nature biotechnology

# Optimized base editors enable efficient editing in cells, organoids and mice

Maria Paz Zafra[1,11], Emma M Schatoff[1,2,11], Alyna Katti[1,3], Miguel Foronda[1], Marco Breinig[4], Anabel Y Schweitzer[4], Amber Simon[1], Teng Han[1,3], Sukanya Goswami[1], Emma Montgomery[1], Jordana Thibado[3], Edward R Kastenhuber[5,6], Francisco J Sánchez-Rivera[5], Junwei Shi[7,8], Christopher R Vakoc[7], Scott W Lowe[5,9], Darjus F Tschaharganeh[4] & Lukas E Dow[1,3,10]

**CRISPR base editing enables the creation of targeted single-base conversions without generating double-stranded breaks. However, the efficiency of current base editors is very low in many cell types. We reengineered the sequences of BE3, BE4Gam, and xBE3 by codon optimization and incorporation of additional nuclear-localization sequences. Our collection of optimized constitutive and inducible base-editing vector systems dramatically improves the efficiency by which single-nucleotide variants can be created. The reengineered base editors enable target modification in a wide range of mouse and human cell lines, and intestinal organoids. We also show that the optimized base editors mediate efficient *in vivo* somatic editing in the liver in adult mice.**

Base editors are hybrid proteins that tether DNA-modifying enzymes to nuclease-defective Cas9 variants. They enable the direct conversion of C to other bases (T, A, or G)[1–4], or A to inosine or G nucleic acids[5,6], thus allowing the creation or repair of disease-associated single-nucleotide variants (SNVs). The BE3 base editor carries a rat APOBEC cytidine deaminase at the N terminus of Cas9n ($Cas9^{D10A}$) and a uracil glycosylase inhibitor (UGI) domain at the C terminus. This construct has been shown to drive targeted C-to-T transitions at nucleotide positions 3–8 of the protospacer (**Fig. 1a**) after transfection of plasmid DNA or ribonuclear particles[7,8].

To enable base editing in difficult-to-transfect cells, we cloned a lentiviral vector for expression from the EF1 short (EF1s) promoter of BE3 linked to a puromycin (puro)-resistance gene via a P2A self-cleaving peptide (pLenti-BE3-P2A-Puro, BE3). Despite efficient production of viral particles and integration of the vector into target cells (**Supplementary Fig. 1**), we were not able to generate puro-resistant cells (**Fig. 1b** and **Supplementary Fig. 1c**). To test whether this result was due to low expression of the BE3-linked Puro cassette, we generated a new lentivirus wherein puro was driven by an independent (PGK) promoter (pLenti-BE3-PGK-Puro). This vector produced equivalent viral titer and target cell integration (**Supplementary Fig. 1**) but, in contrast to BE3-P2A-Puro, enabled effective puro resistance (**Fig. 1b** and **Supplementary Fig. 1c**).

These data suggested that an issue in the production of BE3 protein was limiting effective base editing. During cloning of lentiviral constructs, we observed that the Cas9n DNA sequence in BE3 was not optimized for expression in mammalian cells, and it contained a large number of nonfavored codons (**Supplementary Fig. 2** and **Supplementary Table 1**) and six potential polyadenylation sites (AATAAA or ATTAAA) throughout the cDNA (**Fig. 1c**); we therefore reconstructed the BE3 enzyme by using an extensively optimized Cas9n sequence[9] (**Supplementary Fig. 2**). The resulting construct with a reassembled BE3 sequence ($BE3^{RA}$; hereafter denoted RA) enabled efficient puro selection (**Fig. 1b** and **Supplementary Fig. 1**), markedly increased protein expression (**Fig. 1d**), and, most notably, showed up to 30-fold-higher target C-to-T conversion (**Fig. 1e,f** and **Supplementary Fig. 3a,b**). Although C-to-T editing increased on average 15-fold, the level of unwanted insertions and deletions (indels) or undesired (C-to-A or C-to-G) editing remained low, thus indicating a substantial improvement in the relative fidelity of base editing compared with that of previous versions (**Supplementary Fig. 3c,d**). Notably, we and others[10] observed similar problems in expression of high-fidelity Cas9 (HF1)[11] and altered protospacer-adjacent motif (PAM)-specificity variants[12], which share the same *Cas9* cDNA as BE3. In each case, these problems were corrected by reengineering the construct[10] (**Fig. 1g** and **Supplementary Fig. 4**). The resulting increased expression of the HF1 enzyme ($HF1^{RA}$) dramatically improved on-target DNA cleavage while maintaining little or no off-target activity[13] (**Fig. 1h**).

Nuclear-localization signal (NLS) sequences at the N terminus of Cas9 can improve the efficiency of gene targeting[14]. Indeed, despite

[1]Sandra and Edward Meyer Cancer Center, Department of Medicine, Weill Cornell Medicine, New York, New York, USA. [2]Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional MD–PhD program, New York, New York, USA. [3]Weill Cornell Graduate School of Medical Sciences, Weill Cornell Medicine, New York, New York, USA. [4]Helmholtz-University Group 'Cell Plasticity and Epigenetic Remodeling', German Cancer Research Center (DKFZ) and Institute of Pathology, University Hospital, Heidelberg, Germany. [5]Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [6]Gerstner Sloan Kettering Graduate School of Biomedical Sciences, New York, New York, USA. [7]Cold Spring Harbor Laboratory, New York, New York, USA. [8]Department of Cancer Biology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA. [9]Howard Hughes Medical Institute, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [10]Department of Biochemistry, Weill Cornell Medicine, New York, New York, USA. [11]These authors contributed equally to this work. Correspondence should be addressed to L.E.D. (lud2005@med.cornell.edu).

Received 20 February; accepted 21 June; published online 3 July 2018; doi:10.1038/nbt.4194

