Improved prime editors enable pathogenic allele correction and cancer modelling in adult mice

Pengpeng Liu[1¶], Shun-Qing Liang[2¶], Chunwei Zheng[2], Esther Mintzer[1], Yan G. Zhao[1], Karthikeyan Ponnienselvan[1], Aamir Mir[2], Erik J. Sontheimer[2,5,6], Guangping Gao[3], Terence R. Flotte[3,4], Scot A. Wolfe[1,6*], Wen Xue[2,5,6*]

[1]Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA, USA.

[2]RNA Therapeutics Institute, University of Massachusetts Medical School, Worcester, MA, USA;

[3]Horae Gene Therapy Center, University of Massachusetts Medical School, Worcester, Massachusetts, USA.

[4]Departments of Pediatrics and Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester, MA, USA.

[5]Department of Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA.

[6]Li Weibo Institute for Rare Diseases Research, University of Massachusetts Medical School, Worcester, Massachusetts, USA.

[¶]Equal contribution to this study

Running title: Prime editing in adult mice

Abbreviation: Cas9, clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9. PEs, prime editors (PEs) pegRNA, prime editing guide RNA. PBS, primer binding site. RT, reverse transcriptase

Keywords: Prime editing; Mouse liver; CRISPR; Cas9, Cancer modeling; Alpha-1 antitrypsin deficiency

Conflict of interest: The authors declare no potential conflicts of interest.

***Corresponding author:** Scot Wolfe (scot.wolfe@umassmed.edu) and Wen Xue (wen.xue@umassmed.edu)

Abstract

Prime editors (PEs) mediate genome modification without utilizing double-stranded DNA breaks or exogenous donor DNA as a template. PEs facilitate nucleotide substitutions or local insertions or deletions within the genome based on the template sequence encoded within the prime editing guide RNA (pegRNA). However, the efficacy of prime editing in adult mice has not been established. Here we report an NLS-optimized SpCas9-based prime editor that improves genome editing efficiency in both fluorescent reporter cells and at endogenous loci in cultured cell lines. Using this genome modification system, we could also seed tumor formation through somatic cell editing in the adult mouse. Finally, we successfully utilize dual adeno-associated virus (AAVs) for the delivery of a split-intein prime editor and demonstrate that this system enables the correction of a pathogenic mutation in the mouse liver. Our findings further establish the broad potential of this new genome editing technology for the directed installation of sequence modifications *in vivo*, with important implications for disease modeling and correction.

Introduction

Disease-associated genetic variations, including deletions, insertions and base substitutions, require precise gene correction strategies that are both robust and flexible[1]. Homology-direct repair (HDR) enables precise genome editing through an exogenous donor DNA. However, HDR is inefficient in most therapeutically relevant cell types, especially in post-mitotic cells[2-4]. Base editing enables efficient nucleotide transitions without inducing double-strand breaks (DSBs)[5]. However, targeted nucleotide transversions, deletions and insertions are not easily facilitated by well-established editing systems. In addition, depending on the local sequence context, base editing systems can also convert “bystander” nucleotides within the same editing window, which may be mutagenic, leading to the creation of unproductive or counter-productive alleles.

The prime editor (PE) is a new genome editing tool that can produce template-directed local sequences changes in the genome without the requirement for a DSB or exogenous donor DNA templates [6]. A PE comprises a fusion protein consisting of a catalytically impaired SpCas9 nickase (H840A) fused to an engineered reverse transcriptase (RT). A prime editing guide RNA (pegRNA) targets the PE to the desired genomic sequence and encodes the primer binding site (PBS) and reverse transcriptase template to enable the RT to copy the new genetic information into the target genomic locus [6]. Prime editing enables nucleotide conversion and targeted sequence insertions and deletions based on the RT template sequence encoded within the

69 pegRNA[6]. PE2 harbors five mutations within the M-MLV RT that improve editing efficiency[6].
70 PE3 uses an additional sgRNA to direct SpCas9^{H840A} to nick the non-edited DNA strand to
71 encourage the edited strand to be utilized as a repair template by DNA repair factors, leading to
72 further increases in editing efficiency[6].
73
74 The ability to precisely install or correct pathogenic mutations regardless of their composition
75 makes prime editing an intriguing approach to perform somatic genome editing in model
76 organisms to study disease processes or to utilize for therapeutic applications. Previous studies
77 have used prime editing to recode loci in cultured cells[6], plants[7], stem cells[8, 9], and mouse
78 zygotes[10, 11]. However, PE delivery in adult animals has not yet been described. In this study,
79 we report a nuclear localization signal sequence optimized PE2 (referred to as PE2*) with
80 higher editing efficiency than PE2. We demonstrate that PE2* enables somatic genome editing
81 in the liver of adult mice, where it can correct a pathogenic disease allele or introduce a directed
82 mutation to drive tumor formation to facilitate cancer modeling. The size of a prime editor
83 precludes its packaging in a single AAV vector. We show that dual AAV-mediated delivery of a
84 split-intein prime editor is functional *in vivo* for gene editing in the mouse liver. These data
85 demonstrate the feasibility of employing PE *in vivo* for the targeted, precise alteration of
86 genomic sequence with potential utility both in model organisms and as a therapeutic modality.
87
88 **Results**
89 **Optimized nuclear localization signal sequence composition improves prime editing**
90
91 PEs have the remarkable ability to introduce a variety of different types of sequence alterations
92 into the genome. Their editing efficiency is influenced by a variety of different parameters (PBS
93 length, position of the RT initiation site relative to the desired sequence alteration, composition
94 of the desired sequence alteration, relative position of the alternate strand nick, etc.). Even
95 under optimal conditions the incorporation rate of the desired edit into the genome is incomplete
96 [6]. Previously, we and others have noted that the composition and number of nuclear localization
97 signal (NLS) sequences within a Cas9 effector can influence its efficiency of genome editing[3, 12-
98 16]. The original PE2 contains two SV40 NLS sequences [6], which can be suboptimal for
99 maximizing editing rates by SpCas9 protein-sgRNA complexes[12]. In transient transfection
100 assays of original PE2, we observed incomplete nuclear localization based on
101 immunofluorescence: ~60% of the protein is present in the nucleus in U2OS cells and ~85% is
102 present in the nucleus in HeLa cells; (**Supplementary Fig. S1**). We found that addition of a N-

103 terminal c-Myc NLS[17] and C-terminal variant bipartite SV40 NLS (vBP-SV40)[18] gives rise to
104 nearly complete nuclear localization of the prime editor (PE2*) (**Fig. 1a, Supplementary Fig.**
105 **S1c,d**).
106
107 In parallel, we constructed PE2* frameworks using the orthogonal Staphylococcus aureus
108 (SaCas9) nickase (N580A)[19] to replace the SpCas9 nickase (H840A). To expand the potential
109 targeting range of these alternate prime editors systems, we utilized both the standard SaCas9
110 backbone that recognizes an NNGRRT PAM (SaPE2*)[19] and the SaCas9^{KKH} variant that
111 broadens the targeting to an NNNRRT PAM (Sa^{KKH}PE2*)[20] (**Fig. 1a**). By immunofluorescence
112 staining using the incorporated 3xHA-tag we observed that SaPE2* or Sa^{KKH}PE2* proteins were
113 localized to the nucleus (**Supplementary Fig. S1a,b**).
114
115 To determine whether the observed improvements in nuclear localization translate into
116 increases in editing efficiency, we evaluated the rate of nucleotide conversion for PE2 and PE2*
117 in HEK293T mCherry reporter cells that contains a premature TAG stop codon that prevents
118 translation of a functional protein (**Supplementary Fig. S2a,d**). PE2 and PE2* were
119 programmed with a pegRNA designed to revert the TAG codon to CAG and delivered without
120 and with different nicking sgRNAs. 3 days after transfection, we performed flow cytometry to
121 quantify prime-editing efficiency. PE2* produced a 1.5-1.6 fold increase in editing efficiency
122 (14.3% to 26.4%) relative to PE2 (9.2% to 16.5%; **Fig. 1b**). A compatible PAM is also present
123 for Sa^{KKH}PE2* within the reporter to evaluate nucleotide conversion rates, with which we
124 observed more modest editing efficiencies (1.8% to 4.7%). All PE systems displayed lower
125 editing activity than an adenine base editor system (ABEmax)[13] for restoration of reporter
126 function (**Fig. 1b**).
127
128 Next, we tested the efficiency for generating a targeted deletion using PE2, PE2* and
129 Sa^{KKH}PE2* in an HEK293T reporter line that can quantify both precise deletions and indel
130 formation. This reporter design shares similarities to the traffic light reporter system [21, 22]. A
131 precise deletion of 47bp will remove a sequence insertion disrupting GFP expression, whereas
132 indels that produce a particular reading frame alteration restore mCherry expression
133 (**Supplementary Fig. S2b,d**). PE2* produced a 1.6-1.9 fold increase in the level of precise
134 deletions (5.6%-11.3%) compared to PE2 (3.0%-7.3%; **Fig. 1c**). The relative level of undesired
135 indel formation was roughly proportional to the overall activity levels for PE2 and PE2*. We also

136 observed that Sa^{KKH}PE2* could generate precise 47bp deletion with efficiencies ranging from
137 1.3% to 4.2% (**Fig. 1c**).
138
139 Finally, we tested the efficiency for generating a targeted insertion using PE2, PE2* and
140 Sa^{KKH}PE2* in a different HEK293T reporter line (TLR-MCV1) that can quantify both precise
141 insertions and indel formation[23]. A targeted, precise replacement of 39bp disruption sequence
142 with a 18bp missing sequence element can restore GFP expression, whereas indels that
143 produce a different reading frame alteration restore mCherry expression (**Supplementary Fig.**
144 **S2c,d**). PE2* led to a 1.7 to 2.1-fold increase in the level of precise insertions (5.5%- 11.6%)
145 compared to PE2 (3.2%- 5.5%; **Fig. 1d**). Again, the relative level of undesired indel formation
146 was roughly proportional to the overall activity levels for PE2 and PE2*. We also observed that
147 Sa^{KKH}PE2* could generate 18bp replacement with efficiencies ranging from 1.3% to 4.2% (**Fig.**
148 **1d**). Across all of these reporter systems, we observed that nicking the nonedited strand (PE3
149 format) increased the editing efficiency by 1.5- to 2.4-fold and the indel rate by 0.2 % to 3.3 %
150 compared to pegRNA only in both PE2 and PE2* (**Fig. 1b-d**), consistent with previous
151 observations[6]. Together, these results demonstrate that PE2* performed nucleotide conversion,
152 sequence deletion or insertion more efficiently than PE2. In addition, the new SaCas9-based
153 PEs displayed appreciable genome editing activity.
154
155 **Improved prime editor increases editing efficiency at endogenous loci**

156 Next, we compared the editing efficiency of PE2 and PE2* for the creation of previously
157 described nucleotide substitutions, deletions and insertions at the *EMX1* locus[6]. HEK293T cells
158 were transfected with different prime editors, pegRNAs and different nicking sgRNAs. Genomic
159 DNA was isolated and editing outcomes at each target site were quantified by high-throughput
160 sequence (HTS). PE2* (3.1%- 6.5%) led to an average 1.9-fold increase in the rate of point
161 mutation introduction compared to PE2 (1.5%- 3.7%) (**Fig. 2a**). Targeted 3bp deletions were
162 generated at 1.4 to 2.1-fold higher rate by PE2* (2.1%- 6.0%) than PE2 (1.5%- 2.9%) (**Fig. 2a**).
163 Targeted 6bp insertions were generated at 1.7 to 2.4-fold higher rate by PE2* (2.2%-3.1%) than
164 PE2 (0.9%-1.8%) (**Fig. 2a**). As observed with the various reporter systems, the level of indel
165 formation was roughly proportional to the activity levels of PE2 and PE2*. Together, these
166 observations suggest that PE2* has broadly improved editing efficiency at endogenous loci.
167

