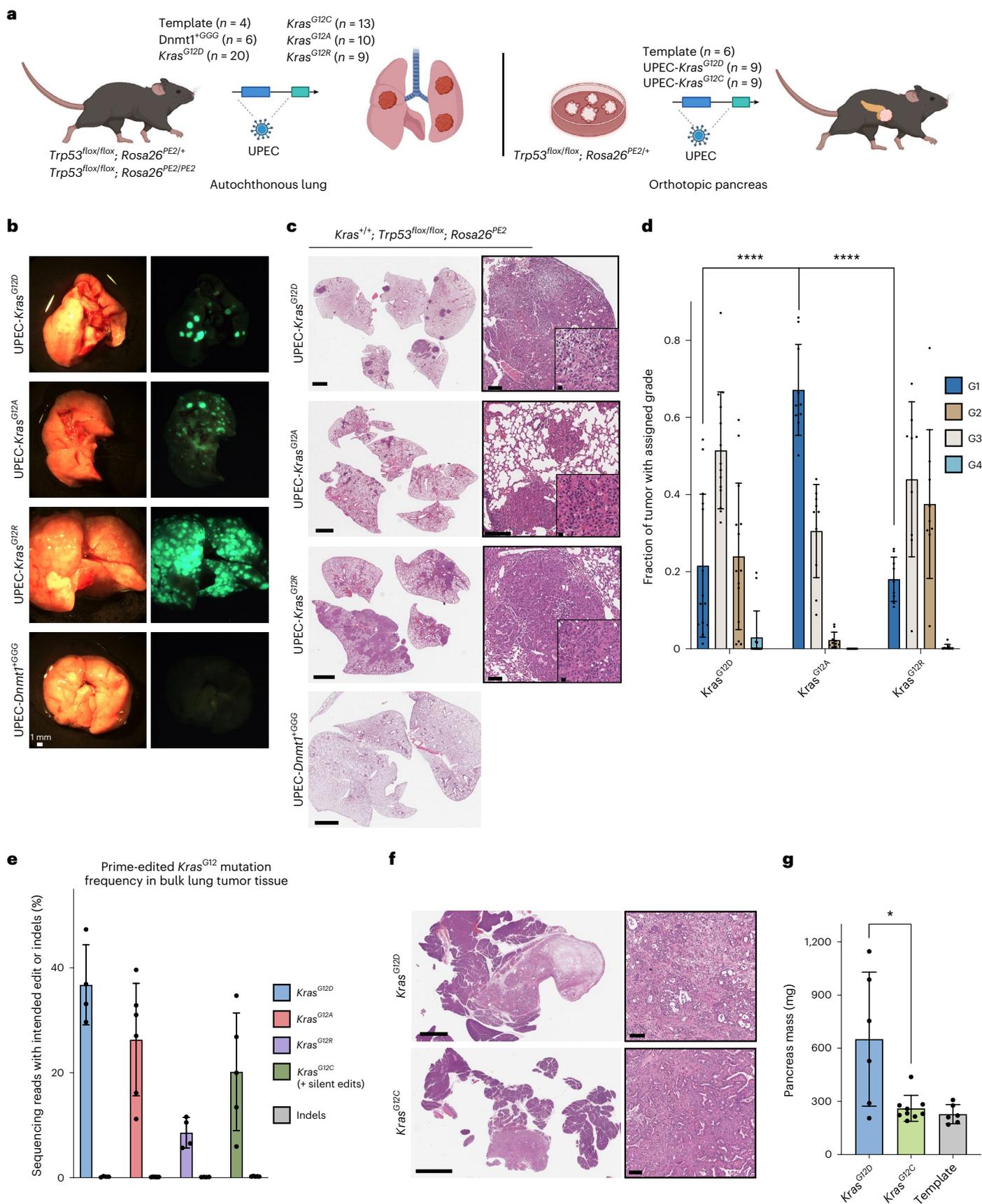
Similar to UPEC-$Kras^{G12D}$ recipients, UPEC-$Kras^{G12A}$ and UPEC-$Kras^{G12R}$ recipients consistently presented multifocal fluorescent lesions driven by on-target prime editing throughout the lung (Fig. 4c–e). However, both UPEC-$Kras^{G12A}$ and UPEC-$Kras^{G12R}$ recipients presented with greater tumor numbers than UPEC-$Kras^{G12D}$ recipients (Fig. 4b and Extended Data Fig. 8b). While this is likely attributable in part to more efficient editing with the $Kras^{G12A}$ and $Kras^{G12R}$ epegRNAs (Fig. 3b), there were also discernible differences in the apparent oncogenic capacity of these mutations. In 8 of 9 UPEC-$Kras^{G12R}$ recipients,

**Fig. 4 | PE GEMMs enable autochthonous and orthotopic modeling of lung and pancreatic cancer. a**, Schematic depicting the design of in vivo experiments. Lung tumors were initiated with lentivirus-encoding UPEC vectors. Pancreatic tumors were initiated by orthotopic transplantation of prime-edited pancreatic organoids. 'Template' refers to the template UPEC vector lacking a pegRNA. **b**, Representative bright-field and fluorescence images of lungs derived from mice infected with the UPEC vector encoding the neutral $Dnmt1^{+GGG}$ pegRNA, $Kras^{G12D}$, $Kras^{G12A}$ or $Kras^{G12R}$ epegRNAs described in Fig. 3b. $Kras^{G12D}$ modeling was performed twice with consistent results. $Kras^{G12A}$ and $Kras^{G12R}$ modeling was performed once, and replicates were consistent with representative images shown. **c**, H&E staining of representative tissue from a control mouse infected with UPEC-$Dnmt1^{+GGG}$ (bottom), and tumor-bearing mice infected with UPEC-$Kras^{G12D}$, UPEC-$Kras^{G12A}$ and UPEC-$Kras^{G12R}$ (top). Scale bars from left to right indicate 2 mm, 100 μm and 20 μm, respectively. **d**, Bar charts indicate the distribution of grades across 16-week lesions from UPEC-$Kras^{G12D}$ ($n = 14$ mice), UPEC-$Kras^{G12A}$ ($n = 10$ mice) or UPEC-$Kras^{G12R}$ ($n = 9$ mice). Data and

error bars indicate the mean and standard deviation of all biological replicates in each condition. Statistical significance was calculated using unpaired, two-tailed t-tests comparing the fraction of grade 1 lesions in $Kras^{G12A}$-driven tumor tissue to $Kras^{G12D}$-driven tumor tissue ($P < 0.0001$) or $Kras^{G12R}$-driven tumor tissue ($P < 0.0001$). ****$P < 0.0001$. **e**, Allele frequencies of $Kras^{G12D}$ ($n = 4$ mice), $Kras^{G12A}$ ($n = 6$ mice), $Kras^{G12R}$ ($n = 4$ mice) (16 weeks) and $Kras^{G12C + silent edits}$ ($n = 5$ mice) (12 weeks) and indel byproducts in bulk lung tumors. Data and error bars indicate the mean and standard deviation across tumors from independent mice. **f**, H&E staining of representative pancreatic adenocarcinomas from a mouse transplanted with $Kras^{G12D}$ organoids (top) and a mouse transplanted with $Kras^{G12C}$ organoids (bottom). Scale bars from left to right indicate 2 mm and 25 μm, respectively. **g**, Mass of pancreata of $Kras^{G12D}$ ($n = 6$ mice), $Kras^{G12C}$ ($n = 9$ mice) or UPEC-template ($n = 6$ mice) organoid transplant recipients measured in milligrams ($n = 6$–9 mice). Data and error bars indicate the mean and standard deviation across tumors from independent mice. Statistical significance was calculated using a two-tailed Mann–Whitney U test ($P = 0.036$). *$P < 0.05$.

the overall tumor burden was substantially higher than the $Kras^{G12A}$ setting (Extended Data Fig. 8a). Furthermore, histopathological analysis revealed that $Kras^{G12R}$ and $Kras^{G12D}$ tumors were of consistently higher grades relative to $Kras^{G12A}$ lesions (Fig. 4d). This is particularly striking given the relative rarity of $KRAS^{G12R}$ in patients with lung cancer (<1% of $KRAS$ mutations; Discussion), although, of note, our data are consistent



**a**

Template (n = 4)    $Kras^{G12C}$ (n = 13)
Dnmt1$^{+GGG}$ (n = 6)    $Kras^{G12A}$ (n = 10)
$Kras^{G12D}$ (n = 20)    $Kras^{G12R}$ (n = 9)

$Trp53^{flox/flox}$; $Rosa26^{PE2/+}$
$Trp53^{flox/flox}$; $Rosa26^{PE2/PE2}$

UPEC

Autochthonous lung

Template (n = 6)
UPEC-$Kras^{G12D}$ (n = 9)
UPEC-$Kras^{G12C}$ (n = 9)

$Trp53^{flox/flox}$; $Rosa26^{PE2/+}$

UPEC

Orthotopic pancreas

**b**

UPEC-$Kras^{G12D}$
UPEC-$Kras^{G12A}$
UPEC-$Kras^{G12R}$
UPEC-$Dnmt1^{+GGG}$

1 mm

**c** $Kras^{+/+}$; $Trp53^{flox/flox}$; $Rosa26^{PE2}$

UPEC-$Kras^{G12D}$
UPEC-$Kras^{G12A}$
UPEC-$Kras^{G12R}$
UPEC-$Dnmt1^{+GGG}$

**d**

**** ****

Fraction of tumor with assigned grade

$Kras^{G12D}$    $Kras^{G12A}$    $Kras^{G12R}$

G1
G2
G3
G4

**e**

Prime-edited $Kras^{G12}$ mutation frequency in bulk lung tumor tissue

Sequencing reads with intended edit or indels (%)

$Kras^{G12D}$
$Kras^{G12A}$
$Kras^{G12R}$
$Kras^{G12C}$ (+ silent edits)
Indels

**f**

$Kras^{G12D}$
$Kras^{G12C}$

**g**

Pancreas mass (mg)

*

$Kras^{G12D}$    $Kras^{G12C}$    Template

with previous study demonstrating that $Kras^{G12R}$ is highly oncogenic in mouse models[15]. Taken together, these results highlight significant allele-specific differences in the oncogenic capacity of different $Kras$ mutations and showcase the utility of PE GEMMs for rapidly discovering such phenotypes.

In contrast to other $Kras$ mutations, only 4 of 13 (31%) UPEC-$Kras^{G12C}$ recipients presented tumors when collected at 19 weeks, likely a consequence of the lower prime-editing efficiency of the $Kras^{G12C}$ epegRNA. Furthermore, deep amplicon sequencing of these tumors occasionally revealed unintentional edits, including an additional silent substitution in codon 11 in one case (Extended Data Fig. 8d). To address this shortcoming, we designed an improved $Kras^{G12C}$ epegRNA encoding MMR-evasive substitutions, which edits at a 3.2-fold higher efficiency (Extended Data Fig. 4b). At 12 weeks, 8 of 9 $Trp53^{flox/flox};Rosa26^{PE2/PE2}$ mice infected with this epegRNA developed multifocal tumor burden (Extended Data Fig. 8b,c). Targeted sequencing confirmed the presence of the multinucleotide substitution encoding $Kras^{G12C}$, without any unintended byproducts (Fig. 4e).

To further test the potential of PE GEMMs for cancer modeling in vivo, we transplanted prime-edited $Kras^{G12D/+};Trp53^{flox/flox};Rosa26^{PE2/+}$ and $Kras^{G12C/+};Trp53^{flox/flox};Rosa26^{PE2/+}$ pancreatic organoids into immunocompetent mice harboring the $Rosa26^{PE2}$ allele (to ensure immunological tolerance[75] to the prime editor enzyme). As controls, we transplanted $Trp53^{flox/flox};Rosa26^{PE2/+}$ organoids infected with the template UPEC vector. Tumors were visible via ultrasound by 5 weeks (Supplementary Fig. 8), and fluorescent tumors that reflected the spectrum of pancreatic neoplasia were observed in 8 of 9 $Kras^{G12D/+}$ recipients by 9 weeks post-transplantation (Fig. 4f and Extended Data Fig. 9a). Notably, only 4 of 9 mice (44%) from the cohort of animals transplanted with $Kras^{G12C/+}$ pancreatic organoids developed lesions. Of the remaining five mice, one developed a high-grade PanIN, while the rest did not develop any lesions. Tumor burden in $Kras^{G12C}$ mice was substantially lower than in $Kras^{G12D}$ mice, as reflected in pancreatic weight measurements (Fig. 4g). These results are consistent with previous observations, suggesting that $Kras^{G12C}$ may be less tumorigenic in the pancreas[58]. Metastases were only observed in $Kras^{G12D}$ recipients (Extended Data Fig. 9a), indicative of a more aggressive phenotype of these tumors. We did not observe tumor formation in control recipients by ultrasound, microscopy or histology, consistent with previous study showing that $Trp53$ knockout alone is insufficient for pancreatic tumorigenesis[47,76].

To model autochthonous pancreatic adenocarcinoma, we adapted a strategy of retrograde pancreatic duct viral delivery[47,77]. We infected $Trp53^{flox/flox};Rosa26^{PE2/+}$ mice with UPEC vectors encoding either $Kras^{G12D}$ or $Kras^{Y96C}$ as a control. Notably, 3 of 4 $Kras^{G12D}$-infected animals developed pancreatic adenocarcinoma, while no tumors were detected in $Kras^{Y96C}$-infected animals (Extended Data Fig. 9e).

## Discussion

Advances in genome editing technologies have accelerated functional genetic studies, yet most approaches to model cancer mutations have relied on Cas9-mediated gene disruption via non-homologous end joining, failing to recapitulate many genetic lesions observed in human cancer. Emerging precision genome editing technologies like base editing and prime editing are poised to fill this gap by allowing the engineering of specific cancer-associated mutations. Nevertheless, the considerable size of base editors and prime editors makes delivery to most tissues and cell types challenging, posing significant limitations for in vivo studies. Previous studies have addressed this using split-prime editor systems that enable prime editing in vivo when delivered by dual adeno-associated virus (AAV) vectors. However, dual-AAV approaches remain hampered by delivery challenges to many tissues and, notably, they can elicit an immune response against the prime editor enzyme[34,78]. The immunogenicity of genome editing reagents delivered exogenously substantially complicates cancer modeling experiments. With these challenges in mind, we developed a PE GEMM capable of rapidly installing a variety of genetic lesions with single nucleotide precision across in vitro, ex vivo or in vivo contexts, as well as in an autochthonous, immunocompetent setting. By expressing the PE2 enzyme endogenously, we bypass the risk of a confounding immune response and substantially expand the capacity to deliver other functional cargo, such as Cre.

We used this model to install a variety of cancer-associated mutations, including transversions, transitions, multinucleotide substitutions and deletions across $Trp53$, $Egfr$ and $Kras$. In the context of our pancreatic orthotopic transplant experiments, we observed that different $Kras$ mutations exhibit variable in vivo tumor-initiating potential, consistent with previous study comparing $Kras^{G12C}$ and $Kras^{G12D}$ autochthonous models in the pancreas[58]. In the lung, we found that $Kras^{G12A}$, $Kras^{G12D}$ and $Kras^{G12R}$ promote efficient but variable tumor formation. Tumor burden differences across genotypes are likely driven in part by variable pegRNA efficiencies, yet we also observed significant differences in the phenotype and grade of tumors when using rationally optimized pegRNAs. For example, $Kras^{G12A}$-driven tumors exhibited a less advanced, more differentiated histopathology than $Kras^{G12R}$ and $Kras^{G12D}$.

