nature biotechnology

ARTICLES

https://doi.org/10.1038/s41587-022-01255-9

A split prime editor with untethered reverse transcriptase and circular RNA template

Bin Liu[1,5], Xiaolong Dong[1,5], Haoyang Cheng[1], Chunwei Zheng[1], Zexiang Chen[1], Tomás C. Rodríguez[1], Shun-Qing Liang[1], Wen Xue[1,2,3,4] ✉ and Erik J. Sontheimer[1,3,4] ✉

Delivery and optimization of prime editors (PEs) have been hampered by their large size and complexity. Although split versions of genome-editing tools can reduce construct size, they require special engineering to tether the binding and catalytic domains. Here we report a split PE (sPE) in which the Cas9 nickase (nCas9) remains untethered from the reverse transcriptase (RT). The sPE showed similar efficiencies in installing precise edits as the parental unsplit PE3 and no increase in insertion–deletion (indel) byproducts. Delivery of sPE to the mouse liver with hydrodynamic injection to modify β-catenin drove tumor formation with similar efficiency as PE3. Delivery with two adeno-associated virus (AAV) vectors corrected the disease-causing mutation in a mouse model of type I tyrosinemia. Similarly, prime editing guide RNAs (pegRNAs) can be split into a single guide RNA (sgRNA) and a circular RNA RT template to increase flexibility and stability. Compared to previous sPEs, ours lacks inteins, protein–protein affinity modules and nuclease-sensitive pegRNA extensions, which increase construct complexity and might reduce efficiency. Our modular system will facilitate the delivery and optimization of PEs.

PEs enable deletion, insertion and base substitution without double-strand breaks[1]. However, the current PE2, PE2* and PEmax effectors (fusions of nCas9 with Moloney murine leukemia virus RT (M-MLV RT))[1–3] are >6.3 kilobases (kb), which is beyond the packaging capacity of an AAV. Production of such a large protein or mRNA in high yield (for ribonucleoprotein (RNP) or RNA delivery) can also be challenging. Although some split strategies have been tested for the delivery of Cas9-associated genome-editing tools[4], including split inteins[5–7] and MS2 (refs. [8–10]) or SunTag[11] tethers, most split approaches are only beginning to be applied to PEs[2,12,13]. These elements add size, molecular complexity and production and delivery burdens to PE systems, and they limit the combinatorial throughput of PE development (that is, the mixing and matching of nuclease and RT components).

pegRNA optimization is also important for effective prime editing. The current pegRNA is a conjugated RNA that consists of an sgRNA and a 3′ extension containing the RT template (RTT) and the primer-binding site (PBS). Despite the simplicity of having an integrated RNA molecule in the PE system, it is prone to RNA misfolding due to the inevitable base pairing between the PBS and the spacer and potential RTT–scaffold interactions. Finally, the 3′-terminal extension in the pegRNA is exposed and susceptible to degradation by nucleases, which may compromise the integrity of the pegRNA. Although 3′-terminal secondary structures improve pegRNA stability[14], further efforts to reduce pegRNA misfolding and instability are needed.

Results

An sPE with untethered RT mediates prime editing. To facilitate PE2 delivery, we initially designed MS2–PE2 and SunTag–PE2 (Extended Data Fig. 1). To construct MS2–PE2, we fused MS2 coat protein (MCP) to the N terminus of M-MLV RT, and we engineered multiple MS2-pegRNAs by incorporating MS2 stem-loops into different positions of the sgRNA. We also tested splitting the PE via the SunTag system by fusing the single-chain variable fragment (scFv) to the N terminus of M-MLV RT. Subsequently, the scFv–RT was recruited by either GCN4–nCas9 or nCas9–GCN4. We designated these two new prime-editing systems as SunTag–PE2 (GCN4–nCas9) and PE2–SunTag (nCas9–GCN4) based on domain order. As controls for both MS2- and SunTag-based split PEs, we also devised a direct sPE system (Fig. 1a,b) in which the respective tethers were omitted. All of these platforms were used in PE3 format, which differs from PE2 only by the inclusion of an additional sgRNA that directs nicking of the unedited strand, thereby biasing repair[1]. We first cotransfected the respective nCas9-, RT-, pegRNA- and nicking sgRNA-expressing plasmids into an HEK293T-derived mCherry reporter lentivector-transduced cell line with a premature TAG stop codon[2] (Extended Data Fig. 2) that can be reverted to wild type, yielding red fluorescence. The most potent MS2- and SunTag-tethered configurations (Extended Data Fig. 2) were comparable in editing efficiency to PE3 (Fig. 1c). Notably, the untethered sPE also generated edited (mCherry$^+$) cells with similar efficiency as the fused (PE3) and tethered (MS2–PE3-1.1 and SunTag–PE3) systems (Fig. 1c). We confirmed these results with a separate set of pegRNAs that drive a 3-nucleotide (nt) insertion in an endogenous locus (*HEK3*). Sanger sequencing revealed that sPE editing efficiency was again very similar to that of PE3 and higher than those of MS2–PE3-1.1 and SunTag–PE3 (Extended Data Fig. 2). A titration of the RT plasmid indicated that it can be transfected at lower levels (Extended Data Fig. 2) with little effect on efficiency with the mCherry reporter.

To test the efficiencies of MS2–PE3-1.1, SunTag–PE3 and sPE with larger insertions or deletions, we tested them alongside PE3 with two 'traffic light reporters' in which GFP is activated by repair of a disruption sequence (an 18-bp replacement of a 39-bp insertion or the removal of a 47-bp insertion; Extended Data Fig. 3)[2].

[1]RNA Therapeutics Institute, University of Massachusetts Chan Medical School, Worcester, MA, USA. [2]Department of Molecular, Cell and Cancer Biology, University of Massachusetts Chan Medical School, Worcester, MA, USA. [3]Li Weibo Institute for Rare Diseases Research, University of Massachusetts Chan Medical School, Worcester, MA, USA. [4]Department of Molecular Medicine, University of Massachusetts Chan Medical School, Worcester, MA, USA. [5]These authors contributed equally: Bin Liu, Xiaolong Dong. ✉e-mail: wen.xue@umassmed.edu; erik.sontheimer@umassmed.edu

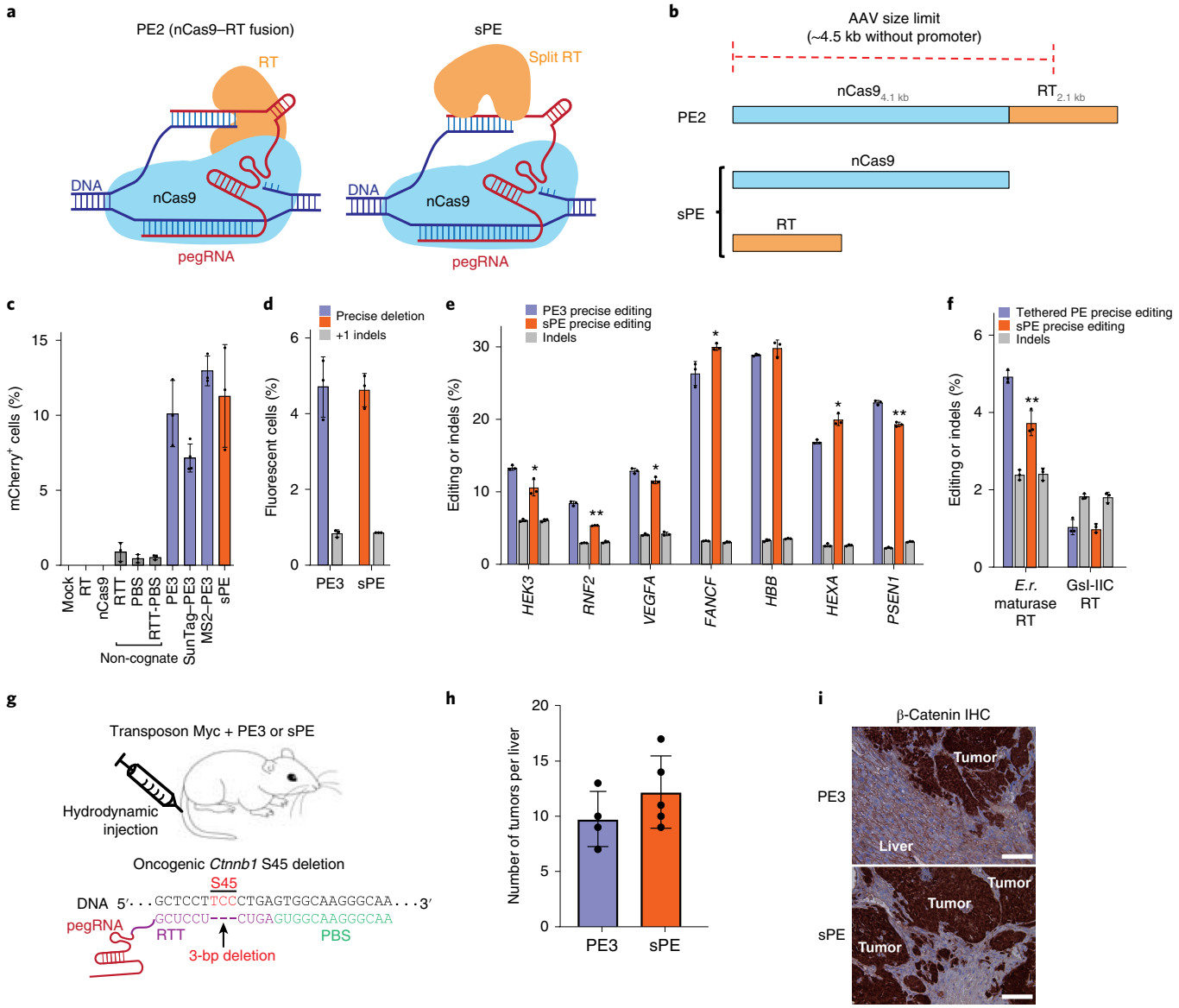


Fig. 1 | sPE enables genome editing in cells and in adult mouse liver. a, PE consists of nCas9 fused to RT. The sPE is expressed from two separate open reading frames. **b**, sPE components will each fit into an AAV vector without inteins. **c**, mCherry reporter cells were transfected with the indicated plasmids. **d**, Prime editing of a green fluorescent protein (GFP) reporter line by deletion of 47 base pairs (bp) for restoring GFP expression using PE3 or sPE. Indels (+1) can restore mCherry expression. **e**, Amplicon sequencing of PE3 and sPE for multiple endogenous sites in HEK293T cells. **f**, Prime editing by alternative RT orthologs at the *FANCF* site by 3-nt substitutions (+2 C to T and +4–5 TG to AC). Human codon-optimized *Eubacterium rectale* (*E.r.*) maturase RT and GsI-IIC RT were cloned into the original PE2 in place of the M-MLV RT. Nicking sgRNAs (PE3 format) were used for all prime editing. Data in **c**–**f** are derived from three or more independent biological replicates (mean ± s.d.); two-tailed unpaired Student's *t*-test; $^{*}P < 0.05$; $^{**}P < 0.01$. **g**, PE induces oncogenic activation of *Ctnnb1* by a 3-bp in-frame deletion. PE3 ($n = 4$) and sPE ($n = 5$) were delivered into mouse liver by hydrodynamic tail vein injection along with the Myc transposon and transposase plasmids. **h**, Tumor numbers in livers were counted 25 d after injection with PE3 ($n = 4$) or sPE ($n = 5$) (mean ± s.d.). **i**, Immunohistochemistry (IHC) staining shows nuclear β-catenin in liver tumors. Each experiment was repeated at least three times independently with similar results; scale bars, 100 μm (×20 objective). Data and error bars indicate the mean and s.d. of three independent biological replicates.

We found that the sPE was just as efficient as PE3 with the precise 47-bp deletion (Fig. 1d) and that both MS2–PE3 and SunTag–PE3 were nearly as efficient as PE3 with the two reporters (Extended Data Fig. 3). More comprehensive evaluations of the editing efficiency of MS2–PE2 and SunTag–PE2 in different reporter lines and multiple endogenous sites revealed that the overall editing efficiency of MS2–PE2 and SunTag–PE2 is generally lower than sPE (Supplementary Note and Extended Data Fig. 4). We therefore chose the sPE system to pursue further.