168 We also compared the editing efficiency of SaPE2* and Sa^{KKH}PE2* for the creation of similar
169 nucleotide substitutions, deletions and insertions at the *EMX1* locus. Notably, SaPE2* installed
170 point mutations at two positions in the *EMX1* locus with editing efficiency from 4.7% to 9.3% and
171 a modest indel rate (0.0%-0.5%). Targeted 3-bp deletion and 6-bp insertion were introduced by
172 SaPE2* with an editing efficiency of 4.1% to 9.4% and 2.7% to 5.5%, respectively. Indel
173 induction generated by SaPE2* ranged from 0.0% to 0.6% (**Fig. 2b**). Overall, Sa^{KKH}PE2*
174 exhibited lower editing efficiency at the *EMX1* locus than SaPE2* with the same set of pegRNAs
175 (typically between 1 and 2%; **Fig. 2c**). Notably, at these loci the editing efficiencies for SaPE2*
176 were similar to the rates obtained with PE2*, suggesting that the SaPE2* platform has the
177 potential to broaden the scope of available prime editing systems.
178
179 A homozygous 32-bp deletion in the *CCR5* gene is associated with resistance to human
180 immunodeficiency virus (HIV-1) infection[24-26]. We evaluated the utility of PE2 and PE2* to
181 generate a large, therapeutically relevant 32-bp deletion within *CCR5* that recapitulates the HIV-
182 1 resistance allele (**Fig. 2d**). We utilized linear amplification to incorporate unique molecular
183 identifiers (UMI) prior to sequencing[27] to avoid PCR amplification bias for the assessment of the
184 deletion rate in the population of treated cells. Both PE2 and PE2* were able to generate the
185 desired 32-bp deletion within the *CCR5* locus in HeLa cells, where the PE2* editor displayed
186 higher deletion rates than PE2 (average 1.4-fold across all conditions) with a maximum
187 efficiency of about 6.0% (**Fig. 2e, Supplementary Fig. S3**). Sa^{KKH}PE2* exhibited lower editing
188 efficiency than PE2 for the generation of the 32-bp deletion, with a maximum efficiency of 2.9%
189 (**Fig. 2e, Supplementary Fig. S3**). Overall, these results demonstrate that PE can introduce a
190 therapeutically relevant deletion in the *CCR5* gene in human cells.
191
192 **Improved PE increases the correction efficiency of a pathogenic mutation *in vivo***
193
194 Alpha-1 antitrypsin deficiency (AATD) is an inherited disorder that is caused by mutations in the
195 Serpin Peptidase Inhibitor Family A member 1 (*SERPINA1*) gene[28]. The E342K mutation (via
196 G•C-to-A•T) in *SERPINA1* (PiZ allele) is the most frequent mutation and causes severe lung
197 and liver disease[28]. Patients with homozygous mutation in *SERPINA1* (PiZZ) have PiZ protein
198 aggregates in hepatocytes and lack of functional AAT protein in the lung. The PiZ transgenic
199 mouse contains 16 copies of the human *SERPINA1* PiZ allele and is a commonly-used mouse
200 model of human AATD [29]. For the correction of the E342K mutation, there are some challenges
201 for the utilization of an adenine base editor: no optimal NGG PAM is present nearby for SpCas9,

202 and in addition to the target adenine there are several other adenines that may also be
203 susceptible to base conversion ("bystander effect")[30]. Consequently, we investigated the
204 utilization of the PE platform to correct this pathogenic mutation. To test the efficiency of
205 different prime editors at this locus, we first evaluated the generation of the pathogenic E342K
206 mutation (via G•C-to-A•T conversion) in wildtype *SERPINA1* in HEK293T cells. A series of
207 different nicking sgRNAs were also evaluated in conjunction with the PEs (PE3 strategy[6]). We
208 observed 1.6 to 3.4-fold increase for G•C-to-A•T base transition in *SERPINA1* by PE2* (6.4%-
209 15.8%) compared to PE2 (1.9%-9.9%; **Fig. 3a**). The average rate of indel generation with a
210 nicking sgRNA slightly increased with PE2* (0.1% -3.8%) compared to PE2 (0.0%-2.2%).
211 Sa^{KKH}PE2* exhibited lower overall editing efficiency for the installation of the E342K mutation
212 (1.1%-4.4%). Indel generation at the target site by Sa^{KKH}PE2* ranged from 0.0% to 1.4% (**Fig.**
213 **3a**).
214
215 Next, we investigated the ability of prime editors to directly correct a pathogenic mutation *in vivo.*
216 Based on the editing results for *SERPINA1* in HEK293T cells, we chose the nicking sgRNA3 for
217 use with the PEs for the *in vivo* experiments. We utilized a different pegRNA designed to revert
218 the E342K mutation (**Fig. 3b)**. A PAM mutation (AGG to AAG) was also included to reduce re-
219 cutting of the locus that introduced a synonymous codon change (**Fig. 3b)**. We introduced PE2
220 or PE2*, pegRNA and nicking sgRNA into the liver of PiZ mice (n=3/group) through
221 hydrodynamic tail vein injection (**Fig. 3c**). Hydrodynamic injection can deliver plasmid DNA to
222 20-30% of hepatocytes [31]. Using a PE2 plasmid encoding an HA-tag, we observed that 19.98
223 ± 0.88% hepatocytes were HA-tag positive (n=4 mice) (**Supplementary Fig. 4a**). 45 days after
224 injection, livers of PiZ mice were collected and DNA was purified for HTS analysis. Notably, we
225 observed 3.1-fold increase for A•T-to-G•C correction in PiZ *SERPINA1* by PE2* (6.7% on
226 average) compared to PE2 (2.1%). The indel rate at the locus was also increased from 0.4% to
227 2.7% (**Fig. 3d**). Interestingly, we observed a low frequency of large deletions between pegRNA
228 and nicking sgRNA (**Supplementary Fig. S4b**). Together, these data demonstrate that prime
229 editors can restore the wild type *SERPINA1* allele thereby enabling pathogenic gene correction
230 in adult mice.
231
232 **Improved PE enhances tumor burden by somatic engineering in the mouse liver**
233
234 *CTNNB1* (β-catenin) is a commonly mutated gene in hepatocellular carcinoma[32].
235 Overexpression of a mutant *Ctnnb1* and Myc oncogene have been used to generate liver

236 cancer models[14]. To explore the potential of prime editors to drive tumor formation *in vivo*, we
237 delivered PE2 or PE2*, a pegRNA used for installation (via C•G-to-T•A) of the oncogenic S45F
238 mutation in *Ctnnb1* (**Fig. 4a**), and a nicking sgRNA to the livers of adult FVB mice (n=4/group)
239 by hydrodynamic tail vein injection (**Fig. 4b**). A MYC transposon and transposase were co-
240 injected to provide a second oncogenic driver necessary for tumor formation in conjunction with
241 the *Ctnnb1* mutation[33]. 25 days after injection, livers of adult mice were collected and tumor
242 nodules on the liver were quantified. PE2-treated animals showed an average 5.5 ± 1.1 tumors
243 per mouse, whereas PE2*-treated mice displayed higher rates of tumor formation, with an
244 average 10.0 ± 2.7 tumors on the liver (**Fig. 4c,d**). Consistent with gain of function of the S45F
245 mutation, liver tumors were positive for nuclear β-Catenin[14] (**Supplementary Fig.5**). Sanger
246 sequence of gDNA from the tumor nodules showed precise conversion of S45F in *Ctnnb1* (**Fig.**
247 **4e**).
248
249 Prime editors afford the opportunity to install other types of mutations within the genome. To
250 assess the feasibility of generating deletions *in vivo*, we designed a pegRNA to delete the S45
251 codon in *Ctnnb1*, which is a previously described oncogenic mutation at this locus [34](**Fig. 4f**).
252 The prime editor (PE2*), pegRNA for *Ctnnb1* S45 deletion and nicking sgRNA plasmids were
253 delivered by hydrodynamic tail-vein injection along with the MYC transposon and transposase
254 plasmids. pegRNA *Ctnnb1* S45 deletion-treated animals showed extensive tumor formation,
255 whereas pegRNA SERPINA1-treated animals did not induce any tumor formation (**Fig. 4g**).
256 Deep sequencing showed that more than 80% of tumor gDNA contained precise editing
257 removing the S45 codon (**Fig. 4h**). Together, these results demonstrate that prime editors can
258 be used for generating tumor models by somatic cell engineering *in vivo*, and that PE2*
259 provides a platform with improved editing activity.
260
261 **Utilizing a split-intein approach for prime editor delivery**
262
263 The ~6.3-kb coding sequence of PE exceeds the ~4.8-kb packaging size limit of AAV[35]. To
264 deliver prime editors with AAVs, we adapted a split Cas9 dual-AAV strategy[36] in which the
265 original PE2 prime editor is divided into an amino-terminal (PE2-N) and carboxy-terminal (PE2-
266 C) segments, which are then reconstituted to full length PE by a *trans*-splicing intein[37] (**Fig. 5a**).
267 To ensure that each prime editor segment is smaller than the AAV packaging size limit, we
268 divided the PE within the SpCas9 amino acid before Ser 714[38]. We generated AAV8 particles
269 encoding the split-intein PE, a nicking sgRNA and a pegRNA (PE3 strategy[6]) to correct the

270 E342K mutation in SERPINA1 (**Fig. 5a**). We then characterized the performance of the split-
271 intein AAV prime editor *in vivo*. PiZ mice were treated by tail-vein injection of a low dose dual
272 AAV8-PE (2×10^{11} viral genome total) (**Fig. 5b**). Livers were harvested at 2 weeks (n=2), 6
273 weeks (n=3) and 10 weeks (n=3) after injection. By targeted deep sequencing, we detected 0.6
274 ± 0.0 % precise editing at 2 weeks. The precise editing efficiency increased significantly to 2.3 ±
275 0.4 % at 6 weeks and 3.1 ± 0.6 % at 10 weeks (**Fig. 5c**). We observed corresponding increases
276 of indel rates at the target site by split-intein AAVs from 0.1 ± 0.0 % (2 weeks) to 0.4 ± 0.1 % (10
277 weeks). Utilization of the UDiTaS unidirectional sequencing approach[39] with locus specific
278 primers for library construction affords the opportunity to assess the rate of large deletions or
279 other types of genomic rearrangements in an unbiased manner. UDiTaS analysis at the
280 SERPINA1 transgene locus revealed primarily precise editing among the modified alleles. As
281 observed by the amplicon deep sequencing, a fraction of the modified alleles contained indels
282 and there were a small number of larger deletions between the two nicking sites or extending
283 beyond these sites, although many of the largest deletions (>100 bp) did not meet the level of
284 statistical significance (**Fig. 5d-e, Supplementary Fig. 6 and Supplementary table 5**). There
285 was also evidence of a very low rate of AAV insertion at the target site (0.014% of total UMIs).
286 Together, these results demonstrate that delivery of split-intein PE by the dual AAV8 enables
287 low rates of precise editing via the PE3 prime editing strategy *in vivo*.
288
289 **Discussion**
290 Prime editing systems potentially provide a powerful approach for the template-directed
291 incorporation of a variety of types of alterations (nucleotide changes, insertions, deletions) into
292 genomic DNA sequence without relying on homology directed repair [6]. In principle, this provides
293 a strategy for the correction of a variety of different disorders, since prime editing should not be
294 dependent on the cell cycle for efficacy as is HDR[2]. Moreover, unlike HDR and MMEJ[40]
295 approaches for precise sequence insertions, prime editing does not require co-delivery of donor
296 DNA. However, the length of the sequence that can be inserted is limited by the length of the
297 encoded pegRNA. While there are many potential advantages to prime editing, the development
298 of prime editing systems is in its initial stages, and many questions with regard to the utility of
299 this system for genome editing remain to be addressed. This study provides an important
300 demonstration of the *in vivo* utility of prime editors for the template-based modification of
301 genomic sequence with implications for improving the utility of disease model systems and for
302 the eventual translation of this tool to the correction of pathogenic disorders.
303

304 The efficacy of genome editing systems is dependent on a number of factors, one of which is
305 the efficiency of nuclear import. In rapidly proliferating cells, the nuclear envelope provides only
306 a modest barrier to entry for genome editing tools. However, in post-mitotic or quiescent cells
307 the nuclear envelope may provide a greater barrier to the entry of Cas9-based systems, such
308 that the number and composition of the NLS sequences can impact editing efficacy [3, 12-16]. By
309 incorporating additional NLSs, we developed an improved PE (PE2*) that increases the
310 efficiency of genome editing across multiple endogenous sites relative to the original PE2.
311 Importantly, the observed improvements in genome editing for PE2* in cell culture translated to
312 increased rates of genome editing *in vivo*. Thus, demonstrating that NLS sequence composition
313 and architecture is an important parameter to consider in the design of prime editing systems to
314 maximize *in vivo* efficacy, as has been observed for other genome editing systems [3, 36].
315
316 This study provides an important proof-of-concept demonstration of the feasibility of *in vivo*
317 editing of somatic cells in mammalian systems by prime editing. We demonstrated the utility of
318 prime editing systems for two different types of applications: the correction of a pathogenic
319 mutation (AATD) and generation of animal cancer model. The ability to precisely correct a
320 pathogenic mutation or install somatic mutations *in vivo* has important implications for both gene
321 therapy and the development of new mouse models to study cancer and other disorders[41].
322 Using hydrodynamic injection, PE2* shows higher editing efficiency (~6%, **Fig. 3**) than
323 previously published HDR (~0.5%)[42] in the mouse liver for installing oncogenic point mutations.
324 Future work will address whether PEs can simultaneously generate cooperating oncogenic point
325 mutations (a list of potential mutations that can be studied is in **Supplementary Table 4**).
326
327 Compared with HDR and base editors, prime editing provides a complementary method for
328 generating precise genome alterations. Although a PE is less efficient than base editors for
329 introducing base transitions, a PE can generate precise nucleotide substitutions in target sites
330 where the neighboring sequence composition could potentially be challenging for precise
331 conversion by base editing systems [6]. Both of our *in vivo* target sites have neighboring
332 nucleotides that would potentially be susceptible to conversion when targeted by a base editor
333 (bystander effects). In the case of the *SERPINA1* locus, these bystander changes would
334 introduce missense mutations that could have undesirable effects.
335
336 Prime editing systems can also produce undesired editing outcomes in some instances[11]. The
337 use of the PE2 system in cell culture systems produces primarily precise edits[6]. The rate of