The significant tumor-initiating potential of $Kras^{G12R}$ is notable, given the rarity of $KRAS^{G12R}$ in patients with non-small cell lung cancer[61], but is consistent with previous results from ref. [15]. Intriguingly, KRAS$^{G12R}$ is known to have substantially impaired GTP hydrolysis relative to other KRAS$^{G12}$ mutants[79]. This property could enhance oncogenicity, yet $KRAS^{G12R}$ is found at low frequency in most solid tumor types, except pancreatic cancer[61]. In pancreatic models, Zafra et al. previously found that $Kras^{G12R}$ mutations exhibit little to no PanIN formation potential in $Trp53^{+/+}$ mice constitutively expressing $Kras^{G12R}$ in the pancreas, while $Kras^{G12D}$ promoted significant PanIN formation in most of the entire organ[58]. In contrast, transplanted $Kras^{G12R};Trp53^{flox/flox}$ organoids generated tumors with efficiency similar to $Kras^{G12D};Trp53^{flox/flox}$ organoids. These findings and our study suggest that mutation-specific properties may subject KRAS$^{G12R}$ to especially potent tumor suppressive mechanisms that are lost in the context of concomitant $Trp53$ knockout specific to the mouse experiments described. This warrants further investigation in the context of other genotypes (for example, $Trp53^{+/+}$) and experiments in which the sequence of mutations is temporally controlled.

We also observed $Kras$ allele-specific responses to mutant-specific targeted therapies. For example, similar to previous studies of KRAS$^{G12C}$ inhibitors[58,80], we found that a KRAS$^{G12D}$ inhibitor, MRTX1133, elicits a more powerful effect on prime-edited $Kras^{G12D}$ pancreatic organoids when combined with the EGFR inhibitor, gefitinib. Several other clinical agents targeting a broader spectrum of oncogene mutations are undergoing clinical evaluation, and sotorasib and adagrasib, two KRAS$^{G12C}$ inhibitors, have now been approved by the Food and Drug Administration[60,62]. PE GEMMs represent ideal systems for rapid interrogation of the effects of targeted therapies in the context of virtually any oncogenic mutation, including secondary resistance mutations like $KRAS^{Y96C}$ that are now being identified in patients. PE GEMMs also enable in vivo interrogation of these mutations in the context of syngeneic and immunocompetent mice. This broad utility for modeling $Kras$ mutations in vivo is critical, as mutant KRAS inhibition has been shown to impact the tumor-immune microenvironment in models of colon cancer[81,82] and may synergize with immune checkpoint blockade in other tissues not yet examined.

Beyond $KRAS$, we demonstrate in pancreatic organoids the precise installation and selection of two $Trp53$ dinucleotide substitutions encoding two mutant amino acid residues frequently observed at the same codon in human pancreatic cancer, as well as out-of-frame multinucleotide deletions at a nearby codon. We observed over 90% editing purity after the selection of all these mutations in vitro. Despite

a high intended edit-to-unintended indel ratio, we also observed an unintended single nucleotide substitution at variable frequency when prime editing *Trp53^R245Q* (Supplementary Fig. 6). We attribute this event to partial homology between the genomic region immediately following the RTT and the few nucleotides in the pegRNA scaffold that are commonly reverse-transcribed and excised during DNA repair, a prime-editing intermediate noted by ref. 51. Such unintended edits could be avoided by using an alternative pegRNA with an RTT ending a few nucleotides up or downstream to eliminate the homology or could be reduced by introducing silent edits that prevent repeated editing of the same target site, as we demonstrated with the epegRNA encoding *Trp53^M240FS-14nt*. This pegRNA is based on the same protospacer as *Trp53^R245Q*, yet has a longer RTT and encodes a deletion that eliminates both the seed and PAM sequences. However, this phenomenon merits additional caution during pegRNA design and may be exacerbated in long-term prime-editing experiments, such as when selecting cell lines over several passages with continuous expression of the prime editor and pegRNA.

The overall editing purity highlights the utility of prime editing for precise engineering of mutations with negligible indel byproducts. This is a key advantage over Cas9 HDR-based approaches, in which the high rate of indel byproducts could dilute intended point mutations in vitro and in vivo. Low editing purity could especially limit the study of specific point mutations in tumor suppressor genes, as unintended indels in these genes can produce frameshift mutations subject to positive selection. This limitation is especially important when considering that many genes, including *TP53*, often harbor point mutations that confer different properties relative to loss-of-function truncations, including gain-of-function effects[68,83–85]. For instance, Schulz-Heddergott et al. demonstrated that *TP53^R248Q* exhibits a gain-of-function effect by hyperactivating the JAK2/STAT3 pathway, leading to more aggressive tumor progression in models of colon cancer[68]. These observations remain largely untested in models of pancreatic cancer in vivo due to a lack of suitable transgenic mouse models and human cell lines[84]. PE GEMMs are poised to fill critical gaps like this by allowing rapid and fine-tuned mutation control in a variety of tissue settings.

Although we did not explore them here, a variety of techniques are available to optimize prime-editing efficiency, such as PE3 and PE3b editing strategies that combine nicking guides to bias DNA repair toward the incorporation of prime-edited nucleotides. Nevertheless, strategies based on single pegRNAs are more straightforward, have better multiplexing capacity because they rarely cause indels and are better suited for high-throughput studies like genetic screens. In general, we found that spacer optimization and testing of up to 15 guides were sufficient to identify epegRNAs suitable for our experiments. We also found that silent or benign MMR-evasive edits close to the intended mutation reliably amplify prime-editing efficiency by several fold, even for epegRNAs with optimized spacer sequences and PBS and RTT lengths. These techniques enabled us to identify epegRNAs that edit with greater than 20% efficiency across several cancer-associated genes. Future users should consider these and other strategies, including the co-delivery of an MLH1 dominant negative gene (PE4/5)[29] or sensor-based pegRNA library approaches[35], to maximize overall prime-editing efficiencies, which may be especially helpful for in vivo applications.

We generally observed negligible off-target activity at computationally predicted loci, including one example with a protospacer identical to the intended target. This result corroborates the high on-target fidelity of prime editing. As established in previous studies[24,29,51], additional homology required for repair using the RT product limits activity at off-target loci. While our results are consistent with previous literature, future studies could employ whole-genome sequencing to fully characterize off-target prime editing beyond a limited number of prioritized loci.

While we focused on installing somatic cancer driver mutations, we anticipate that PE GEMMs could be employed for broader applications. In principle, germline *Rosa26^PE2* alleles could be used to construct heritable mutations by modifying zygotes with pegRNAs encoding known drivers of inherited disease. We also envision sophisticated tumor modeling with the insertion of custom neoepitopes and other functional genetic sequences. These applications would enable investigators to address key questions in cancer genetics, immunology and diverse genetic diseases while reducing the need to generate, genotype and otherwise maintain traditional GEMMs. Finally, the combination of multiple epegRNAs in the context of a modified UPEC vector or LNP formulation should enable autochthonous generation of tumors defined by custom sets of multiple driver mutations in wild-type prime editor mice. This would enable increasingly complex studies of cooperating driver mutations. With these capabilities, PE GEMMs can provide a rapid preclinical avenue to complement both fundamental and clinical investigations aimed at treating cancer with precision treatment paradigms.

## Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-023-01783-y.

## References

1. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
2. Garraway, L. A. & Lander, E. S. Lessons from the cancer genome. *Cell* **153**, 17–37 (2013).
3. Hyman, D. M., Taylor, B. S. & Baselga, J. Implementing genome-driven oncology. *Cell* **168**, 584–599 (2017).
4. Chang, M. T. et al. Accelerating discovery of functional mutant alleles in cancer. *Cancer Discov.* **8**, 174–183 (2018).
5. Hong, D. S. et al. KRASG12C inhibition with sotorasib in advanced solid tumors. *N. Engl. J. Med.* **383**, 1207–1217 (2020).
6. Lynch, T. J. et al. Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
7. Hill, W., Caswell, D. R. & Swanton, C. Capturing cancer evolution using genetically engineered mouse models (GEMMs). *Trends Cell Biol.* **31**, 1007–1018 (2021).
8. Kersten, K., de Visser, K. E., van Miltenburg, M. H. & Jonkers, J. Genetically engineered mouse models in oncology research and cancer medicine. *EMBO Mol. Med.* **9**, 137–153 (2017).
9. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
10. Platt, R. J. et al. CRISPR–Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014).
11. Annunziato, S. et al. In situ CRISPR–Cas9 base editing for the development of genetically engineered mouse models of breast cancer. *EMBO J.* **39**, e102169 (2020).
12. Han, T. et al. R-Spondin chromosome rearrangements drive Wnt-dependent tumour initiation and maintenance in the intestine. *Nat. Commun.* **8**, 15945 (2017).
13. Dow, L. E. et al. Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**, 390–394 (2015).
14. Sánchez-Rivera, F. J. et al. Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* **516**, 428–431 (2014).
15. Winters, I. P. et al. Multiplexed in vivo homology-directed repair and tumor barcoding enables parallel quantification of Kras variant oncogenicity. *Nat. Commun.* **8**, 2053 (2017).

16. Leibowitz, M. L. et al. Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat. Genet.* **53**, 895–905 (2021).

17. Blair, L. M. et al. Oncogenic context shapes the fitness landscape of tumor suppression. Preprint at *bioRxiv* https://doi.org/10.1101/2022.10.24.511787 (2022).

18. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).

19. Kurt, I. C. et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2021).

20. Koblan, L. W. et al. Efficient C·G-to-G·C base editors developed using CRISPRi screens, target-library analysis, and machine learning. *Nat. Biotechnol.* **39**, 1414–1425 (2021).

21. Lam, D. K. et al. Improved cytosine base editors generated from TadA variants. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01611-9 (2023).

22. Tong, H. et al. Programmable A-to-Y base editing by fusing an adenine base editor with an N-methylpurine DNA glycosylase. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01595-6 (2023).

23. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).

24. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).

25. Liu, P. et al. Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat. Commun.* **12**, 2121 (2021).

26. Gainor, J. F. et al. Molecular mechanisms of resistance to first- and second-generation ALK inhibitors in ALK-rearranged lung cancer. *Cancer Discov.* **6**, 1118–1133 (2016).

27. Gainor, J. F. et al. Patterns of metastatic spread and mechanisms of resistance to crizotinib in ROS1-positive non-small-cell lung cancer. *JCO Precis. Oncol.* **2017**, PO.17.00063 (2017).

28. Kobayashi, S. et al. EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **352**, 786–792 (2005).

29. Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652 (2021).

30. Canté-Barrett, K. et al. Lentiviral gene transfer into human and murine hematopoietic stem cells: size matters. *BMC Res. Notes* **9**, 312 (2016).

31. Zinn, E. & Vandenberghe, L. H. Adeno-associated virus: fit to serve. *Curr. Opin. Virol.* **8**, 90–97 (2014).

32. Wang, D. et al. Adenovirus-mediated somatic genome editing of Pten by CRISPR/Cas9 in mouse liver in spite of Cas9-specific immune responses. *Hum. Gene Ther.* **26**, 432–442 (2015).

33. Annunziato, S. et al. Modeling invasive lobular breast carcinoma by CRISPR/Cas9-mediated somatic genome editing of the mammary gland. *Genes Dev.* **30**, 1470–1480 (2016).

34. Böck, D. et al. In vivo prime editing of a metabolic liver disease in mice. *Sci. Transl. Med.* **14**, eabl9238 (2022).

35. Sánchez-Rivera, F. J. et al. Base editing sensor libraries for high-throughput engineering and functional analysis of cancer-associated single nucleotide variants. *Nat. Biotechnol.* **40**, 862–873 (2022).

36. Erwood, S. et al. Saturation variant interpretation using CRISPR prime editing. *Nat. Biotechnol.* **40**, 885–895 (2022).

37. Chen, L. et al. Programmable C:G to G:C genome editing with CRISPR-Cas9-directed base excision repair proteins. *Nat. Commun.* **12**, 1384 (2021).

38. Grünewald, J. et al. A dual-deaminase CRISPR base editor enables concurrent adenine and cytosine editing. *Nat. Biotechnol.* **38**, 861–864 (2020).

39. Yuan, T. et al. Optimization of C-to-G base editors with sequence context preference predictable by machine learning methods. *Nat. Commun.* **12**, 4902 (2021).

40. Anzalone, A. V. et al. Programmable deletion, replacement, integration and inversion of large DNA sequences with twin prime editing. *Nat. Biotechnol.* **40**, 731–740 (2022).

41. Jiang, T., Zhang, X.-O., Weng, Z. & Xue, W. Deletion and replacement of long genomic sequences using prime editing. *Nat. Biotechnol.* **40**, 227–234 (2022).

42. Bult, C. J. et al. Mouse Genome Database (MGD) 2019. *Nucleic Acids Res.* **47**, D801–D806 (2019).

43. Mouse Genome Informatics. Mouse Genome Database (MGD). https://www.informatics.jax.org/index.shtml (2022).

44. Shaner, N. C. et al. A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat. Methods* **10**, 407–409 (2013).

45. 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).

46. 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. USA* **117**, 513–521 (2020).

47. Freed-Pastor, W. A. et al. The CD155/TIGIT axis promotes and maintains immune evasion in neoantigen-expressing pancreatic cancer. *Cancer Cell* **39**, 1342–1360 (2021).

48. Schlake, T. & Bode, J. Use of mutated FLP recognition target (FRT) sites for the exchange of expression cassettes at defined chromosomal loci. *Biochemistry* **33**, 12746–12751 (1994).

49. Bindels, D. S. et al. mScarlet: a bright monomeric red fluorescent protein for cellular imaging. *Nat. Methods* **14**, 53–56 (2017).

50. Vassilev, L. T. et al. In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* **303**, 844–848 (2004).

51. Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* **40**, 402–410 (2022).

52. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).

53. Qiu, M. et al. Lipid nanoparticle-mediated codelivery of Cas9 mRNA and single-guide RNA achieves liver-specific in vivo genome editing of Angptl3. *Proc. Natl Acad. Sci. USA* **118**, e2020401118 (2021).

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

55. el Marjou, F. et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39**, 186–193 (2004).

56. Kim, H. K. et al. Predicting the efficiency of prime editing guide RNAs in human cells. *Nat. Biotechnol.* **39**, 198–206 (2021).

57. Drost, J. et al. Sequential cancer mutations in cultured human intestinal stem cells. *Nature* **521**, 43–47 (2015).

58. Zafra, M. P. et al. An in vivo *Kras* allelic series reveals distinct phenotypes of common oncogenic variants. *Cancer Discov.* **10**, 1654–1671 (2020).

59. Strickler, J. H. et al. First data for sotorasib in patients with pancreatic cancer with KRAS p.G12C mutation: a phase I/II study evaluating efficacy and safety. *J. Clin. Oncol.* **40**, 360490 (2022).

60. Hofmann, M. H., Gerlach, D., Misale, S., Petronczki, M. & Kraut, N. Expanding the reach of precision oncology by drugging all KRAS mutants. *Cancer Discov.* **12**, 924–937 (2022).

61. Waters, A. M. & Der, C. J. KRAS: the critical driver and therapeutic target for pancreatic cancer. *Cold Spring Harb. Perspect. Med.* **8**, a031435 (2018).

62. Wang, X. et al. Identification of MRTX1133, a noncovalent, potent, and selective KRASG12D inhibitor. *J. Med. Chem.* **65**, 3123–3133 (2022).

63. Awad, M. M. et al. Acquired resistance to KRASG12C inhibition in cancer. *N. Engl. J. Med.* **384**, 2382–2393 (2021).

64. Pao, W. et al. Acquired resistance of lung adenocarcinomas to gefitinib or erlotinib is associated with a second mutation in the EGFR kinase domain. *PLoS Med.* **2**, e73 (2005).