Next, we compared sPE and PE3 for editing at multiple endogenous sites. sPE achieved comparable levels of editing as PE3 (up to 30%; Fig. 1e). To test the interchangeability of the RT component, we replaced the M-MLV RT in the PE3 and in the sPE with two human codon-optimized bacterial RTs (Fig. 1f and Extended Data Fig. 5), *E.r.* maturase RT (MarathonRT)[15] and GsI-IIC RT (TGIRT III)[16,17], both of which are compact in size compared to M-MLV RT. PEs with these RTs successfully worked in tandem with pegRNAs that install 3-nt substitutions at *FANCF* and *VEGFA* sites, albeit with reduced

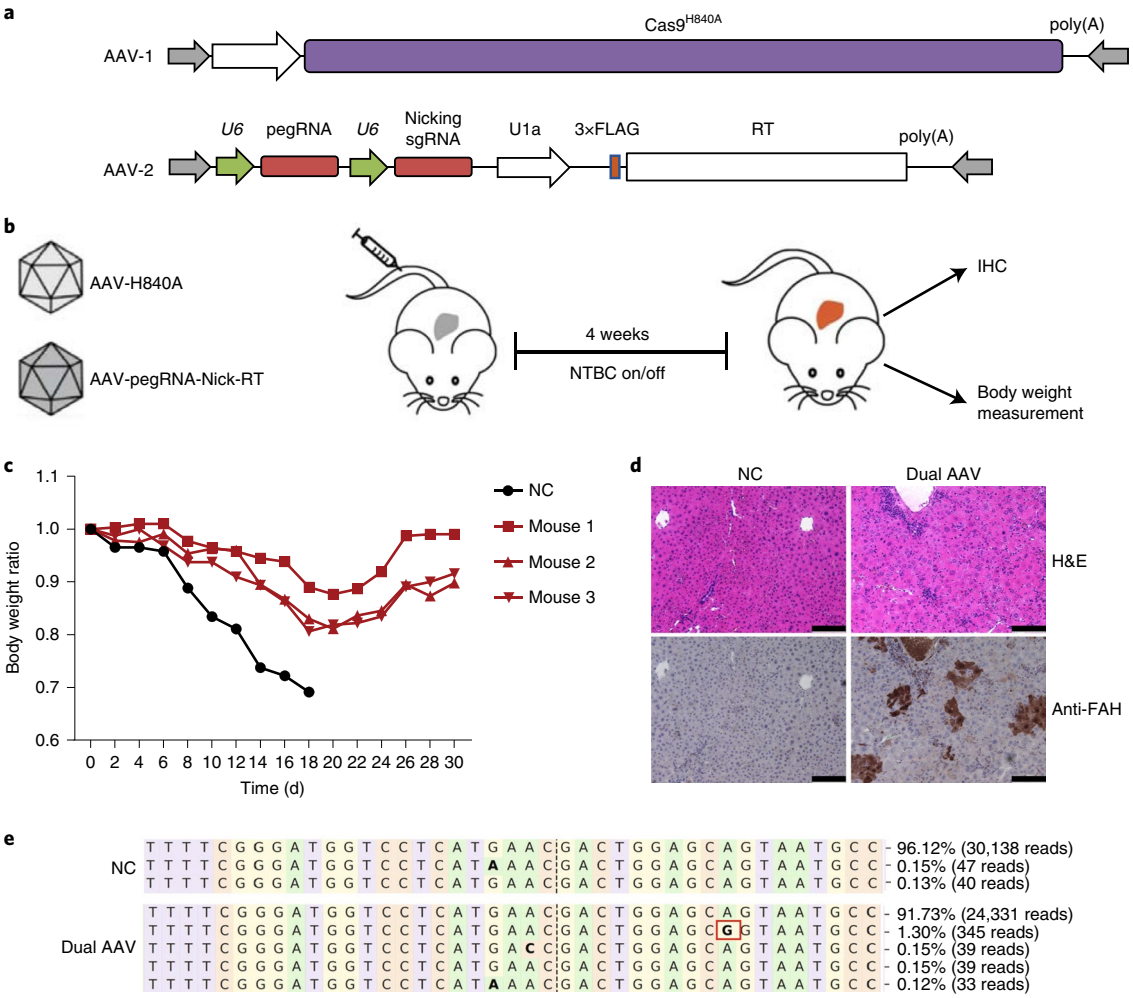


Fig. 2 | sPE dual AAV rescues weight loss in *Fah*-mutant mice. a, Schematic of the sPE dual AAV PE. Cas9^{H840A} and M-MLV RT were subcloned into two AAV8 vectors. **b**, Schematic of the in vivo experiments. The dual AAV8 sPE (1×10^{12} viral genomes each) was delivered to 6-week-old *Fah*-mutant mice via tail vein injection. **c**, Dual AAV rescued body weight loss after NTBC withdrawal. Body weight ratio is normalized to day 0 of NTBC withdrawal. Mice 1–3 are three mice treated with dual AAV; NC, non-treated control. **d**, Hematoxylin and eosin (H&E) staining and FAH IHC staining ($n=3$ mice; 24 d without NTBC); scale bars, 100 μm (×20 objective). **e**, Amplicon sequencing from representative animals in **d** using liver genomic DNA. The corrected 'G' by sPE is boxed in red.

efficiency compared to the PE2 effector (Fig. 1f and Extended Data Fig. 5). Notably, the two RT orthologs also mediated editing in the sPE design with various efficiencies (Fig. 1f and Extended Data Fig. 5). These results indicate that alternative RTs are applicable to the sPE system. Future efforts to improve thermostability and processivity of RT, facilitated by the sPE format, may enable more efficient PE designs. We also generated two PE2 mutants carrying either Cas9^{D10A} (dCas9-RT) or RT$^{\Delta YVDD}$ (nCas9-mutRT) mutations. Both were deficient in prime editing at the *FANCF* locus (Extended Data Fig. 6), as expected. Cotransfection of nCas9-mutRT and RT constructs resulted in successful prime editing, whereas cotransfection of nCas9-mutRT and dCas9-RT did not (Extended Data Fig. 6). These data suggest that, in contrast to sPE, nCas9 and RT from separate PE2 proteins do not function *in trans* at a given nick site, possibly due to local molecular clashes. Together, these data suggest that sPE has similar capabilities as PE2 in prime editing in mammalian cells.

Given the efficient genome editing by sPE in cells, we further evaluated sPE in vivo. As precise microhomology-independent and homology-directed repair (HDR)-independent deletion is a unique ability of prime editing, we used sPE to delete the S45 codon in *Ctnnb1* (β-catenin) to drive tumor formation in adult FVB mice via hydrodynamic tail vein injection (Fig. 1g). Four weeks after injection, PE3 rendered an average of 9.75 ± 2.5 tumors per mouse ($n=4$), whereas the sPE induced 12.2 ± 3.3 tumors per mouse ($n=5$; Fig. 1h and Extended Data Fig. 7). IHC staining confirmed the oncogenic activation of β-catenin in the liver tumors (Fig. 1i), and amplicon deep sequencing confirmed the precise 3-bp deletion in *Ctnnb1* in both PE3- and sPE-treated groups (Extended Data Fig. 7). Therefore, the sPE enables precise and highly efficient prime editing in vivo.

To test whether modular PE systems are compatible with AAV delivery, we cloned dual-AAV vectors of sPE (Fig. 2a). The first AAV expresses nCas9, and the second AAV expresses RT, a pegRNA and a nicking sgRNA described previously[18]. Both vectors are within the ~4.8-kb packaging capacity of AAV. We chose a mouse model of tyrosinemia type I caused by a loss-of-function G-to-A mutation of fumarylacetoacetate hydrolase (FAH; Fig. 2b). Loss of FAH, a key enzyme of the tyrosine catabolic pathway[19,20], leads to hepatocyte toxin accumulation and liver damage. *Fah*-mutant mice need to be kept on 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC)[20]-supplemented water to prevent

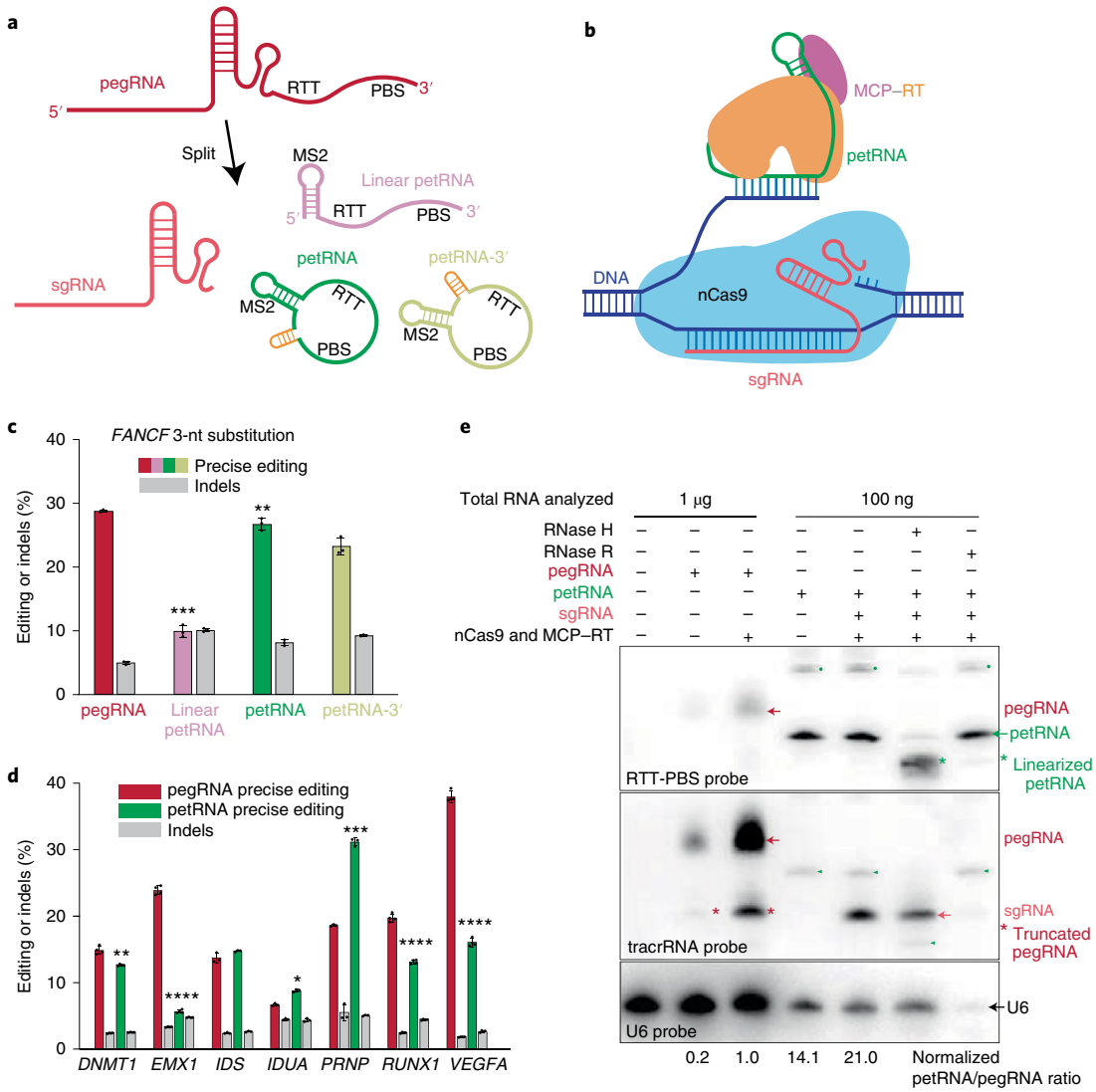


Fig. 3 | Prime editing by modular RNA components. a, Illustration of petRNA. The RTT-PBS sequence was arranged with an MS2 hairpin on the 5′ side in an independent RNA with or without circularization via a ribozyme-enabled pathway to give rise to petRNA and its linear counterpart (linear petRNA), respectively. petRNA-3′ was also tested. The petRNA acts at the nick site in lieu of the pegRNA PBS-RTT region to recruit an MCP-fused RT and serves as RT template. **b**, Model of petRNA function in PE. **c**, PE efficiency of petRNA at the *FANCF* locus. Plasmids expressing the pegRNA or modular petRNA + sgRNA were cotransfected with nCas9 and MCP–RT. PE introduces a 3-nt substitution (+2 C to T and +4–5 TG to AC). **d**, Validation of petRNA function at seven endogenous loci. Nicking sgRNAs were used for all prime editing. Data are derived from three or more independent experiments (mean ± s.d.) unless otherwise indicated; two-tailed unpaired Student's *t*-test; $^{*}P < 0.05$; $^{**}P < 0.01$; $^{***}P < 0.001$; $^{****}P < 0.0001$. **e**, Representative Northern blot demonstrating the integrity and abundance of pegRNAs and petRNAs (*FANCF* 3-nt substitution) 72 h after transfection in HEK293T cells. Total RNAs containing petRNA were loaded in 10× lower amounts than pegRNA groups to allow for better comparison between RNA signal. ^{32}P-labeled probes were used to hybridize to RTT-PBS (top), trans-activating CRISPR RNA (tracrRNA) (middle) or U6 RNA (bottom). Putative cleavage or truncation products are indicated by asterisks. U6 snRNA, sgRNA, pegRNA and petRNA bands are indicated by color-matched arrows. Circular byproducts of the petRNA (top) are indicated by green dots. A non-specific band in the middle is indicated by green triangles. Relative pegRNA and petRNA band intensities (RTT-PBS probe) are normalized to U6 and are shown at the bottom. For the last two samples, RNase H and RNase R demonstrated the circular nature of the petRNA. Three biological repeats were performed with similar results. Data and error bars indicate the mean and s.d. of three independent biological replicates.

body weight loss. In the *Fah*-mutant mice, intravenous injection of dual AAVs rescued body weight loss (Fig. 2c). Analysis of additional animals killed 24 d after NTBC withdrawal revealed FAH^{+} hepatocytes (Fig. 2d) and molecular confirmation of precision editing via deep sequencing (Fig. 2e). Editing efficiency by this delivery approach was relatively modest (1.3%), pending further pegRNA and nicking sgRNA optimization.