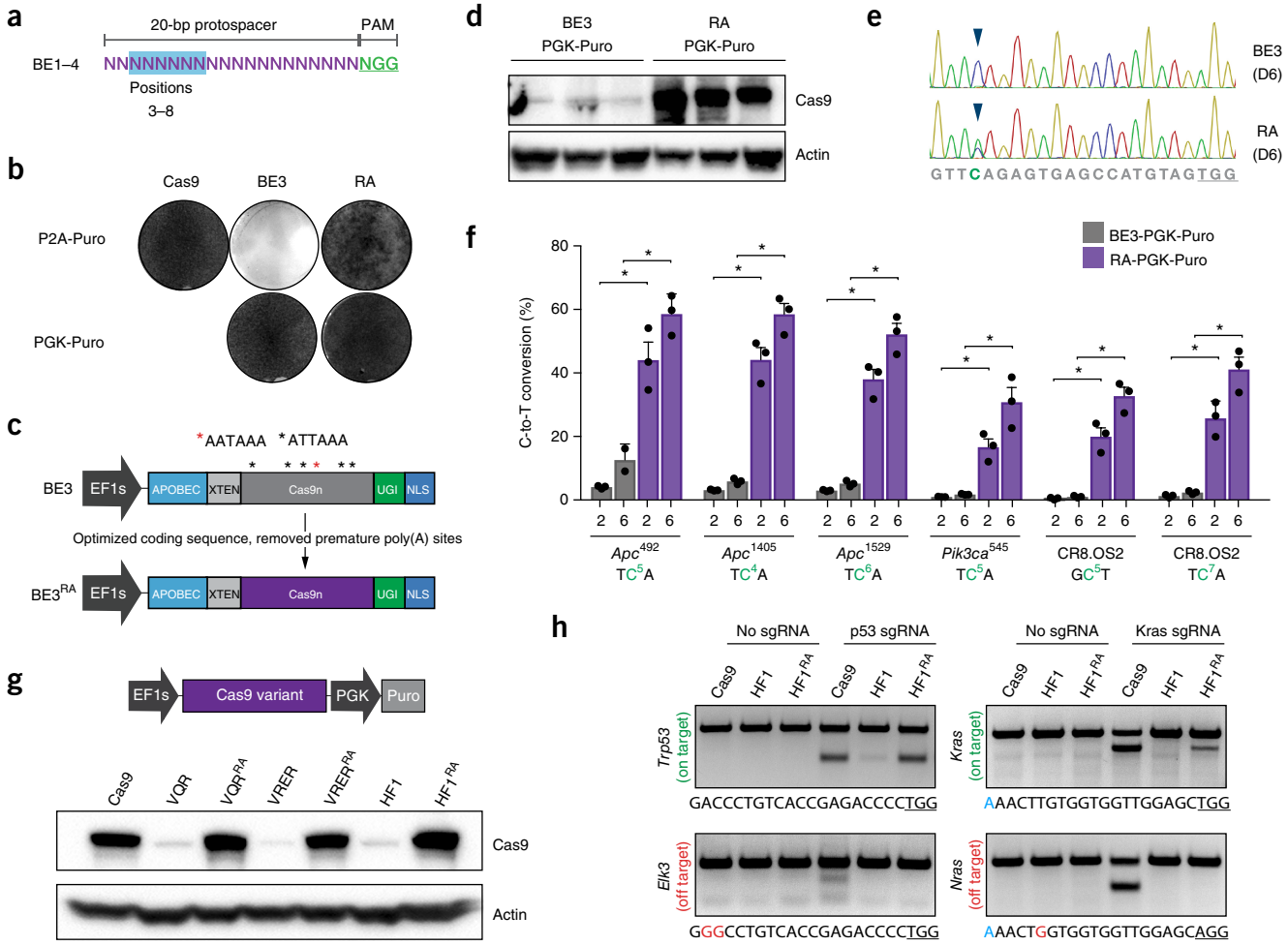


**Figure 1** Optimizing the coding sequence of BE3 improves protein expression and target base editing. (**a**) Schematic depiction of the canonical region of target base editing. Positions 3–8 (highlighted in blue) within the protospacer are susceptible to C-to-T conversion by BE3. The PAM is shown in green. (**b**) Giemsa-stained NIH/3T3 cells after transduction with the indicated lentiviruses and selection in puromycin for 6 d. Representative of similar results from three independent experiments; additional data in **Supplementary Figure 1**. (**c**) Schematic representation of original BE3 (top) and codon-optimized RA sequences (bottom). (**d**) Cas9 immunoblot of independently derived NIH/3T3 lines transduced with BE3 or RA constructs ($n = 3$). β-actin, loading control. (**e**) Sanger-sequencing chromatogram showing the target region of the *Apc*$^{1405}$ sgRNA. Arrowheads highlight a C at position 4 that shows dramatically increased editing by RA 6 d after sgRNA transduction. Representative of similar results from three independent experiments; additional data in **f**. (**f**) Frequency of target C-to-T editing across five different sgRNA targets, 2 d and 6 d after sgRNA transduction, as indicated. CR8.OS2 targets a nongenic region on mouse chromosome 8 (ref. 13). Graphs show mean values. Error bars, s.d. ($n = 3$ biologically independent samples); $^*P < 0.05$ between groups, by one-way analysis of variance (ANOVA) with Sidak's multiple-comparison test. (**g**) Western blot showing expression of original and optimized HF1- and PAM-variant Cas9 proteins. Representative of similar results from three independent blots. (**h**) T7 endonuclease assays on *Trp53* and *Kras* target sites, and off-target sites (*Elk3* and *Nras*), showing that reassembled HF1 (HF1$^{RA}$) improves on-target activity while maintaining little to no off-target cutting. Genomic target sites for each region are shown below. Notably, the slightly decreased on-target activity of HF1$^{RA}$ at the *Kras* site may be due to the G–A mismatch at position 1 of the protospacer (highlighted blue). The experiment was performed twice with similar results.

the presence of a C-terminal NLS (**Fig. 2a**), RA protein was largely excluded from the nucleus (**Fig. 2b**). We tested two different N-terminal positions for the NLS in case the inclusion of these sequences in one location might have interfered with APOBEC function: (i) with a FLAG epitope tag at the N terminus (FNLS) and (ii) within the XTEN linker that bridges APOBEC and Cas9n (2X) (**Fig. 2a** and **Supplementary Fig. 5a**). Whereas 2X showed no obvious increase in nuclear targeting compared with that of RA, FNLS protein was more evenly distributed through the nucleus and cytoplasm (**Fig. 2b**).

In transfection-based assays, FNLS improved editing approximately twofold across multiple target positions and single guide RNAs (sgRNAs) (**Supplementary Fig. 5b**). In contrast, 2X did not alter editing within the normal target window but substantially increased the range of editing of C nucleotides at positions 10 and 11 in the protospacer (**Fig. 2c** and **Supplementary Fig. 5b,c**); the expanded range was not attributable solely to the increased length of the linker (**Supplementary Fig. 5c**). We next generated codon-optimized 2X-P2A-Puro and FNLS-P2A-Puro lentiviral vectors and transduced mouse NIH/3T3 cells (**Supplementary Fig. 6**). Two days after sgRNA transduction, FNLS-expressing cells showed greater than 50% C-to-T conversion for all sgRNAs tested (**Supplementary Fig. 7a**), and by day six, 80–95% of all target C nucleotides were converted (**Fig. 2d**). In contrast, at that time point, only one of five sgRNAs showed >80% editing with RA (**Fig. 2d**). On average, FNLS increased editing by 35% compared with RA and by up to 50-fold compared with the original BE3 construct (**Fig. 2d**), and it produced fewer indels and undesired