65. Freed-Pastor, W. A. & Prives, C. Mutant p53: one name, many proteins. *Genes Dev.* **26**, 1268–1286 (2012).

66. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics sata. *Cancer Discov.* **2**, 401–404 (2012).

67. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).

68. Schulz-Heddergott, R. et al. Therapeutic ablation of gain-of-function mutant p53 in colorectal cancer inhibits Stat3-mediated tumor growth and invasion. *Cancer Cell* **34**, 298–314 (2018).

69. Barta, J. A., Pauley, K., Kossenkov, A. V. & McMahon, S. B. The lung-enriched p53 mutants V157F and R158L/P regulate a gain of function transcriptome in lung cancer. *Carcinogenesis* **41**, 67–77 (2020).

70. Shakya, R. et al. Mutant p53 upregulates alpha-1 antitrypsin expression and promotes invasion in lung cancer. *Oncogene* **36**, 4469–4480 (2017).

71. Jackson, E. L. et al. The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res.* **65**, 10280–10288 (2005).

72. Winslow, M. M. et al. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* **473**, 101–104 (2011).

73. LaFave, L. M. et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* **38**, 212–228 (2020).

74. 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).

75. Aoyama, N. et al. Transgenic mice that accept luciferase- or GFP-expressing syngeneic tumor cells at high efficiencies. *Genes Cells* **23**, 580–589 (2018).

76. 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).

77. 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).

78. Liu, B. et al. A split prime editor with untethered reverse transcriptase and circular RNA template. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-022-01255-9 (2022).

79. Hunter, J. C. et al. Biochemical and structural analysis of common cancer-associated KRAS mutations. *Mol. Cancer Res.* **13**, 1325–1335 (2015).

80. Amodio, V. et al. EGFR blockade reverts resistance to KRASG12C inhibition in colorectal cancer. *Cancer Discov.* **10**, 1129–1139 (2020).

81. Canon, J. et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* **575**, 217–223 (2019).

82. Kemp, S. B. et al. Efficacy of a small-molecule inhibitor of KrasG12D in immunocompetent models of pancreatic cancer. *Cancer Discov.* **13**, 298–311 (2023).

83. Escobar-Hoyos, L. F. et al. Altered RNA splicing by mutant p53 activates oncogenic RAS signaling in pancreatic cancer. *Cancer Cell* **38**, 198–211 (2020).

84. Klemke, L. et al. The gain-of-function p53 R248W mutant promotes migration by STAT3 deregulation in human pancreatic cancer cells. *Front. Oncol.* **11**, 642603 (2021).

85. Schoenfeld, A. J. et al. The genomic landscape of SMARCA4 alterations and associations with outcomes in patients with lung cancer. *Clin. Cancer Res.* **26**, 5701–5708 (2020).

## Methods

### Analysis of prime and base editor capabilities for modeling cancer-associated mutations

We constructed a Python-based computational pipeline to compare the abilities of prime and base editors to model cancer-associated mutations. Data were retrieved from MSK-IMPACT datasets[35].

### Analysis of cancer mutations incorporated in transgenic mouse models

We used the MouseMine tool from the Mouse Genome Informatics database[42,43] to obtain a comprehensive list of published transgenic alleles. We initiated our search using the mammalian phenotype code MP:0002006 ('neoplasm') to retrieve all mouse models related to the study of cancer. We then modified the search with the following parameters: 'allele type', 'mutations (name)', 'alleles (name and molecular note and attribute string)' and 'subjects (synonyms → names)'. We then filtered the results to retain only allele types annotated as 'targeted', 'transgenic' or 'endonuclease mediated'.

After exporting these data (Supplementary Table 2), we identified the 100 most frequent SNVs present in the MSK-IMPACT dataset. We then manually cross-referenced these two lists to identify available models representing specific mutations. In cases where models were absent in the MouseMine list, we performed a manual literature search to confirm the absence of models in the published literature. Using this approach, we designated for each mutation (1) whether any transgenic allele exists that can be used to model cancer in mice and (2) whether any existing models enable selective expression in a tissue of interest (for example, through Cre recombinase-induced removal of an LSL cassette).

### Design and cloning of the Cre-inducible prime editor allele

The PE2-P2A-mNG *Rosa26* targeting vector was generated with a backbone formed via BstBI and AscI restriction enzyme digestion of the *Sp*Cas9-NLS-P2A-EGFP *Rosa26* targeting vector[10,45]. A fragment encoding the PE2 enzyme was generated by PCR amplification from the pCMV-PE2 plasmid obtained from Addgene[24] (132775), and a fragment containing the P2A-mNG sequence was amplified from a plasmid encoding Cre-P2A-mNG. Two additional fragments containing WPRE-pA-PGK (Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element-poly(A)-PGK promoter) and a neomycin resistance gene (NeoR-pA) were PCR-amplified from the *Sp*Cas9-NLS-P2A-EGFP vector. An FRT3 site was installed by incorporating overlapping portions of this motif into the PCR primers. All primers used are listed and described in Supplementary Table 5. A five-part Gibson assembly reaction generated the final targeting vector using these components[86].

### Embryonic stem cell targeting, validation and chimera generation

P4*, a C57BL/6J *Kras*^+/+^;*Trp53*^flox/flox^ (P) mouse embryonic stem (ES) cell line, was generated by crossing a hormone-primed C57Bl6J *Trp53*^flox/flox^ female with a C57Bl/6J *Kras*^LSL-G12D^;*Trp53*^flox/flox^ male. At 3.5 d after coitum, blastocysts were flushed from the uterus, isolated and cultured individually on a mouse embryonic fibroblast (MEF) feeder layer. After 5–7 d in culture, the outgrown inner cell mass was isolated, trypsinized and replated on a fresh MEF layer. ES cell lines were genotyped for *Kras*^LSL-G12D^, *Trp53*^flox/flox^ and Zfy (Y-chromosome specific).

Notably, 36 μg of the prime editor targeting vector (R26–CAGG-LoxStopLox-Cas9^(H840A)^-MMLVRT-P2A-mNeonGreen-WPRE-bHGpA; PGK-Neo-PGKpA) was linearized with PvuI, phenol/CHCl$_3$ extracted, and then ethanol precipitated. After resuspending the DNA in 150 μl of PBS, it was mixed with $3 \times 10^6$ P4* ES cells in 650 μl of PBS in a 4-mm electroporation cuvette. The cell–DNA mixture was pulsed once in a BioRad Genepulsar 2 (600 V and 25 μF) followed by replating of the cells on irradiated MEFs. After 48 h, the ES cell cultures were placed under selection with Geneticin (GIBCO) at 350 μg ml$^{-1}$. A total of 45 colonies were manually picked using a stereomicroscope. Each clone was expanded and evaluated for correct integration by PCR with primers spanning the 5′ homology arm. Eleven PCR-positive clones were further evaluated using southern blot analysis. Briefly, genomic DNA was digested with EcoRV-HF (NEB) overnight. Digestions were electrophoresed on 0.7% agarose gels and blotted to Amersham Hybond XL nylon membranes (GE Healthcare). Samples were probed with $^{32}$P-labeled 'Rosa26 3′ 'external' and Cas9 'internal' probes applied in Church buffer (probe sequences available on request).

Correctly targeted clones verified by both PCR and southern blot analysis were injected into albino C57BL/6J blastocysts. High-degree chimeras (visually assessed by coat color percentage) from the 100C7 and 100C8 ES cell clones successfully transmitted the prime editor allele through the germline.

### Nucleofection of Neuro-2a cells and genomic DNA preparation

To evaluate spacers near the genetic locus encoding G12 in *Kras*, Neuro-2a cells were nucleofected using the SF Cell Line 4D-Nucleofector X Kit (Lonza) with $2 \times 10^5$ cells per sample (program DS-137). Notably, 800 ng of SpCas9-expressing plasmid and 200 ng of single guide RNA (sgRNA)-expressing plasmid were used according to the manufacturer's protocol. Three days following nucleofection, the cells were washed with PBS after removing the media and then lysed by the addition of 150 μl of freshly prepared lysis buffer (10 mM Tris–HCl, pH 8 at 23 °C; 0.05% SDS; 25 μg ml$^{-1}$ of proteinase K (Qiagen)). The *Kras* amplicon was amplified from the genomic DNA samples, sequenced on an Illumina MiSeq and analyzed with CRISPResso2 (ref. 87) for indel quantification[37].

### pegRNA design and cloning

pegRNAs were designed in part using the pegRNA design tool, Prime Design[88]. In some cases (for example, editing at *Kras*^G12D^), CRISPR sgRNAs were tested before pegRNA design to select spacers that exhibited the highest level of Cas9 activity. For some designs, the trimmed evopreQ$_1$ motif was included to form epegRNAs and optimize editing efficiency within a limited cohort of initial candidates[51]. pegRNAs and their sequences are provided in Supplementary Table 3.

All pegRNAs were tested within the context of UPEC or hU6-RFP/UPEmS vectors. All pegRNA-expressing vectors were assembled via Golden Gate Assembly[89] using the uncut template plasmid and three annealed oligo pairs consisting of the spacer sequence, the scaffold and the 3′ extension, all with compatible overhangs. Assembly was facilitated using the Golden Gate Assembly Kit (BsmBI-v2) from New England BioLabs.

The UPEmS template vector was generated via Gibson assembly of three insert fragments and a linearized backbone. Two fragments were formed by PCR amplification from the 'pU6 pegRNA GG acceptor' plasmid (Addgene plasmid, 132777)[24]. Specifically, the hU6 promoter was amplified using primers modified to install a BsmBI recognition site and the pAF Gibson adapter sequence on either side of the promoter (pAF-hU6-BsmBI), and the RFP component was also amplified in part using a primer that installed another BsmBI recognition site (forming BsmBI-RFP-BsmBI-pAR/gBF). A third fragment, gAR/pBF-EFS-mScarlet-gBR, was amplified from a separate lentiviral plasmid containing U6-sgRNA-EFS-mScarlet. All fragments were designed to contain compatible overhangs for Gibson assembly. All vectors with detailed maps and sequences will be deposited into Addgene.

The UPEC template plasmid (hU6-RFP-EF-1α-Cre) was developed by Gibson assembly of two insert fragments and the same backbone used to clone pUPEmS. The pBF-EF-1alpha-Cre-gBR fragment was generated using pBF and gBR PCR primers targeting the pUSEC (U6-sgRNA-EF-1alpha-Cre) vector[86,90]. The pAF-U6-RFP-gAR fragment was amplified from the UPEmS vector.

## Generation of tail-tip-derived *Rosa26*[PE2/+] fibroblasts

To generate *Rosa26*[PE2] cell lines for convenient testing of pegRNAs, a 2-cm piece was excised from the tail tip of an anesthetized, 3.5-week-old male. The sample was sprayed with ethanol and then dipped in PBS several times. A lengthwise incision was made, and the outside skin and hair were removed. The sample was then incubated at 37 °C in digestion buffer comprised of 5 ml DMEM, 25 µl penicillin-streptomycin, 5 µl Amphotericin B, 10 µl DNase (40 U ml$^{-1}$; −20 °C; 1:500), 50 µl collagenase (100 mg ml$^{-1}$; 1:100) and 50 µl CaCl$_2$ (36 mM; 1:100). Samples were then washed twice with PBS, and dissociated chunks were added to a 6-cm dish. Additional media containing Amphotericin B was added the following day.

## HEK293 and fibroblast cell culture conditions

HEK293, split-PE2 3TZ and tail-tip-derived *Rosa26*[PE2/+] fibroblast cells were cultured in standard media consisting of Dulbecco's Modified Eagle's Medium (DMEM) (Corning), penicillin-streptomycin and 10% (vol/vol) FBS. All cultured cells were incubated at 37 °C and 5% CO$_2$.

## Pancreatic ductal organoid culture

Pancreata from mice of the desired genotype were dissected manually and minced with a razor blade. Pancreas tissue was then dissociated by 20 min of gentle agitation in pancreas digestion buffer (1× PBS (Corning), 125 U ml$^{-1}$ collagenase IV (Worthington)) at 37 °C. Tissue suspensions were then strained through 70 µM filters, washed with 1× PBS, and pelleted with slow deceleration by centrifugation. Cells were resuspended in 100% Matrigel (Corning) and plated as 50 µl domes into 24-well plates (GenClone). Upon solidification of domes, cells were cultured in organoid complete media[47], or alternatively, in a complete medium as follows: AdDMEM/F-12 medium supplemented with HEPES (1×, Invitrogen), GlutaMAX (1×, Invitrogen), penicillin/streptomycin (1×, VWR), B27 (1×, Invitrogen), R-Spondin1-Conditioned Medium (10% vol/vol), A83-01 (0.5 µM, Tocris), mouse epidermal growth factor (mEGF; 0.05 µg ml$^{-1}$, PeproTech), Fibroblast Growth Factor 10 (FGF-10; 0.1 µg ml$^{-1}$, PeproTech), Gastrin I (0.01 µM, Tocris), recombinant mouse Noggin (0.1 µg ml$^{-1}$, PeproTech), N-acetyl-L-cysteine (1.25 mM, Sigma-Aldrich), nicotinamide (10 mM, Sigma-Aldrich) and Y-27632 (10.5 µM, Cayman Chemical Company).

Organoids were passaged using TrypLE Express (Life Technologies) for Matrigel digestion for 15–30 min at 37 °C. Organoids were infected at a high multiplicity of infection to ensure 100% recombination. Briefly, concentrated lentivirus (either diluted 1:9 or undiluted) was introduced to cell suspensions at the time of passage. For *Trp53*[flox/flox] lines, Nutlin-3a was added to organoid media (10 µM; Sigma-Aldrich) to ensure the purification of recombined organoids. For prime-edited organoids harboring *Kras*[G12D] or *Kras*[G12C] mutations, organoids were cultured in the presence of 1 µM Gefitinib in full organoid media (Cayman) to select for the intended edit. Sotorasib (Selleck) was added to media at 1, 2 and 5 µM. MRTX1133 (MedChem) was added to the media at 2 µM or 5 µM. Prime-edited mutations were confirmed by deep amplicon sequencing of organoids several days after the initial infection with lentivirus, and then again after several passages under treatment with the drug. For the selection of transgene-containing cells from chimera-derived pancreatic organoids, organoids were treated with 800 µg ml$^{-1}$ of Geneticin (GIBCO).

Organoid viability and proliferation were quantified using the alamarBlue HS Cell Viability Reagent (Thermo Fisher Scientific). Viability reagent was directly added to organoid culture at 1/10 media volume. After 24 h, 200 µl of reagent-containing media was removed and assayed in replication in a Tecan Infinite Pro m200 using the manufacturer's parameters.

## Lung organoid culture

Lung organoids were derived from 8–20-week-old mice[91]. Fresh lung tissue was transferred into 500 µl disase and minced. Digestion buffer of 3–5 ml containing advanced DMEM/F-12, penicillin–streptomycin,

Amphotericin B, 1 mg ml$^{-1}$ Collagenase (Sigma, C9407-500MG), 40 U ml$^{-1}$ DNase I (Roche, 10104159001), 5 µM HEPES and 0.36 mM CaCl$_2$ was added for a 20–60-min incubation at 37 °C in a rotating oven. The resulting suspension was incubated in 1 ml ACK Lysis Buffer (Thermo Fisher Scientific, A1049201) for 3–5 min at room temperature to lyse red blood cells. Samples were then washed two times with fluorescence-activated cell sorting (FACS) buffer (1× PBS with 1 mM EDTA and 0.1% BSA) and filtered through 40 µm mesh. Samples were resuspended in 150 µl FACS buffer, and CD45 cells were depleted using the EasySep Mouse CD45 Positive Selection kit (STEMCELL Technologies, 18945). Cells were stained with anti-mouse CD31-APC (1:500; Biolegend, 102507), CD45-APC (1:500; BD Biosciences, 559864), EpCAM-PE (1:500; Biolegend, 118206) and MHCII-APC-eFluor-780 (1:500; Thermo Fisher Scientific, 47-5321-82). The suspensions were then sorted for DAPI$^-$, CD31$^-$, CD45$^-$, EpCAM$^+$ and MHCII$^+$ cells, visualized using BD FACS Diva v8. Approximately 20,000 sorted AT2 cells were mixed with Growth Factor Reduced Matrigel (Corning) at a ratio of 1:9 and seeded onto multiwell plates as 20 µl drops. The drops were incubated at 37 °C for 15 min to solidify and then overlaid with F$^7$NHCS medium supplemented with Y-27632 (Cayman). For passaging, matrigel drops were dissolved in TrypLE Express (Sigma, 12604-013) and incubated at 37 °C for 7–15 min. The organoid suspensions were then dissociated into single cells by vigorous pipetting, washed twice, resuspended in 1× PBS and plated as described above.