Given the success of the untethered sPE effector, we considered analogous opportunities for modular organization of the RNA components (Extended Data Fig. 8). We split the pegRNA into sgRNA and a separate RTT-PBS sequence (Fig. 3a). An MS2 aptamer was added to the RTT-PBS sequence to tether it to an MCP–RT fusion protein. Inspired by a previously reported method to increase small RNA abundance[21,22], we further engineered the RTT-PBS into a circular form, referred to as a prime editing template RNA (petRNA), via a ribozyme- and RtcB-mediated pathway for comparison to the linear petRNA (Extended Data Fig. 8). The petRNA and the 3′-MS2 version of the petRNA (petRNA-3′)

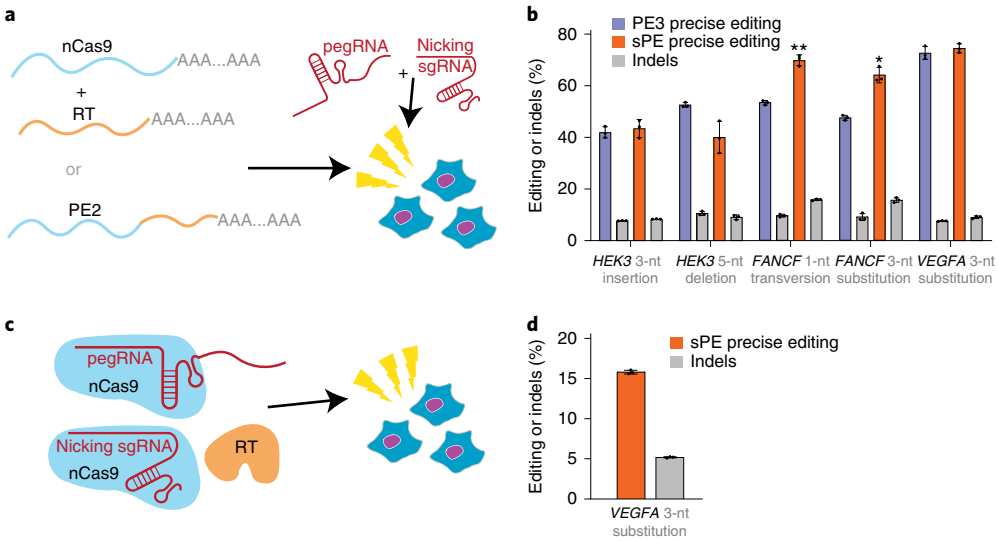


Fig. 4 | Effective prime editing by sPE through mRNA and RNP nucleofection. a, Illustration of prime editing by PE3 and sPE through mRNA nucleofection. The mRNAs were coelectroporated with a pegRNA and a nicking sgRNA into HEK293T cells. **b**, Amplicon sequencing results showing prime editing efficiency of mRNA-based sPE compared to PE3 at various endogenous loci ($n = 3$); two-tailed unpaired Student's t-test; $^{*}P < 0.05$; $^{**}P < 0.01$. **c**, Illustration of sPE RNP nucleofection. The pegRNA and the nicking sgRNA were preincubated with nCas9 before being coelectroporated with the NLS-RT protein into HEK293T cells. **d**, Amplicon sequencing results showing the efficiency of prime editing by RNP-based sPE at the *VEGFA* locus. Data and error bars indicate the mean and s.d. of three independent biological replicates.

exhibited comparable editing efficiency as that of the pegRNA at the *FANCF* site (Fig. 3b,c). The petRNA (that is, circularized) was more efficient than the linear petRNA (Fig. 3c), suggesting that the expected stability and abundance of the RTT-PBS RNA is positively correlated with the editing efficiency. The petRNA design was validated at seven additional endogenous sites (Fig. 3d). At a subset of endogenous loci, editing by petRNAs was lower than that by pegRNA counterparts (Fig. 3d), suggesting that the current petRNA design will benefit from further optimization for some genomic sites. We also observed that the petRNA supported prime editing activity with unaltered RT but with an efficiency lower than that observed with the MCP–RT fusion (Extended Data Fig. 8), consistent with a benefit from MS2–MCP tethering.

To compare the abundance and integrity of pegRNAs and petRNAs in cells, we transfected plasmids expressing nCas9, MCP–RT, pegRNA, petRNA and sgRNA in HEK293T cells and performed Northern blot analyses (Fig. 3e). Using an RTT-PBS probe, we observed that the petRNA accumulates to higher levels than pegRNA (top, lanes 3 and 5) and, unlike the pegRNA, does not require effector coexpression for stability (Fig. 3e). The circular nature of the petRNA was demonstrated by its susceptibility to specific RNase H cleavage and its resistance to RNase R (lanes 6 and 7). Using a probe targeting sequences in the tracrRNA, we observed that the pegRNA is subject to truncation, presumably at its exposed 3′ end, consistent with the conclusions of another recent study[14]. These data indicate the circular petRNA's superiority over the pegRNA design in RNA integrity and abundance, both paramount to the efficiency and fidelity of prime editing outcomes.

To further validate the modularity of the system, we used a site within the *FANCF* locus where SpyCas9 and SauCas9 can target a common spacer sequence, and where a shared petRNA could be tested in combination with both effectors and their corresponding sgRNAs. Both nSpyCas9 and nSauCas9 and their sgRNAs, in combination with MCP–RT and the petRNA, successfully installed a 1-nt transversion (Extended Data Fig. 9). These results demonstrate that the use of separate RNA modules for nicking and carriage of genetic information also allowed for the use of alternative nickases while maintaining satisfactory editing efficiency. The potential use of orthologous nCas9s or non-CRISPR–Cas effectors may liberate the system from the restriction of the NGG PAM and provide more flexibility to prime editing experimental design. In addition, both sPE and petRNA exhibited only background levels of off-target editing at known SpyCas9 off-target sites (*FANCF* and *HEK3*), similar to observations with the canonical PE3 system (Extended Data Fig. 9).

sPE delivery in cells by mRNA or protein. Delivery of two smaller, modular sPE mRNAs would be expected to improve stability, flexibility, scalability and efficiency relative to those of full-length PE2-like mRNAs that are 6–7 kb in length. To test whether sPE effectors can be delivered as mRNAs in cultured cells, we flanked the coding sequences of nCas9, M-MLV RT and PE2 by a capped 5′-untranslated region (5′-UTR) and a 3′-UTR followed by a poly(A) tract. The in vitro transcribed mRNAs (Extended Data Fig. 10) were coelectroporated with a pegRNA and a nicking sgRNA into HEK293T cells (Fig. 4a). Deep sequencing showed that sPE mRNAs mediate 40.1–74.5% prime editing at five endogenous sites in HEK293T cells (Fig. 4b). The sPE efficiencies are comparable to those of the full-length PE2 mRNA (Fig. 4b). To further test whether sPE can be delivered as an RNP, we purified a nuclear localization signal (NLS)-containing M-MLV RT protein (Extended Data Fig. 10). The pegRNA and the nicking sgRNA were preincubated with nCas9 protein before being mixed with the RT protein for coelectroporation into HEK293T cells (Fig. 4c). We observed 16.3% prime editing at the *VEGFA* locus, rendering a 3-nt substitution (Fig. 4d). These results indicate that mRNA and RNP delivery of sPE is feasible, although extensive future work will be needed to optimize relative doses of the nCas9, RT and guide/template components as well as many other parameters.

Discussion

In summary, this study demonstrates that a modular PE system enables versatile and efficient genome editing in mammalian cells and in adult mouse liver. Our data showed that the sPE enables comparable genome editing as the PE2 effector. The sPE offers advantages over other split methods, including its simplicity for AAV delivery. The sPE will fit in a dual AAV setup in which sufficient

space for additional modifications or control elements can be accommodated, such as using modified and controllable RT and engineered or multiplexed sgRNAs/pegRNAs. The limited editing efficiency seen thus far with the dual AAV-delivered sPE system is not drastically different from that seen previously with split intein PE3 (2–3%) at a different locus[2]. Nonetheless, further work will be needed to optimize AAV sPE editing efficiency, such as testing stronger promoters and screening more efficient pegRNAs and nicking sgRNAs. In addition, the reduced size of the individual nickase and RT compared to PE2 may improve the yields of mRNA and/or protein production for non-viral delivery. The small size of sPE mRNA may also facilitate packaging into lipid nanoparticles and efficient entry into cells. Similarly, RNP compositions comprising modular sPE proteins may provide manufacturing advantages compared to full-length PE2 protein. Finally, the modularity of the sPE system will accelerate the mix-and-match combinatorial exploration of nuclease and RT components of PE systems. Thus, the sPE approach promises to facilitate new implementations of prime editing.

Furthermore, our study indicates that, perhaps unexpectedly, the RT component can engage the RNA–DNA hybrid at the Cas9 nicking site without direct fusion to Cas9. Future structural studies may shed light on the nature of PE effector engagement. Lastly, despite recent reports of modest PE off-target effects[23], further analyses will be needed to determine whether the RT (tethered or untethered) could process endogenous RNA–RNA or RNA–DNA hybrids and induce potentially undesirable genomic integration events[24].

By separating the PBS-RTT from the editing guide, petRNAs may not only provide superior stability but also enable combinatorial or tiling approaches to identify highly efficient editing designs. We anticipate that circular RNAs produced in vitro may be useful for prime editing template delivery as well. In summary, modular PEs promise to facilitate effective and versatile in vivo delivery of PE for precise genome editing.

Online content

Any methods, additional references, Nature Research 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/s41587-022-01255-9.