338 precise edits can be increased through the use of an additional nicking sgRNA (PE3 strategy),
339 but this also results in the production of a low rate of indels within the genome[6, 11]. Prime editing
340 in plant protoplasts also produces a fraction of undesired editing outcomes when employing
341 either the PE2 and PE3 strategy[7]. Interestingly, in mouse zygotes the PE3 strategy produces
342 alleles containing the desired edit, but a large fraction also harbor deletions of various sizes
343 between the target site and the nicking site [11]. Reassuringly, for our *in vivo* PE3 editing the
344 majority of modified alleles contain the intended product without additional modifications. Similar
345 to the observation in cell culture experiments, we observe a small fraction of the alleles that
346 contain unintended changes (indels). A small proportion of these (~6.6% modified alleles)
347 contain deletions between the pegRNA and nicking sgRNA sites. Employing a sequential
348 nicking strategy (PE3b) approach instead of PE3 may reduce the indel rates *in vivo*, as
349 observed in other systems[6, 7], when an overlapping nicking RNA can be designed at the prime
350 editor target site.
351
352 We established a dual-AAV system to deliver split-intein prime editors that produces precise
353 editing *in vivo* following a single injection. While we focused on liver editing, the dual AAV-
354 mediated prime editor developed in this study should be applicable to other organ systems.
355 Because AATD requires a high level of gene correction to generate sufficient wildtype protein
356 for disease amelioration[28], our initial AAV PE design is suboptimal for AATD treatment. Future
357 work will be required to increase PE efficiency *in vivo*. The modest editing efficiency of our AAV
358 system may be due in part to the low vector dose ($2x10^{11}$ vg/kg total). In addition, some aspects
359 of our design of split-PE AAV vectors could be improved. We retained the original PE2
360 architecture in our split-intein design due to its more compact size for vector packaging –
361 substitution of the SV40 NLSs with more efficient bipartite SV40[3] or c-myc NLSs[17] may improve
362 *in vivo* activity. Optimizing the position of intein insertion within the PE may also improve activity,
363 as may alteration of the AAV vector layout and promoters that are utilized [36]. The more compact
364 SaPE2* or Sa^{KKH}PE2* systems described herein may also have utility in the context of AAV
365 delivery once they are further optimized. In addition, the development of nanoparticle-mediated
366 delivery or direct RNP delivery could also provide alternate avenues for translation of this tool
367 for *in vivo* therapeutic application without the constraints of AAV cargo capacity. Once the
368 efficiency of prime editing systems *in vivo* is improved, it will also be important to explore in
369 greater depth the frequency of off-target editing before these systems are transitioned to gene
370 therapy applications. In summary, our findings demonstrate the broad potential scope of PEs for

371 *in vivo* genome editing, both for disease gene correction and the development of new cancer
372 models in adult animals.
373
374 **Methods**
375 **Generation of plasmids**
376 To generate pegRNA expression plasmids, PCR products including spacer sequences, scaffold
377 sequences and 3' extension sequences were amplified with indicated primers (Supplementary
378 table 2) using Phusion master mix (ThermoFisher Scientifc) or Q5 High-Fidelity enzyme
379 (NewEnglandBioLabs), which were subsequently cloned into a custom vector (Supplementary
380 sequence 2, BfuAI and EcoR I digested) by the Gibson assembly method (NEB). To generate
381 nicking sgRNA expression plasmids, annealed oligos were cloned into BfuAI-digested vector or
382 pmd264 vector (Supplementary sequence 2). Sequences of all pegRNA and nicking sgRNAs
383 are listed in Supplementary Table 1. PE2* was generated through Gibson assembly, by
384 combining SpyCas9(H840A) and the M-MLV RT from PE2 [6] with additional NLS sequences by
385 PCR and inserting into a NotI/PmeI-digested pCMV-PE2 backbone (Supplementary sequence
386 3). SaPE2* and Sa^{KKH}PE2* were generated through Gibson assembly, by combining the
387 following three DNA fragments: (i) PCR amplified M-MLV RT with additional NLS sequences
388 from PE2, (ii) a NotI/PmeI-digested PE2 backbone, (iii) a SaCas9 N580A nickase or a Sa^{KKH}-
389 Cas9 nickase (supplementary sequence 3). All plasmids used for *in vitro* experiments were
390 purified using Midiprep kit including endotoxin removal step (Qiagen). pCMV-PE2 was a gift
391 from David Liu (Addgene plasmid # 132775) [6]. AAV-PE-N was generated through Gibson
392 assembly, by combining the following five DNA fragments: (i) gBlock pegRNA driven by U6, (ii)
393 gBlock nicking sgRNA driven by U6, (iii) PCR amplified N-terminal PE2 (amino acid 1-713 of
394 SpCas9 H840A), (iv) gBlock split-intein N terminal, (v) a KpnI/SacI-digested AAV backbone[43].
395 AAV-PE-C was generated through Gibson assembly, by combining the following four DNA
396 fragments: (i) gBlock split-intein C terminal, (ii) PCR amplified C-terminal PE2 (amino acid 714-
397 1368 of SpCas9 H840A) and M-MLV RT from PE2, (iii) gBlock β-globin poly(A) signal, (iv) a
398 KpnI/NotI-digested AAV backbone.
399
400 **AAV vector production**
401
402 AAV vectors (AAV8 capsids) were packaged at the Viral Vector Core of the Horae Gene
403 Therapy Center at the University of Massachusetts Medical School. Vector titers were
404 determined by gel electrophoresis followed by silver staining and qPCR.
405

406 **Generation reporter cells and Cell culture conditions**

407 HEK293T cells were purchased from ATCC, and cells were maintained in Dulbecco's Modified
408 Eagle's Medium supplemented with 10% FBS. For generation of the mCherry reporter and
409 47bp-insertion TLR reporter cells, we created single-copy reporter cells using the Invitrogen Flp-
410 In system. Briefly, Flp-In 293T cells were maintained in DMEM, 10% Fetal bovine serum (FBS),
411 1% pen-strep and 100 µg/ml Zeocin. $1x10^6$ Flp-In 293T cells were plated in a 6 well plate 24
412 hours before transfection. On the day of transfection, the cells were washed and fresh media
413 without Zeocin was added. The plasmid coding for FLP recombinase and the mCherry reporter
414 or 47bp-insertion TLR reporter plasmid were transfected into the cells at a 9:1 ratio using
415 polyfect (QIAGEN) according to the manufacturer's protocol (900ng mCherry reporter or 47bp-
416 insertion TLR reporter plasmid and 100ng FLP recombinase plasmid to make 1 µg plasmid in
417 total). 48 hours following transfection, the cells were washed and split into a 10cm dish with
418 fresh media. 100 µg/ml of hygromycin was used to select for cells that contained an integration
419 of the reporter plasmid. Two weeks post selection, hygromycin resistant foci were pooled and
420 propagated for cryopreservation and further experiments. The construction an characterization
421 of the TLR-MCV1 reporter cells was described in [23]. All cell types were maintained at 37°C and
422 5% CO_2 and were tested negative for mycoplasma.

423

424 **Cell culture transfection/electroporation and DNA preparation**

425 For transfection-based editing experiments in HEK293T cells or HEK293T reporter cells, cells
426 were plated 100,000 per well on a 48-well plate. 24 hours later, the cells were co-transfected
427 with 540ng of prime editor plasmid, 270ng of pegRNA plasmid and 90ng of Nicking sgRNA
428 plasmid. Lipofectamine 2000 (Invitrogen) was used for the transfection according to the
429 manufacturer's instructions. FACS analysis was performed 3 days after transfection in
430 HEK293T reporter cells. To detect editing efficiency in endogenous genomic loci, HEK293T
431 cells were cultured for 3 days after transfection, and genomic DNA was isolated using QIAamp
432 DNA mini kit (QIAGEN) according to the manufacturer's instructions. For *CCR5* PE editing, the
433 same amount of plasmid were delivered (540ng of prime editor plasmid, 270ng of pegRNA
434 plasmid and 90ng of Nicking sgRNA plasmid), where 2×10^5 HeLa cells were treated per
435 electroporation using the Neon® TransfectionSystem 10 L Kit (Thermo Fisher Scientific) with
436 the recommended electroporation parameters: Pulse voltage (1350 v), Pulse width (10 ms),
437 Pulse number (3).

438

439 **Fluorescent reporter assay**

440 48 h post-transfection cells are trypsinized and harvested into a microcentrifuge tube. Cells are
441 centrifuged at 500×g for 2 min, washed once with 1× PBS, recentrifuged at 500×g for 2 min and
442 resuspended in 1× PBS for flow cytometry (Becton Dickonson FACScan). 10,000 events were
443 counted from each sample for FACS analysis. Experiments were performed in six replicates on
444 different days. The data are reported as mean values with error bars indicating SD.
445
446 **Immunofluorescence and immunohistochemistry**
447 HeLa and U2OS cells are transfected in six-well format via Lipofectamine 2000 (Invitrogen)
448 using the manufacturer's suggested protocol with 300 ng each PE expression plasmid and 150
449 ng of each pegRNA expression plasmid on a cover slip. 48 h following transfection, transfection
450 media was removed, cells were washed with 1× PBS and fixed with 4% formaldehyde in 1×
451 PBS for 15 min at room temperature. Following blocking (blocking solution: 2% BSA, 0.3%
452 Triton X-100, within 1× PBS), samples were stained with mouse antihemagglutinin (Sigma,
453 H9658, 1:500), and Alexa 488 donkey anti-mouse IgG (H+L; Invitrogen, A-21202, 1:2000),
454 sequentially. VECTASHIELD mounting medium with DAPI (Vector Laboratories, H-1200) was
455 used to stain the nuclei and to mount the samples on the slide. Images were taken with ZEISS
456 LSM 710 Confocal Microscope System. We used Fiji[44] and the "CMCI-EMBL" plugin to calculate
457 the signal from the nuclear or cytoplasm comparments as previously described[45]
458 (https://github.com/miura/NucleusRimIntensityMeasurementsV2/). We used 60 to 80 cells for
459 each treatment group to determine the nuclear to cytoplasmic ratio for each PE construct, which
460 was calculated as the intensity value from nucleus divided by the intensity value from cytoplasm
461 and nucleus.
462 IHC staining was performed as previously described[42] using beta-catenin antibody (BD, 610154,
463 1:100) or HA tag (CST, 3724, 1:400).
464
465 **Animal studies**
466 All animal experiments were authorized by the Institutional Animal Care and Use Committee
467 (IACUC) at UMASS medical school. No animals were excluded from the analyses. All prime
468 editor plasmids were prepared by EndoFreeMaxi kit (Qiagen) and were delivered through
469 hydrodynamic tail-vein injection. For PiZ correction, eight-week-old PiZ mice were injected with
470 2.3ml 0.9% saline containing 30μg PE2 or PE2* (n=3), 15μg pegRNA (*SERPINA1*) and 5μg
471 Nicking sgRNA 3. For cancer model generation, FVB/NJ (Strain #001800) was purchased from
472 Jackson Laboratories. Each FVB mouse (n=4) was injected with 2.3ml 0.9% saline containing
473 30μg PE2 or PE2*, 15μg pegRNA (*Ctnnb1*), 15μg Nicking sgRNA 2 (*Ctnnb1*), 5μg pT3 EF1a-