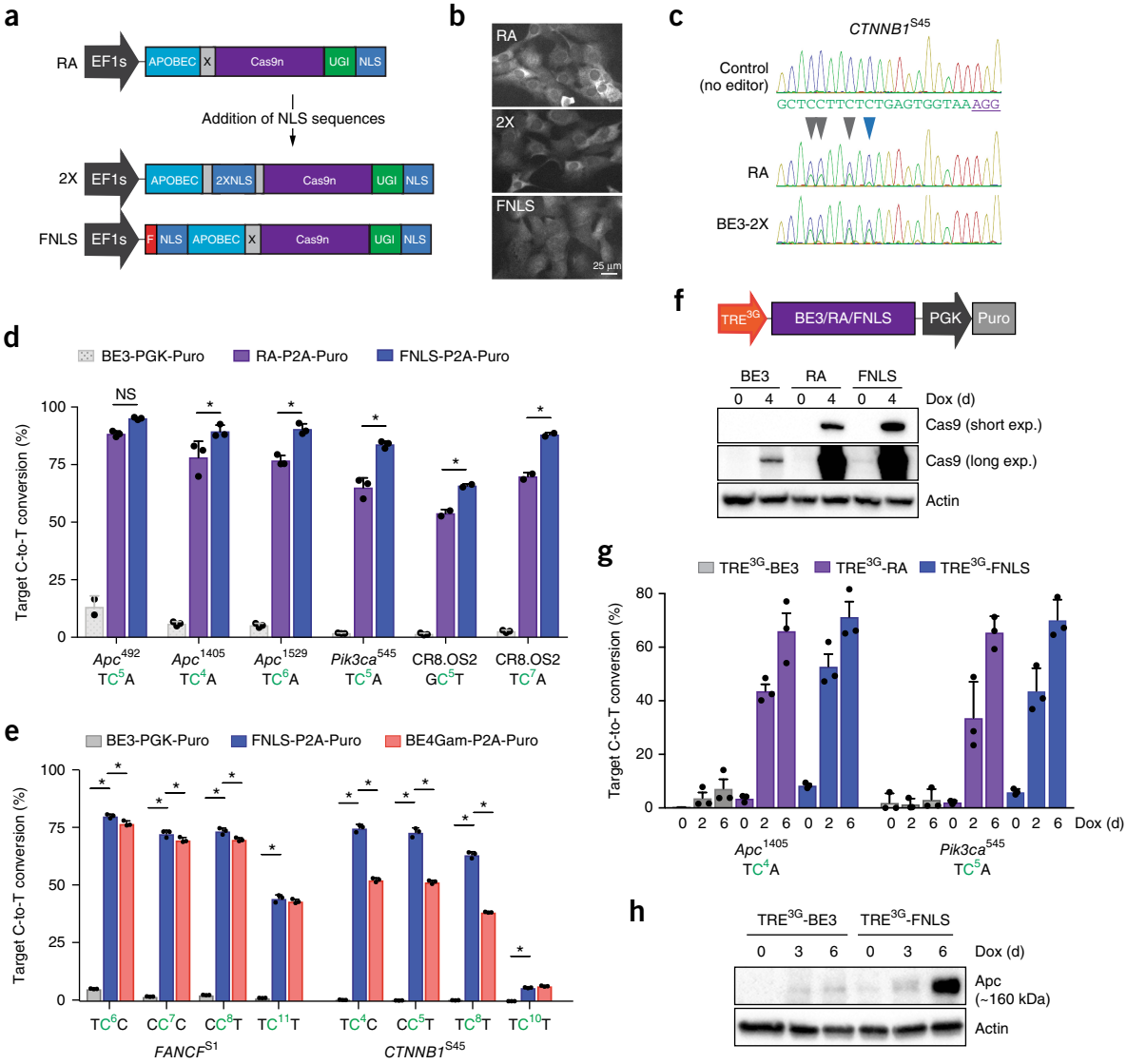


**Figure 2** N-terminal NLS sequences increase the range and potency of target base editing. (**a**) Schematic representation of RA enzyme (top) and two new variants carrying NLS sequences within the XTEN linker (2X) or at the N terminus (FNLS). (**b**) Immunofluorescence staining of Cas9 in NIH/3T3 cells expressing RA, 2X, or FNLS. The experiment was repeated twice with similar results. (**c**) Sanger-sequencing chromatogram showing increased editing of the C at position 10 (blue arrowhead) within the protospacer of a *CTNNB1*$^{S45}$ sgRNA. (**d**) Frequency (%) of C-to-T conversion in NIH/3T3 cells transduced with RA- or FNLS-P2A-Puro lentiviral vectors 6 d after introduction of different sgRNAs, as indicated. Editing in BE3-PGK-Puro cells (from **Fig. 1e**) is shown for comparison. (**e**) Frequency (%) of C-to-T conversion in PC9 cells transduced with BE3-PGK-Puro, FNLS, or BE4Gam$^{RA}$-P2A-Puro lentiviral vectors 6 d after introduction of different sgRNAs, as indicated. In **d** and **e**, graphs show mean values. Error bars, s.e.m. ($n = 3$ biologically independent samples); $^*P < 0.05$ between groups, by two-way ANOVA with Tukey's correction for multiple testing; NS, not significant. (**f**) Schematic representation of dox-inducible BE3 lentiviral construct and immunoblot of Cas9 in transduced and selected NIH/3T3 cells treated with dox (1 µg/ml) for 4 d or left untreated (0 d), as indicated. Blotting was performed twice with similar results. Exp., exposure. (**g**) Frequency (%) of C-to-T conversion in NIH/3T3 cells transduced with TRE$^{3G}$-BE3, TRE$^{3G}$-RA, or TRE$^{3G}$-FNLS, and sgRNA lentiviral vectors, 0, 2, and 6 d after dox treatment. Graph shows mean values. Error bars, s.e.m. ($n = 3$ biologically independent experiments); $^*P < 0.05$ between groups, by two-way ANOVA with Tukey's correction for multiple testing. (**h**) Immunoblot showing induction of truncated (~160 kDa) Apc product after target editing in NIH/3T3 cells expressing BE3 or FNLS. Blotting was performed twice with similar results.

(C-to-A and C-to-G) edits compared with RA (**Supplementary Fig. 7b,c**). To confirm that the reengineered enzymes were active in multiple cell types, we transduced three different human cancer cell lines (PC9, H23, and DLD1) and measured editing at *FANCF* and *CTNNB1* target sites. Although the absolute editing efficiency varied, FNLS increased target C-to-T conversion 15- to 150-fold within the expected window (positions 3–8 bp) (**Fig. 2e** and **Supplementary Fig. 8a**). Indels and undesired edits were elevated in each of the cancer lines compared with 3T3 cells but were decreased through use of an optimized version of the second-generation editor BE4Gam[15] (**Supplementary Figs. 8b** and **9**). The improved efficiency also increased editing at predicted off-target sites, although the overall level of off-target editing remained low (**Supplementary Fig. 10**). As predicted from transfection experiments, the 2X construct did not alter the overall efficiency of the enzyme but significantly extended the range of editing in both mouse and human cells (**Supplementary Fig. 11**).

To provide a temporally controlled system for base editing, we generated (TRE$^{3G}$) doxycycline (dox)-inducible constructs (**Fig. 2f**).