## Generation of a split-PE2 fibroblast cell line

A cell line based on mouse 3TZs cells was developed to test *Trp53*-targeted pegRNAs on a *Trp53*[+/+] background. Two plasmids containing halves of the PE2 enzyme and distinct antibiotic resistance genes were generated via Gibson assembly. The split intein-based constructs described in refs. [24,92] were used to enable post-translational splicing of the intein motifs and subsequent joining of the halves to form the full PE2 enzyme. Specifically, the N-terminal half of PE2 (the first 573 amino acids of the Cas9 nickase joined to the Npu N-intein) was PCR-amplified from the U6-DNMT1-hSynapsin-PE2-N-terminal-P2A-EGFP-KASH-lenti plasmid (Addgene, 135955) and then cloned into a puromycin resistance gene-containing backbone. A blasticidin resistance gene-containing backbone was assembled into a second vector with a PCR-amplified DNA fragment encoding the C-terminal half of PE2 (Npu C-intein joined to the remaining C-terminal half of PE2), amplified from the hSynapsin-PE2-C-terminal-lenti plasmid (Addgene, 135956). The two constructs were incorporated into lentiviruses, which were used to transduce mouse 3TZ fibroblast cells, followed by selection with up to 10 µg m$^{-1}$ of puromycin and 20 µg ml$^{-1}$ of blasticidin.

## Production of lentivirus and transduction

Lentivirus was produced by transfection of the expression vector into 293FS* cells along with psPAX2 (psPAX2 was a gift from Didier Trono−Addgene plasmid, 12260; http://n2t.net/addgene:12260; RRID:Addgene_12260) and pMD2.G (pMD2.G was a gift from Didier Trono−Addgene plasmid, 12259; http://n2t.net/addgene:12259; RRID:Addgene_12259) packaging plasmids at a 4:3:1 ratio using polyethylenimine or Mirus transfection reagent. A volume of 1 ml of small-scale viral supernatant was added directly to 1 × 10$^5$ cells at seeding in a six-well plate (Corning) for transduction. Small-scale transductions were supplemented with polybrene (10 mg ml$^{-1}$, 1:1,000; Sigma). Concentrated large-scale lentivirus and small-scale viruses were stored at −80 °C if not used immediately. Generally, cell lines were infected with small-scale virus, while organoids were infected with large-scale virus. Quantification of lentiviral titer was performed using a GFP Cre reporter 3TZ cell line[14].

## Intratracheal delivery of lentivirus into the lung

Mice were anesthetized in an isoflurane chamber. A total of 6 × 10$^4$ transducing units (TU) or 1 × 10$^5$ TU of lentivirus containing UPEC vectors encoding pegRNAs and Cre recombinase were injected intratracheally

into *Rosa26*[PE2] mice[93]. Mice were sex and age-matched within 4 weeks across experimental arms.

## Orthotopic transplantation of pancreatic organoids

Animals were anesthetized with isoflurane, the left abdominal side was depilated with Nair and the surgical region was disinfected with Chloraprep swabstick (BD). A small incision (~1.5 cm) was made in the left subcostal area, and the spleen and pancreas were exteriorized with ring forceps. The organoid suspension (containing $1 \times 10^5$ organoid cells in 100 μl of 50% PBS + 50% Matrigel) was injected using a 30-gauge needle into the pancreatic parenchyma parallel to the main pancreatic artery. The pancreas and spleen were gently internalized, and the peritoneal and skin layers were sutured independently using a 4/0 PGA suture and a 4/0 silk suture, respectively (AD Surgical). All mice received preoperative analgesia Buprenorphine Sustained-Release (Bup-SR; 0.5 mg kg$^{-1}$) and were followed postoperatively for any signs of distress. Organoid/Matrigel mixtures were kept on ice throughout the whole procedure to avoid solidification. For orthotopic transplantation, syngeneic C57BL/6J *Rosa26*[PE2] mice (aged 6–17 weeks) were used as recipients. Male pancreatic organoids were only transplanted into male recipients.

## Autochthonous pancreatic tumor modeling

Retrograde pancreatic duct infection with lentivirus was modified from previously reported techniques[77]. The ventral abdomen was depilated (using Nair) 1 d before surgery. Animals were anesthetized with isoflurane. The surgical area was disinfected with betadine/isopropanol and a 2- to 3-cm incision was made in the anterior abdomen. A subsequent vertical incision was made through the abdominal wall, securing the incision edges with a Colibri retractor. A Nikon stereomicroscope was used to visualize the pancreas, common bile duct and sphincter of Oddi. The common bile duct and cystic duct were gently separated from the portal vein and hepatic artery using blunt dissection with Moria forceps. A microclip was placed over the common bile duct to prevent the influx of the viral particles into the liver or gallbladder. A 30-gauge needle was used to cannulate the common bile duct at the level of the sphincter of Oddi, and 150 μl virus was injected over 30 s. After injection and removal of instruments, the peritoneum was closed using running 5-0 Vicryl sutures. The abdominal wall and fascia were closed using simple interrupted 5-0 Vicryl sutures. Animals were administered postoperative sustained-release Buprenorphine (Bup-SR) and were monitored postoperatively for signs of discomfort or distress. For retrograde pancreatic ductal installation, male mice (aged 8–20 weeks) and female mice (aged 8–20 weeks) were transduced with 500,000 TU in serum-free media (Opti-MEM; Gibco).

## Lipid nanoparticle (LNP) formulation and injection

LNPs were formulated with modifications from an existing protocol[53]. Dnmt1[+GGG] Synthetic pegRNA (1 mg) was ordered from Agilent Technologies. The first three and last three nucleotides were modified with 2′O-methyl groups. The first three and last three nucleotide bonds were phosphorothioate-modified bonds. Both modifications were made to increase the stability of the guide. Cre mRNA was obtained from Trilink. A weight ratio of 1:7.5 total mRNA ionizable lipid was used for LNP formulation, with a 1:2 ratio of Cre mRNA: pegRNA. The aqueous phase was prepared with 25 mM sodium acetate (pH 5.2), Cre mRNA and pegRNA solution. Two organic phase preparations were made by adding an ionizable lipid (Lipid 10 or 306-O12B) to cholesterol (Sigma-Aldrich), DOPC (Avanti) and DMG-PEG (Sunbright) stock solutions in 100% ethanol, at a 50:38.5:10:1.5 molar ratio. Nanoparticles were prepared by combining the organic and aqueous phases at a 1:3 ratio and assembled using a NanoAssemblr (Precision Nanosystems). LNPs were dialyzed for 4 h against PBS in Thermo Fisher Scientific Slide-A-Lyzer dialysis Cassettes (3.5 K MWCO). LNPs were kept on ice before animal dosing. Mice were administered a maximal dose of 60 μg total RNA via tail vein injection, corresponding to roughly 200 μl per mouse.

## Animal studies

All mouse experiments described in this study were approved by the Massachusetts Institute of Technology Institutional Animal Care and Use Committee (IACUC) (institutional animal welfare assurance, A-3125-01). For *Vilin-cre*[ERT2];*Rosa26*[PE2/+] animals, Tamoxifen was administered in the diet (Envigo, TD.130860) for 2 weeks before tissue collection. Mice aged between 7 and 20 weeks old were chosen for in vivo experiments. Mice of both sexes were used for autochthonous lung tumor initiation, and male mice were chosen for orthotopic pancreatic organoid experiments as the transplanted organoid line was male-derived. Mice were assessed for morbidity according to guidelines set by the MIT Division of Comparative Medicine and were humanely killed before natural expiration.

## Ultrasound imaging

Animals were anesthetized with isoflurane and the left subcostal region of animals was depilated with Nair. Animals were imaged with a Vevo3100/LAZRX ultrasound and photoacoustic imaging system (Fujifilm Visualsonics). Anesthetized animals were positioned supine and angled on an imaging platform for visualization of peritoneal organs. Landmark organs including the kidney and spleen were first identified before imaging. A thin layer of ultrasound gel was applied over the depilated region of the abdomen. The transducer (VisualSonics 550S) was positioned above the abdomen and set at the scanning midpoint of the healthy pancreas or tumor. Approximately 1 cm of scanning area was used to capture the entirety of pancreas tumors, using a z-slick thickness of 0.04 mm. Ultrasound scans were uploaded to Vevo Lab Software, from which representative images were exported.

## Rodent μCT

Mice were anesthetized with isoflurane (3%, then maintained at 2.0–2.5% in oxygen−VetEquip) and scanned in a prone position using a Skyscan 1276 (Bruker) with the following parameters: 100 kVp source voltage, 200 μA current, 0.5 mm aluminum X-ray filter, 108 ms exposure time and 0.65-degree rotational step size over 360 degrees in a continuous rotation. With 4 × 4 detector binning, the nominal pixel size after reconstruction (Bruker NRecon software) was 40.16 microns. Data were visualized using ImageJ.

## Histology, immunohistochemistry and immunofluorescence

Pancreata from control and tumor-bearing animals were manually dissected from the peritoneal cavity after they were killed. Tumor-bearing lung was flushed with 1× PBS and separated into separate lobes. Tissue was fixed in Zinc Formalin overnight, transferred to 70% ethanol and then embedded in paraffin. Hematoxylin and eosin (H&E) staining was performed and digitally scanned images of H&E slides were obtained with an Aperio ScanScope at ×20 magnification. Histologic quantification of tumor grade was performed by an automated deep neural network available through Aiforia image analysis software with the nsclc_v25 algorithm.

For IHC, slides were incubated at 4 °C overnight with the following antibodies: anti-NKX2-1 (1:1,000; Abcam, ab76013; RRID:AB_1310784), anti-SFTPC (1:5,000; Millipore Sigma, AB3786; RRID:AB_91588) and anti-HMGA2 (1:1,000; Cell Signaling Technologies, 8179S; RRID:AB_11178942). ImPRESS Anti-Rabbit Horseradish Peroxidase and DAB Peroxidase Substrate Kits (Vector) were used to develop slides. Tissues were counterstained with hematoxylin. Slides were digitally scanned and analyzed using QuPath[94].

For IF, slides were incubated at 4 °C with anti-Cas9 (E7M1H, 1:100, CST 19526). Horse anti-rabbit secondary (AF488, 1:400) was used. All slides were counterstained with DAPI (1:20,000) and imaged using a Nikon 80 Eclipse 80i fluorescence microscope using ×10 and ×20 objectives and an attached Andor camera.

## Immunoblotting

Pancreatic organoids were dissociated with TrypLE for 30 min at 37 °C, washed with 6× PBS and then lysed in cell lysis buffer (RIPA with 100× HALT protease and phosphatase inhibitors). Blots were incubated with primary antibody (p53 clone 1C12, Cell Signaling Technology (CST), β-Actin clone 13E5, CST, 1:5,000) overnight at 4 ° imaged on a ChemiDoc Gel Imaging System (BioRad).

## DNA sequencing and analysis of genomic DNA samples

Target loci were amplified from genomic DNA using PCR primers listed and described in Supplementary Table 5. Amplicons were then purified using either agarose gel extraction or using a QIAquick PCR purification kit (Qiagen). Purified amplicons were typically then submitted to the Massachusetts General Hospital Center for Computational and Integrative Biology's DNA Core for next-generation sequencing (samples prepared according to guidelines provided for the CRISPR Sequencing service).

Amplicons prepared for evaluating prime-editing efficiency of the initial *Trp53*[245]- and *Kras96*-targeted pegRNAs were given unique Illumina TruSeq barcodes for pooled sequencing. Barcoded PCR products were pooled and purified by electrophoresis with a 2% agarose gel using a Gel Extraction Kit (QIAGEN), eluting with 30 µl $H_2O$. DNA concentration was quantified using a Qubit dsDNA High Sensitivity Assay Kit (Thermo Fisher Scientific) and sequenced on an Illumina MiSeq instrument (single-end read, 250–300 cycles) according to the manufacturer's protocols.

Sequencing reads were aligned to reference amplicons and analyzed using the deep sequencing analysis program, CRISPResso2 (ref. [87]), V2.2.6. CRISPResso2 parameters employed for each target are described in Supplementary Note. Prime-editing efficiency was calculated as the percentage of reads aligning to the prime-edited amplicon (excluding indels) relative to all reads aligning to both the prime-edited and reference amplicons (including indels). Only reads with an average Phred score of ≥30 were considered. Indel percentages were calculated in a similar fashion using the total number of indel-bearing reads designated as 'discarded' by CRISPResso2. For experiments involving pegRNAs that alter multiple nucleotides, the allele frequency tables output by CRISPResso2 were consulted to confirm that the majority of prime-edited reads contained all of the intended nucleotide alterations.

## Sequencing and analysis of off-target loci

Off-target loci were identified using Cas-OFFinder[52]. For each of the four indicated protospacers (Supplementary Table 4), all off-target sites with three or fewer mismatches relative to the target protospacer were identified. Bulges were not permitted. These results were then filtered down to up to four off-target loci per protospacer. First, loci were prioritized by selecting candidates with the lowest number of protospacer mismatches. When loci contained the same number of protospacer mismatches, off-target sites were ranked by the lowest number of primer binding site mismatches. Finally, if loci contained the same number of both protospacer and primer binding site mismatches, the difference in DNA melting temperature between the mismatched target and the original primer binding site sequence was computed. This calculation was performed with the OligoAnalyzer Tool, version 3.1, from Integrated DNA Technologies[95] with default parameters. Loci with the smallest difference in melting temperature compared to the non-mismatched strand were then prioritized. The off-target loci identified by this analysis are described in Supplementary Table 4.

Next-generation sequencing was performed for amplicons generated from each off-target locus. Off-target prime editing was assessed by aligning sequencing reads to off-target amplicons using CRISPResso2 in batch mode[87]. The parameters '-w 20' and '-q 30' were used in all cases, along with the corresponding off-target protospacer

as the guide RNA input sequence. Prime editing was then assessed using an approach described in ref. [29]. For each sample, the 'Nucleotide_percentage_summary' file output by CRISPResso2 was used to compare the DNA sequence immediately downstream of the nick site to the sequence encoded by the pegRNA. The first mismatched nucleotide site was then examined to quantify the percentage of reads bearing the allele encoded by the pegRNA. For each off-target locus, this percentage of prime-edited reads was compared between samples cultured in the presence of the examined pegRNA versus samples cultured in the presence of an irrelevant pegRNA. All off-target samples were sequenced within the same MiSeq run to ensure a similar sequencing error rate between tested and control groups.

## Mutation frequency estimates from cBioPortal

Somatic mutation frequencies of *TP53* in human pancreatic cancer were estimated using cBioPortal[66,67] for the following four patient cohorts: CPTAC, TCGA (Firehose Legacy), QCMG and ICGC. Non-Small Cell Lung Cancer[96] (TRACERx, NEJM and Nature 2017) and Pan-Lung Cancer[97] (TCGA, Nat Genet 2016) cohorts were used for estimates for lung adenocarcinoma.

## Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

*Rosa26*[PE2] mice on wild-type (JAX stock JR037953) and *Trp53*[flox/flox] (JAX stock JR037954; ref. [98]) backgrounds are available from the Jackson Laboratory. Plasmids will be made available through Addgene upon publication. Amplicon sequencing data have been deposited in the SRA repository under accession PRJNA951647 (ref. [99]). All other materials and data, including *Rosa26*[PE2] cell and organoid lines, are available from the corresponding author upon reasonable request. Source data are provided in this paper.

## Code availability

The pipeline, all related scripts, and intermediate data needed to reproduce our results are available at https://github.com/samgould2/prime-vs-base-editing (2022; ref. [100]).

## References

86. Gibson, D. G. et al. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
87. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
88. Hsu, J. Y. et al. PrimeDesign software for rapid and simplified design of prime editing guide RNAs. *Nat. Commun.* **12**, 1034 (2021).
89. Engler, C., Kandzia, R. & Marillonnet, S. A one pot, one step, precision cloning method with high throughput capability. *PLoS ONE* **3**, e3647 (2008).
90. Akama-Garren, E. H. et al. A modular assembly platform for rapid generation of DNA constructs. *Sci. Rep.* **6**, 16836 (2016).
91. Naranjo, S. et al. Modeling diverse genetic subtypes of lung adenocarcinoma with a next-generation alveolar type 2 organoid platform. *Genes Dev.* **36**, 936–949 (2022).
92. Zettler, J., Schütz, V. & Mootz, H. D. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett.* **583**, 909–914 (2009).
93. 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).
94. Bankhead, P. et al. QuPath: open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).

95.  Owczarzy, R. et al. IDT SciTools: a suite for analysis and design of nucleic acid oligomers. *Nucleic Acids Res.* **36**, W163–W169 (2008).

96.  Jamal-Hanjani, M. et al. Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).

97.  Campbell, J. D. et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **48**, 607–616 (2016).

98.  Marino, S., Vooijs, M., van Der Gulden, H., Jonkers, J. & Berns, A. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes Dev.* **14**, 994–1004 (2000).

99.  Ely, Z. A. et. al. A prime-editor mouse to model a broad spectrum of somatic mutations in vivo. Sequence Read Archive. Datasets. https://www.ncbi.nlm.nih.gov/sra/?term=PRJNA951647 (2023).

100.  Gould, S. Prime-vs-base-editing. Github. https://github.com/samgould2/prime-vs-base-editing (2022).

## Author contributions

Z.A.E. and T.J. conceived of the study. Z.A.E., N.M.A. and T.J. designed the experiments. N.M.A. conducted all autochthonous lung modeling. S.N. and Z.A.E. designed and constructed all DNA vectors described, including the transgene cassette. Z.A.E. designed pegRNAs, with contributions from N.M.A., P.B.R., J.R.D. and K.H.; G.A.N., S.I.G. and F.J.S.R. designed the computational pipeline to analyze patient mutation data; and S.I.G. conducted related bioinformatics analyses. W.A.F.-P., G.C.J. and Z.K. conducted pancreatic orthotopic transplant experiments. L.L. and G.C.J. conducted autochthonous pancreatic modeling. N.M.A., J.M.K., B.L.H. and G.A.N. conducted LNP experiments. C.M.C. and S.N. conducted experiments involving lung organoids. N.M.A., Z.A.E., N.B.P. and W.A.F.-P. conducted in vitro experiments using pancreatic organoids. W.M.R. conducted mESC targeting and chimera generation. K.L.M. provided animal husbandry expertise and conceptual advice. S.N., G.A.N., P.B.R., A.V.A., W.A.F.-P., P.M.K.W., K.L.M., S.L.S., F.J.S.R. and D.R.L. provided conceptual advice. Z.A.E., N.M.A., S.I.G., F.J.S.R. and T.J. wrote the manuscript with input from all authors.

## Competing interests

T.J. is a member of the Board of Directors of Amgen and Thermo Fisher Scientific, and a co-founder of Dragonfly Therapeutics and T2 Biosystems. T.J. serves on the Scientific Advisory Board of Dragonfly Therapeutics, SQZ Biotech, and Skyhawk Therapeutics. T.J. is also the President of Break Through Cancer. None of these affiliations represent a conflict of interest with respect to the design or execution of this study or the interpretation of data presented in this manuscript. His laboratory currently receives funding from the Johnson & Johnson Lung Cancer Initiative, but this funding did not support the research described in this manuscript. D.R.L. is a consultant for Prime Medicine, Beam Therapeutics, Pairwise Plants, Chroma Medicine, and Nvelop Therapeutics, companies that use or deliver genome editing or genome engineering agents, and owns equity in these companies. K.H. and A.V.A. are currently employees of Prime Medicine. The remaining authors declare no competing interests.

## Additional information

**Extended data** is available for this paper at https://doi.org/10.1038/s41587-023-01783-y.

**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41587-023-01783-y.

**Correspondence and requests for materials** should be addressed to Tyler Jacks.

**Peer review information** *Nature Biotechnology* thanks Andrea Ventura and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at www.nature.com/reprints.
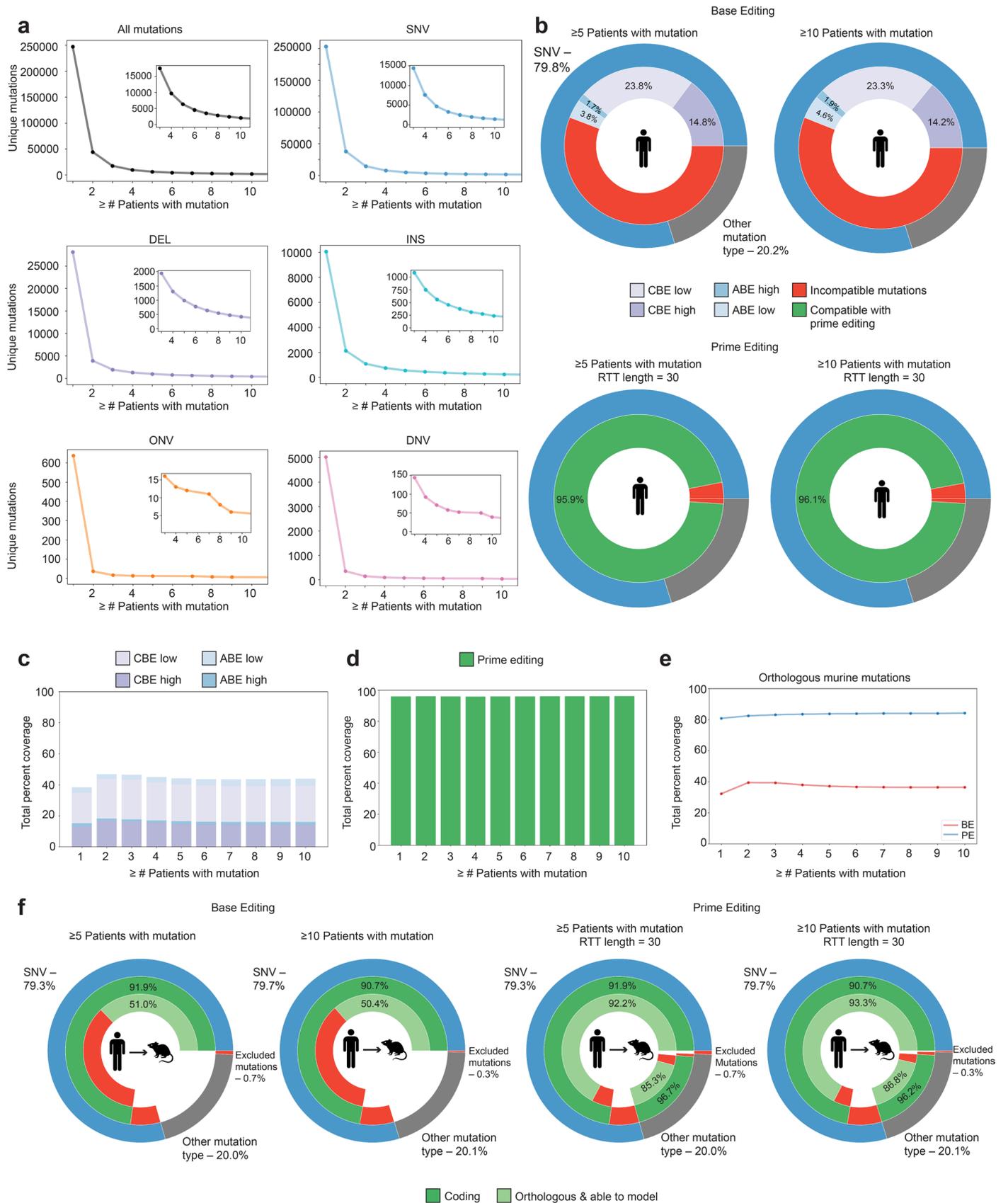
**Extended Data Fig. 1 | See next page for caption.**

**Extended Data Fig. 1 | Prime editing has a greater capacity to model recurrent cancer-associated mutations than base editing. a** Quantification of mutations that are detected in multiple patients in cancer-associated genes, depicted for each mutant variant type (Single nucleotide variants = SNV, deletions = DEL, insertions = INS, di-nucleotide variants = DNV, oligo-nucleotide variants = ONV). The y-axis in all plots indicates the total number of unique mutations per variant type. **b** Quantification of recurrent mutations potentially amenable to modeling by a base editor with an NGG PAM (top) or a prime editor with an NGG PAM and a 30 base pair RT template (bottom). The columns show results considering mutations that occur in ≥5 patients (left) or ≥10 patients (right). 'CBE/ABE high' indicates that the SNV falls in the high efficiency editing window (position +4 to +8 in the protospacer), while 'CBE/ABE low' indicates the SNV falls within the protospacer but outside the high efficiency window. The data include SNVs (blue outer circle; 78.8% of mutations) and other mutation types (gray outer circle; 20.2% of mutations). All calculations assume a base or prime editor that recognizes only NGG PAMs. **c** Total percentage of recurrent mutations amenable to modeling by a base editor with an NGG PAM, quantified at multiple thresholds of mutation frequency, from mutations that occur ≥1 patient to those that occur ≥10 patients. **d** Total percentage of recurrent mutations amenable to modeling by a prime editor with an NGG PAM and a 30 base pair RT template at multiple thresholds of mutation frequency, from mutations that occur ≥1 patient to those that occur ≥10 patients. **e** Capabilities of prime versus base editing to model orthologous mutations in mice at multiple thresholds of mutation frequency, from ≥1 to ≥10 patients with each mutation. Prime editing can model approximately double the number of orthologous mutations in mice at all thresholds of mutation frequency. Base editing (BE) shown in red. Prime editing (PE) shown in blue. **f** Quantification of the ability of base editing (left) and prime editing (right) to model orthologous mutations in mice for mutations that occur in ≥5 or ≥10 patients.
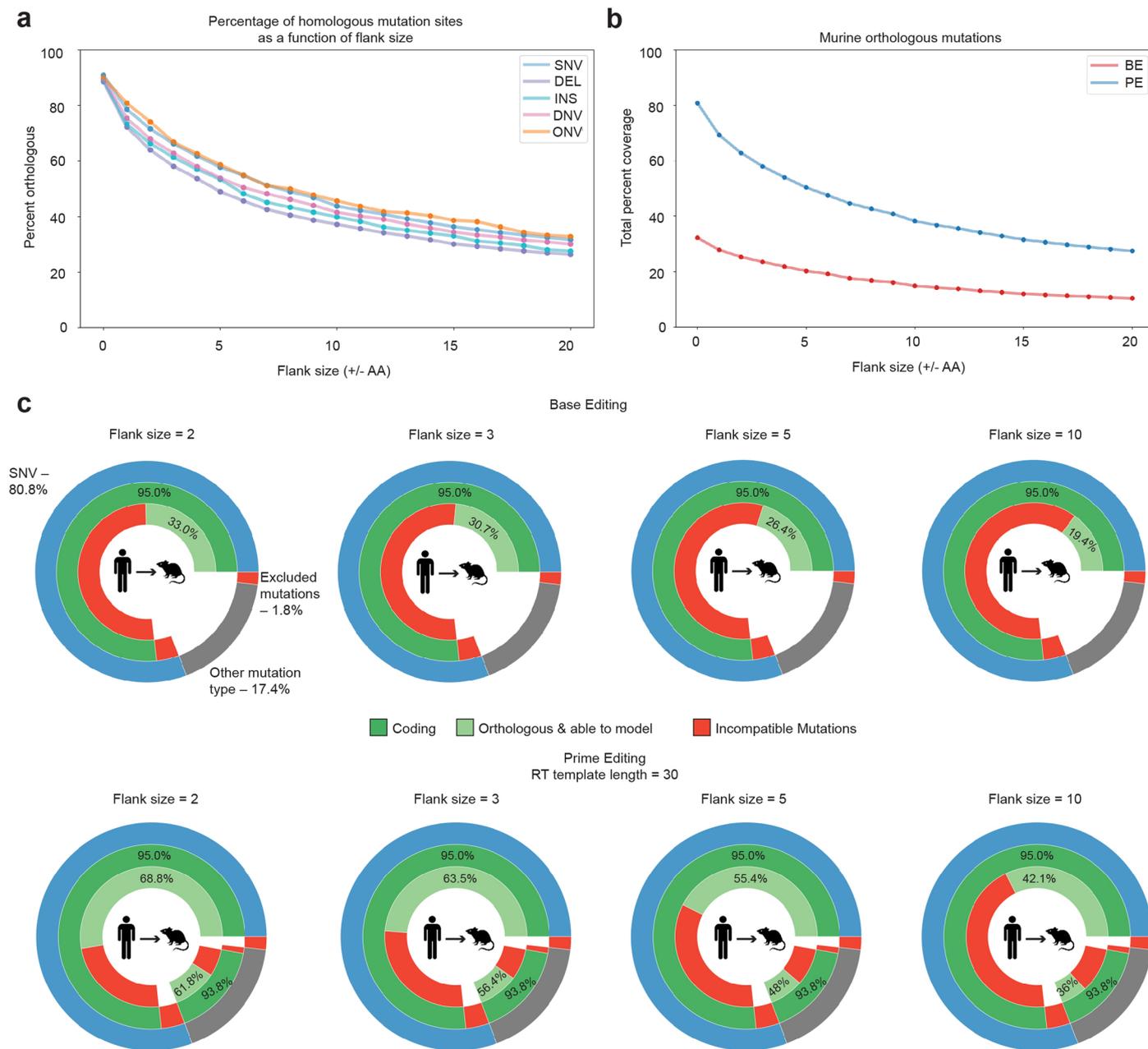
**Extended Data Fig. 2 | Prime editing enables modeling a broader scope of cancer-associated mutations from residues conserved in mice at various homology stringencies. a** The percentage of mutations, categorized by variant type, that fall in a region of homology as a function of flank size. Flank size is defined as the number of amino acids on either side of the mutant codon that must match between the human and mouse orthologs for the mutated codon to be considered orthologous (for example, a flank size of two would mean that a total of five amino acids – two upstream and two downstream of the mutant codon – would need to match in the human and mouse protein to be considered orthologous.). **b** Capabilities of prime editing (blue) and base editing (red) to model mutations that derive from a wild-type amino acid residue conserved in mice as a function of conserved flank size (that is, the stringency of homology). All calculations assume a base or prime editor that recognizes only NGG PAMs. **c** Quantification of the ability of base editing (top) and prime editing (bottom) to model orthologous mutations in mice at different flank size values (2, 3, 5, and 10 from left to right). These plots correspond with the data points in panel (b).
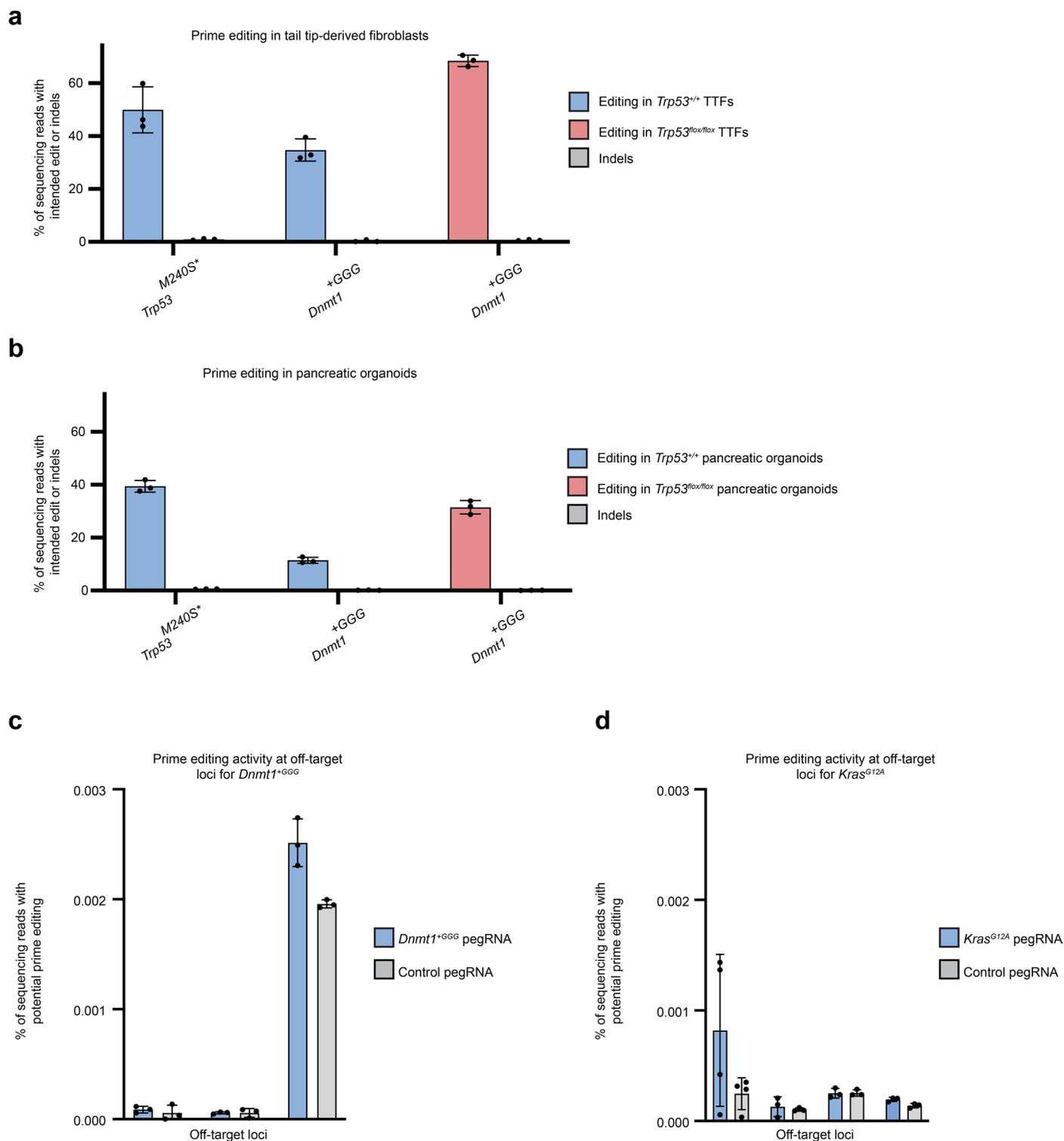
**Extended Data Fig. 3 | Prime editing in *Trp53*-proficient cells and off-target activity at computationally predicted loci. a** Prime editing in *Trp53*$^{flox/flox}$; *Rosa26*$^{PE2/+}$ versus *Trp53*$^{+/+}$;*Rosa26*$^{PE2/+}$ TTFs across several indicated loci. Data and error bars indicate the mean and standard deviation of three independent transductions. **b** Prime editing in *Trp53*$^{flox/flox}$;*Rosa26*$^{PE2/+}$ versus *Trp53*$^{+/+}$; *Rosa26*$^{PE2/+}$ pancreatic across several indicated loci. Note that the editing data for *Dnmt1*$^{+GGG}$ in *Trp53*$^{flox/flox}$ organoids are the same depicted in Fig. 2f. Data and error bars indicate the mean and standard deviation of three independent transductions. **c** Quantification of off-target prime editing for a pegRNA targeted