Received: 23 June 2021; Accepted: 8 February 2022;
Published online: 04 April 2022

References

1. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
2. Liu, P. et al. Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice. *Nat. Commun.* **12**, 2121 (2021).
3. Chen, P. J. et al. Enhanced prime editing systems by manipulating cellular determinants of editing outcomes. *Cell* **184**, 5635–5652 (2021).
4. Wang, D., Zhang, F. & Gao, G. CRISPR-based therapeutic genome editing: strategies and in vivo delivery by AAV vectors. *Cell* **181**, 136–150 (2020).
5. Truong, D. J. et al. Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res.* **43**, 6450–6458 (2015).
6. Maji, B. et al. Multidimensional chemical control of CRISPR–Cas9. *Nat. Chem. Biol.* **13**, 9–11 (2017).
7. Liu, K. I. et al. A chemical-inducible CRISPR–Cas9 system for rapid control of genome editing. *Nat. Chem. Biol.* **12**, 980–987 (2016).
8. Li, C. et al. SWISS: multiplexed orthogonal genome editing in plants with a Cas9 nickase and engineered CRISPR RNA scaffolds. *Genome Biol.* **21**, 141 (2020).
9. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature* **517**, 583–588 (2015).
10. Wang, Y. et al. sgBE: a structure-guided design of sgRNA architecture specifies base editing window and enables simultaneous conversion of cytosine and adenosine. *Genome Biol.* **21**, 222 (2020).
11. Jiang, W. et al. BE-PLUS: a new base editing tool with broadened editing window and enhanced fidelity. *Cell Res.* **28**, 855–861 (2018).
12. Jang, H. et al. Application of prime editing to the correction of mutations and phenotypes in adult mice with liver and eye diseases. *Nat. Biomed. Eng.* **6**, 181–194 (2021).
13. Zhi, S. et al. Dual-AAV delivering split prime editor system for in vivo genome editing. *Mol Ther.* **30**, 283–294 (2022).
14. Nelson, J. W. et al. Engineered pegRNAs improve prime editing efficiency. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-021-01039-7 (2021).
15. Zhao, C., Liu, F. & Pyle, A. M. An ultraprocessive, accurate reverse transcriptase encoded by a metazoan group II intron. *RNA* **24**, 183–195 (2018).
16. Mohr, S. et al. Thermostable group II intron reverse transcriptase fusion proteins and their use in cDNA synthesis and next-generation RNA sequencing. *RNA* **19**, 958–970 (2013).
17. Stamos, J. L., Lentzsch, A. M. & Lambowitz, A. M. Structure of a thermostable group II intron reverse transcriptase with template-primer and its functional and evolutionary implications. *Mol. Cell* **68**, 926–939 (2017).
18. Kim, Y. et al. Adenine base editing and prime editing of chemically derived hepatic progenitors rescue genetic liver disease. *Cell Stem Cell* **28**, 1614–1624 (2021).
19. Azuma, H. et al. Robust expansion of human hepatocytes in *Fah*$^{-/-}$/*Rag2*$^{-/-}$/*Il2rg*$^{-/-}$ mice. *Nat. Biotechnol.* **25**, 903–910 (2007).
20. Paulk, N. K. et al. Adeno-associated virus gene repair corrects a mouse model of hereditary tyrosinemia in vivo. *Hepatology* **51**, 1200–1208 (2010).
21. Litke, J. L. & Jaffrey, S. R. Highly efficient expression of circular RNA aptamers in cells using autocatalytic transcripts. *Nat. Biotechnol.* **37**, 667–675 (2019).
22. Roth, A. et al. A widespread self-cleaving ribozyme class is revealed by bioinformatics. *Nat. Chem. Biol.* **10**, 56–60 (2014).
23. Jin, S. et al. Genome-wide specificity of prime editors in plants. *Nat. Biotechnol.* **39**, 1292–1299 (2021).
24. Zhang, L. et al. Reverse-transcribed SARS-CoV-2 RNA can integrate into the genome of cultured human cells and can be expressed in patient-derived tissues. *Proc. Natl Acad. Sci. USA* **118**, e2105968118 (2021).

Methods

Plasmids. pegRNA expression plasmids were constructed by using a custom vector (BfuAI- and EcoRI-digested) that has been described elsewhere[2]. The gBlock fragments with or without MS2 sequences were synthesized by Integrated DNA Technologies, followed by Gibson assembly using Gibson Assembly Master Mix (New England Biolabs). Colonies were selected and confirmed by Sanger sequencing using a commercial human U6 primer. Nicking sgRNA plasmids were generated by annealing oligonucleotides and inserting them into the pMD217 vector digested by BfuAI (Supplementary Table 1).

The MCP and M-MLV RT have been described elsewhere[9,25]. Briefly, the sequence of M-MLV RT was derived from a PE2 construct via PCR. The vector was prepared by excising nCas9 through dual digestion of PE2 plasmids with NotI and KpnI. Two gBlock gene fragments were synthesized for fusing MCP and M-MLV RT partially together with either a 32-amino acid linker or an NLS linker followed by Gibson assembly. The ligation mixtures were transformed into competent HB101 cells. The colonies were selected and confirmed by Sanger sequencing.

The scFv-RT plasmids were constructed by replacing nCas9 with the scFv fragment in PE2. The scFv sequence was derived from Addgene 60904 via PCR[26] followed by Gibson assembly. The 3×Flag-RT plasmid was constructed by excising nCas9. The 3×Flag sequence was derived from Addgene 80456 via PCR, and the DNA fragments were assembled by Gibson assembly.

Gene fragments of alternative RTs along with their bridging fragments were synthesized by Genewiz. PEs with alternative RTs were constructed by Gibson Assembly with PE2 digested by EcoRI and BsmI. Split alternative RT plasmids were modified from the 3×Flag-RT plasmid by digesting with NotI and EcoRI followed by Gibson Assembly with appropriate gene fragments.

The ribozyme-flanked RNA expression plasmids were constructed by replacing the pegRNA sequence from the U6-driven plasmid using Gibson Assembly with synthesized ribozyme fragments and a fragment containing the transcript to be circularized flanked by the hairpin sequences for circularization (Supplementary Sequences).

Plasmids were purified using a Miniprep kit (QIAGEN) or a ZymoPURE II Plasmids Midiprep kit for in vitro experiments. Plasmids were purified by using a Maxiprep kit (QIAGEN) for in vivo experiments.

Cell culture, transfection and genomic DNA isolation. HEK293T cells acquired from ATCC (CRL-3216) were cultured in Dulbecco's modified Eagle medium (DMEM) with 10% fetal bovine serum (FBS) and 1% penicillin/streptomycin. The mCherry and GFP reporter lines have been described elsewhere[8]. Cells were cultured at 37 °C and 5% CO_2.

Cell transfection was performed according to the manufacturer's instructions for the Lipofectamine 3000 reagent (Invitrogen, L3000015). Briefly, 1×10^5 cells were seeded per well in a 12-well plate overnight. Cells were transfected using 3 µl of Lipofectamine 3000 and P3000 (2 µl μg^{-1} DNA). For each well, 330 ng of pegRNA, 110 ng of nicking sgRNA and 1 µg of PE2 plasmids were used. The same amounts of plasmids were used for sPE groups. Seventy-two hours after transfection, cells were collected and lysed using 100 µl of Quick extraction buffer (Lucigen). Subsequently, the lysate was incubated on a thermocycler at 65 °C for 15 min and 98 °C for 5 min.

Mouse genomic DNA was isolated using a PureLink Genomic DNA Mini kit (Thermo Fisher) according to the manufacturer's protocol.

Nucleofection. For nucleofections of mRNA and RNP, the Neon electroporation system was used. pegRNAs and nicking sgRNAs were ordered from Integrated DNA Technologies as Alt-R gRNA with a chemical modification profile consistent with a previous study[27]. For mRNA nucleofection, 1 µg of each mRNA, 120 pmol of pegRNA, 40 pmol of nicking sgRNA and 50,000 HEK293T cells were mixed in Buffer R and electroporated using 10-µl Neon tips using the following electroporation parameters: 1,150 V, 20 ms, two pulses. After electroporation, cells were plated in prewarmed 48-well plates with DMEM containing 10% FBS and incubated for 72 h before analysis. For RNP nucleofection, the same conditions were used except that the mRNAs were replaced by 61 pmol of nCas9 protein (Alt-R Spy Cas9 H840A Nickase V3, purchased from Integrated DNA Technologies) and 150 pmol of NLS–RT protein produced in-house. Alt-R Custom Guide RNAs are listed in Supplementary Table 2.

Sanger sequencing and analysis using EditR. PCR amplification was performed around the target locus using Phusion Flash PCR Master Mix (Thermo Fisher) and specific primers. Sanger sequencing was performed by Genewiz. The results were quantified using EditR[28].

Deep sequencing and data analysis. Sequencing library preparation was done as described previously[1]. Briefly, for the first round of PCR, specific primers carrying Illumina forward and reverse adapters (Supplementary Table 3) were used for amplifying the genomic sites of interest with Phusion Hot Start II PCR Master Mix. For the second round of PCR, primers containing unique Illumina barcodes were used. PCR reactions were performed using the following parameters: 98 °C for 10 s, 20 cycles of 98 °C for 1 s, 55 °C for 5 s and 72 °C for 6 s, followed by 72 °C for 2 min as a final extension. The DNA products of second-round PCR were collected and purified by gel purification using the QIAquick Gel Extraction kit (Qiagen). DNA concentration was determined by Qubit dsDNA HS assay. Subsequently, the library was sequenced on an Illumina MiniSeq following the manufacturer's protocols.

MiniSeq sequencing reads were demultiplexed by bcl2fastq (Illumina). Prime editing efficiency was determined by aligning amplicon reads to a reference sequence using CRISPResso2 (ref. [29]) using HDR mode with a quantification window ranging from 3 bp before the pegRNA cut site to 1 bp after the nicking sgRNA cut site. Only reads with an average quality score ≥35 and a minimum homology of 90% are included in the analysis. The percentage of 'HDR' in the CRISPResso report represented the percentage of desired editing, and the percentage of indels was calculated as the sum of the percentages of 'Ambiguous', 'Imperfect HDR' and 'NHEJ' groups in the report.

Off-target analysis has been described elsewhere[1]. On-target and off-target genomic loci were amplified using PCR with primer sequences listed in Supplementary Table 3 and sequenced on an Illumina MiniSeq.

In vitro transcription. To construct the in vitro transcription templates, a CleanCap Reagent AG-compatible T7 promoter (TAATACGACTCACTATAAG) and a 5′-UTR were inserted at the 5′ end of the Kozak sequence of the coding sequence in the mammalian expression vectors for nCas9, M-MLV RT and PE2. Additionally, a 3′-UTR, a 110-nt poly(A) tract and a restriction site (BsmBI) were inserted after the stop codon. Plasmids were completely linearized using BsmBI (New England Biolabs) before being used in the in vitro transcription, which was performed at 37 °C using a HiScrib T7 High Yield RNA Synthesis kit (New England Biolabs) with the addition of CleanCap Reagent AG (Trilink Biotechnologies) for Cap1 structure and with a 100% replacement of UTP by N1-Methylpseudo-UTP (Trilink Biotechnologies). The reaction was terminated after 4 h by a 15-min incubation with DNase I (New England Biolabs). The RNA was then purified using a Monarch RNA Cleanup kit (New England Biolabs).

Flow cytometry analysis. Flow cytometry analysis was performed on day 3 after transfection. mCherry or GFP reporter lines were collected after PBS washing and 0.25% trypsin digestion, followed by recentrifuging at 300*g* for 5 min and resuspension in PBS with 2% FBS. The proportions of GFP^- and/or $mCherry^+$ cells were quantified using flow cytometry (MACSQuant VYB). Data were analyzed by FlowJo v10 software. FACS gating examples for reporter cells are shown in Extended Data Fig. 10.

NLS–RT purification. For bacterial expression of the NLS–RT protein, the coding sequence of M-MLV RT was flanked by one SV40 NLS at each end and cloned into a pET28a vector, transformed into BL21 (DE3) Rosetta competent cells (Novagen) and selected on LB agar plates containing 50 µg ml^{-1} kanamycin and 17 µg ml^{-1} chloramphenicol. One liter of LB + kanamycin + chloramphenicol medium was inoculated at 37 °C with 4 ml of overnight culture from a single colony. The culture was induced with 1 mM IPTG at an optical density at 600 nm (OD_{600}) of 0.8 for 3 h at 37 °C. The pellet was washed with 1× PBS and snap-frozen using liquid nitrogen.

For purification of NLS–RT protein, the cell pellet was resuspended in Lysis Buffer (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM imidazole, 1 mM DTT and 0.1% Triton X-100) and incubated on ice for 30 min with lysozyme added to a final concentration of 1 mg ml^{-1}. The cells were lysed by using an EpiShea Probe Sonicator (Active Motif) and cleared by centrifugation at 20,000*g* for 30 min at 4 °C before being loaded onto a column with Ni-NTA resin preequilibrated with Wash Buffer I (50 mM Tris-HCl (pH 8.0), 500 mM NaCl, 10 mM imidazole, 1 mM DTT and 0.1% Triton X-100). The resin was washed with 10 column volumes of Wash Buffer I and 10 column volumes of Wash Buffer II (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM imidazole and 1 mM DTT) before the protein was eluted using elution buffer (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 250 mM imidazole and 1 mM DTT). The eluate was dialyzed against Buffer A (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.1 mM EDTA and 5 mM DTT) overnight before being loaded onto a HiTrap SP HP cation exchange chromatography column (Cytiva). The column was washed with 10 column volumes of Buffer A before the protein was eluted with a 0–100% gradient of Buffer B (50 mM Tris-HCl (pH 8.0), 1 M NaCl, 0.1 mM EDTA and 5 mM DTT). The peak fractions were pooled and buffer exchanged with Storage Buffer (50 mM Tris-HCl (pH 8.0), 100 mM NaCl, 0.1 mM EDTA, 5 mM DTT and 30% glycerol).