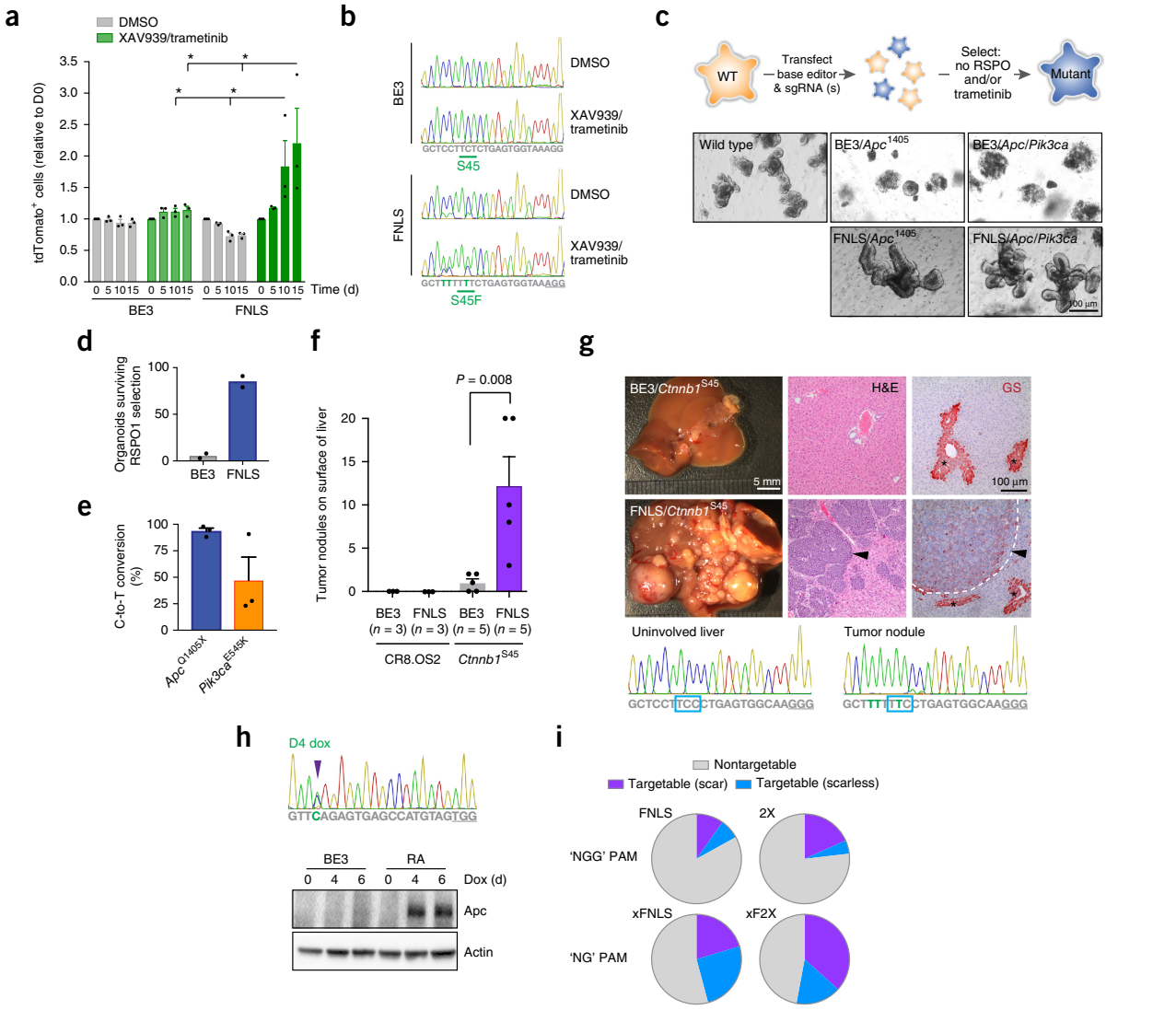


**Figure 3** Optimized enzymes induce efficient base editing in a wide range of cell systems. (**a**) Graph showing relative abundance of tdTomato-positive (sgRNA-expressing) cells in BE3 and FNLS-transduced DLD1 cells, after treatment with DMSO control or XAV939 (1 µM) and trametinib (10 nM). Bars in each case represent serial passages every 5 d, starting at day 0. Graphs show mean values. Error bars, s.e.m. ($n = 3$ biologically independent samples); $^*P < 0.05$ between groups, by two-way ANOVA with Tukey's correction for multiple testing. (**b**) Chromatograms showing sequencing of the *CTNNB1*$^{S45}$ target site in BE3 and FNLS cells, treated with DMSO (top) or XAV939/trametinib (bottom). Chromatogram shows representative of sequencing of three independent samples with similar results. Drug-treated cells show enrichment of the S45F mutation, thus suggesting that this mutation provides an advantage in XAV939/trametinib-treated populations. (**c**) Schematic representation of the process of editing and selection in intestinal organoids. Images show wild-type (WT) mouse small intestinal organoids after editor/sgRNA transfection and selection by RSPO1 withdrawal (6 d). Only FNLS-transfected organoids show consistent outgrowth of large budding organoids in the absence of RSPO1. Images are representative of three independent experiments with similar results. Transfection with tandem sgRNAs targeting *Apc* and *Pik3ca* drives the generation of compound mutant organoids that survive RSPO1 withdrawal and treatment with 25 nM trametinib (additional data in **Supplementary Fig. 13**). (**d**) Number of viable organoids 6 d after RSPO1 withdrawal. Graphs show mean values ($n = 2$ biologically independent samples). (**e**) Mean frequency of *Apc*$^{Q1405X}$ and *Pik3ca*$^{E545K}$ mutations in intestinal organoids after selection in RSPO1-free medium, but no selection in trametinib. Error bars, s.e.m. ($n = 3$ independent transfections). (**f**) Mean number of visible tumor nodules counted in the livers of mice 4 weeks after hydrodynamic delivery of BE3 or FNLS, a mouse *Ctnnb1*$^{S45}$ sgRNA and *Sleeping Beauty* transposon–based *Myc* cDNA. Error bars, s.e.m., $n = 3$–5 biologically independent animals, as indicated; significant differences between groups were calculated with a one-way ANOVA with Tukey's correction for multiple testing. (**g**) Representative images of tumor burden after editing of *Ctnnb1* with FNLS and BE3. Right, hematoxylin and eosin (H&E) staining and immunohistochemical staining for GS (red stain) of representative sections of livers from BE3- and FNLS-transfected mice. Asterisks highlight pericentral hepatocytes staining positively for GS. Arrowheads indicate tumors within the liver in FNLS-transfected mice. Images are representative of five independent samples, with similar results. Bottom, Sanger sequencing from uninvolved liver and a tumor nodule from an FNLS/*Ctnnb1*$^{S45}$ sgRNA-transfected mice, showing near-complete editing of the *Ctnnb1* locus in tumor cells. BE3 tumor nodules were too few and too small to dissect and perform sequencing. (**h**) Sanger-sequencing chromatogram showing editing of *Apc* in embryonic stem cells after 4 d of treatment with dox (1 µg/ml) and immunoblot showing induction of the expected truncated allele of *Apc* in RA-expressing cells but not in BE3 cells. Blotting was performed twice with similar results. (**i**) Pie charts indicating the theoretical number of recurrent cancer-associated mutations that could be modeled with FNLS or 2X ('NGG' PAM) or xFNLS and xF2X ('NG' PAM) constructs. Purple indicates sites where only the target C would be affected (scarless); blue indicates sites where creation of the desired mutation would probably be accompanied by additional C-to-T alterations (scar). An editing window of positions 4–8 (for FNLS and xFNLS) and 4–11 (for 2X and xF2X) is assumed. Details in Online Methods.

As expected, dox treatment drove strong induction of RA and FNLS, but limited expression of the original BE3 construct (**Fig. 2f**). Using sgRNAs targeting *Apc* and *Pik3ca*, we observed a time-dependent generation of target missense (*Pik3ca*$^{E545K}$) and nonsense (*Apc*$^{Q1405X}$) mutations (**Fig. 2g**). In agreement with earlier observations, both RA and FNLS dramatically increased editing efficiency compared with that of the original BE3 enzyme (**Fig. 2g**), which for *Apc*$^{1405}$ led to production of a truncated Apc protein (**Fig. 2h**).

Together, these data demonstrate that our optimized enzymes increase the range (2X) and efficiency (FNLS) of targeted base editing. To demonstrate the utility and effects of the improved editors, we set out to engineer a series of precise and functional genetic changes in different model systems: human cancer cells, intestinal organoids, mouse embryonic stem cells, and mouse hepatocytes *in vivo*.

DLD1 colorectal cancer cells are sensitive to combined inhibition of tankyrase and MEK[16,17], but WNT-activating mutations in *CTNNB1* are predicted to bypass this response[18]. Hence, we cultured DLD1 cells carrying sgRNAs targeting the *CTNNB1*$^{S45}$ or *FANCF*$^{S1}$ codons in the presence of inhibitors of tankyrase (XAV939; 1 μM) and MEK (trametinib; 10 nM) and tracked tdTomato-positive, sgRNA-expressing cells over time (**Supplementary Fig. 12**). At treatment initiation, cells expressing RA, 2X, and FNLS, but not BE3, showed efficient editing (40–50%) at the *FANCF* control site and showed *CTNNB1*$^{S45F}$ mutations at a frequency of 12–18% (**Supplementary Fig. 8a**). In the presence of inhibitors, *CTNNB1* sgRNA-transduced cells (expressing RA, 2X, or FNLS, but not the original BE3) outcompeted the nontransduced population (**Fig. 3a** and **Supplementary Fig. 12b**), and inhibitor-treated cells, but not control dimethylsulfoxide (DMSO)-treated cells, showed enrichment in the expected S45F alteration (**Fig. 3b**). Together, these data imply that editor-induced *CTNNB1*$^{S45F}$ mutations are functional and enable resistance to upstream WNT suppression by tankyrase inhibitors.

Truncating *Apc* mutations are the most common genetic events observed in human colorectal cancers[19], and they drive WNT- and R-Spondin (RSPO)-independent proliferation. To engineer *Apc* truncations, we cotransfected intestinal organoids with either BE3 or FNLS, and the *Apc*$^{1405}$ sgRNA (**Fig. 3c**). FNLS-transfected cultures showed a tenfold higher outgrowth of RSPO1-independent organoids than BE3-transfected cells (**Fig. 3d**) and carried a high frequency of targeted *Apc* editing (>97%) (**Fig. 3e**) with less than 1% indels. Co-delivery of two tandem-arrayed sgRNAs (*Apc*$^{1405}$ and *Pik3ca*$^{545}$) produced *Apc*$^{Q1405X}$; *Pik3ca*$^{E545K}$ double-mutant organoids (**Fig. 3c,e**) that were able to survive and expand in the presence of a MEK inhibitor (trametinib; 25 nM) (**Supplementary Fig. 13a,b**), as has been described for homology directed repair–generated *PIK3CA*$^{E545K}$ mutations in human organoids[20].

In hepatocellular carcinoma, *CTNNB1* mutations are the primary mechanism of WNT-driven tumorigenesis. To explore the potential of base editors to drive tumor formation *in vivo*, we introduced BE3 or FNLS, a mouse *Ctnnb1*$^{S45}$ sgRNA and *Myc* cDNA to the livers of adult mice via hydrodynamic transfection. After 4 weeks, three of five BE3-transfected animals showed one or two small tumor nodules on the liver, whereas FNLS-transfected mice showed a dramatically higher disease burden, and all mice (five of five) carried multiple tumors (**Fig. 3f**). The tumors resembled hepatocellular carcinoma with a trabecular and solid growth pattern, and showed upregulation of the WNT target glutamine synthetase[21] (GS; **Fig. 3g**). The tumor nodules showed near-complete editing of the *Ctnnb1* locus, creating activating S45F mutations (**Fig. 3g**).

An alternate approach to *in vivo* somatic base editing is the generation of temporally regulated transgenic strains, which enables the manipulation of tissues and cell types that cannot be easily transfected *in vivo* and avoids the potential immunogenicity of exogenous Cas9 delivery[22,23]. Accordingly, we generated TRE-inducible, knock-in mouse embryonic stem cells. We chose RA for targeting mouse embryonic stem cells, because we observed low-level 'leaky' editing in 3T3 cells carrying TRE$^{3G}$-FNLS lentivirus (**Fig. 2g**). TRE-RA cells showed efficient dox-dependent C-to-T conversion and generation of the predicted mutant alleles (**Fig. 3h** and **Supplementary Fig. 13c**). Together, these data show that optimized RA and FNLS constructs offer a flexible and efficient platform to engineer directed somatic alterations in animals.