to *Dnmt1*. Off-targets represent three computationally predicted loci for tail-tip fibroblasts cultured for more than one month in the presence of the *Dnmt1*$^{+GGG}$ pegRNA or a control pegRNA. Data and error bars indicate the mean and standard deviation of three independent transductions. Off-target locus details are described in Supplementary Table 4. **d** Quantification of off-target prime editing at four computationally predicted loci for tail-tip fibroblasts cultured with a *Kras*$^{G12A}$ epegRNA or a control pegRNA. Data and error bars indicate the mean and standard deviation of three independent transductions.
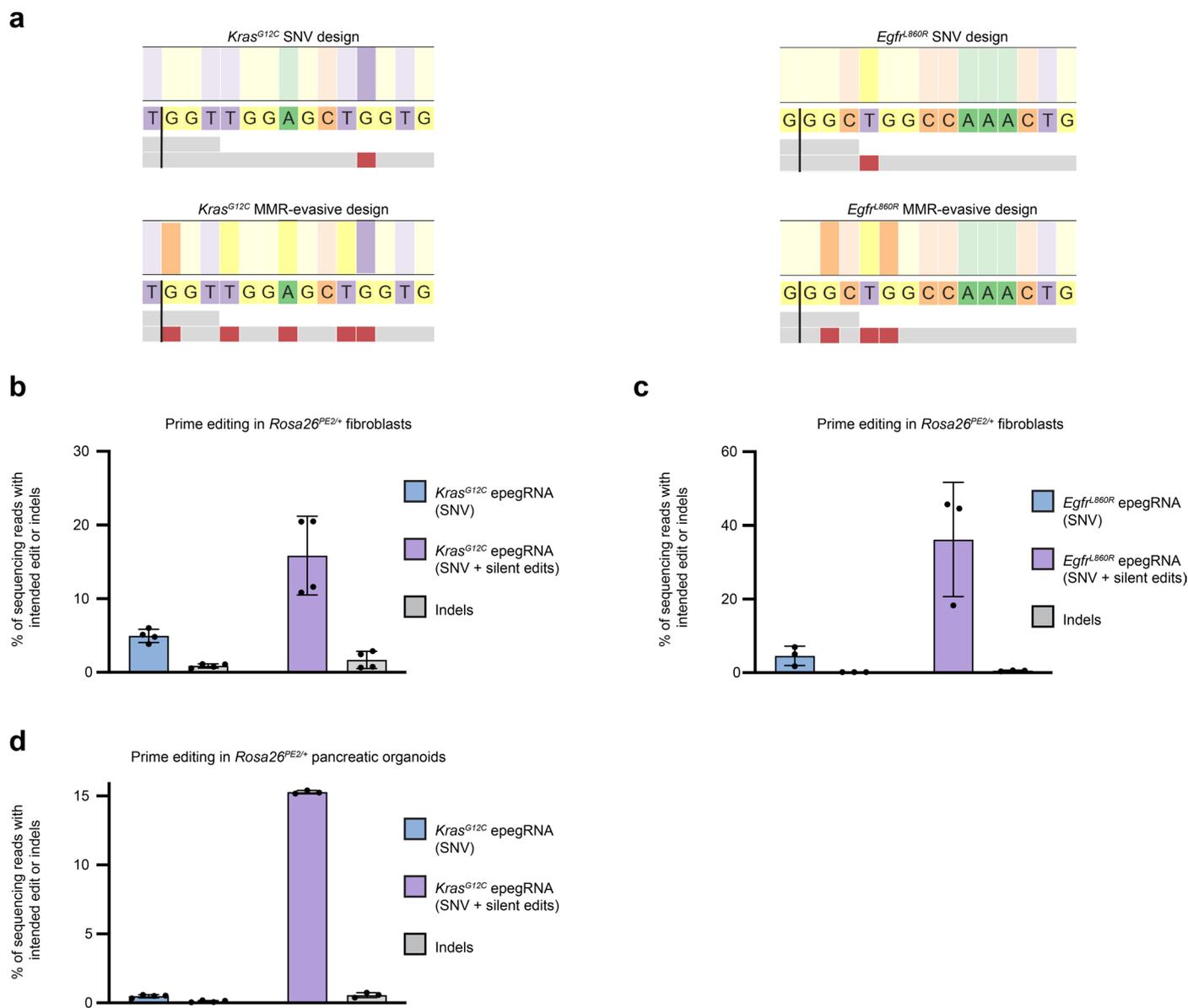
**Extended Data Fig. 4 | Mismatch repair-evasive edits amplify prime editing efficiency across multiple cancer-associated loci. a** Design of epegRNAs encoding *Egfr*<sup>L860R</sup> (right) or *Kras*<sup>G12C</sup> (left) with (bottom) or without (top) silent mutations intended to promote MMR evasion. The black line on the sequence schematics indicates the Cas9 nicking site. **b** Editing efficiency and indel byproduct frequency of epegRNAs encoding *Kras*<sup>G12C</sup> with or without silent mutations in tail tip-derived fibroblasts. Note that the data for *Kras*<sup>G12C</sup> without silent edits is also featured in Fig. 3. Data and error bars indicate the mean and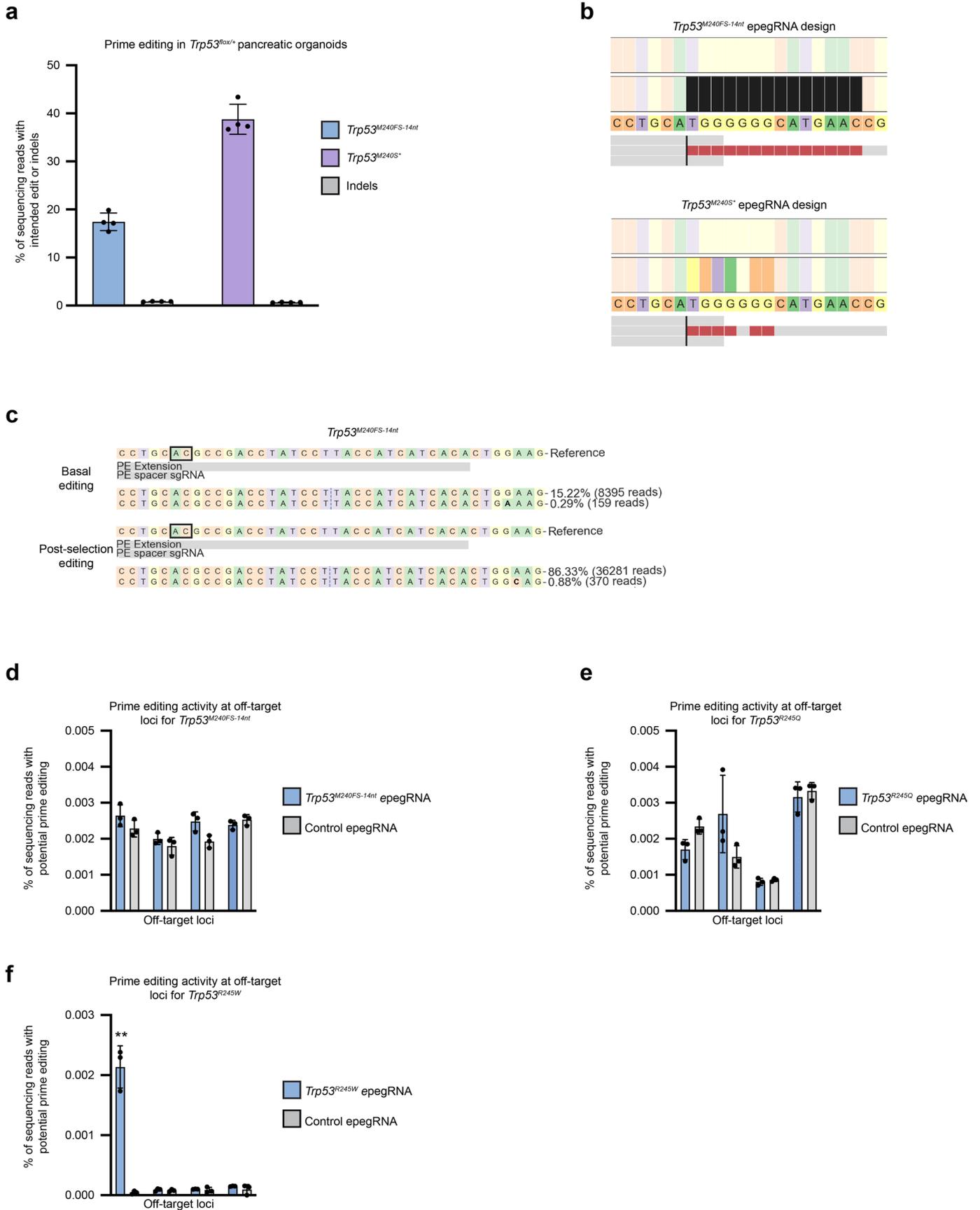 standard deviation of three independent transductions. **c** Editing efficiency and indel byproduct frequency of epegRNAs encoding *Egfr*<sup>L860R</sup> with or without silent mutations. Data and error bars indicate the mean and standard deviation of three independent transductions. **d** Editing efficiency and indel byproduct frequency of epegRNAs encoding *Kras*<sup>G12C</sup> with or without silent mutations in pancreatic organoids. Note that the data for *Kras*<sup>G12C</sup> without silent edits is also featured in Fig. 3. Data and error bars indicate the mean and standard deviation of three independent transductions.
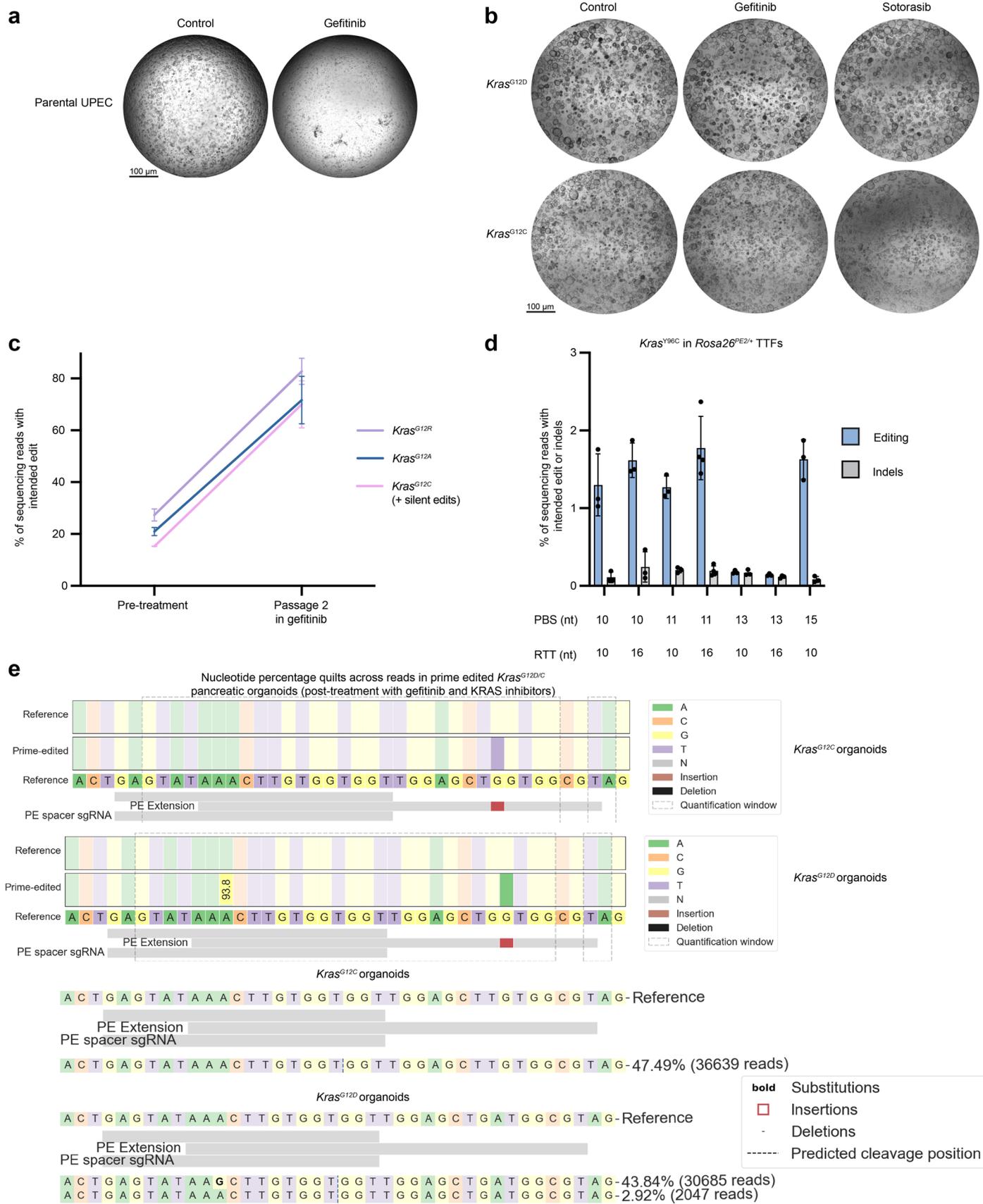
**Extended Data Fig. 5 | See next page for caption.**