Northern blot. Plasmids expressing nCas9, MCP–RT, pegRNA (*FANCF* 3-nt substitution), petRNA (*FANCF* 3-nt substitution) and sgRNA (*FANCF*) were lipofected into HEK293T cells as described above. Seventy-two hours after transfection, total RNA was extracted by using a Monarch Total RNA Miniprep kit (New England Biolabs). For RNase H treatment, 100 ng of total RNA from cells expressing the sgRNA, petRNA, nCas9 and MCP–RT was incubated with 5 pmol RNase H DNA oligonucleotide (Supplementary Table 4), 5 U of RNase H (New England Biolabs) and 1× RNase H Reaction Buffer (New England Biolabs) at 37 °C for 20 min; for RNase R treatment, 100 ng of the same total RNA was incubated with 10 U of RNase R (Lucigen) and 1× RNase R Reaction Buffer (Lucigen) at 37 °C for 30 min.

One microgram of total RNA from cells expressing the pegRNA, 100 ng of total RNA from cells expressing the petRNA or all of the RNase-treated samples were loaded onto a 10% (wt/vol) polyacrylamide mini gel containing 8 M urea. After gel electrophoresis, RNAs were transferred onto a Hybond-N+ membrane

(Amersham) using a semi-dry transfer apparatus, followed by UV cross-linking in a Stratalinker 1800 UV Crosslinker using the 'auto cross-link' function. The membrane was prehybridized with ULTRAhyb-Oligo hybridization buffer (Invitrogen) at 42 °C for 10 min before a ^{32}P-labeled probe (Supplementary Table 4) was added for overnight incubation at 42 °C. At the end of the hybridization, the membrane was washed twice with 2× SSC and 0.1% SDS for a total of 60 min at 42 °C. Finally, the membrane was exposed to a phosphor screen (GE Healthcare) overnight before the image of the blot was acquired by a Typhoon FLA 9000 Imaging System (GE Healthcare). Quantitation was performed using ImageJ.

Animals. All mouse studies were approved by the Institutional Animal Care and Use Committee (IACUC) at University of Massachusetts Chan Medical School (PROTO202000051). All plasmids were prepared using an Endo-Free-Maxi kit (Qiagen) and delivered by hydrodynamic tail vein injection. For cancer model generation, 8-week-old FVB/NJ mice (strain 001800) were injected with 30 μg of PE2 or split nCas9 and RT, 15 μg of pegRNA, 15 μg of sgRNA nicking, 5 μg of pT3 EF1a-MYC (Addgene, 92046) and 1 μg of CMV-SB10 (Addgene, 24551) via the tail vein. The maximal tumor size/burden permitted by the University of Massachusetts Chan Medical School IACUC is 1.5 cm, which was not exceeded in this study.

AAV vectors (AAV8 capsids) were packaged and produced at the Viral Vector Core of the Horae Gene Therapy Center, University of Massachusetts Medical School. Viral titers were measured by gel electrophoresis, silver staining and droplet digital PCR. For AAV injection, 1×10^{12} genome copies of AAV-Cas9^{H840A} and 1×10^{12} genome copies of AAV-U6-pegRNA-U6-sgRNA-M-MLV-RT were resuspended in 200 μl of 0.9% NaCl and administered to 8- to 10-week-old female *Fah*$^{mut/mut}$ mice via tail vein injection.

Histology and IHC. The procedure of IHC staining has been described previously[30]. Briefly, livers were fixed with 4% formalin overnight, embedded with paraffin and sectioned at 5 μm, followed by H&E staining for pathology. Liver sections were dewaxed, rehydrated and stained according to previous IHC protocols[31]. The following antibodies were used: β-catenin (BD, 610154; 1:200) and anti-FAH (Abcam, 1:400). Images were captured by Leica DMi8 microscopy.

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

Data availability

A Reporting Summary for this article is available as a supplementary information file. The raw DNA sequencing data are available at the NCBI Sequence Read Archive database under project number PRJNA802843. Plasmids for mammalian expression of MS2-pegRNA, petRNA, alternative PEs and split RTs as well as for bacterial expression of recombinant NLS-M-MLV RT have been deposited to Addgene for distribution. Source data are provided with this paper.

References

25. Peabody, D. S. The RNA binding site of bacteriophage MS2 coat protein. *EMBO J.* **12**, 595–600 (1993).
26. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
27. Petri, K. et al. CRISPR prime editing with ribonucleoprotein complexes in zebrafish and primary human cells. *Nat. Biotechnol.* **40**, 189–193 (2022).
28. Kluesner, M. G. et al. EditR: a method to quantify base editing from Sanger sequencing. *CRISPR J* **1**, 239–250 (2018).
29. Clement, K. et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nat. Biotechnol.* **37**, 224–226 (2019).
30. Xue, W. et al. Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656–660 (2007).
31. Xue, W. et al. CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* **514**, 380–385 (2014).

Acknowledgements

We thank S. Wolfe, P. Zamore and members of the Xue and Sontheimer labs for helpful discussions. We thank Y. Liu in the University of Massachusetts Chan Medical School Morphology Core and G. Gao, Q. Su and J. Xie in the University of Massachusetts Chan Medical School Viral Vector Core for support. W.X. was supported by grants from the National Institutes of Health (DP2HL137167, P01HL131471, P01HL158506 and UG3HL147367), American Cancer Society (129056-RSG-16-093) and the Cystic Fibrosis Foundation. X.D. and E.J.S. acknowledge support from the Leducq Foundation Transatlantic Network of Excellence Program.

Author contributions

B.L., X.D., W.X. and E.J.S. conceptualized the project and designed experiments. B.L., X.D., C.Z., S.-Q.L. and Z.C. conducted molecular biological experiments. B.L. performed mouse work. B.L., X.D., H.C. and T.C.R. conducted high-throughput sequencing and bioinformatic analyses. B.L., X.D., W.X. and E.J.S. interpreted the data and wrote the paper.

Competing interests

E.J.S. is a co-founder and Scientific Advisory Board member of Intellia Therapeutics and a Scientific Advisory Board member at Tessera Therapeutics. The University of Massachusetts Chan Medical School has filed a patent application on this work. The authors declare no competing interests.

Additional information

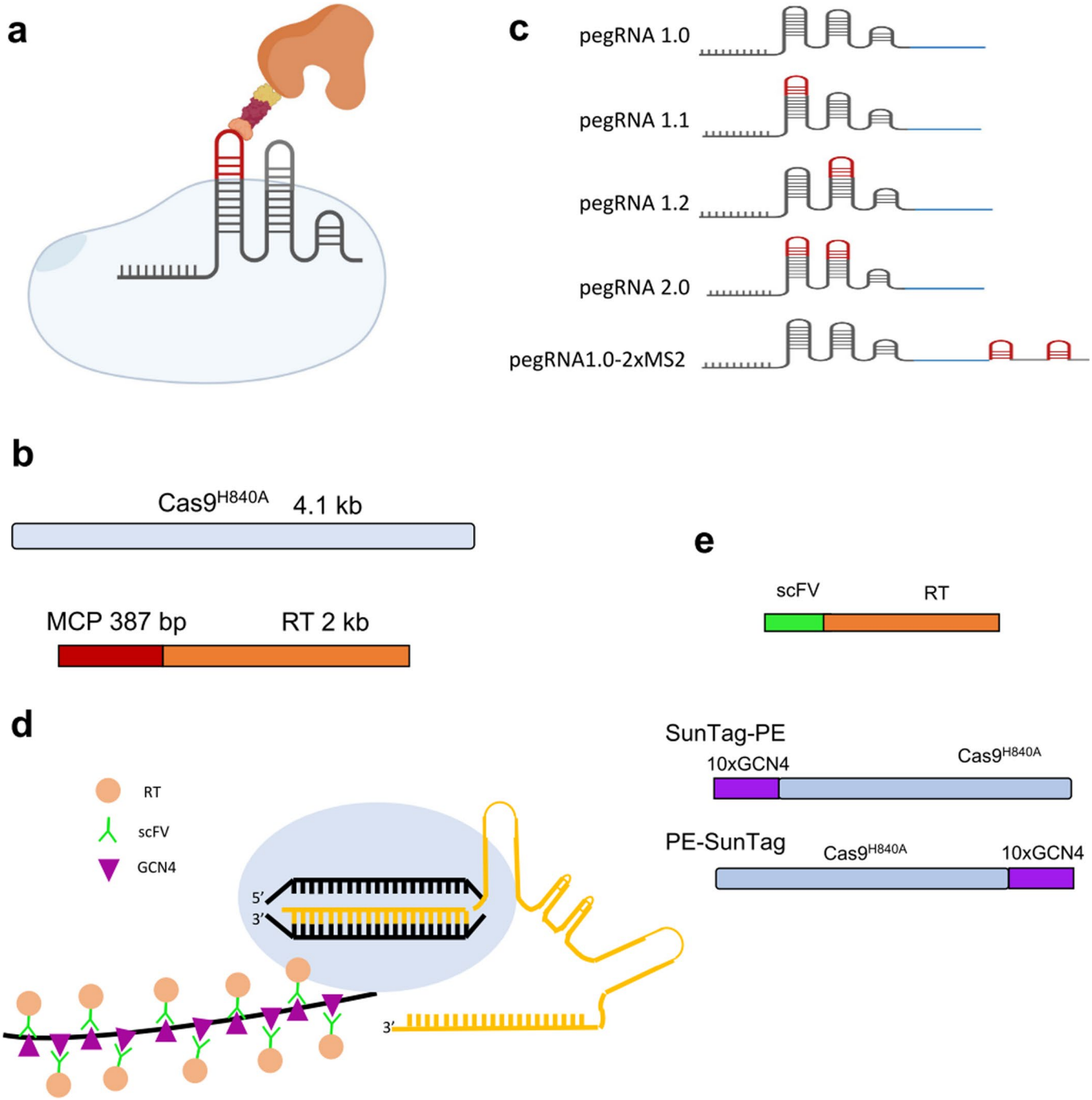
Extended data is available for this paper at https://doi.org/10.1038/s41587-022-01255-9.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-022-01255-9.

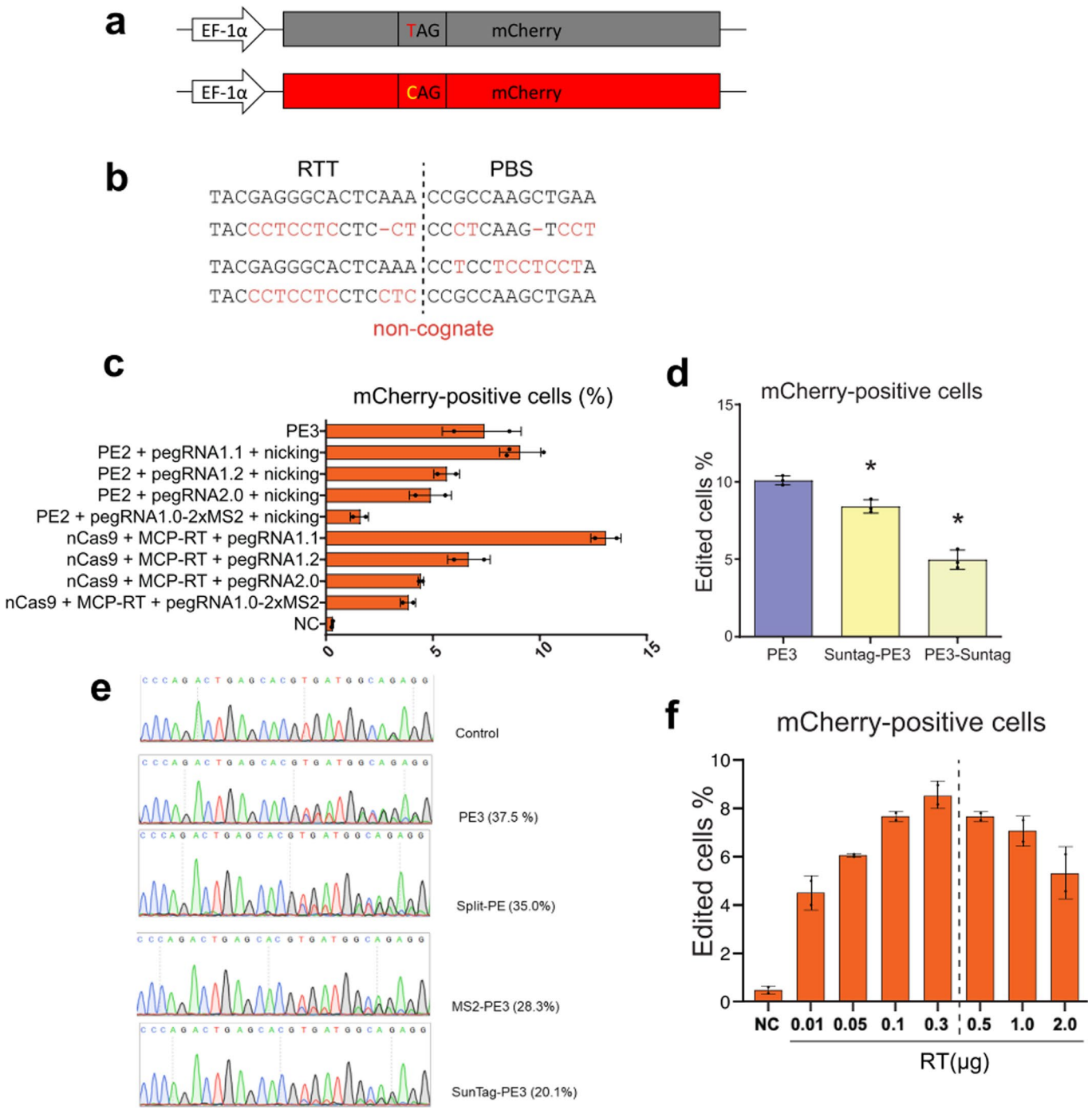
Correspondence and requests for materials should be addressed to Wen Xue or Erik J. Sontheimer.