474 MYC (a gift from Xin Chen, Addgene plasmid # 92046)[33] and 1μg CMV-SB10 (a gift from Perry
475 Hackett, Addgene plasmid # 24551).
476
477 **Deep sequencing and data analysis**
478 Library construction for deep sequencing is modified from our previous report[27]. Briefly, 72 h
479 after transfection or electroporation, cells were harvested and genomic DNA was extracted with
480 GenElute Mammalian Genomic DNA Miniprep Kit (Sigma). Genomic loci spanning the target
481 and off-target sites were PCR amplified with locus-specific primers carrying tails complementary
482 to the Truseq adapters (Supplementary Data 3). 50 ng input genomic DNA was PCR amplified
483 with Q5 High-Fidelity DNA Polymerase (New England Biolabs): (98 °C, 15 s; 67 °C 25 s; 72 °C
484 20 s) ×30 cycles. For the construction of the CCR5 UMI-based library, 50 ng input genomic
485 DNA was first linearly pre-amplified with 10 nM final concentration 5p-CCR5_UMI primer using
486 the Q5 High-Fidelity DNA Polymerase (New England Biolabs): (98 °C, 60 s; 67 °C, 25 s; 72 °C,
487 20 s) × 10 cycles. In the same reaction mix, 500 nM final concentration 5p-DS_constant and 3p-
488 CCR5_DS primers were added for further amplification (98 °C, 60 s; 67 °C, 25 s; 72 °C, 20 s)
489 for 30 cycles. Next, 0.1 μl of each PCR reaction was amplified with index-containing primers to
490 reconstitute the TruSeq adaptors using the Q5 High-Fidelity DNA Polymerase (New England
491 Biolabs): (98 °C, 15 s; 67 °C, 25 s; 72 °C, 20 s) x10 cycles. Equal amounts of the PCR products
492 from each experimental condition (identified by different indices) were pooled and gel purified.
493 The purified library was deep sequenced using a paired-end 150 bp Illumina MiniSeq run.
494
495 MiniSeq data analysis was done as previously reported [27]. First, the quality of paired-end
496 sequencing reads (R1 and R2 fastq files) was assessed using FastQC
497 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw paired-end reads were
498 combined using paired end read merger (PEAR) (PMID: 24142950) to generate single merged
499 high-quality full-length reads. Reads were then filtered by quality (using Filter FASTQC (PMID:
500 20562416)) to remove those with a mean PHRED quality score under 30 and a minimum per
501 base score under 24. Each group of reads was then aligned to a corresponding reference
502 sequence using BWA (version 0.7.5) and SAMtools (version 0.1.19). To determine indel
503 frequency, size and distribution, all edited reads from each experimental replicate were
504 combined and aligned, as described above. Indel types and frequencies were then cataloged in
505 a text output format at each base using bam-readcount (https://github.com/genome/bam-
506 readcount). For each treatment group, the average background indel frequencies (based on
507 indel type, position and frequency) of the triplicate negative control group were subtracted to

508 obtain the precise editing and indel frequencies for each group. The fraction of precise editing is
509 calculated as sequencing reads with the desired allele editing/ all reads for the target locus. The
510 results were concatenated and loaded into GraphPad Prism 8.4 for data visualization.

511

512 **Tn5 tagmentation and library preparation for UDiTaS**

513

514 For tagmentation, transposome was assembled as previously described[39] using purified Tn5
515 protein and oligonucleotides purchased from IDT. 200ng of genomic DNA was incubated with
516 2ul of assembled transposome at 55 degree for 7 mins, and the product was cleaned up (20ul)
517 with a Zymo column (Zymo Research, #D4013). Tagmented DNA was used for the 1st PCR
518 using PlatinumTM SuperFi DNA polymerase (Thermo) with i5 primer and gene specific primers
519 (Supplementary table 3). Two different libraries were prepared for gDNA from each mouse with
520 different combinations of primers (i5+Locus_F [UDiTaS], i5+Locus_R [UDiTaS]). The i7 index
521 was added in the 2nd PCR and the PCR product was cleaned up with Ampure XP SPRI beads
522 (Agencourt, 0.9X reaction volume). Completed libraries were quantified by Tapestation and
523 Qubit (Agilent), pooled with equal mole and sequenced with 150 bp paired-end reads on an
524 Illumina MiniSeq instrument.

525

526 **UDiTaS data analysis**

527 The analysis pipeline was built using python code. Briefly, the analysis steps are as follows:

528 i) Demultiplexing. Raw BCL files were converted and demultiplexed using the appropriate
529 sequencing barcodes, allowing up to one mismatch in each barcode. Unique molecular
530 identifiers (UMIs) for each read were extracted for further downstream analysis.

531 ii) Trimming. Remove 3′ adapters using cutadapt, version 3.0
532 http://journal.embnet.org/index.php/embnetjournal/article/view/200/479

533 iii) Create reference sequence based the UDiTaS locus-specific primer position and AAV
534 plasmid map separately. Build index files for the reference using bowtie2-index[46], version
535 2.4.0.

536 iv) Alignment analysis. Paired reads were then globally aligned (end-to-end mode) to all the
537 reference amplicons using bowtie2's very sensitive parameter. Finally, samtools[47] (version
538 0.1.19) was used to create and index sorted bam files. Paired-end reads covering a window
539 between pegRNA targeting site and nicking sgRNA targeting sites were extracted and the
540 total number of unique UMIs were counted. Precise editing or small indels were analyzed as

541 previously described[27]. Pindel[48] (version 0.2.5b8) was used to detect breakpoints of large
542 deletions. Raw sequencing reads that align to the reference sequence were collapse to a
543 single read by common UMI and categorized as an exemplar for each UMI to a specific
544 category—for example, Wild Type, precise editing, small indel/substitution and Large
545 Deletions. Then the number of UMIs assigned per category is determined to define the ratio
546 of each event.

547 v) AAV integration. Extract the unmapped reads that did not locally align to the AAV/plasmid in
548 steps 3 and 4 using bedtools bamtofastq. With bowtie2, index the AAV plasmid sequence
549 and then do a local alignment of the reads. Of the reads that locally align to the AAV plasmid,
550 first filter out those reads which are directly adjacent to the UDiTaS primer (on read 2) and
551 do not contain any target locus sequence. This removes reads that are due to false priming.
552 Of the remaining reads, collapse these by UMI and count the UMIs. Classify the exemplar
553 read for each UMI as 'AAV/Plasmid Integrations'.

554

555 The software used for the above analysis is available at:
556 https://github.com/locusliu/GUIDESeq-Preprocess_from_Demultiplexing_to_Analysis
557 https://github.com/editasmedicine/uditas. https://github.com/ericdanner/REPlacE_Analysis.
558 and https://github.com/locusliu/PCR_Amplicon_target_deep_seq/blob/master/CRESA-lpp.py.

559 **Statistical analysis**

560 The fold changes of editing (precise editing or indels) are calculated between the corresponding
561 groups: pegRNA_only between PE2 and PE2*, or with specific Nicking sgRNA between PE2
562 and PE2*. All raw data is listed in Supplementary Table 6 and statistical analyses were
563 performed using GraphPad Prism 8.4. Sample size was not pre-determined by statistical
564 methods, but rather, based on preliminary data. Group allocation was performed randomly. In all
565 studies, data represent biological replicates (n) and are depicted as mean ± s.d. as indicated in
566 the figure legends. Comparison of mean values was conducted with one-way ANOVA with
567 Tukey's multiple comparisons test, as indicated in the figure legends. In all analyses, P values <
568 0.05 were considered statistically significant.

569

570 **Data availability**

571 Illumina Sequencing data have been submitted to the Sequence Read Archive. These datasets
572 are available under BioProject Accession number PRJNA496245.The authors declare that all
573 other data supporting the findings of this study are available within the paper and its

Supplementary Information files or upon reasonable request. Backbone plasmids used for pegRNA and sgRNA cloning are available from addgene.

Acknowledgements

We thank C. Mello and P. Zamore for helpful discussions. We thank Y. Liu, J. Xie, and Q. Su in the UMass Morphology and Viral Vector Cores for support. W.X. was supported by grants from the National Institutes of Health (DP2HL137167, P01HL131471 and UG3HL147367), American Cancer Society (129056-RSG-16-093), the Lung Cancer Research Foundation, and the Cystic Fibrosis Foundation. P.Z., E.M., K.P. and S.A.W. were supported in part by the National Institutes of Health (R01GM115911, and UG3TR002668) and the Rett Syndrome Research Trust. A.M. and E.J.S. were supported by UG3TR002668.

Author contributions

PL and SQL performed the experiments, analyzed data and wrote the manuscript. CZ, EM, YGZ, KP, and AM performed experiments. GG, EJS, TRF, SAW, and WX supervised the study and wrote the manuscript with all co-authors.

Competing interests

The authors declare that they have no competing interests.

Supplementary data accompanies this paper

596 **Figure Legends:**

597 **Fig. 1 Improved NLS composition enhances prime editing efficiency**. (a) Schematic
598 representation of the original primer editor (PE2) and optimized prime editor (PE2*) carrying
599 additional NLS sequences at the N-terminus and C-terminus. The M-MLV reverse transcriptase
600 in PE2 was fused to a SaCas9 nickase (SaPE2*) or a SaCas9^{KKH} nickase (Sa^{KKH}PE2*) with the
601 same NLS composition to develop orthogonal prime editors. vBP-SV40 NLS = variant BP-SV40
602 NLS. (b) Diagram of the A•T-to-G•C transition required to convert a stop codon to GLN to
603 restore function to an mCherry reporter in HEK293T cells (top). Frequencies of targeted A•T-to-
604 G•C transition by different prime editors (PE2, PE2* and Sa^{KKH}PE2*) were quantified by flow
605 cytometry (bottom). (c) Diagram of the deletion reporter in HEK293T cells containing a broken
606 GFP with 47-bp insertion, P2A, and out-of-frame mCherry (top). A targeted, precise deletion of
607 47bp will restore GFP expression, whereas indels that create a particular reading frame
608 alteration produce mCherry expression. Frequencies of precise deletion (GFP+) and indel
609 (mCherry+) introduced by different prime editors (PE2, PE2*, and Sa^{KKH}PE2*) were quantified
610 by flow cytometry (bottom). (d) Diagram of the insertion reporter in HEK293 cells containing a
611 broken GFP with 39-bp insertion, T2A and mCherry (top). A targeted, precise insertion of 18bp
612 that substitutes for a disrupting sequence can restore GFP expression, whereas indels that
613 create a particular reading frame alteration produce mCherry expression. Frequencies of
614 targeted 18-bp replacement and indel generation by different prime editors (PE2, PE2*, and
615 Sa^{KKH}PE2*) were quantified by flow cytometry (bottom). All expression vectors were delivered
616 by transient transfection. The presence of sgRNAs to promote nicking of the complementary
617 strand is indicated in each figure legend. Results were obtained from six independent
618 experiments and presented as mean ± SD. $^{**}P<0.01$, $^{***}P<0.001$, $^{****}P<0.0001$ by one-way
619 ANOVA with Tukey's multiple comparisons test between each PE2 and PE2* using the same
620 nicking sgRNA.

621

622 **Fig. 2 Improved PE2* increases editing efficiency at endogenous loci**. (a) Comparison of
623 editing efficiency for nucleotide substitution, targeted 3-bp deletion, and 6-bp insertion with PE2
624 and PE2* at *EMX1* locus in HEK293T cells. Indels broadly indicates mutations to the
625 endogenous sequence that do not result in the desired sequence alteration. (b-c) Editing
626 efficiency for nucleotide substitution, targeted 3-bp deletion, and 6-bp insertion with SaPE2* (b)
627 and Sa^{KKH}PE2* (c) at *EMX1* locus in HEK293T cells. (d) Sequence of CCR5 locus and pegRNA
628 used for the 32bp deletion. Two mutations in red were included to demonstrate that sequence
629 collapse was not a function of nuclease-induced microhomology mediated deletion and to

630 reduce re-cutting of deletion allele. Bottom panel shows the alignment of pegRNA with the
631 *CCR5* sense strand. (e) Comparison of efficiency for generating a targeted 32-bp deletion with
632 PE2, PE2*, and Sa^{KKH}PE2* within *CCR5* in HeLa cells. All expression vectors were delivered by
633 transient transfection. The presence of sgRNAs to promote nicking of the complementary strand
634 is indicated in each figure panel. Results were obtained from three independent experiments
635 and presented as mean ± SD. **P*<0.05, ***P*<0.01, ****P*<0.001, *****P*<0.0001 by one-way
636 ANOVA with Tukey's multiple comparisons test. ns, not significant.
637
638 **Fig. 3 Improved PE2* increases the correction efficiency of a pathogenic mutation *in vivo*.**
639 (a) Installation (via G•C-to-A•T) of the pathogenic *SERPINA1* E342K mutation in HEK293T cells
640 using PE2, PE2*, and Sa^{KKH}PE2*. Editing efficiencies reflect sequencing reads which contain
641 the desired edit. The presence of sgRNAs to promote nicking of the complementary strand is
642 indicated on the x-axis. Results were obtained from three independent experiments and are
643 presented as mean ± SD. (b) pegRNA used for correction (via A•T-to-G•C) of the E342K
644 mutation includes a spacer sequence, a sgRNA scaffold, a RT template including edited bases
645 (red) and a primer-binding site (PBS). A PAM mutation (AGG to AAG) was introduced to reduce
646 re-cutting of the locus that results in a synonymous codon change. (c) Schematic overview of
647 correction strategy of the *SERPINA1* E342K mutation in PiZ transgenic mouse model of AATD.
648 Prime editor, pegRNA and nicking sgRNA plasmid were delivered by hydrodynamic tail-vein
649 injection. (d) Comparison of the efficiency of K342E correction and indels in mouse livers in PE2
650 or PE2* treatment groups. Precise editing is defined as the fraction of sequencing reads with
651 both A to G prime editing and synonymous PAM modification. Results were obtained from three
652 mice and presented as mean ± SD. ***P*<0.01, ****P*<0.001, *****P*<0.0001 by one-way ANOVA
653 with Tukey's multiple comparisons test.
654
655 **Fig. 4 Generating mouse cancer models using improved PE2***. (a) pegRNA used for
656 installation (via C•G-to-T•A) of the oncogenic S45F in *Ctnnb1* in mouse liver. (b) Schematic
657 overview of the somatic cell editing strategy to drive tumor formation. Prime editor (PE2 or
658 PE2*), pegRNA for *Ctnnb1* S45F and nicking sgRNA plasmids were delivered by hydrodynamic
659 tail-vein injection along with the MYC transposon and transposase plasmids. (c) Representative
660 images of tumor burden in mouse liver with PE2 or PE2*. (d) Tumor numbers in the livers of
661 mice 25 days after injection with PE2 or PE2*. Control group was pegRNA only. (e) Sanger
662 sequencing from normal liver and representative tumors. The dashed box denotes C to T editing
663 in tumors. **P*<0.05 by one-way ANOVA with Tukey's multiple comparisons test. (f) Schematic

664 of *Ctnnb1* S45 deletion strategy using PE2* (S45del). pegRNA used for 3bp deletion (TCC) is
665 shown. (g) PE2* treatment leads to oncogenic activation of *Ctnnb1*. Prime editor (PE2*),
666 pegRNA (*Ctnnb1* S45del or SERPINA1) and nicking sgRNA plasmids were delivered by
667 hydrodynamic tail-vein injection along with the MYC transposon and transposase plasmids.
668 Mice treated with the peg*Ctnnb1* S45del (n=4) displayed a large number of liver tumors
669 whereas mice treated with pegSERPINA1 as a control displayed no noticeable oncogenic
670 lesions. beta-Catenin (CTNNB1) IHC staining was performed. Scale bars: 100 μm (20X lens). (h)
671 Prime editing efficiency and indels determined by targeted deep sequencing in control liver and
672 representative tumors.