Base editing is arguably the most direct approach to engineer disease-associated SNVs in model systems. To estimate the number of cancer-related SNVs that could potentially be modeled with Cas9-mediated base editing, we analyzed MSK-IMPACT targeted deep sequencing of more than 22,000 tumors and defined a list of 2,696 recurrent mutations (observed in at least four individual patients). With a conservative base-editing window of positions 4–8 (FNLS) and 4–11 (2X), we estimate that ~17% of cancer-associated SNVs could be engineered with FNLS, and ~23% could be engineered by exploiting the expanded range of the 2X construct. Of these, approximately 40% could be generated without any collateral editing (or 'scar') at nontarget C nucleotides (**Fig. 3i**). In principle, through use of Cas9 variants with less restrictive PAM requirements (for example, xCas9)[24], more than 50% of all mutations could be created (**Fig. 3i**). To that end, we produced optimized xFNLS and xF2X constructs that enable more efficient base editing than the published xBE3 construct (**Supplementary Fig. 14**). Notably, the xCas9-derived base editors showed lower on-target activity for both sgRNAs and cell lines tested (**Supplementary Fig. 14b,c**). Further work will be required to assess the contexts in which these newer variants enable efficient base editing.

Here, by optimizing protein expression and nuclear targeting, we developed a range of potent base-editing and Cas9 enzymes that dramatically improve DNA editing across multiple *in vitro* and *in vivo* model systems. These tools, along with similar optimized versions for A-base editors[25,26], should enable the rapid generation of targeted SNVs in a variety of cell systems *in vitro* and *in vivo* and should be key to implementing base editing in genetic screens, in which high efficiency is essential. Moreover, the improved protein expression of our reengineered enzymes should substantially enhance therapeutic approaches that rely on delivery of mRNA molecules[27], whereas enhanced nuclear targeting will probably improve the delivery and/or activity of ribonuclear particles[14]. In all, we expect that the toolkit described herein will make base editing a feasible and accessible option for a wide range of research and therapeutic applications.

## METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

### ACKNOWLEDGMENTS

This work was supported by a project grant from the NIH/NCI (CA195787-01), a U54 grant from the NIH/NCI (U54OD020355), a project grant from the Starr Cancer Consortium (I10-0095), a Research Scholar Award from the American Cancer Society (RSG-17-202-01), and a Stand Up to Cancer Colorectal Cancer Dream Team Translational Research Grant (SU2C-AACR-DT22-17). Stand Up to Cancer is a program of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, a scientific partner of SU2C. M.P.Z. is supported in part by National Cancer

Institute (NCI) grant NIH T32 CA203702. E.M.S. was supported by a Medical Scientist Training Program grant from the National Institute of General Medical Sciences of the NIH under award number T32GM07739 to the Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional MD–PhD Program and an F31 Award from the NCI/NIH under grant number 1 F31 CA224800-01. E.R.K. is supported by an F31 NRSA predoctoral fellowship from the NCI/NIH under award number F31CA192835. F.J.S.-R. was supported by the MSKCC TROT program (5T32CA160001) and is supported as an HHMI Hanna Gray Fellow. S.W.L. is supported as the Geoffrey Beene Chair of Cancer Biology and as an Investigator of the Howard Hughes Medical Institute. D.F.T. is supported by the Helmholtz Association (VH-NG-1114) and by the German Research Foundation (DFG) project B05, SFB/TR 209 'Liver Cancer'. L.E.D. was supported by a K22 Career Development Award from the NCI/NIH (CA 181280-01). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. We thank H. Varmus (Weill Cornell Medicine) for providing cells.

## AUTHOR CONTRIBUTIONS

M.P.Z. and E.M.S. performed experiments, analyzed data, and wrote the paper. A.K., M.F., A.S., S.G., E.M., T.H., J.T., and F.J.S.-R. performed experiments and analyzed data. M.B. and A.Y.S. performed and analyzed *in vivo* experiments. D.F.T. designed and supervised *in vivo* experiments. E.R.K. performed computational analysis of MSKCC IMPACT data. J.S., S.W.L., and C.R.V. supplied critical previously unreported reagents. L.E.D. performed and supervised experiments, analyzed data, and wrote the paper.

## COMPETING INTERESTS

The authors declare no competing interests.



1. 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).
2. Nishida, K. *et al.* Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
3. Hess, G.T. *et al.* Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
4. Ma, Y. *et al.* Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).
5. Gaudelli, N.M. *et al.* Programmable base editing of A.T to G.C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
6. Cox, D.B.T. *et al.* RNA editing with CRISPR-Cas13. *Science* **358**, 1019–1027 (2017).
7. Rees, H.A. *et al.* Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).
8. Kim, K. *et al.* Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
9. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
10. Kim, S., Bae, T., Hwang, J. & Kim, J.S. Rescue of high-specificity Cas9 variants using sgRNAs with matched 5′ nucleotides. *Genome Biol.* **18**, 218 (2017).
11. Kleinstiver, B.P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).
12. Kleinstiver, B.P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).
13. Dow, L.E. *et al.* Inducible in vivo genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**, 390–394 (2015).
14. Staahl, B.T. *et al.* Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nat. Biotechnol.* **35**, 431–434 (2017).
15. Komor, A.C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
16. Huang, S.M. *et al.* Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature* **461**, 614–620 (2009).
17. Schoumacher, M. *et al.* Inhibiting Tankyrases sensitizes KRAS-mutant cancer cells to MEK inhibitors via FGFR2 feedback signaling. *Cancer Res.* **74**, 3294–3305 (2014).
18. Mashima, T. *et al.* mTOR signaling mediates resistance to tankyrase inhibitors in Wnt-driven colorectal cancer. *Oncotarget* **8**, 47902–47915 (2017).
19. Cancer Genome Atlas Network. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330–337 (2012).
20. Matano, M. *et al.* Modeling colorectal cancer using CRISPR-Cas9-mediated engineering of human intestinal organoids. *Nat. Med.* **21**, 256–262 (2015).
21. Cadoret, A. *et al.* New targets of beta-catenin signaling in the liver are involved in the glutamine metabolism. *Oncogene* **21**, 8293–8301 (2002).
22. 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).
23. 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).
24. Hu, J.H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
25. Koblan, L.W. *et al.* Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* https://doi.org/10.1038/nbt.4172 (2018).
26. Ryu, S.M. *et al.* Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat. Biotechnol.* **36**, 536–539 (2018).
27. Yin, H. *et al.* Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat. Biotechnol.* **35**, 1179–1187 (2017).