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

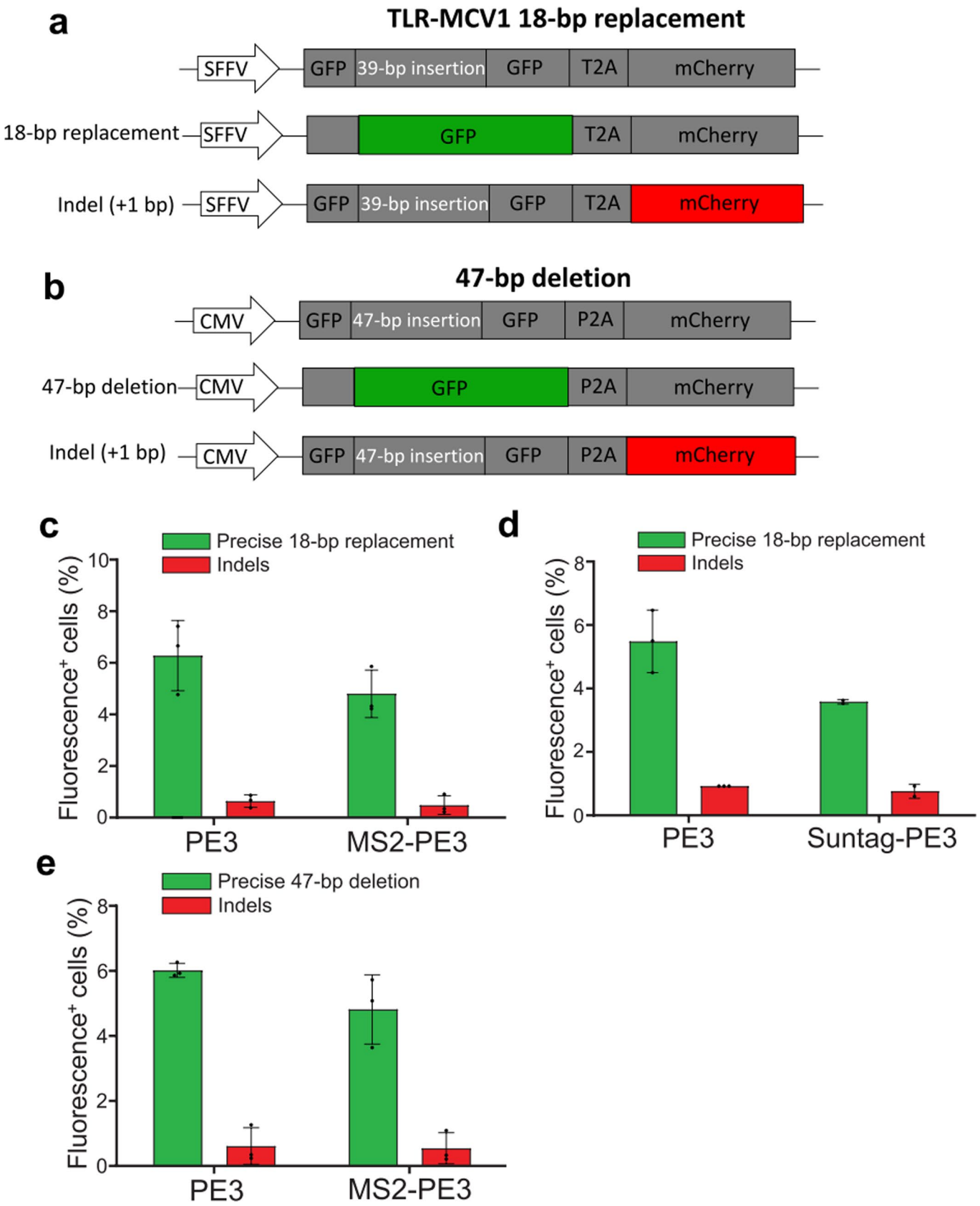
Reprints and permissions information is available at www.nature.com/reprints.



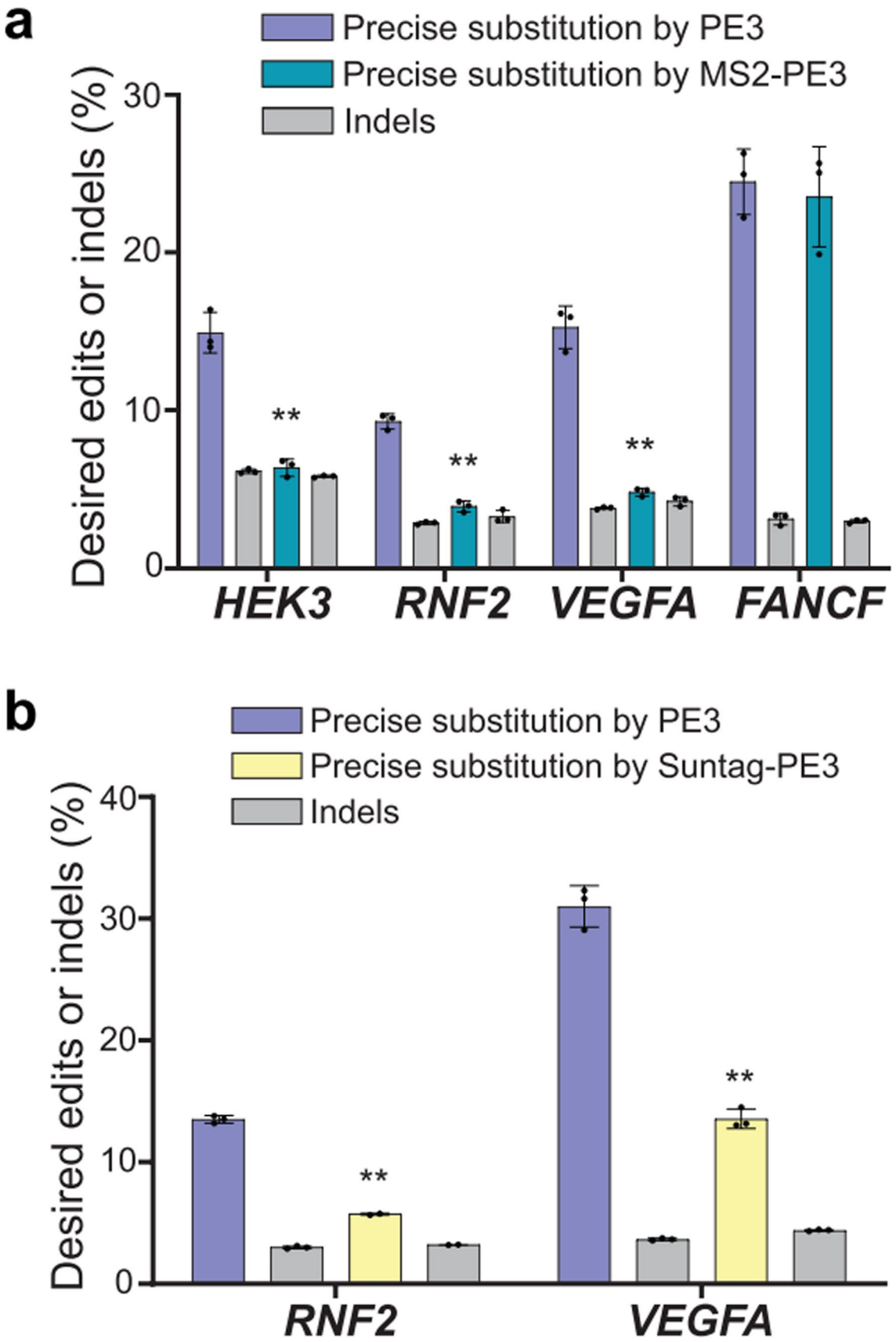
Extended Data Fig. 1 | MS2-PE2 and SunTag-PE2 design. a, Schematic overview of MS2-PE2. The MS2 coat protein (MCP) was fused to the N terminus of M-MLV reverse transcriptase to enable recruitment by the MS2-pegRNAs. **b**, Sizes of Cas9 nickase and MCP-RT ORFs. **c**, Engineered MS2-pegRNAs with MS2 sequences appended into distinct sgRNA stem-loops, or onto the 3′ terminus. **d**, Schematic overview of SunTag-PE2. **e**, Schematics of scFv-RT and GCN4-Cas9 nickase. The scFv was fused to the N terminus of M-MLV RT (top). The 10xGCN4 epitope was fused to either the N terminus (SunTag-PE) or the C terminus of SpyCas9^{H840A} (PE-SunTag).



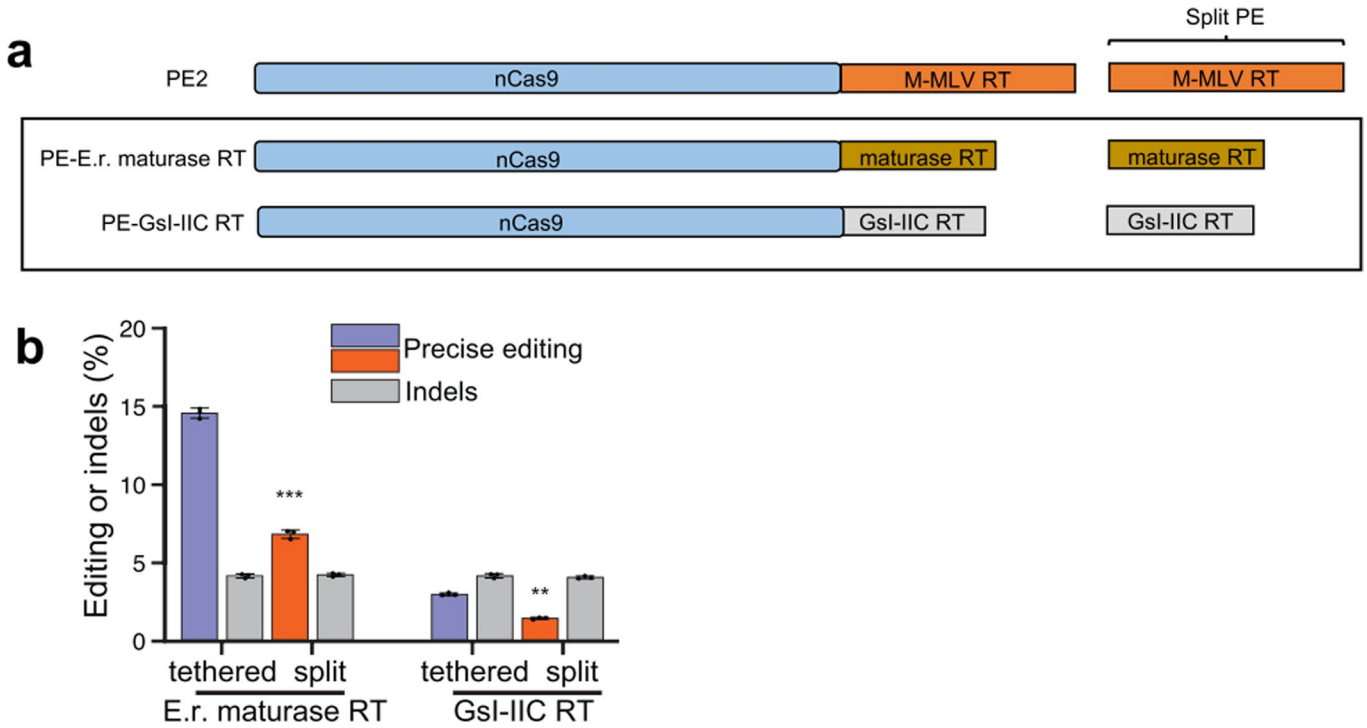
Extended Data Fig. 2 | Split-PE, SunTag-PE3 and MS2-PE3 tested in an mCherry reporter line and an endogenous locus. a, A diagram of the mCherry reporter line that functions by converting a premature stop codon. **b**, Sequences of RTT and PBS, non-cognate (PBS + RTT), non-cognate PBS, and non-cognate RTT for the mCherry reporter line. **c**, Multiple MS2-pegRNAs tested in mCherry reporter cell lines. The pegRNA with MS2 on the repeat/anti-repeat stem-loop (pegRNA-1.1) has the highest editing efficiency (higher even than that of the original PE3) in this mCherry reporter line (n = 2). Therefore, the pegRNA1.1-Cas9^{H840A}-MCP-RT system was designated as MS2-PE3. **d**, SunTag-PE3 and PE3-SunTag were tested in the mCherry reporter cell line. Two-tailed unpaired Student's t-test: *$P < 0.05$ (n = 3). **e**, Sanger sequencing and EditR quantification of PE3, Split PE, SunTag-PE3 and MS2-PE3 by installing "CTT" at *HEK3* sites in HEK293T cells. All plasmids were transfected at the same molar ratio. Genomic DNAs were isolated 72 h post transfection. **f**, Dose dependence of the RT-encoding plasmid. One microgram of H840A plasmid was co-transfected with plasmids encoding additional sPE components [pegRNA (0.3 μg), nicking sgRNA (0.1 μg), and RT (0.01-2 μg)] per well in a 12-well plate (n = 2). Data and error bars indicate the mean and standard deviation of two or three independent biological replicates, as indicated.