673 **Fig. 5 Systemic injection of dual AAV8 split-intein prime editor achieves pathogenic**
674 **mutation correction in PiZ mice. (a)** Schematic of split-intein dual AAV prime editor. Full-
675 length primer editor (original PE2[6]) was reconstituted from two PE2 fragments employing the
676 Npu DNAE split intein[37]. C, carboxy terminal; N, amino terminal. (b) Schematic of the *in vivo*
677 dual AAV8 prime editor injection experiments. Dual AAV8 split-intein prime editor (2×10^{11} vg
678 total) was delivered to six-week-old PiZ mice by tail-vein injection. Livers were harvested at 2
679 (n=2), 4 (n=3) and 10 (n=3) weeks after injection and the genomic DNA was isolated for
680 sequencing. (c) Prime editing efficiency of K342E correction and indels determined by targeted
681 deep sequencing in mouse livers of dual AAV-treated mice. Precise editing is defined as the
682 fraction of sequencing reads with both A to G prime editing and synonymous PAM modification.
683 Results were obtained from two (2 weeks) or three mice (6 and 10 weeks) and presented as
684 mean ± SD. $^{**}P<0.01$, $^{***}P<0.001$ by one-way ANOVA with Tukey's multiple comparisons test.
685 (d) Composition of edited alleles at *SERPINA1* by UDiTaS analysis. Circle plot shows the
686 fraction of edits that are precise (intended base conversion), small indels (<50bp) or substitution,
687 deletions between pegRNA and nicking sgRNA sites (<100bp), large deletions (>100bp), and
688 AAV fragment insertion. Numbers are average of 3 mice in 10 week treated cohort. (e) The
689 statistically significant large deletion sequences detected by UDiTaS in the 10 week treated
690 cohort are displayed as bars spanning the sequence that is deleted (a representative liver of
691 n=3 mice). Positions of the pegRNA and nicking sgRNA are indicated by dotted lines and the
692 approximate positions of the locus-specific UDiTaS primers are indicated by arrows below the
693 bar chart. The deletion size and number of UMIs associated with each deletion are indicated to
694 the right of each bar.

695

696 **Supplementary Fig. 1 Optimized prime editor increases the degree of nuclear import.** (a)
697 Representative immunofluorescence images of prime editors transfected into U2OS cells
698 immunostained with a HA tag antibody to visualize subcellular localization. PE2 contains two
699 SV40 NLSs, whereas the PE2* variant include an extra N-terminal c-Myc NLS and a C-terminal
700 variant bipartite SV40 NLS (vBPSV40). DNA was stained with DAPI. Scale bar, 100μm. (b) The
701 nuclear/cytoplasm ratio of different prime editors as determined by confocal microscopy. (c-d)
702 Representative immunofluorescence images and quantification in HeLa cells. ***P<0.001 by
703 one-way ANOVA with Tukey's multiple comparisons test.
704
705 **Supplementary Fig. 2 Prime editing in reporter cells by PE2 and PE2*.** (a) Sequence of
706 reporter locus and pegRNA used for repair (via A•T-to-G•C transition) of mCherry reporter in
707 HEK293T cells. Bar above the cDNA indicates the stop codon with the target "t" for conversion
708 indicated in red. Two additional silent mutations are included to discourage recutting of the
709 repaired DNA sequence (b) Sequence of reporter and pegRNA used for generation of a 47bp
710 deletion to restore function of the GFP reporter in HEK293T cells. The bars above the cDNA
711 indicates three nucleotide blocks that correspond to codons in the GFP reporter. (c) Sequence
712 of reporter and pegRNA used for replacement of a 18bp element to restore function of the GFP
713 reporter. The bars above the cDNA indicates three nucleotide blocks that correspond to codons
714 in the GFP reporter. (d) Representative images of HEK293T reporter cells transfected with
715 control, ABEmax or PE2*. Scale bar, 400μm.
716
717 **Supplementary Fig.3 Sequencing analysis of the *CCR5* prime editing by PE2, PE2*, or**
718 **Sa^{KKH}PE2*.**
719 The percentage of most common sequences in PE2, PE2* and Sa^{KKH}PE2* transfected cells is
720 shown on the right (representative of n=3, determined by Illumina sequencing). The PE target
721 site is underlined. The PAM sequences are indicated in light blue for SpCas9 and green for
722 SaCas9^{KKH}. The box denotes the 32bp sequence deleted in CCR5delta32. Two "t" mutations in
723 red were included in the PE2 and PE2* RT template (Fig. 2d) to demonstrate that sequence
724 collapse was not a function of nuclease-induced microhomology mediated deletion and to
725 reduce re-cutting of deletion allele. Deleted bases are indicated by dashes and inserted bases
726 are in blue.
727
728 **Supplementary Fig.4 Sequencing analysis of *SERPINA1* editing in the liver of PiZ mouse.**

729 (a) Evaluating prime editor expression in mouse liver. Left panel: FVB mice were injected with
730 30μg of control vector or PE2 plasmid with HA-tag. Livers were harvested at day 2 and IHC
731 staining were performed with an HA-tag antibody. Representative IHC images are shown. Scale
732 bars: 100 μm (20X lens). Right panel: Quantification of HA^+ cells. Numbers are mean ± sem
733 (n=4 mice).
734 (b) The percentage of most common sequences in the liver of PE2 and PE2*-treated mice is
735 shown on the right (representative liver of n=3, determined by Illumina sequencing). The PE
736 target site is underlined. The PAM sequences are in light blue. Nucleotide substitutions are
737 labeled in red. Deleted bases are indicated by dashes. Inserted bases are shown in blue/lower
738 case.
739
740 **Supplementary Fig.5 Liver tumors are positive for nuclear beta-Catenin**. (a),
741 Representative H&E and beta-catenin IHC staining in PE2 or PE2*-induced S45F tumors in Fig.
742 4. Scale bars: 100 μm (20X lens).
743
744 **Supplementary Fig. 6 Sequencing analysis of the *Ctnnb1* and *SERPINA1* prime editing by**
745 **PE2 or PE2*.**
746 (a), The percentage of most common sequences in the liver of dual AAV-treated PiZ mice
747 determined by locus amplification is shown on the right (representative liver of n=3). A portion of
748 the PE target site is underlined. The PAM sequences are in light blue. Nucleotide substitutions
749 are labeled in red. Deleted bases are indicated by dashes. (b), Length distribution of precise
750 editing and other indels at *SERPINA1 (K342E correction)* and *Ctnnb1* target sites by UDiTaS.
751 Also included are sequence modifications that may be associated with pegRNA scaffold
752 insertions[7]. A representative liver is shown (n=3 mice).
753
754 **References:**

755 1. Porteus, M.H. A New Class of Medicines through DNA Editing. *N Engl J Med* **380**, 947-
756 959 (2019).
757 2. Panier, S. & Boulton, S.J. Double-strand break repair: 53BP1 comes into focus. *Nat Rev*
758 *Mol Cell Biol* **15**, 7-18 (2014).
759 3. Suzuki, K. et al. In vivo genome editing via CRISPR/Cas9 mediated homology-
760 independent targeted integration. *Nature* **540**, 144-149 (2016).
761 4. Pickar-Oliver, A. & Gersbach, C.A. The next generation of CRISPR-Cas technologies
762 and applications. *Nat Rev Mol Cell Biol* **20**, 490-507 (2019).
763 5. Rees, H.A. & Liu, D.R. Base editing: precision chemistry on the genome and
764 transcriptome of living cells. *Nature Reviews Genetics* **19**, 770-788 (2018).

765 6. Anzalone, A.V. et al. Search-and-replace genome editing without double-strand breaks or
766 donor DNA. *Nature* **576**, 149–157 (2019).
767 7. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat Biotechnol* **38**, 582-585
768 (2020).
769 8. Schene, I.F. et al. Prime editing for functional repair in patient-derived disease models.
770 *Nature Communications* **11**, 5352 (2020).
771 9. Geurts, M.H. et al. Evaluating CRISPR-based Prime Editing for cancer modeling and
772 CFTR repair in intestinal organoids. *bioRxiv*, 2020.2010.2005.325837 (2020).
773 10. Liu, Y. et al. Efficient generation of mouse models with the prime editing system. *Cell*
774 *Discov* **6**, 27 (2020).
775 11. Aida, T. et al. Prime editing primarily induces undesired outcomes in mice. *bioRxiv*,
776 2020.2008.2006.239723 (2020).
777 12. Wu, Y. et al. Highly efficient therapeutic gene editing of human hematopoietic stem
778 cells. *Nat Med* **25**, 776-783 (2019).
779 13. Koblan, L.W. et al. Improving cytidine and adenine base editors by expression
780 optimization and ancestral reconstruction. *Nat Biotechnol* (2018).
781 14. Zafra, M.P. et al. Optimized base editors enable efficient editing in cells, organoids and
782 mice. *Nat Biotechnol* (2018).
783 15. Song, C.-Q. et al. Adenine base editing in an adult mouse model of tyrosinaemia. *Nature*
784 *Biomedical Engineering* **4**, 125-130 (2020).
785 16. Walton, R.T., Christie, K.A., Whittaker, M.N. & Kleinstiver, B.P. Unconstrained genome
786 targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* (2020).
787 17. Ray, M., Tang, R., Jiang, Z. & Rotello, V.M. Quantitative tracking of protein trafficking
788 to the nucleus using cytosolic protein delivery by nanoparticle-stabilized nanocapsules.
789 *Bioconjugate chemistry* **26**, 1004-1007 (2015).
790 18. Makkerh, J.P., Dingwall, C. & Laskey, R.A. Comparative mutagenesis of nuclear
791 localization signals reveals the importance of neutral and acidic amino acids. *Curr Biol* **6**,
792 1025-1027 (1996).
793 19. Ran, F.A. et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**,
794 186-191 (2015).
795 20. Kleinstiver, B.P. et al. Broadening the targeting range of Staphylococcus aureus
796 CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* **33**, 1293-1298 (2015).
797 21. Certo, M.T. et al. Tracking genome engineering outcome at individual DNA breakpoints.
798 *Nature methods* **8**, 671-676 (2011).
799 22. Mir, A. et al. Heavily and fully modified RNAs guide efficient SpyCas9-mediated
800 genome editing. *Nat Commun* **9**, 2641 (2018).
801 23. Iyer, S. et al. Efficient Homology-directed Repair with Circular ssDNA Donors. *bioRxiv*,
802 864199 (2019).
803 24. Dean, M. et al. Genetic restriction of HIV-1 infection and progression to AIDS by a
804 deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study,
805 Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco
806 City Cohort, ALIVE Study. *Science* **273**, 1856-1862 (1996).
807 25. Liu, R. et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some
808 multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367-377 (1996).
809 26. Samson, M. et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant
810 alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722-725 (1996).