**Extended Data Fig. 5 | Treatment of prime edited *Kras*[G12D] and *Kras*[G12C] pancreatic organoids with EGFR and KRAS inhibitors in the presence or absence of a secondary resistance mutation. a** Bright-field images of unedited *Rosa26*[PE2/+] pancreatic organoids treated with gefitinib after infection with the template UPEC vector (that is, lacking a pegRNA). Scale bars: 100 μm. **b** Additional bright-field images supplementing Fig. 3d and depicting prime edited *Kras*[G12D] and *Kras*[G12C] organoids treated either with DMSO, gefitinib, or sotorasib. Scale bars: 100 μm. **c** Allele frequencies of *Kras*[G12R], *Kras*[G12A], or *Kras*[G12C+silent edits] mutations in pancreatic organoids before and after two passages of treatment with gefitinib (1 μM). Data and error bars indicate the mean and standard deviation of three independent transductions pre- and post-treatment.
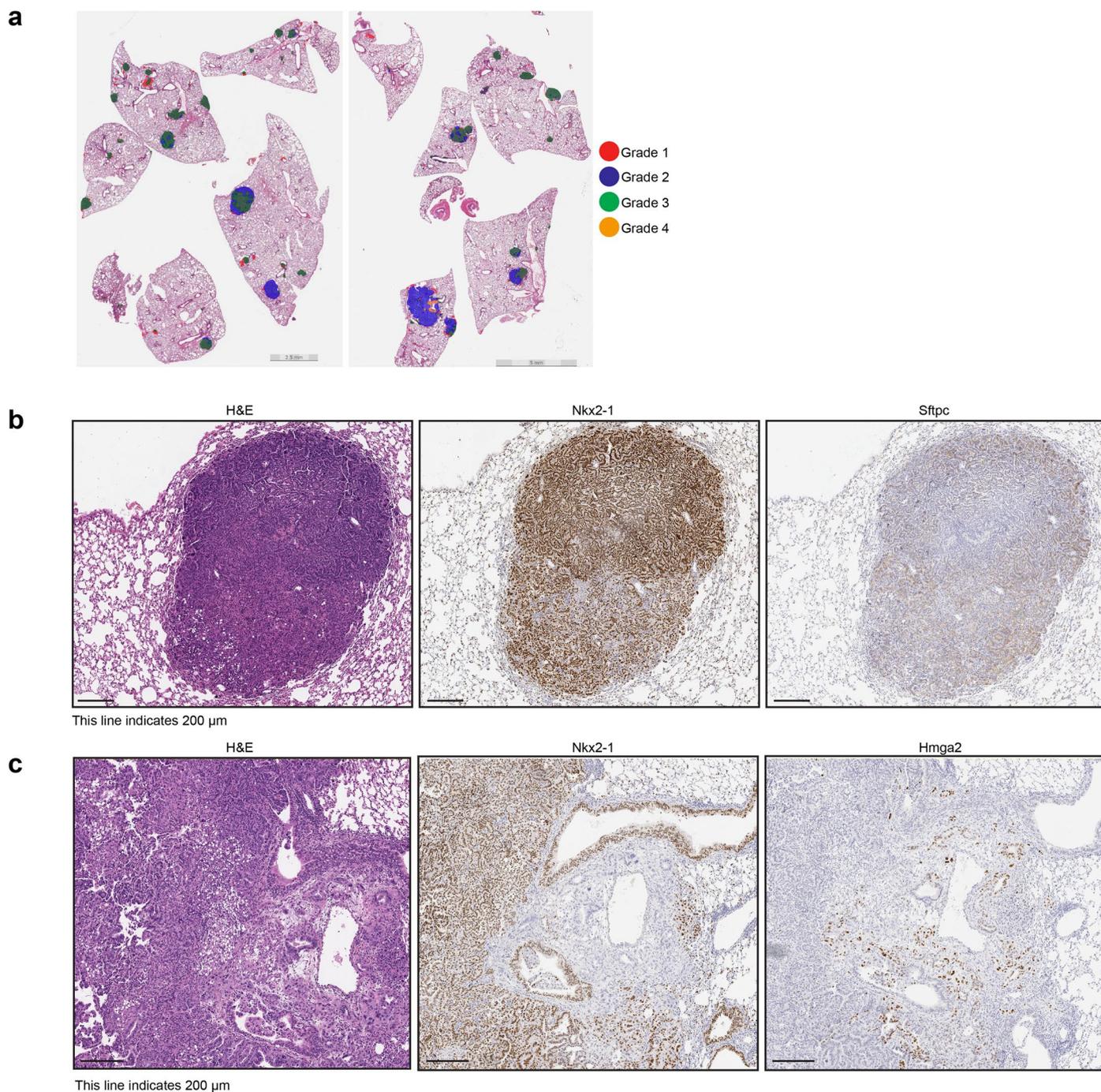
Note that pre-treatment replicate data is also represented in Extended Data Fig. 4 (*Kras*[G12C+silent edits]) and Fig. 3 (*Kras*[G12R] and *Kras*[G12A]). **d** Editing efficiency and indel frequencies of epegRNAs templating the *Kras*[Y96C] transversion in *Rosa26*[PE2/+] TTFs. Data and error bars indicate the mean and standard deviation of three independent transductions. **e** Nucleotide percentage quilts and allele frequency tables output by CRISPResso2 depicting sequencing reads derived from prime edited *Kras*[G12C] and *Kras*[G12D] pancreatic organoids. Prime edited alleles were previously selected with gefitinib, and these sequencing data derive from organoids after treatment with gefitinib in combination with mutant KRAS inhibitors.

**a**



Prime editing in *Trp53^flox/+^* pancreatic organoids

**b**



*Trp53^M240FS-14nt^* epegRNA design

*Trp53^M240S*^* epegRNA design

**c**



*Trp53^M240FS-14nt^*

**d**



Prime editing activity at off-target loci for *Trp53^M240FS-14nt^*

**e**



Prime editing activity at off-target loci for *Trp53^R245Q^*

**f**



Prime editing activity at off-target loci for *Trp53^R245W^*
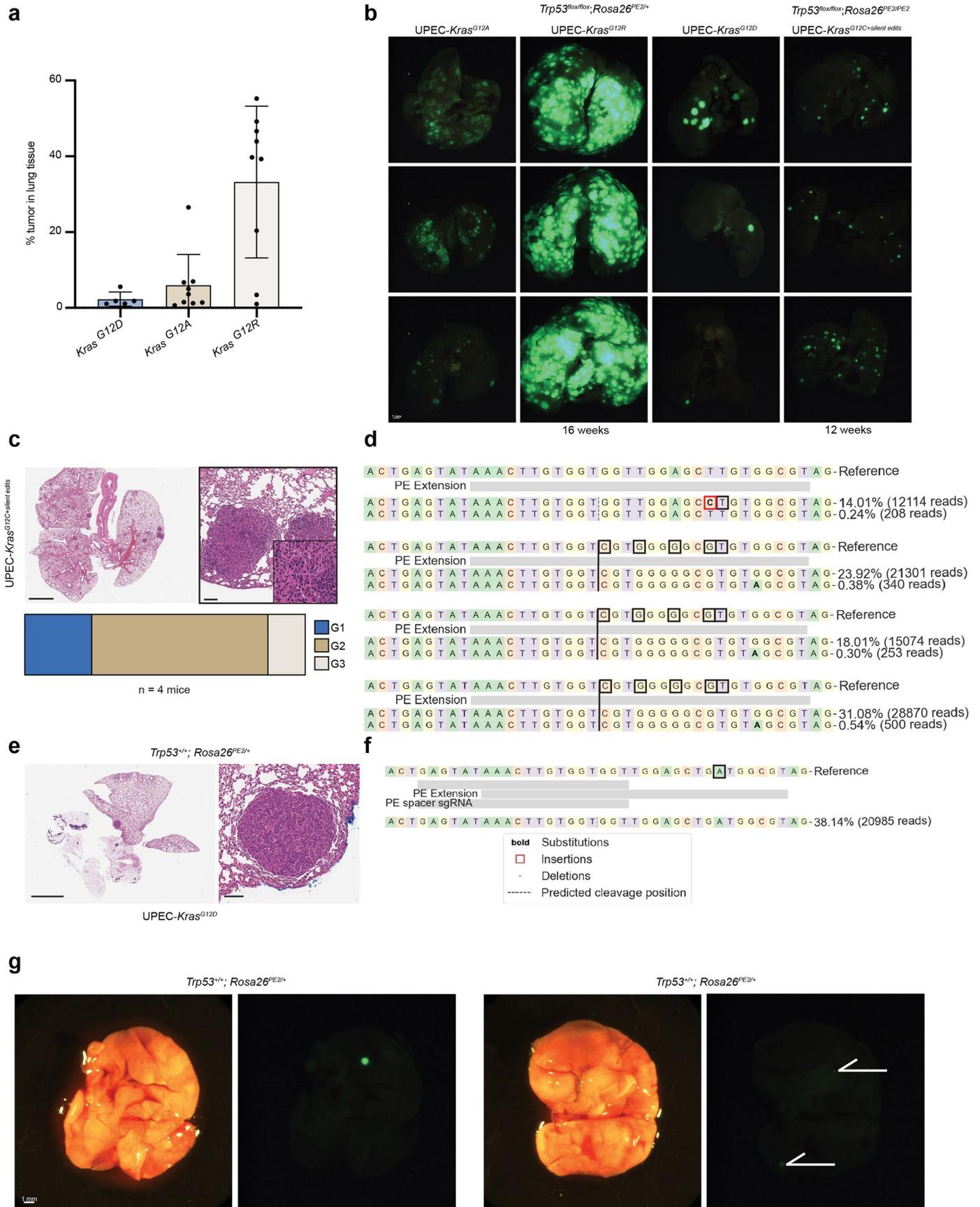
**Extended Data Fig. 6 | See next page for caption.**

**Extended Data Fig. 6 | On- and off-target prime editing at _Trp53_ in pancreatic organoids. a** Editing efficiency and indel byproduct frequency of epegRNAs encoding _Trp53_$^{M240FS\text{-}14nt}$ (n = 4 independent transductions) or _Trp53_$^{M240S^*}$ (n = 4 independent transductions) in _Trp53_$^{flox/+}$ pancreatic organoids. Data and error bars indicate the mean and standard deviation across independent transductions. **b** Design of the epegRNAs for _Trp53_$^{M240FS\text{-}14nt}$ or _Trp53_$^{M240S^*}$. The black line below the sequence schematics indicates the Cas9 nicking site. These epegRNAs were designed to both evade MMR and modify the seed and PAM sequences. **c** CRISPResso allele frequency plots of prime edited reads for _Trp53_$^{M240FS\text{-}14nt}$ before and after nutlin selection in pancreatic organoids. **d** Quantification of off-target prime editing for an epegRNA encoding _Trp53_$^{M240FS\text{-}14nt}$. Off-targets represent four computationally predicted loci for pancreatic organoids cultured for more than one month in the presence of the _Trp53_$^{M240FS\text{-}14nt}$ epegRNA and nutlin-3a or a control pegRNA. Data and

error bars indicate the mean and standard deviation of three independent transductions. **e** Quantification of off-target prime editing for an epegRNA encoding _Trp53_$^{R245Q}$. Off-targets represent four computationally predicted loci for pancreatic organoids cultured for more than one month in the presence of the _Trp53_$^{R245Q}$ epegRNA and nutlin-3a or a control pegRNA. Data and error bars indicate the mean and standard deviation of three independent transductions. **f** Quantification of off-target prime editing for an epegRNA encoding _Trp53_$^{R245W}$. Off-targets represent four computationally predicted loci for pancreatic organoids cultured for more than one month in the presence of the _Trp53_$^{R245W}$ epegRNA and nutlin-3a or a control pegRNA. Data and error bars indicate the mean and standard deviation of three independent transductions. Statistical significance was calculated using an unpaired, one-tailed Welch's $t$-test ($P = 0.0045$ for the leftmost off-target site).

**a**



- Grade 1
- Grade 2
- Grade 3
- Grade 4

**b**



H&E | Nkx2-1 | Sftpc

This line indicates 200 μm

**c**



H&E | Nkx2-1 | Hmga2

This line indicates 200 μm

**Extended Data Fig. 7 | Prime edited tumors recapitulate histological features of gold standard GEMMs of lung adenocarcinoma. a** Aiforia grade analysis of 16-week tumor-bearing lungs initiated by *in vivo* prime editing (red = Grade 1, green = Grade 2, blue = Grade 3, orange = Grade 4). **b** H&E and immunohistochemical (IHC) staining of a representative Grade 2 tumor 16 weeks post-initiation. IHC staining was performed with antibodies specific to Nkx2.1, a lung lineage transcription factor, or Surfactant Protein C (Sftpc), a marker of AT2 cells. Scale bars: 200 μm. **c** H&E and IHC staining of a Grade 3/4 tumor 16 weeks post-initiation. IHC staining was performed with antibodies specific to NKX2-1 or HMGA2, a lung embryonic chromatin regulator. Scale bars: 200 μm.