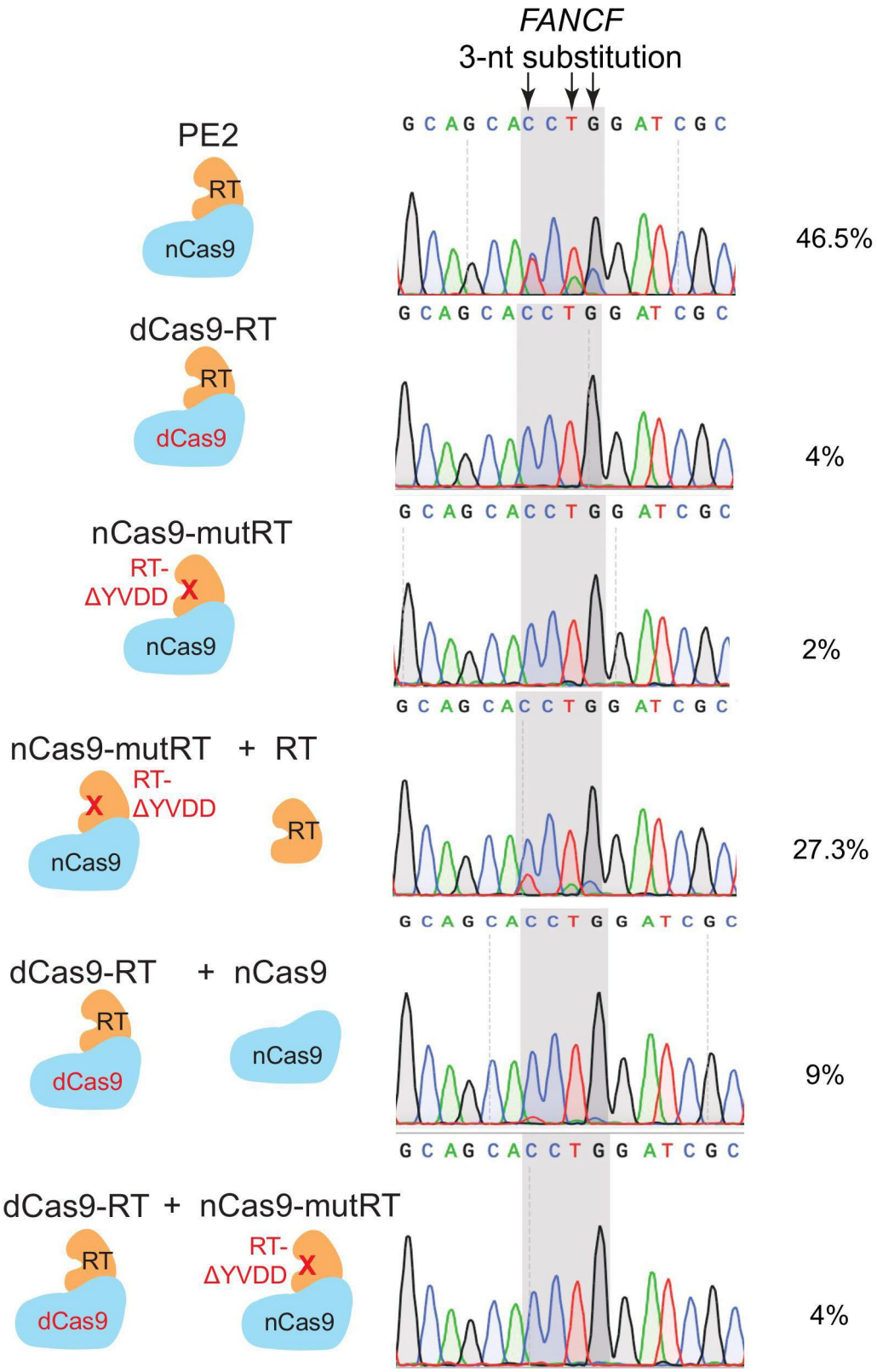
Extended Data Fig. 3 | SunTag-PE3 and MS2-PE3 tested in reporter lines. a, A diagram of the GFP reporter line that is activated by precise insertion of 18 bp (in place of a 39-bp non-functional sequence). Indels (+1) can restore mCherry expression. **b**, A diagram of the GFP reporter line that is activated by deletion of 47 bp; indels (+1) can restore mCherry expression. **c**, MS2-PE3 was tested in the GFP reporter line shown in panel **a** (n = 3). **d**, SunTag-PE3 was tested in the GFP reporter line shown in **a** (n = 3). **e**, MS2-PE3 was tested in the GFP reporter line shown in **b** (n = 3). Data and error bars indicate the mean and standard deviation of three independent biological replicates.



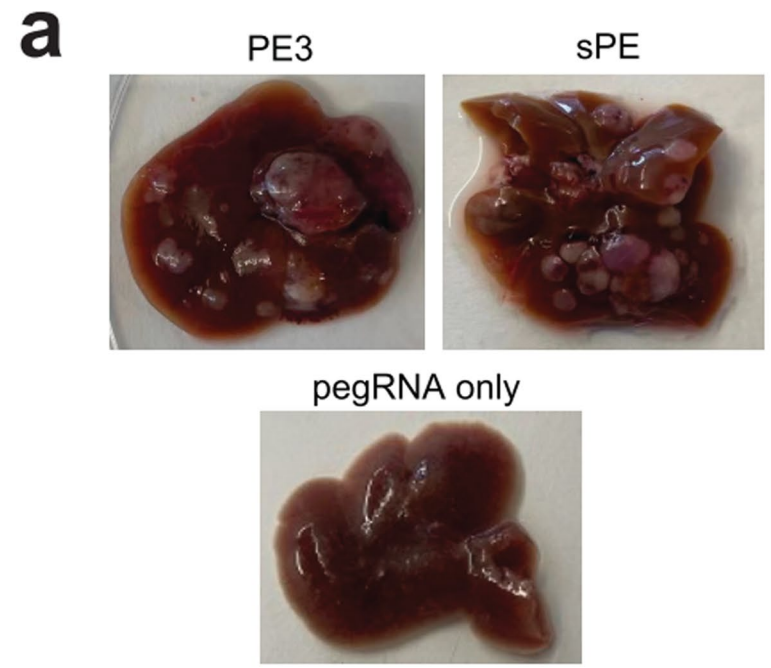
Extended Data Fig. 4 | Amplicon sequencing of MS2-PE3 and SunTag-PE3 at multiple endogenous sites. a, MS2-PE3 for editing by 1-bp substitution at multiple endogenous loci, including *HEK3*, *RNF2*, *VEGFA*, and *FANCF* in HEK293T cells. **b**, SunTag-PE3 for *RNF2* and *VEGFR* editing to generate a 1-bp substitution in HEK293T cells. Two-tailed unpaired Student's t-test: *$P < 0.05$, **$P < 0.01$. Data and error bars indicate the mean and standard deviation of three independent biological replicates.

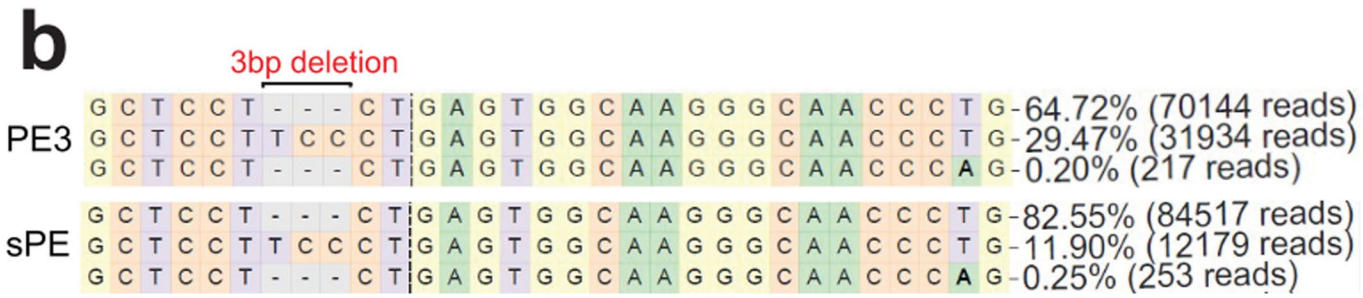


Extended Data Fig. 5 | Prime editing by alternative reverse transcriptases. a, Illustration of prime editors with alternative RTs. Human codon-optimized *E.r.* maturase RT and GsI-IIC RT were cloned into the original PE2 in place of the M-MLV RT. **b**, Prime editing by alternative RT orthologs at the *VEGFA* site by 3-nt substitutions (+2 G to C and +4-5 GG to CT). Two-tailed unpaired Student's t-test: *P < 0.05, **P < 0.01, ***P < 0.001. Data and error bars indicate the mean and standard deviation of three independent biological replicates.