811 27. Bolukbasi, M.F. et al. Orthogonal Cas9-Cas9 chimeras provide a versatile platform for
812 genome editing. *Nat Commun* **9**, 4856 (2018).
813 28. Loring, H.S. & Flotte, T.R. Current status of gene therapy for alpha-1 antitrypsin
814 deficiency. *Expert Opin Biol Ther* **15**, 329-336 (2015).
815 29. Carlson, J.A. et al. Accumulation of PiZ alpha 1-antitrypsin causes liver damage in
816 transgenic mice. *J Clin Invest* **83**, 1183-1190 (1989).
817 30. Gaudelli, N.M. et al. Programmable base editing of A*T to G*C in genomic DNA
818 without DNA cleavage. *Nature* **551**, 464-471 (2017).
819 31. Liu, F., Song, Y. & Liu, D. Hydrodynamics-based transfection in animals by systemic
820 administration of plasmid DNA. *Gene Ther* **6**, 1258-1266 (1999).
821 32. Zucman-Rossi, J., Villanueva, A., Nault, J.C. & Llovet, J.M. Genetic landscape and
822 biomarkers of hepatocellular carcinoma. *Gastroenterology* **149**, 1226-1239 (2015).
823 33. Liu, P. et al. A functional mammalian target of rapamycin complex 1 signaling is
824 indispensable for c-Myc-driven hepatocarcinogenesis. *Hepatology* **66**, 167-181 (2017).
825 34. Marquardt, J.U., Andersen, J.B. & Thorgeirsson, S.S. Functional and genetic
826 deconstruction of the cellular origin in liver cancer. *Nat Rev Cancer* **15**, 653-667 (2015).
827 35. Wang, D.A.-O., Tai, P.A.-O. & Gao, G.A.-O. Adeno-associated virus vector as a
828 platform for gene therapy delivery.
829 36. Levy, J.M. et al. Cytosine and adenine base editing of the brain, liver, retina, heart and
830 skeletal muscle of mice via adeno-associated viruses. *Nat Biomed Eng* **4**, 97-110 (2020).
831 37. Zettler, J., Schütz, V. & Mootz, H.D. The naturally split Npu DnaE intein exhibits an
832 extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett* **583**, 909-914
833 (2009).
834 38. Wright, A.V. et al. Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci*
835 *U S A* **112**, 2984-2989 (2015).
836 39. Giannoukos, G. et al. UDiTaS™ , a genome editing detection method for indels and
837 genome rearrangements. *BMC genomics* **19**, 212 (2018).
838 40. Yao, X. et al. CRISPR/Cas9 - Mediated Precise Targeted Integration In Vivo Using a
839 Double Cut Donor with Short Homology Arms. *EBioMedicine* **20**, 19-26 (2017).
840 41. Maddalo, D. et al. In vivo engineering of oncogenic chromosomal rearrangements with
841 the CRISPR/Cas9 system. *Nature* **516** (2014).
842 42. Xue, W. et al. CRISPR-mediated direct mutation of cancer genes in the mouse liver.
843 *Nature* **514**, 380-385 (2014).
844 43. Yin, H. et al. Therapeutic genome editing by combined viral and non-viral delivery of
845 CRISPR system components in vivo. *Nat Biotechnol* **34**, 328-333 (2016).
846 44. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat*
847 *Methods* **9**, 676-682 (2012).
848 45. Miura, K. Measurements of Intensity Dynamics at the Periphery of the Nucleus. In:
849 Miura K., Sladoje N. (eds) Bioimage Data Analysis Workflows. Learning Materials in
850 Biosciences. Springer, Cham. (2020).
851 46. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
852 **9**, 357-359 (2012).
853 47. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
854 2078-2079 (2009).

855 48. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth
856 approach to detect break points of large deletions and medium sized insertions from
857 paired-end short reads. *Bioinformatics* **25**, 2865-2871 (2009).
858

Figure 1

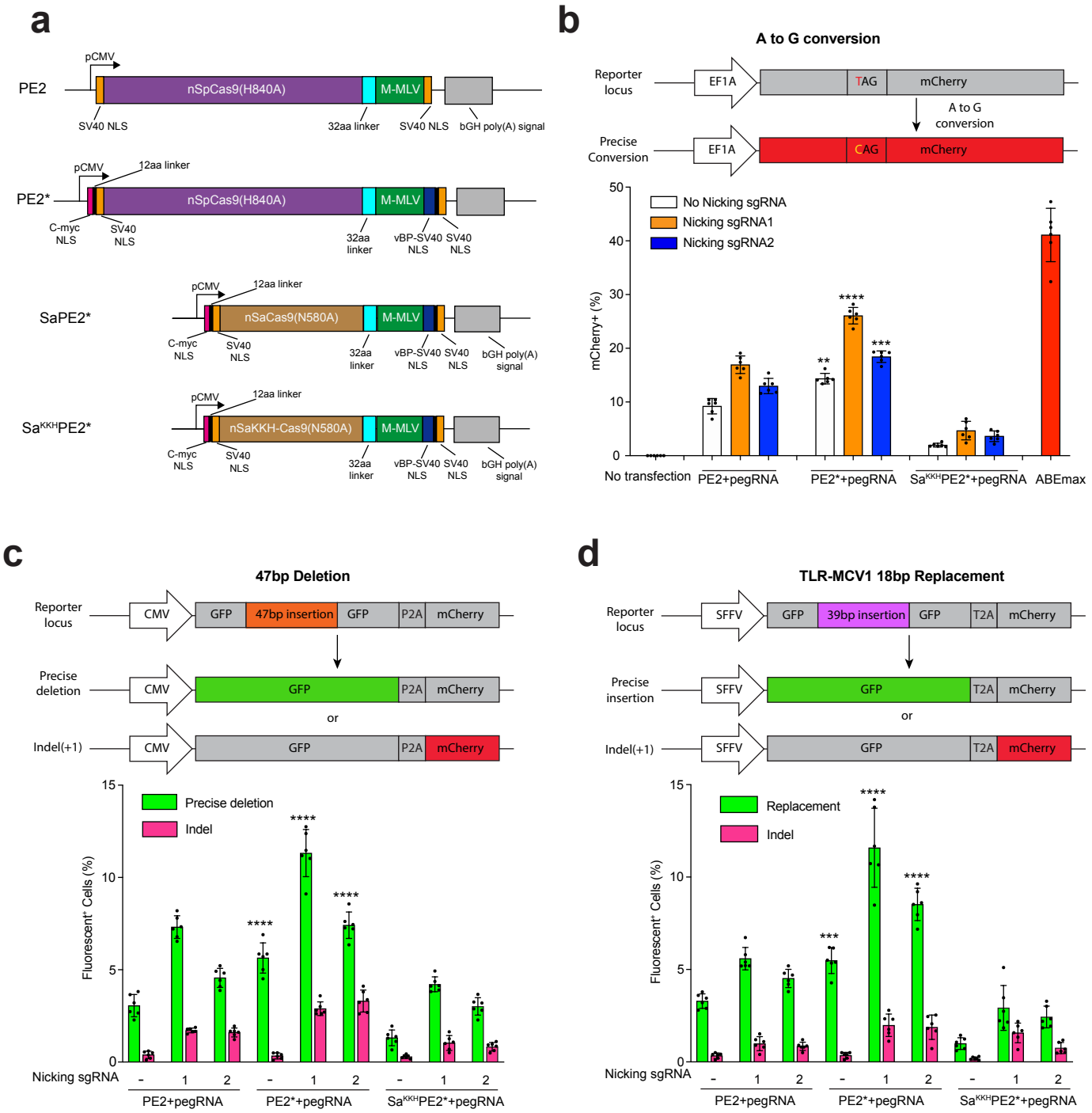

a
PE2
pCMV
nSpCas9(H840A)
M-MLV
SV40 NLS
32aa linker
SV40 NLS
bGH poly(A) signal
PE2*
12aa linker
C-myc NLS
SV40 NLS
vBP-SV40 NLS
SaPE2*
nSaCas9(N580A)
SaKKHPE2*
nSaKKH-Cas9(N580A)
b
A to G conversion
Reporter locus
EF1A
TAG
mCherry
A to G conversion
Precise Conversion
CAG
No Nicking sgRNA
Nicking sgRNA1
Nicking sgRNA2
mCherry+ (%)
No transfection
PE2+pegRNA
PE2*+pegRNA
SaKKHPE2*+pegRNA
ABEmax
c
47bp Deletion
Reporter locus
CMV
GFP
47bp insertion
P2A
mCherry
Precise deletion
or
Indel(+1)
Precise deletion
Indel
Fluorescent+ Cells (%)
Nicking sgRNA
d
TLR-MCV1 18bp Replacement
SFFV
39bp insertion
T2A
Precise insertion
Replacement

Figure 2

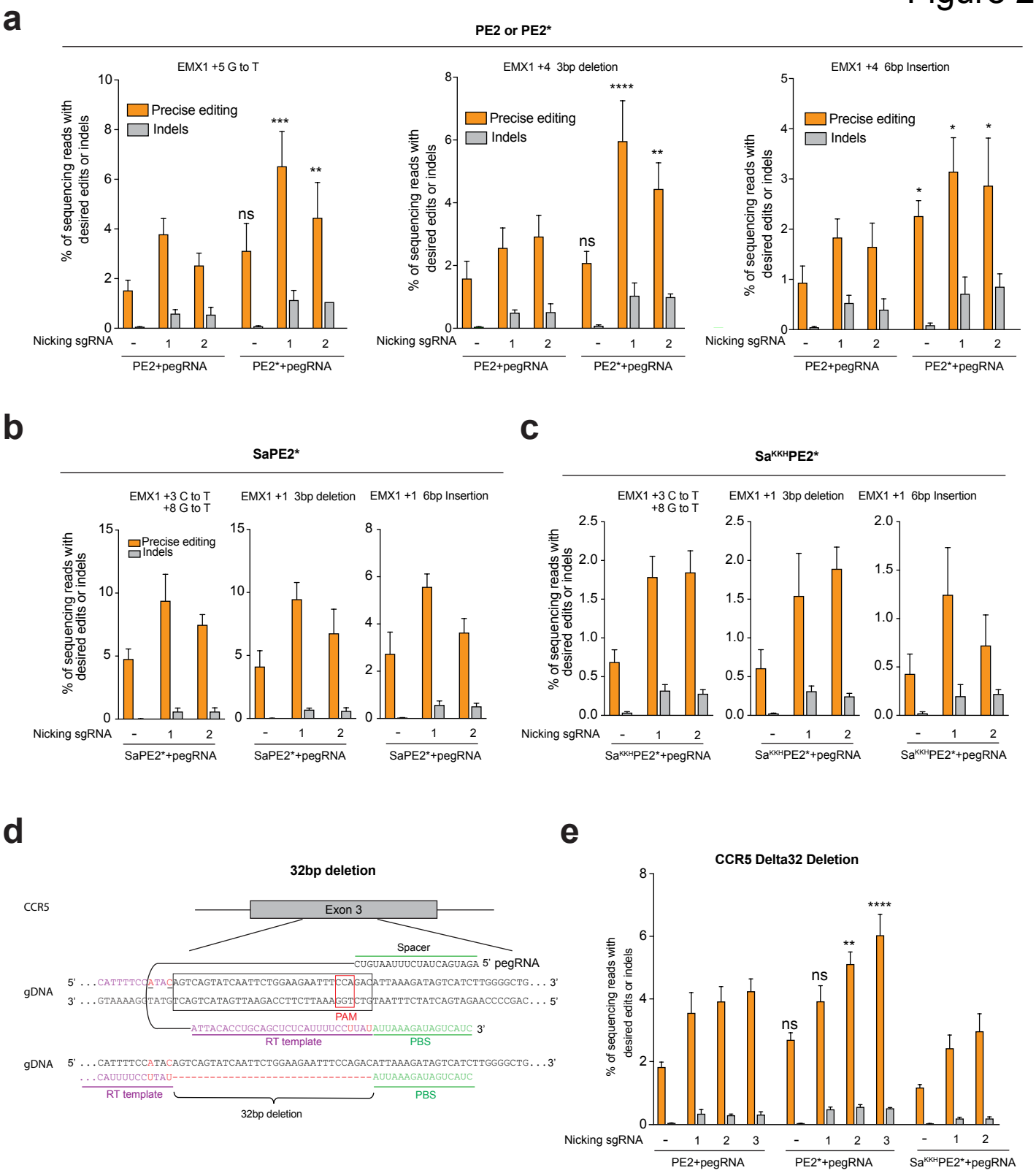

a
PE2 or PE2*
EMX1 +5 G to T
EMX1 +4 3bp deletion
EMX1 +4 6bp Insertion
Precise editing
Indels
% of sequencing reads with desired edits or indels
Nicking sgRNA
PE2+pegRNA
PE2*+pegRNA
b
SaPE2*
EMX1 +3 C to T +8 G to T
EMX1 +1 3bp deletion
EMX1 +1 6bp Insertion
SaPE2*+pegRNA
c
SaKKHPE2*
SaKKHPE2*+pegRNA
d
32bp deletion
CCR5
Exon 3
Spacer
CUGUAAUUUCUAUCAGUAGA 5' pegRNA
gDNA
5' ...CATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTG... 3'
3' ...GTAAAAGGTATGTCAGTCATAGTTAAGACCTTCTTAAAGGTCTGTAATTTCTATCAGTAGAACCCCGAC... 5'
PAM
ATTACACCUGCAGCUCUCAUUUUCCUUAUAUUAAAGAUAGUCAUC 3'
RT template
PBS
5' ...CATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTG...3'
...CAUUUUCCUTAU----------AUUAAAGAUAGUCAUC
32bp deletion
e
CCR5 Delta32 Deletion
ns
PE2+pegRNA
PE2*+pegRNA
SaKKHPE2*+pegRNA