**a**



**b**



**c**



**d**



**e**



**f**



**g**



**Extended Data Fig. 8 | See next page for caption.**

**Extended Data Fig. 8 | Autochthonous modeling of diverse *Kras* variants in the lung. a** Tumor burden quantification using Aiforia grading algorithm for *Kras*^*G12D* (n = 5 mice), *Kras*^*G12A* (n = 9 mice), and *Kras*^*G12R* (n = 9 mice), tumors at 16 weeks. Tumor burden is defined as the ratio of tumor area (Grades 1–4) over total lung area. Data and error bars indicate the mean and standard deviation of five to nine biological replicates. **b** Representative fluorescence imaging of lungs from *Trp53*^*flox/flox*;*Rosa26*^*PE2/+* and *Trp53*^*flox/flox*;*Rosa26*^*PE2/PE2* mice infected with UPEC-*Kras*^*G12A* (n = 10), UPEC-*Kras*^*G12R* (n = 9), UPEC-*Kras*^*G12D* (n = 20), all harvested at 16 weeks, as well as mice infected with UPEC-*Kras*^*G12C+silent edits* harvested at 12 weeks. **c** H&E staining of representative tissue from a mouse infected with UPEC-*Kras*^*G12C+silent edits*. Tissue was collected at 12 weeks. Scale bars indicate 2 mm, 100 μm, and 20 μm respectively. Grading analysis indicates proportion of Grade 1, 2, and 3 lesions from four representative animals. **d** CRISPResso allele frequency plots depicting prime edited reads derived from lung tumors from a *Trp53*^*flox/flox*;*Rosa26*^*PE2/PE2* mouse infected with UPEC-*Kras*^*G12C* (top plot) and three different *Trp53*^*flox/flox*;*Rosa26*^*PE2/PE2* mice infected with UPEC-*Kras*^*G12C+silent edits*

(bottom three plots). The black line on the sequence schematics indicates the Cas9 nicking site. The intended mutations are indicated by black boxes. The red box in the top plot indicates an unintended, silent *Kras*^*A11A* substitution. The percentages indicated represent the percentage of all analyzed reads (edited + unedited reads) that bear this sequence. **e** H&E staining of representative tissue from a *Trp53*^*+/+*;*Rosa26*^*PE2/+* mouse infected with UPEC-*Kras*^*G12D*. H&E indicates histopathology consistent with a low grade adenoma. Scale bars indicate 2 mm (left) and 100 μm (right). **f** CRISPResso allele frequency plot depicting prime edited reads derived from an adenoma from a *Trp53*^*+/+*;*Rosa26*^*PE2/+* mouse infected with UPEC-*Kras*^*G12D*. The intended mutation is indicated by a black box. Note that the percentage indicated represents the percentage of all analyzed reads (including unedited reads) that bear this sequence. **g** Representative bright-field and fluorescence imaging of adenomas from two *Trp53*^*+/+*;*Rosa26*^*PE2/+* mice infected with UPEC-*Kras*^*G12D* (harvested at 10 weeks). White arrows indicate smaller fluorescent adenomas.
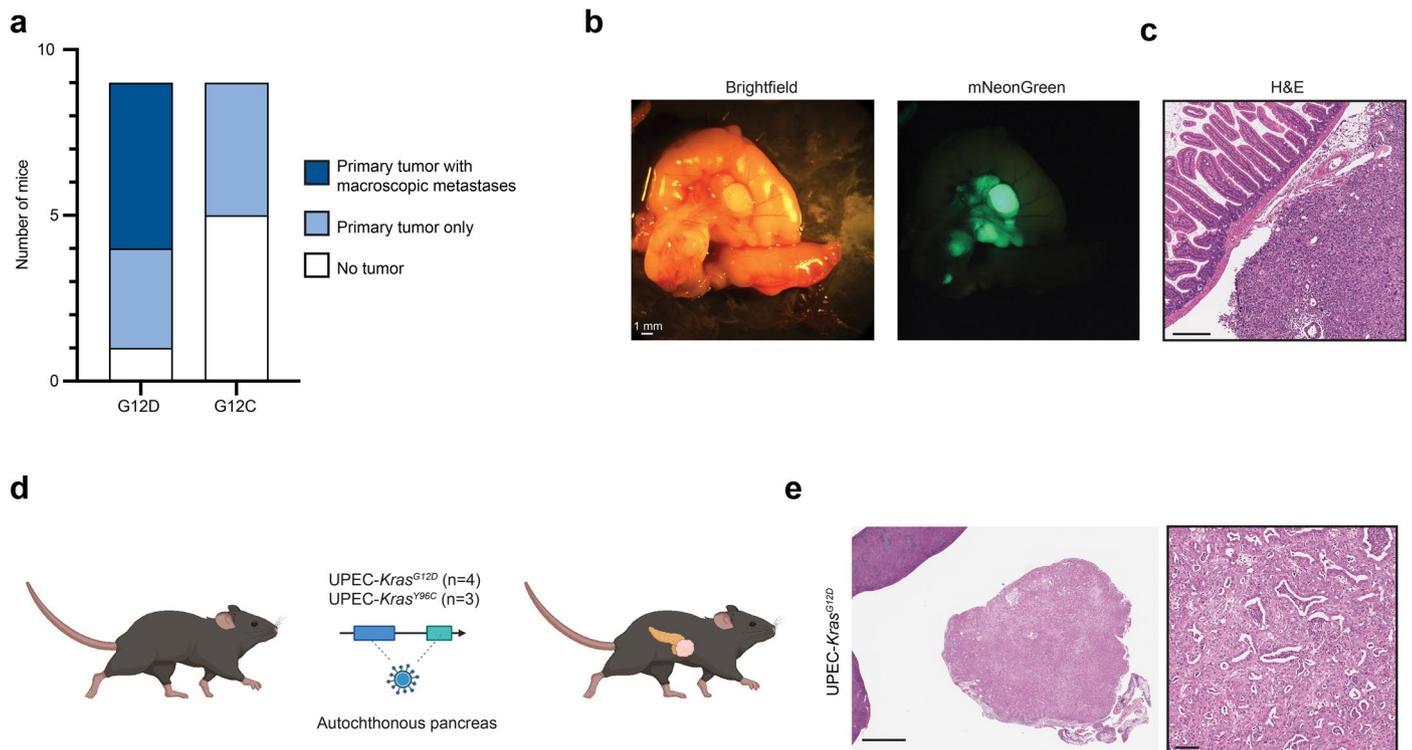
**Extended Data Fig. 9 | Modeling of pancreatic cancer through organoid transplantation and lentiviral initiation. a** Stacked bar plots indicating incidence of tumors and metastases in animals orthotopically transplanted with UPEC-*Kras*^G12D (n = 9) and UPEC-*Kras*^G12C (n = 9) prime-edited pancreatic organoids. **b** Bright-field and fluorescent imaging of an omental metastasis and surrounding tissue in an animal transplanted with UPEC-*Kras*^G12D prime-edited pancreatic organoids. **c** H&E staining of a representative omental metastasis.

Scale bar indicates 100 μm. **d** Experimental workflow for autochthonous pancreatic tumor modeling through retrograde pancreatic ductal infection of lentivirus templating *Kras*^G12D (n = 4) or *Kras*^Y96C (n = 3) pegRNAs. **e** Representative H&E staining of autochthonous pancreatic adenocarcinoma initiated through pancreatic ductal infection of UPEC-*Kras*^G12D. Scale bars indicate 1 mm and 50 μm, respectively.

# nature portfolio

Corresponding author(s): Tyler Jacks

Last updated by author(s): April 2, 2023

# Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

## Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

| n/a | Confirmed | |
|---|---|---|
| ☐ | ☒ | The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement |
| ☐ | ☒ | A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ | The statistical test(s) used AND whether they are one- or two-sided<br>*Only common tests should be described solely by name; describe more complex techniques in the Methods section.* |
| ☒ | ☐ | A description of all covariates tested |
| ☒ | ☐ | A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| ☐ | ☒ | A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| ☐ | ☒ | For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted<br>*Give P values as exact values whenever suitable.* |
| ☒ | ☐ | For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| ☒ | ☐ | For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| ☒ | ☐ | Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated |

*Our web collection on statistics for biologists contains articles on many of the points above.*

## Software and code

Policy information about availability of computer code

| Data collection | Illumina MiSeq, Guava easyCyte, BD FACSCelesta, LSR-II, Fortessa I, Tecan Infinite Pro m200. Scripts used to perform prime editing and base-editing analysis on MSK IMPACT data are detailed at https://github.com/samgould2/ prime-vs-base-editing. The pipeline, all related scripts, and intermediate data needed to reproduce our results are available. CRISPResso2 commands for editing analyses are described in the Supplementary Note and Methods. |
|---|---|
| Data analysis | CRISPResso2, V2.2.6, FlowJo v10.4.2, BD FACS Diva v8, Qupath v.0.2.3, Graphpad PRISM v9.0, ImageJ v.2.1.0/1.53c, Lung adenocarcinoma tumor burden quantification was performed using a previously published convoluted neural network developed by Aiforia (Aiforia Technologies Oy). A list of cancer-associated GEMMs was queried using MouseMine (MGI 6.22). Some schematics (Extended Data 9, Main Figures 2, 3. Supplementary Figures 3, 7) were made with Biorender (version number not available). Mean and standard deviation were calculated using Graphpad PRISM v9.0. |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

## Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

All amplicon sequencing data described in this manuscript are deposited at SRA under BioProject ID PRJNA951647. Plasmids will be made available at Addgene. Mice are available through the Jackson Laboratory. MouseMine data was available through the Mouse Genome Database (MGI 6.22)

## Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

| | |
|---|---|
| Reporting on sex and gender | N/A |
| Population characteristics | N/A |
| Recruitment | N/A |
| Ethics oversight | N/A |

Note that full information on the approval of the study protocol must also be provided in the manuscript.

# Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences          ☐ Behavioural & social sciences          ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

# Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

| | |
|---|---|
| Sample size | Sequencing experiments were performed with transduction triplicate (n=3) unless otherwise specified to ensure reproducibility. Animal study sample sizes are listed in figure legends. All animal experiments were performed in at least triplicate (n=3). For larger comparative studies, animal experiment sizes were determined by mouse availability and based on practice from previous reports (Westcott et al. Nature Cancer 2021). |
| Data exclusions | Data were not excluded from the analyses |
| Replication | All sequencing experiments were performed with three independent transduction replicates unless otherwise specified. Comparative mouse experiments including autochthonous lung and orthotopic pancreas experiments were done with at least n=5 biological replicates. Proof-of-principle mouse experiments, e.g. lentiviral initiation of autochthonous PDAC or p53-proficient editing in the lung were performed with at least n=3 biological replicates. All attempts at replication were successful. |
| Randomization | Randomization was not appropriate to animal experiments as mice were age-, litter-, and sex-matched to the best of our efforts to ensure consistency. |
| Blinding | Blinding was not explicitly performed. In comparative animal studies, tumor burden and grade quantification was performed in an unbiased fashion using a convoluted neural network developed by Aiforia (TM). For other analyses of tumors (e.g. quantification of prime edited sequencing reads), blinding was not applicable as they were not subject to subjective investigator interpretation. |

# Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

## Materials & experimental systems

| n/a | Involved in the study |
|-----|----------------------|
| ☐ | ☒ Antibodies |
| ☐ | ☒ Eukaryotic cell lines |
| ☒ | ☐ Palaeontology and archaeology |
| ☐ | ☒ Animals and other organisms |
| ☒ | ☐ Clinical data |
| ☒ | ☐ Dual use research of concern |

## Methods

| n/a | Involved in the study |
|-----|----------------------|
| ☒ | ☐ ChIP-seq |
| ☐ | ☒ Flow cytometry |
| ☒ | ☐ MRI-based neuroimaging |

# Antibodies

| Antibodies used | Immunohistochemistry: anti-NKX2-1 (1:1000, abcam, Cat# ab76013; RRID:AB_1310784), anti-SFTPC (1:5000, Millipore Sigma, Cat# AB3786; RRID:AB_91588), and anti-HMGA2 (1:1000, Cell Signaling Technologies, Cat# 8179S; RRID:AB_11178942) <br> Flow Cytometry: <br> CD31-APC (1:500, Biolegend, 102507) <br> EpCAM-PE (1:500, Biolegend, 118206) <br> MHCII-APC-eFluor-780 (1:500, Thermo, 47-5321-82) <br> Immunofluorescence: anti-Cas9 <br> (E7M1H, 1:100, CST Cat#19526) <br> Western Blotting: (p53 clone 1C12, Cell Signaling Technology <br> (CST)), β-Actin (clone 13E5, CST, 1:5000) <br> Anti-rabbit IgG HRP-linked antibody (#7074 Cell Signaling Technology) |
|---|---|
| Validation | Antibodies were purchased and used according to manufacturer recommendations. Per manufacturers, anti-NKX2-1 was validated on normal mouse lung tissue, anti-SFPTC was validated on mouse lung epithelial cells, anti-HMGA2 was validated on human colon carcinoma cells, anti-Cas9 was validated on brain tissue from Cas9-expressing mice, anti-p53 was validated on HT-29 cells. Biolegend antibodies are validated by staining of relevant tissues per manufacturer website. Anti-MHCII was validated by staining of splenocytes from C57Bl/6 mice. No additional validation was performed. |

# Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

| Cell line source(s) | Organoid and fibroblast lines were developed by the authors using the pancreata or lung tissue of mice. Neuro2 and HEK293 were ordered from ATCC. |
|---|---|
| Authentication | Authentication of genotypes of relevant organoid lines was performed by PCR genotyping. HEK-293 cells were not authenticated. |
| Mycoplasma contamination | Cell lines tested negative for Mycoplasma through the High-Throughput Core at the Koch Institute |
| Commonly misidentified lines <br> (See ICLAC register) | No ICLAC registered cell lines were used in this study. HEK293 cells (not HEK) were used for lentivirus production. |

# Animals and other research organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research, and Sex and Gender in Research

| Laboratory animals | Mus musculus, C57Bl/6, mice from 6-20 weeks were used for tumor initiation studies until endpoint, relevant genotypes include: <br> LSL-KrasG12D C57Bl/6J <br> Trp53(fl); C57Bl/6 <br> ROSA26PE (this manuscript); C57Bl/6 <br> Pdx1-Cre; C57Bl/6 <br> Villin-CreERT2; Mixed C57Bl/6-129Sv <br><br> Mice were housed in a facility with a 12-hour light/12-hour dark cycles with temperatures between 68-72F and 30-70% humidity. |
|---|---|
| Wild animals | No wild animals were used in this study. |
| Reporting on sex | Animal experiments were sex-matched to the best of our abilities with the exception of orthotopic organoid experiments. The donor organoid line was derived from a male animal, necessitating the use of only male recipient animals. |
| Field-collected samples | No field collected samples were used in this study |
| Ethics oversight | The MIT Institutional Animal Care and Use Committee oversees animal research performed in the Jacks lab. The approved protocol number is 0720-066-23. |

# Flow Cytometry

## Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

## Methodology

| | |
|---|---|
| Sample preparation | Sample organoids or adherent cells were trypsinized with TrypLE Express (Life Technologies) or 0.25% Trypsin-EDTA (Invitrogen) and washed 2X in PBS. Samples were analyzed on the flow cytometer in a buffer of 5% FBS in PBS. |
| Instrument | Samples were analyzed using a BD FACSCelesta, LSR-II, or Fortessa I. |
| Software | Sample data were visualized using BD FACS Diva and analyzed using FlowJo v10 |
| Cell population abundance | Cell population abundance is available in the gating strategy in Supplementary Figure 9 for the relevant flow cytometry experiment. To determine GFP+ of transduced Rosa26PE2 organoids, untransduced organoids were used as a negative control to draw gates for GFP positivity. |
| Gating strategy | Single cells were gated on FSC/SSC, the GREEN channel (Guava easyCyte) was used to quantify mNeonGreen positivity of pancreatic organoids. |

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.