## ONLINE METHODS

**Cloning.** All primers, Ultramers, and gBlocks used for cloning are listed in **Supplementary Tables 2–5**. pCMV-BE3-2X (CMV-2X) and pCMV-BE3-FNLS were generated through Gibson assembly, by combining an XmaI-digested (2X) or NotI-digested (FNLS) pCMV-BE3 backbone with DNA Ultramers (BE3-2X NLS or T7-FLAG-NLS). Double-stranded DNA from Ultramers was generated by PCR amplification with primers XTEN-NLS_F/XTEN-NLS_R and T7-FLAG_F/T7-FLAG_R. pLenti-BE3-PGK-Puro (LBPP) was generated through Gibson assembly, by combining the following four DNA fragments: (i) PCR-amplified EF1s promoter (FSR-19/FSR-20), (ii) PCR-amplified BE3 cDNA (FSR-114/FSR-115), (iii) PCR-amplified PGK-Puro cassette (FSR-16/FSR-17), and (iv) BsrGI/PmeI-digested pLL3-based lentiviral backbone. pLenti-BE3$^{RA}$-PGK-Puro (LRPP) was generated through Gibson assembly, by combining a PCR-amplified BE3$^{RA}$ cDNA (BE3$^{RA}$-PGKPuro_F/BE3$^{RA}$-PGKPuro_R) and an NheI/AvrII-digested BE3-PGK-Puro backbone. pLenti-FNLS-PGK-Puro (LFPP) was generated by restriction cloning of a FLAG-NLS-APOBEC BamHI (blunt)/EcoRI-digested fragment into an NheI (blunt)/EcoRI-digested pLenti-BE3$^{RA}$-PGK-Puro backbone. pLenti-BE3$^{RA}$-P2A-Puro (LR2P) was generated through Gibson assembly, by combining the following four DNA fragments: (i) PCR-amplified APOBEC-XTEN cDNA (BE3$^{RA}$_APOBEC_F/BE3$^{RA}$_XTEN_R), (ii) PCR-amplified Cas9n (BE3$^{RA}$_Cas9n_F/BE3$^{RA}$_Cas9n_R), (iii) PCR-amplified UGI (BE3$^{RA}$_UGI_F/BE3$^{RA}$_UGI_R), and (iv) BamHI/NheI-digested pLenti-Cas9-P2A-Puro viral backbone. Some wobble positions were altered within the UGI (SGGS) linker to avoid complications during Gibson assembly because of an identical region downstream of UGI. pLenti-FNLS-P2A-Puro (LF2P) was generated by restriction cloning of a PCR-amplified (BamHI-FLAG_F/APOBEC-RI_R) BamHI/EcoRI-digested FLAG-NLS-APOBEC fragment into a BamHI/EcoRI-digested pLenti-BE3$^{RA}$-P2A-Puro backbone. pLenti-2X-P2A-Puro (LX2P) was generated through Gibson assembly, by combining a PCR-amplified APOBEC-2XNLS fragment (BE3$^{RA}$_APOBEC_F/BE3$^{RA}$_XTEN_R) and a BamHI/XmaI-digested pLenti-BE3$^{RA}$-P2A-Puro backbone. pLenti-TRE$^{3G}$-BE3-PGK-Puro (L3BP) was generated through Gibson assembly, by combining a PCR-amplified TRE$^{3G}$ promoter (3G_F/3G_R) and APOBEC fragment (APOBEC_F/BE3$^{RA}$_XTEN_R) with an XmaI-digested pLenti-BE3-PGK-Puro backbone. pLenti-TRE$^{3G}$-BE3$^{RA}$-PGK-Puro (L3RP) was generated through Gibson assembly, by combining a PCR-amplified TRE$^{3G}$ promoter (3G_F/3G_R) and APOBEC fragments (APOBEC_F/BE3$^{RA}$_XTEN_R) with an XmaI-digested pLenti-BE3$^{RA}$-PGK-Puro backbone. pLenti-TRE$^{3G}$-FNLS-PGK-Puro (L3FP) was generated through Gibson assembly, by combining a PCR-amplified TRE$^{3G}$ promoter (3G_F/3G_R) and FNLS-APOBEC fragments (FNLS-APOBEC_F/BE3$^{RA}$_XTEN_R) with an XmaI-digested pLenti-BE3$^{RA}$-PGK-Puro backbone. pCol1a1-TRE-BE3 (cTBE3) was generated through Gibson assembly, by combining a PCR-amplified BE3 cDNA (cTRE_BE3_F/cTRE_BE3_R) with an EcoRI-digested pCol1a1-TRE backbone. pCol1a1-TRE-BE3$^{RA}$ (cTBE3$^{RA}$) was generated through a two-step strategy involving (i) Gibson assembly to introduce a PCR-amplified UGI fragment (UGI_F/UGI_R) into a XhoI-digested pCol1a1-TRE-Cas9n backbone (Col1a1-TRE-Cas9n-UGI) and (ii) restriction cloning of a PCR-amplified, XhoI/EcoRV-digested APOBEC-XTEN-Cas9n (APOBEC_F2/APOBEC_R2) fragment into an EcoRV-digested Col1a1-TRE-Cas9n-UGI backbone. pLenti-U6-sgRNA-tdTomato-P2A-Blas (LRT2B) was generated through Gibson assembly, by combining a PCR-amplified EFs-tdTomato-P2A-blasticidin fragment (pLRT2B_EFs_F/pLRT2B_WPRE_R) with an XhoI/BsrGI-digested pLenti-U6-sgRNA-GFP (LRG) backbone. pLenti-VQR-P2A-Puro (LQ2P), pLenti-VRER-P2A-Puro (LER2P), and pLenti-HF1-P2A-Puro (LH2P) were generated through Gibson assembly, by combining PCR-amplified Cas9 variants (from Addgene stocks 65771, 65773, and 72247, respectively; primers KJ_Cas9_F/KJ_Cas9_R) with a BamHI/NheI-digested pLenti-P2A-Puro backbone. pLenti-VQR$^{RA}$-P2A-Puro (LQR2P), pLenti-VRER$^{RA}$-P2A-Puro (LERR2P), and pLenti-HF1$^{RA}$-P2A-Puro (LHR2P) were generated through Gibson assembly, by combining one of two PCR-amplified regions of the 3′ half of Cas9 (Cas9_RA_5F/Cas9_RA_5R or Cas9_RA_3F/Cas9_RA_3R), with gBlock fragments containing the appropriate point mutations (VQR_GB, VRER_GB, or HF1_GB) and an EcoRV/NheI-digested pLenti-Cas9-P2A-Puro backbone. pLenti-xCas9RA-P2A-Puro, pLenti-xFNLS-P2A-Puro, pLenti-xF2X-P2A-Puro, and pLenti-xBE4Gam-P2A-Puro were generated through Gibson assembly of four PCR-amplified regions (EF1s_xCas9_AF × xCas9_AR; xCas9_BF × xCas9_BR; xCas9_CF × xCas9_CR; and xCas9_DF × xCas9_DR) and a BamHI/NheI-digested pLenti-Cas9-P2A-Puro backbone. All constructs described above are schematized in **Supplementary Figure 15**.

**Cell culture, transfection, and transduction.** *Culture.* HEK293T (ATCC CRL-3216) and DLD1 (ATCC CCL-221) cells were maintained in Dulbecco's Modified Eagle's Medium (Corning) supplemented with 10% (vol/vol) FBS, at 37° and 5% $CO_2$. PC9 (obtained from H. Varmus) and NCI-H23 (ATCC CRL-5800) cells were maintained in RMPI-1640 medium supplemented with 10% (vol/vol) FBS, at 37° and 5% $CO_2$. NIH/3T3 (ATCC CRL-1658) cells were maintained in Dulbecco's Modified Eagle's Medium (Corning) supplemented with 10% (vol/vol) bovine calf serum. Mouse KH2 embryonic stem cells were maintained on irradiated MEF feeders in M15 medium containing LIF, as previously described[28].

*Transfection.* For transfection-based editing experiments in HEK293Ts, cells were seeded on a 12-well plate at 80% confluence and cotransfected with 750 ng of base editor, 750 ng of sgRNA expression plasmid, and 4.5 µl of polyethylenimine (1 mg/ml). Cells were harvested for genomic DNA 3 d after transfection. For virus production, HEK293T cells were plated in a six-well plate and transfected 12 h later (at 95% confluence) with a prepared mix in DMEM (with no supplements) containing 2.5 µg of lentiviral backbone, 1.25 µg of PAX2, 1.25 µg of VSV-G, and 15 µl of polyethylenimine (1 mg/ml). 36 h after transfection, the medium was replaced with target cell collection medium, and supernatants were harvested every 8–12 h up to 72 h after transfection. ESC *col1a1*-targeting constructs were introduced via nucleofection in 16-well strips, with buffer P3 (Lonza V4XP-3032) in a 4D Nucleofector with X-unit attachment (Lonza). Two days after nucleofection, cells were treated with medium containing 150 µg/ml hygromycin B, and individual surviving clones were picked after 9–10 d of selection. Two days after clones were picked, hygromycin was removed from the medium, and cells were cultured in M15 thereafter. To confirm integration at the *col1a1* locus, we used a multiplex *col1a1* PCR[28].

*Transduction.* $7.5 \times 10^4$ NIH/3T3, DLD1, PC9, and H23 cells were plated on six-well plates. 24 h after plating, cells were transduced with viral supernatants in the presence of polybrene (8 µg/µl). Two days after transduction, cells were selected in puromycin (2 µg/ml) or blasticidin S (4 µg/ml). 500,000 ESCs were plated in six-well plates on gelatin and spinoculated (90 min, 32 °C, 2,100 r.p.m.) with 150 µl of concentrated lentiviral particles (with 100 mg/ml polyethylene glycol, Sigma Aldrich P4338) in 1 ml of medium containing polybrene (8 µg/µl). After centrifugation, the medium was replaced.

*Fluorescence competitive proliferation assays.* DLD1 cells expressing BE3, RA, 2X, or FNLS were transduced with LRT2B-CTNNB1$^{S45}$ or LRT2B-FANCF$^{S1}$, selected with blasticidin for 4 d, and mixed at defined proportions with parental cells. $5 \times 10^4$ mixed cells were seeded in 96-well plates and treated with DMSO or 1 µM XAV939 plus 10 nM trametinib every 48 h, and the remaining tdTomato-positive cells were tracked every 5 d by flow cytometry with a BD-Accuri C6 cytometer.