Figure 3

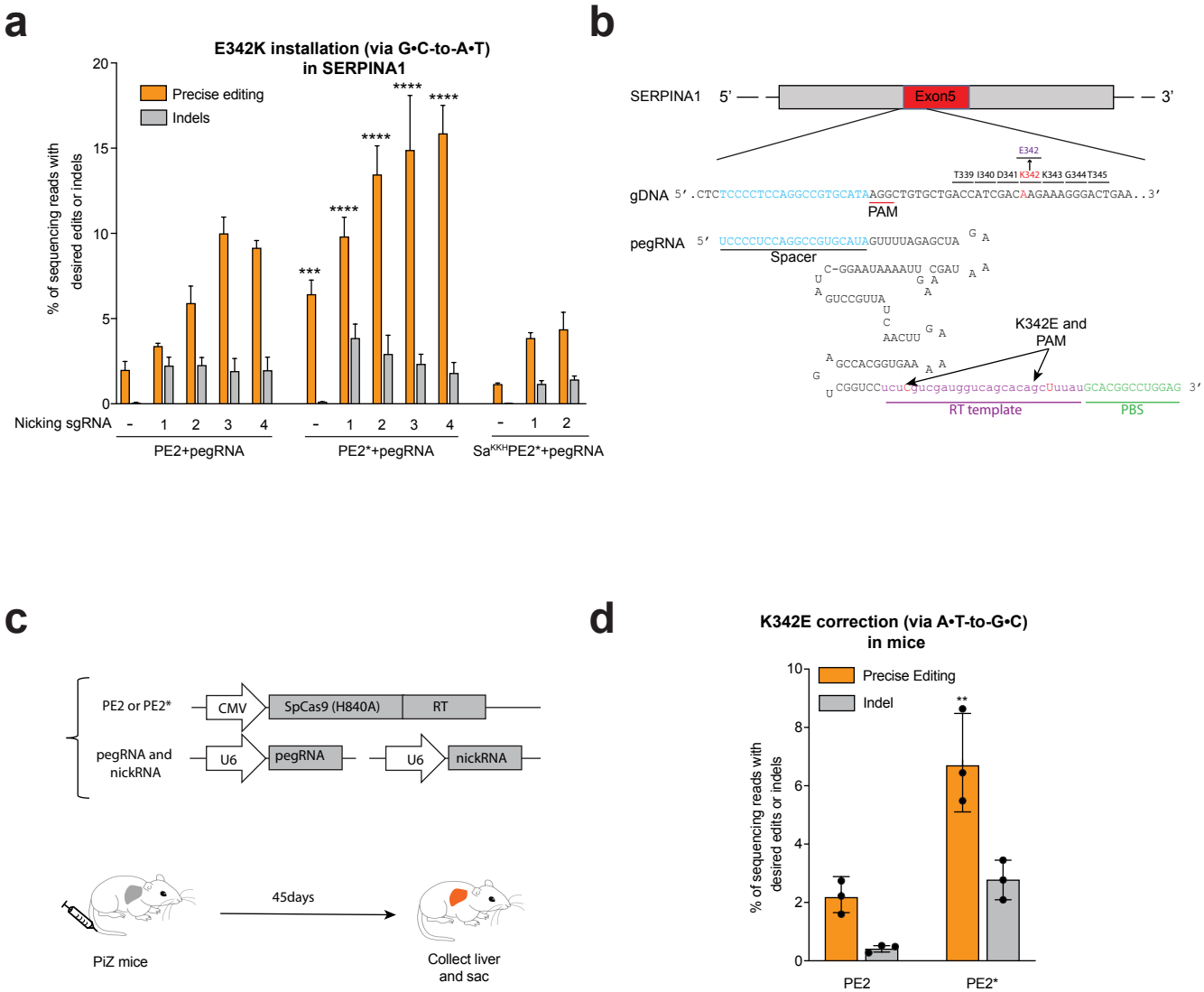

a
E342K installation (via G•C-to-A•T) in SERPINA1
Precise editing
Indels
% of sequencing reads with desired edits or indels
Nicking sgRNA
PE2+pegRNA
PE2*+pegRNA
Sa^KKH PE2*+pegRNA
b
SERPINA1
Exon5
gDNA
PAM
pegRNA
Spacer
K342E and PAM
RT template
PBS
c
PE2 or PE2*
CMV
SpCas9 (H840A)
RT
pegRNA and nickRNA
U6
pegRNA
nickRNA
45days
PiZ mice
Collect liver and sac
d
K342E correction (via A•T-to-G•C) in mice
Precise Editing
Indel
% of sequencing reads with desired edits or indels
PE2
PE2*

Figure 4

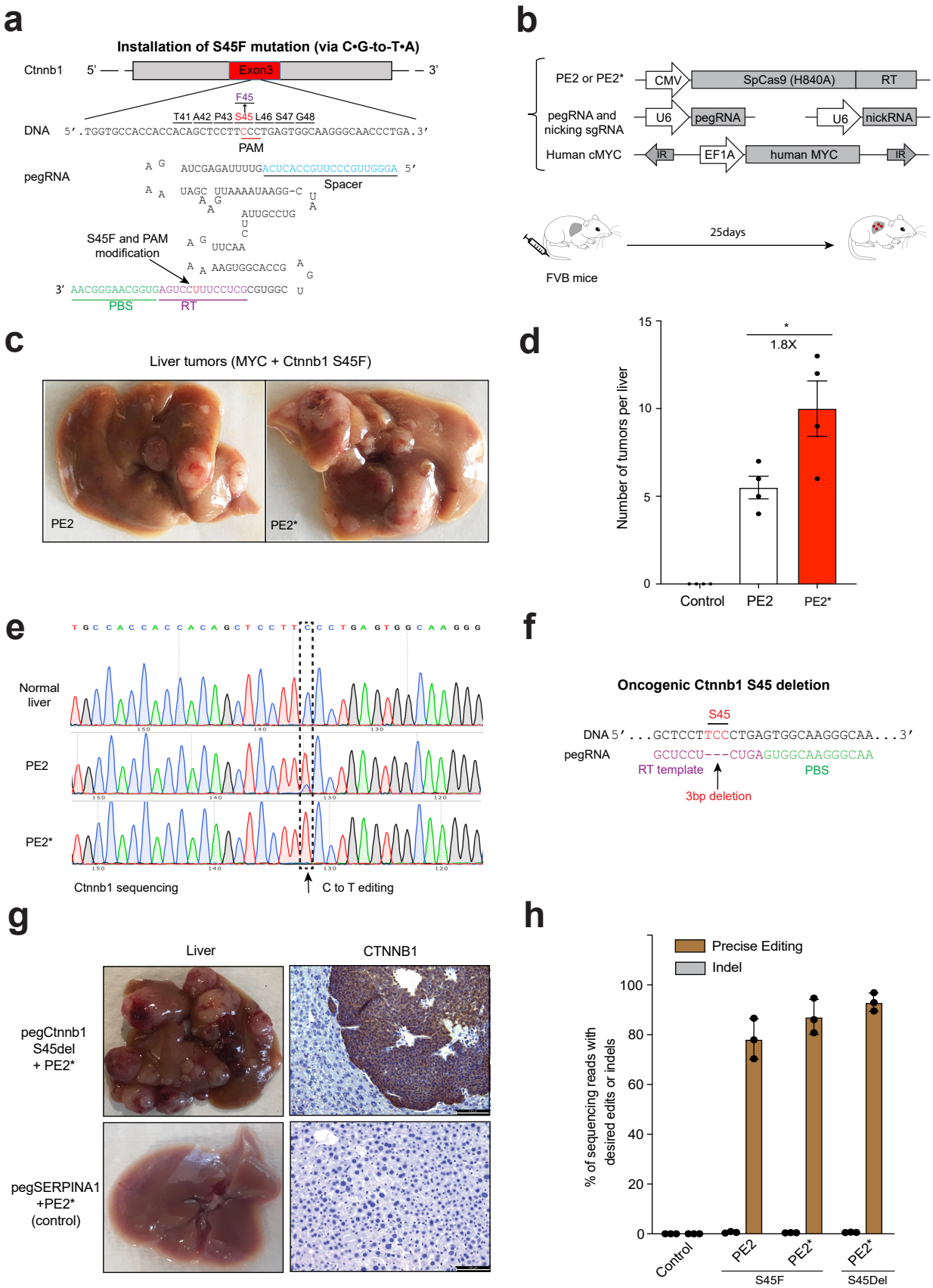

a
Installation of S45F mutation (via C•G-to-T•A)
Ctnnb1
5'
Exon3
3'
F45
T41 A42 P43 S45 L46 S47 G48
DNA
5'.TGGTGCCACCACCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGA.3'
PAM
pegRNA
AUCGAGAUUUUGACUCACCGUUCCCGUUGGGA 5'
Spacer
UAGC UUAAAAUAAGG-C
AUUGCCUG
UUCAA
AAGUGGCACCG
S45F and PAM modification
3' AACGGGAACGGUGAGUCCUUUCCUCGCGUGGC
PBS
RT
b
PE2 or PE2*
CMV
SpCas9 (H840A)
RT
pegRNA and nicking sgRNA
U6
pegRNA
U6
nickRNA
Human cMYC
IR
EF1A
human MYC
IR
25days
FVB mice
c
Liver tumors (MYC + Ctnnb1 S45F)
PE2
PE2*
d
Number of tumors per liver
15
10
5
0
*
1.8X
Control
PE2
PE2*
e
Normal liver
PE2
PE2*
Ctnnb1 sequencing
C to T editing
f
Oncogenic Ctnnb1 S45 deletion
S45
DNA 5'...GCTCCTTCCCTGAGTGGCAAGGGCAA...3'
pegRNA
GCUCCU---CUGAGUGGCAAGGGCAA
RT template
PBS
3bp deletion
g
Liver
CTNNB1
pegCtnnb1 S45del + PE2*
pegSERPINA1 +PE2* (control)
h
Precise Editing
Indel
% of sequencing reads with desired edits or indels
100
80
60
40
20
0
Control
PE2
PE2*
PE2*
S45F
S45Del