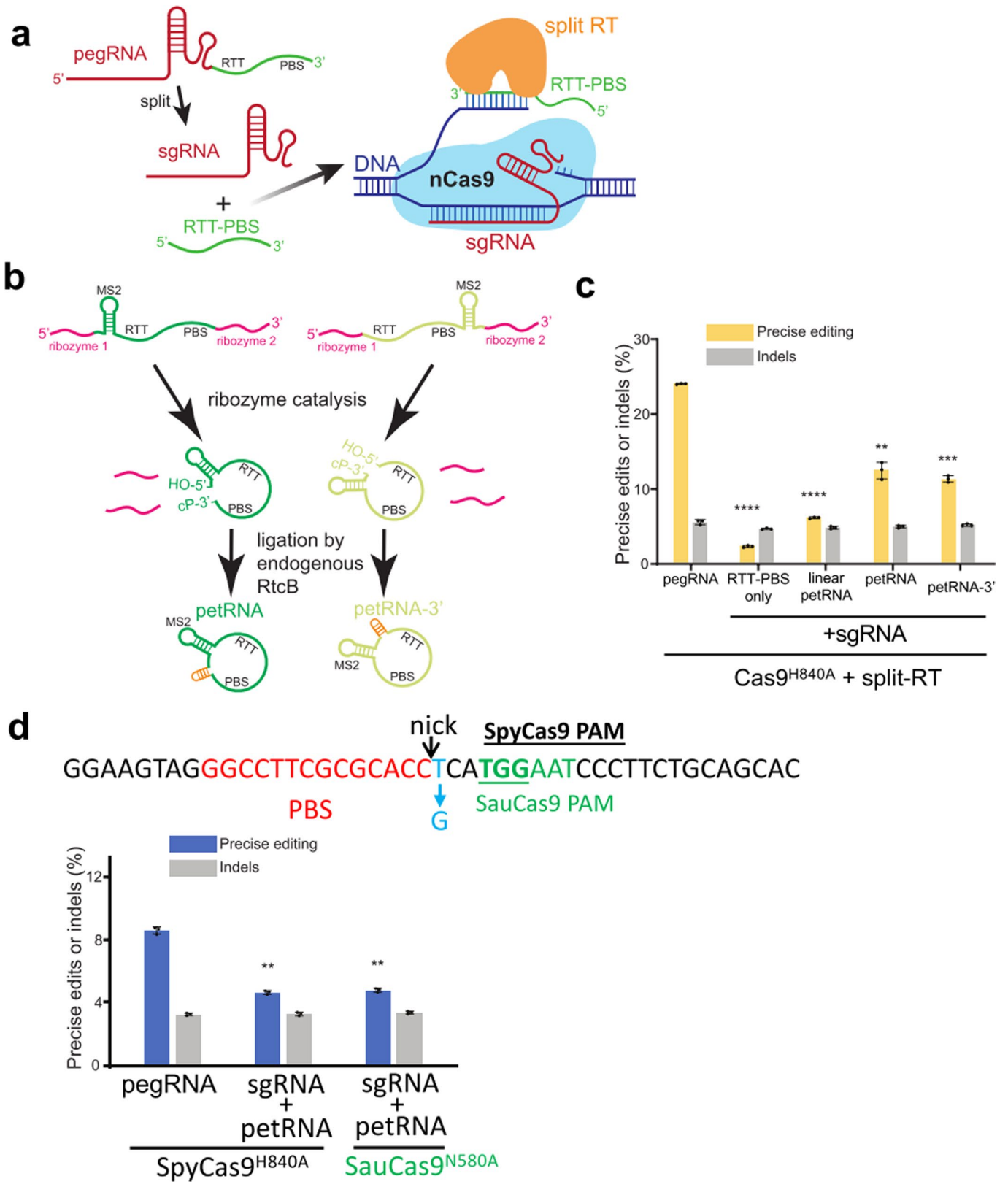


Extended Data Fig. 6 | Prime editing using mutant PE2 and sPE components. Representative Sanger sequencing traces from prime editing experiments using mutant PE2 and sPE components. HEK293T cells were transfected with indicated plasmids, along with others encoding pegRNA and nicking sgRNA. Prime editing introduces a 3-nt substitution at the *FANCF* locus (+2 C to T and +4-5 TG to AC). The experiment was repeated two times. Sanger sequencing traces were analyzed by EditR.





Extended Data Fig. 7 | Split PE2 enables genome editing in adult mice. a, Representative images of tumors in liver with PE3 or split PE. Control group was pegRNA-injected only. **b**, Amplicon sequencing from representative animals using genomic DNA isolated from tumors.



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

Extended Data Fig. 8 | Prime editing by separate RNA modules. a, Schematic of proof-of-concept experiment on delivering the RT template separately. The 3′ extension of the pegRNA (the RTT-PBS sequence) was removed from the 3′ of the tracrRNA scaffold and provided separately under the control of a U6 promoter. An sgRNA plasmid was co-transfected to carry out the nicking event in conjunction with the nCas9. **b**, Illustration of the circularization pathway to generate petRNAs. **c**, PE efficiency by modular RNA components at the *FANCF* locus introducing a 3-nt substitution (+2 C to T and +4-5 TG to AC). Plasmids expressing RNAs were co-transfected with $Cas9^{H840A}$ and the split RT, which lacks the MCP domain. Nicking sgRNAs were used for all prime editing. **d**, Validation of petRNA adaptability to an alternative nickase. The petRNA was designed to target a site at the *FANCF* locus where SpyCas9 and SauCas9 nickases share the same nick and thus a single petRNA guide/primer/template sequence. The petRNA and the MCP-RT were co-transfected with plasmids encoding $SpyCas9^{H840A}$-sgRNA or $SauCas9^{N580A}$. Nicking sgRNAs were used for all prime editing. Two-tailed unpaired Student's t-test: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, ****$P < 0.0001$. Data and error bars indicate the mean and standard deviation of three independent biological replicates.

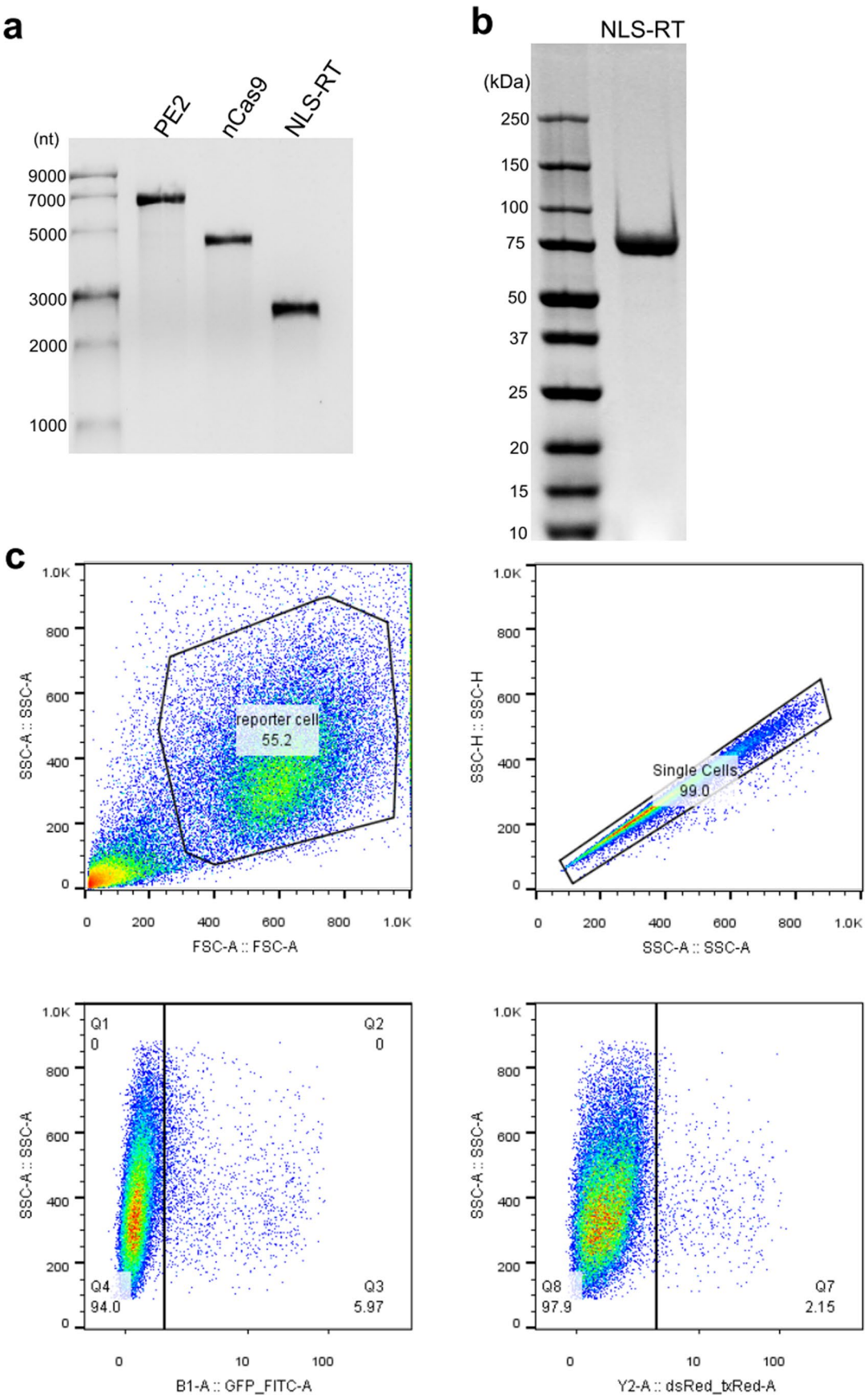
FANCF

	(%) PE3	sPE3	sPE-pet
On-target	26.32±1.35	29.98±0.37	26.82±0.96
Off-target 1	<0.1	<0.1	<0.1
Off-target 2	<0.1	<0.1	<0.1
Off-target 3	<0.1	<0.1	<0.1
Off-target 4	<0.1	<0.1	<0.1

HEK4

	(%) PE3	sPE3	sPE-pet
On-target	5.36±0.11	4.92±0.07	4.57±0.38
Off-target 1	<0.1	<0.1	<0.1
Off-target 2	<0.1	<0.1	<0.1
Off-target 3	<0.1	<0.1	<0.1

Extended Data Fig. 9 | Comparison of PE, sPE, and petRNA off-target effects at known Cas9 off-target sites of FANCF and HEK4 using deep sequencing. On-target edits are shown in red and off-target edits are shown in green. Data and error bars indicate the mean and standard deviation of three independent biological replicates.



Extended Data Fig. 10 | *In vitro* transcribed mRNA and purified RT protein used in the nucleofections, and FACS gating strategy. a, Denaturing agarose gel analysis of the mRNAs produced in-house. The coding sequences of nCas9, MMLV-RT or PE2 were flanked by a capped 5′ UTR and a 3′ UTR, followed by a 110-nt poly(A) tract. **b**, SDS-PAGE analysis of the purified MMLV-RT protein. **C**, FACS gating examples for reporter cells.

nature research

Corresponding author(s): Erik Sontheimer, Wen Xue

Last updated by author(s): Mar 25, 2022

Reporting Summary

Nature Research 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 Research 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 *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 *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 MiniSeq
Data analysis	CRISPResso V2.0.32 was used to analyze deep data for quantifying editing efficiency. FlowJo software V10 was used for flow cytometry analysis. Frequency, mean, and standard deviations were calculated using GraphPad Prism 8.

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

A reporting summary for this article is available as a Supplementary Information file. The raw gel images underlying Fig 3 and Supplementary Fig 10 are provided as a Source Data File and an additional supplementary data file, respectively. The raw DNA sequencing data are available at the NCBI Sequence Read Archive database under PRJNA802843.

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	1x105 cells were used for editing in culture system for Lipo3000 transfection. 5x104 cells were used for editing in culture system for electroporation. All cell samples were evaluated in at least biological triplicates (n = 3) to ensure the reproductability. Our previous editing studies have shown that this sample size and replications are sufficient to ensure reproducibility (Song et al, Nature Biomedical Engineering,2019 and Jiang et al, Ibraheim et al, Nature Communications, 2021). For animal experiment, we described the size in the specific figure legend. The size is determined based on the availability of the mice and previous reports (Song et al, Nature Biomedical Engineering).
Data exclusions	No data was excluded.
Replication	All the culture-related experiments were done in biological triplicate in culture cells, n=3 ,on different days (every three days). All attempts at replication were successful, and standard deviations were in the expected ranges.
Randomization	After seeding cell into 12-well plate, we randomly decided which well is for experiment group or control group. For mouse experiment, we randomly decide the mice treated for PE3 or sPE and for short or long-term studies.
Blinding	It is not applied to molecular and cell experiments. All mouse work are blind.

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
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Antibodies

Antibodies used	beta-catenin antibody (BD, 610154) IHC 1:100. Fumarylacetoacetate hydrolase antibody (ab83770, Abcam Inc), IHC 1:400
Validation	The specificity of the beta-catenin antibody has previously been confirmed showing nuclear beta-Catenin specifically in beta-Catenin mutant cells in mouse liver (Xue et al, Nature, 2014). The specificity of the anti-Fah antibody has previously been confirmed (Yin et al, Nature Biotech, 2016).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T cells, HEK293T-TLR/mCherry cells
Authentication	HEK293T (ATCC) cells were valiated by supplier (ATCC) by STR analysis. And HEK293T-TLR cells were valiated using specific primers to amplify the inserted cassette. And the PCR products were analyzed by Sanger Sequencing.
Mycoplasma contamination	All cell lines tested negative for mycoplasma contamination

Commonly misidentified lines (See ICLAC register)	The cell lines used in this article are not in the list of misidentified lines.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Fah PM mice was a mouse model of tyrosinemia. 9-week-old female mice were used in this study. Temperature of 65-75°F (~18-23° C) with 40-60% humidity are kept in the mouse room. A 14-hour light/10-hour dark cycle.
Wild animals	No wild animals were used in the study.
Field-collected samples	No field-collected samples were used in the study
Ethics oversight	All animal study protocols were approved by the UMass IACUC.

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

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	Post-editing HEK293T-TLR/mCherry cells were trypsinized, resuspended in 200ul PBS with 1% FBS, and directly analyzed by flow cytometry.
Instrument	mCherry analysis is analyzed by MACSQuant VYB Flow Cytometer.
Software	All data were analyzed by FlowJo10.0 software
Cell population abundance	mCherry positive rates w=mCherry positive cell number / total live cell number; GFP positive rates w=GFP positive cell number / total live cell number
Gating strategy	The cells were first gated based on FSC/SSC and FSCA/FSCH to select for live single cells. Unedited cells were employed as negative control for gating mCherry/GFP signal.

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