**Organoid isolation, culture, and transfection.** Organoid isolation was performed as previously described[29,30]. Briefly, 15 cm of the proximal small intestine was removed, flushed, and washed with cold PBS. The intestine was then cut into 5-mm pieces and placed into 10 ml cold 5 mM EDTA-PBS and vigorously resuspended with a 10-ml pipette. The supernatant was aspirated and replaced with 10 ml EDTA and placed at 4 °C on a benchtop roller for 10 min. This procedure was then repeated a second time for 30 min. The supernatant was aspirated, and then 10 ml of cold PBS was added to the intestine, and samples were resuspended with a 10-ml pipette. After this 10-ml PBS-containing crypt fraction was collected, the procedure was repeated, and each successive fraction was collected and examined under a microscope for the presence of intact intestinal crypts and the absence of villi. The 10-ml fraction was then mixed with 10 ml DMEM basal medium (Advanced DMEM F/12 containing pen/strep, glutamine, and 1 mM *N*-acetylcysteine (Sigma Aldrich A9165-SG)) containing 10 U/ml DNase I (Roche 04716728001), and filtered through a 100-µm filter. Samples were then filtered through a 70-µm filter into an FBS (1 ml)-coated tube and spun at 1,200 r.p.m. for 3 min. The supernatant was aspirated, and the cell

pellets (purified crypts) were resuspended in basal medium, mixed 1:10 with Growth Factor Reduced Matrigel (BD 354230), and plated in multiple wells of a 48-well plate. After polymerization for 15 min at 37 °C, 250 µl of small intestinal organoid growth medium (basal medium containing 50 ng/ml EGF (Invitrogen PMG8043), 100 ng/ml Noggin (Peprotech 250-38), and R-spondin (conditioned medium) was then laid on top of the Matrigel.

*Maintenance*. The medium on organoids was changed every 2 d, and organoids were passaged 1:4 every 5–7 d. For passaging, the growth medium was removed, and the Matrigel was resuspended in cold PBS and transferred to a 15-ml conical tube. The organoids were mechanically disassociated with a p1000 or a p200 pipette, through pipetting 50–100 times. 7 ml of cold PBS was added to the tube and pipetted 20 times to fully wash the cells. The cells were then centrifuged at 1,000 r.p.m. for 5 min, and the supernatant was aspirated. Cells were then resuspended in GFR Matrigel and replated as above. For freezing, after spinning, the cells were resuspended in basal medium containing 10% FBS and 10% DMSO and stored in liquid nitrogen indefinitely.

*Transfection*. Mouse small intestinal organoids were cultured in medium containing CHIR99021 (5 µM) and Y-27632 (10 µM) for 2 d before transfection. Cell suspensions were produced by dissociating organoids with TrypLE express (Invitrogen 12604) for 5 min at 37 °C. After trypsinization, cell clusters in 300 µl transfection medium were combined with 100 µl DMEM/F12/Lipofectamine2000 (Invitrogen 11668)/DNA mixture (97 µl/2 µl/1 µg) and transferred into a 48-well culture plate. The plate was centrifuged at 600*g* at 32 °C for 60 min, then incubated another 6 h at 37 °C. The cell clusters were spun down and plated in Matrigel. For selection of organoids with *Apc* mutations, exogenous RSPO1 was withdrawn 2–3 d after transfection. For selection of *Pik3ca* alterations, organoids were cultured in medium containing trametinib (25 nM) for 1 week.

**Hydrodynamic delivery.** All animal experiments were authorized by the regional board, Karlsruhe, Germany (animal permit number G178/16) or the Institutional Animal Care and Use Committee (IACUC) at Weill Cornell Medicine (2014-0038). Eight-week-old C57Bl6/N mice (Charles River) were injected with 0.9% sterile sodium chloride solution containing 20 µg pLenti-BE3-P2A-Puro or pLenti-FNLS-P2A-Puro, 10 µg of the respective sgRNA vector, and 5 µg pT3 EF1a-myc, as well as 1 µg CMV-SB13. The total injection volume corresponded to 20% of each mouse's body weight and was injected into the lateral tail vein in 5–7 s. No animals were excluded from the analyses; the investigators were not blinded during the analyses.

**Lentiviral titer assay.** Lentiviral titers were calculated with a quantitative PCR–based kit (LV900 Applied Biological Materials), according to the manufacturer's instructions. Briefly, 2 µl of unconcentrated viral supernatant was lysed for 3 min at room temperature, and the crude lysate was used to perform qPCR amplification. The concentration of viral particles was calculated as described in the protocol at http://www.abmgood.com/High-Titer-Lentivirus-Calculation.html.

**Flow cytometry.** TdTomato protein abundance was measured by calculating the mean fluorescence intensity after analysis on a BD Accuri C6 flow cytometer. The experiments described represent three independent viral transductions, each at a different MOI, to account for any effects of gene dosage.

**Genomic-DNA isolation.** Cells were lysed in genomic lysis buffer (10 mM Tris, pH 7.5, 10 mM EDTA, 0.5% SDS, and 400 µg/ml proteinase K) for at least 2 h at 55 °C. After proteinase K heat inactivation at 95 °C for 15 min, 0.5 volume of 5 M NaCl was added, and samples were centrifuged for 10 min at 15,000 r.p.m. Supernatants were mixed with one volume of isopropanol, and DNA precipitates were washed in 70% EtOH before resuspension in 10 mM Tris, pH 8.0.

**Puro copy-number assays.** For quantification of lentiviral integrations in transduced cells, we used a custom-designed TaqMan copy-number assay (Invitrogen) to detect the *Pac* (*puroR*) gene. Amplification was conducted on a QuantStudio 6 Real-Time PCR system (Applied Biosystems), with TaqMan master mix reagent (Applied Biosystems) and specific primers and probe (forward, 5′-GCGGTGTTCGCCGAGAT; reverse, 5′-GAGGCCTTCCATCTGTTGCT; probe (FAM), CCGGGAACCGCTCAACTC).

**Protein analysis.** DLD1, PC9, and 3T3 cells were scraped from a confluent well of a six-well plate in 100 µl RIPA buffer, then centrifuged at 4 °C at 13,000 r.p.m. to collect protein lysates. DLD1 cells were pelleted from a confluent well of a six-well plate at 1,000 r.p.m. for 4 min, resuspended in 200 µl RIPA buffer, then centrifuged at 4 °C at 13,000 r.p.m. to collect protein lysates. Organoids were collected from a confluent well of a 12-well plate (~100 µl Matrigel) in 200 µl Cell Recovery Solution (Corning 354253), incubated on ice for 20 min, then pelleted at 300*g* for 5 min. The pellet was then resuspended in 20 µl RIPA buffer and centrifuged at 4 °C at 13,000 r.p.m. to collect protein lysates. ESCs were collected at the indicated time points and filtered through a 40-µm cell strainer (Fisher Scientific) to remove feeders, then pelleted at 1,000 r.p.m. for 4 min and resuspended in 100 µl RIPA buffer. Samples were centrifuged at 4 °C at 13,000 r.p.m. to collect protein lysates. Antibodies to the following proteins were used for western blot analyses: Cas9 (BioLegend 844301), actin (Abcam ab49900), and Apc (Millipore MABC202).

**Immunofluorescence staining and microscopy.** $2 \times 10^4$ editor-expressing 3T3 cells were plated in a chamber slide. 24 h later, cells were washed in PBS and fixed in PBS, 4% PFA solution for 20 min at RT and incubated in permeabilization buffer (PBS, 0.5% Triton X-100) for 10 min on ice. Then cells were stained with anti-Cas9 (BioLegend 844301) at 4 °C overnight. Donkey anti-mouse Alexa 594 (Thermo Fisher Scientific A21203) was used as a secondary antibody.

**Immunohistochemistry.** Slides containing 3-µm-thick liver sections were deparaffinized and rehydrated with a descending graded alcohol series. For antigen retrieval, slides were cooked in sodium citrate buffer, pH 6.0, in a pressure cooker for 8 min. Subsequently, endogenous HRP was blocked for 10 min in 3% $H_2O_2$. Slides were blocked with in PBS containing 5% BSA for 1 h before incubation with the primary antibody (anti–mouse GS, BD BD610517) overnight (1:200 dilution in PBS, 5% BSA). Slides were washed three times, and staining was visualized with a DAKO Real Detection System (DAKO K5003) according to the manufacturer's instructions.

**PCR amplification for MiSeq.** Target genomic regions of interest were amplified by PCR with the primer pairs listed in **Supplementary Table 4**. PCR was performed with Herculase II Fusion DNA polymerase (Agilent 600675) according to the manufacturer's instructions with 200 ng of genomic DNA as a template, under the following PCR conditions: 95 °C, 2 min; 95 °C, 20 s → 58 °C, 20 s → 72 °C, 30 s for 34 cycles; and 72 °C, 3 min. PCR products were column purified (Qiagen) for analysis through Sanger sequencing or MiSeq.

**Mutation detection by T7 assays.** Cas9-induced mutations were detected with T7 endonuclease I (NEB). Briefly, an approximately 500-bp region surrounding the expected mutation site was PCR-amplified with Herculase II (Agilent 600675). PCR products were column purified (Qiagen) and subjected to a series of melt–anneal temperature cycles with annealing temperatures gradually lowered in each successive cycle. T7 endonuclease I was then added to selectively digest heteroduplex DNA. Digest products were visualized on a 2.5% agarose gel.

**Off-target predictions.** sgRNA-dependent off-target mutations were predicted from a previous publication[31] or with the 'Cas-OFFinder' prediction tool (http://www.rgenome.net/cas-offinder/)[32]. Sites were prioritized as the most likely to show off-target editing if they contained the fewest mismatches, and those mismatches were clustered toward the 5′ end of the sgRNA.

**DNA-library preparation and MiSeq.** DNA-library preparation and sequencing reactions were conducted at GENEWIZ. An NEB NextUltra DNA Library Preparation kit was used according to the manufacturer's recommendations (Illumina). Adaptor-ligated DNA was indexed and enriched through limited-cycle PCR. The DNA library was validated with a TapeStation (Agilent) and was quantified with a Qubit 2.0 fluorometer. The DNA library was quantified through real-time PCR (Applied Biosystems). The DNA library was loaded


doi:10.1038/nbt.4194

on an Illumina MiSeq instrument according to the manufacturer's instructions (Illumina). Sequencing was performed with a 2 × 150 paired-end configuration. Image analysis and base calling were conducted in MiSeq Control Software on a MiSeq instrument and verified independently with a custom workflow in Geneious R11.

**Identification of recurrent cancer-associated mutations.** With MSK-IMPACT targeted deep sequencing of 473 cancer-relevant genes across 22,647 patient samples, we identified recurrent somatic variants present in four or more individual samples. This procedure generated a list of 2,696 somatic missense, nonsense, and splice-site mutations. The flanking sequences around each mutation were retrieved and queried for the presence of a relevant PAM (NGG for FNLS and 2X; NG for xFNLS and xF2X) within a specified distance downstream of the target C nucleotide, with the following packages (implemented in *R*, https://cran.r-project.org/): *Bioconductor*, *BSgenome*, and *Biostrings*. For G-to-A mutations, the reverse-complement strand was examined. Target C (or G) nucleotides were considered 'editable' if they were within positions 4–8 of the protospacer (for FNLS and xFNLS) or positions 4–11 (for 2X and xF2X). The presence of a nontargeted C in the editing window was noted, and editable mutations were parsed into those in which only the target C was edited (scarless) and those in which an additional C was predicted to be altered (scar).

**Statistics.** All statistical tests used throughout the manuscript are indicated in the appropriate figure legends. In general, to compare two conditions, a two-sided Student's *t* test was used, assuming unequal variance between samples. In most cases, analyses were performed with one-way or two-way ANOVA, with Tukey's correction for multiple comparisons. Unless otherwise stated, each replicate represents a biologically independent experiment, i.e., an independent cell transfection, independently transduced cell line, or independent animal. Results of all statistical tests are available in **Supplementary Table 6**.

**Life Sciences Reporting Summary.** Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

**Data availability.** Data have been deposited in the NCBI database under accession number SRP151111. All lentiviral and expression vectors described have been deposited at Addgene for distribution.


28. Dow, L.E. *et al*. A pipeline for the generation of shRNA transgenic mice. *Nat. Protoc.* **7**, 374–393 (2012).
29. Han, T. *et al*. R-Spondin chromosome rearrangements drive Wnt-dependent tumour initiation and maintenance in the intestine. *Nat. Commun.* **8**, 15945 (2017).
30. O'Rourke, K.P. *et al*. Transplantation of engineered organoids enables rapid generation of metastatic mouse models of colorectal cancer. *Nat. Biotechnol.* **35**, 577–582 (2017).
31. Tsai, S.Q. *et al*. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
32. 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).

nature research

Corresponding author(s): Lukas Dow

# Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Please do not complete any field with "not applicable" or n/a. Refer to the help text for what text to use if an item is not relevant to your study.
For final submission: please carefully check your responses for accuracy; you will not be able to make changes later.

## Experimental design

1. Sample size

Describe how sample size was determined.

Three independent replicates were performed for most in vitro experiments, based on experience with similar previous studies. For in vivo work, where variance is increased, we used 5 mice per experimental group. This provides the power (0.7) to detect an effect size of 2 (2x increase or decrease), at a probability level of 0.05.

2. Data exclusions

Describe any data exclusions.

No data were excluded

3. Replication

Describe the measures taken to verify the reproducibility of the experimental findings.

Experiments were performed on three independent occasions. For in vitro, transduction-based experiment that make up the majority of this manuscript, this means three independent transduction / selection and target sequencing. All attempts at replication were successful

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

No randomization was performed

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Investigators were not blinded during collection or analysis

Note: all in vivo studies must report how sample size was determined and whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| n/a | Confirmed |
|---|---|
| ☐ | ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| ☐ | ☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| ☐ | ☒ A statement indicating how many times each experiment was replicated |
| ☐ | ☒ 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 any assumptions or corrections, such as an adjustment for multiple comparisons |
| ☐ | ☒ Test values indicating whether an effect is present<br>*Provide confidence intervals or give results of significance tests (e.g. P values) as exact values whenever appropriate and with effect sizes noted.* |
| ☐ | ☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| ☐ | ☒ Clearly defined error bars in all relevant figure captions (with explicit mention of central tendency and variation) |

*See the web collection on statistics for biologists for further resources and guidance.*







Nature Biotechnology: doi:10.1038/nbt.4194

## Software

Policy information about availability of computer code

7. Software

| | |
|---|---|
| Describe the software used to analyze the data in this study. | Statistical calculations were performed in Prism. Sequence analysis was performed on MiSeq control software, and independently validated using Geneious R11. Additional software packages used for the analysis of recurrent cancer-associated mutations: R (version 3.2.1), Bioconductor (version 2.30.0), Bsgenome (version 1.38.0), Biostrings (version 2.38.4). Each of these is open-access. |

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

## Materials and reagents

Policy information about availability of materials

8. Materials availability

| | |
|---|---|
| Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a third party. | All vectors described in the manuscript have been deposited at Addgene for ease of distribution |

9. Antibodies

| | |
|---|---|
| Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species). | The specificity of the Cas9 antibody used was confirmed by including non-Cas9 expressing cells as a negative control (Supplementary Figure 5). The specificity of the Apc antibody used has previous been confirmed using shRNA-mediated knockdown of Apc in mouse tissue (Dow et al, Cell, 2015). Specificity of the Glutamine Synthetase antibody is demonstrated in Figure 3, where it marks pericentral hepatocytes. Specificity of the Alexa594 secondary antibody was confirmed by performing a 'secondary only' staining control - no staining was observed. |

10. Eukaryotic cell lines

| | |
|---|---|
| a. State the source of each eukaryotic cell line used. | 293T, NIH3T3, H23, and DLD1 cells were purchased from ATCC. PC9 cells were obtained from Harold Varmus' laboratory. KH2 mESCs were derived by the Jaenisch laboratory in 2006 (Beard et al, 2006) and have been maintained by us since then. Irradiated MEF feeders were derived from DR4 embryos. Primary organoids were generated directly from C57Bl/6 mice. |
| b. Describe the method of cell line authentication used. | All lines were purchased directly from ATCC and frozen at early passage (P2), thus they did not require additional authentication |
| c. Report whether the cell lines were tested for mycoplasma contamination. | Yes. All lines tested negative for mycoplasma by PCR |
| d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use. | No commonly misidentified cell lines were used |

## Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

| | |
|---|---|
| Provide all relevant details on animals and/or animal-derived materials used in the study. | For hydrodynamic liver transfections, 8 week old, female C57Bl/6N mice (Charles River) were used, and sacrificed 4 weeks following injection. |

Policy information about studies involving human research participants

12. Description of human research participants

| | |
|---|---|
| Describe the covariate-relevant population characteristics of the human research participants. | The study did not involve human subjects |





Nature Biotechnology: doi:10.1038/nbt.4194