Figure 5

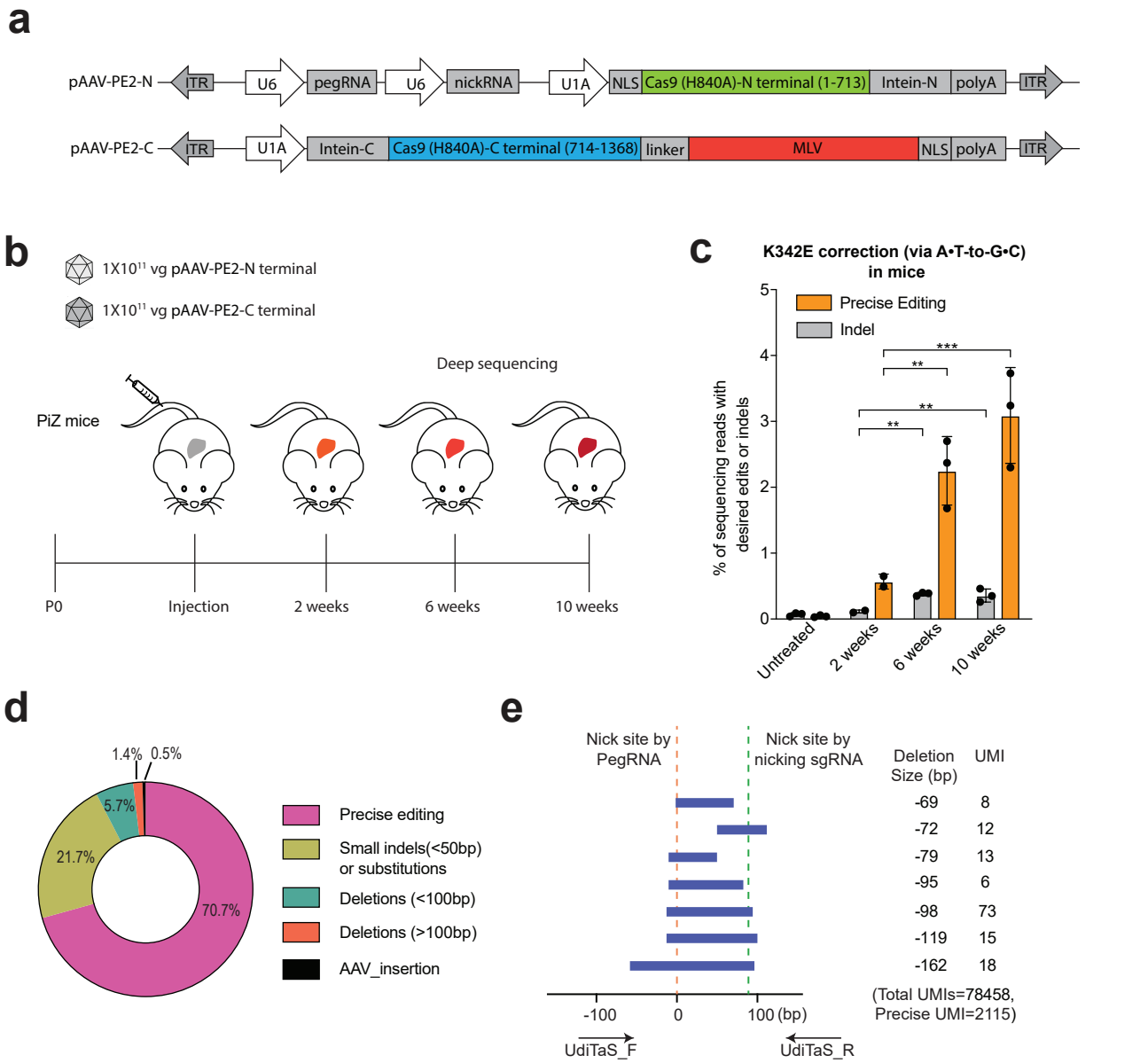

a
pAAV-PE2-N
ITR
U6
pegRNA
U6
nickRNA
U1A
NLS
Cas9 (H840A)-N terminal (1-713)
Intein-N
polyA
ITR
pAAV-PE2-C
ITR
U1A
Intein-C
Cas9 (H840A)-C terminal (714-1368)
linker
MLV
NLS
polyA
ITR
b
1X10¹¹ vg pAAV-PE2-N terminal
1X10¹¹ vg pAAV-PE2-C terminal
Deep sequencing
PiZ mice
P0
Injection
2 weeks
6 weeks
10 weeks
c
K342E correction (via A•T-to-G•C) in mice
Precise Editing
Indel
% of sequencing reads with desired edits or indels
Untreated
2 weeks
6 weeks
10 weeks
d
1.4%
0.5%
5.7%
21.7%
70.7%
Precise editing
Small indels(<50bp) or substitutions
Deletions (<100bp)
Deletions (>100bp)
AAV_insertion
e
Nick site by PegRNA
Nick site by nicking sgRNA
Deletion Size (bp)
UMI
-69 8
-72 12
-79 13
-95 6
-98 73
-119 15
-162 18
(Total UMIs=78458, Precise UMI=2115)
-100
0
100 (bp)
UdiTaS_F
UdiTaS_R

Figure S1

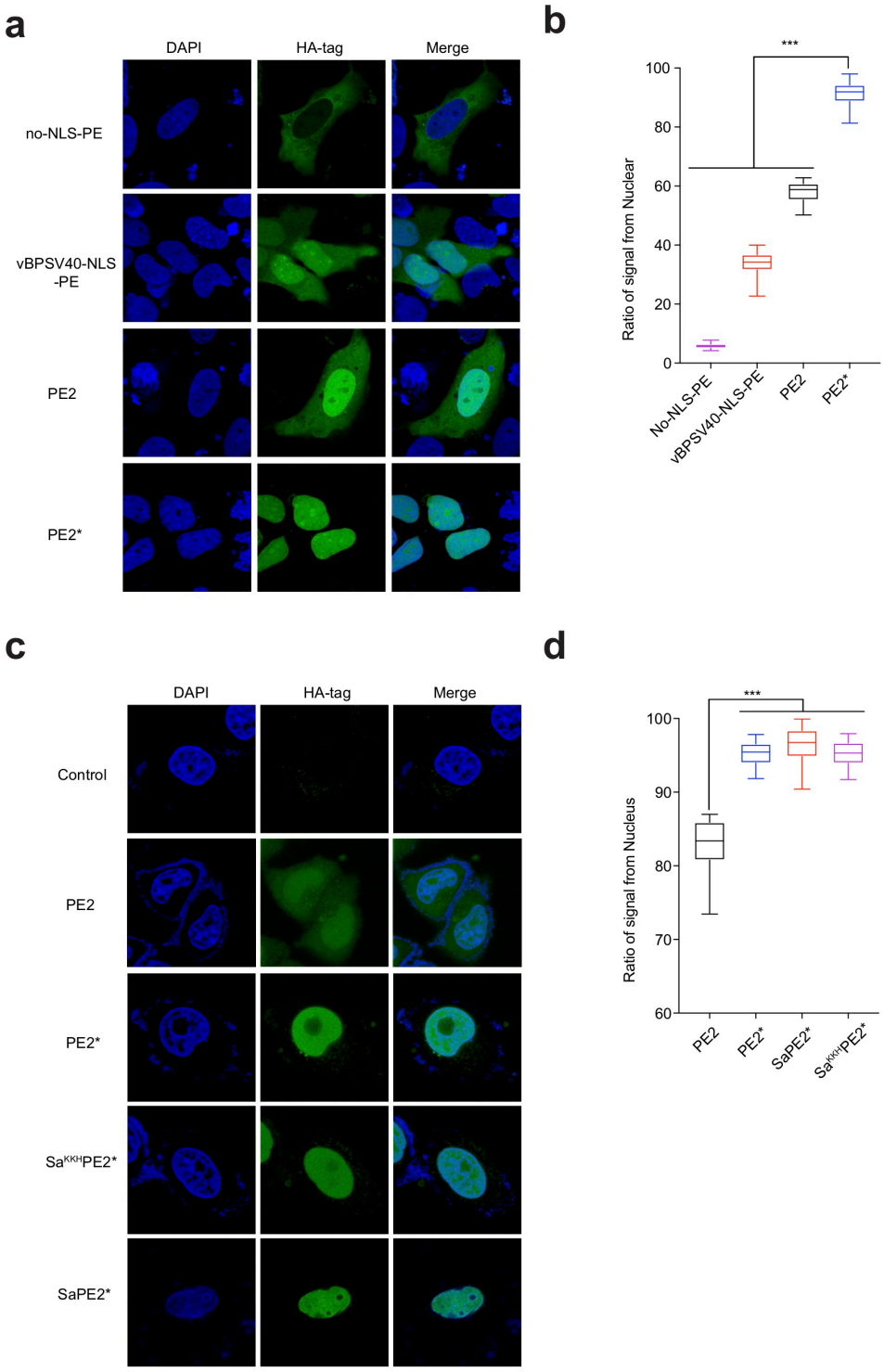

a
DAPI
HA-tag
Merge
no-NLS-PE
vBPSV40-NLS
-PE
PE2
PE2*
b
***
Ratio of signal from Nuclear
100
80
60
40
20
0
No-NLS-PE
vBPSV40-NLS-PE
PE2
PE2*
c
DAPI
HA-tag
Merge
Control
PE2
PE2*
SaKKHPE2*
SaPE2*
d
***
Ratio of signal from Nucleus
100
90
80
70
60
PE2
PE2*
SaPE2*
SaKKHPE2*

Figure S2

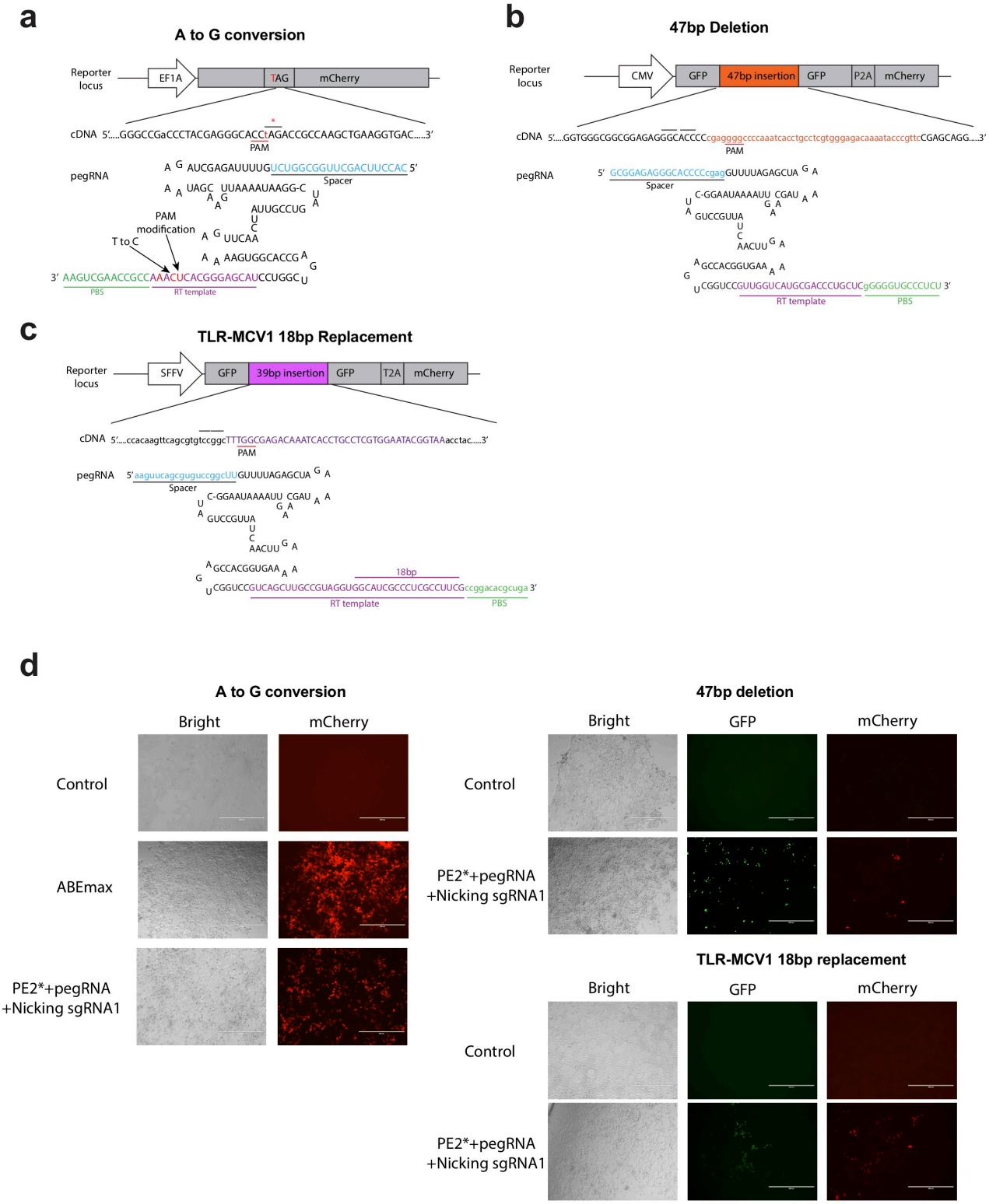

a
A to G conversion
Reporter locus
EF1A
TAG
mCherry
cDNA 5'....GGGCCGaCCCTACGAGGGCACCtAGACCGCCAAGCTGAAGGTGAC.....3'
PAM
pegRNA
AUCGAGAUUUUGUCUGGCGGUUCGACUUCCAC 5'
Spacer
UAGC UUAAAAUAAGG-C
AUUGCCUG
UUCAA
AAGUGGCACCG
PAM modification
T to C
3' AAGUCGAACCGCCAAACUCACGGGAGCAUCCUGGC
PBS
RT template
b
47bp Deletion
Reporter locus
CMV
GFP
47bp insertion
GFP
P2A
mCherry
cDNA 5'....GGTGGGCGGCGGAGAGGGCACCCCcgaggggccccaaatcacctgcctcgtgggagacaaaatacccgttcCGAGCAGG.....3'
PAM
pegRNA
5' GCGGAGAGGGCACCCCcgagGUUUUAGAGCUA
Spacer
C-GGAAUAAAAUU CGAU
GUCCGUUA
AACUU
GCCACGGUGAA
CGGUCCGUUGGUCAUGCGACCCUGCUCgGGGGUGCCCUCU 3'
RT template
PBS
c
TLR-MCV1 18bp Replacement
Reporter locus
SFFV
GFP
39bp insertion
GFP
T2A
mCherry
cDNA 5'....ccacaagttcagcgtgtccggcTTTGGCGAGACAAATCACCTGCCTCGTGGAATACGGTAAacctac.....3'
PAM
pegRNA
5'aaguucagcgugucccggcUUGUUUUAGAGCUA
Spacer
C-GGAAUAAAAUU CGAU
GUCCGUUA
AACUU
GCCACGGUGAA
18bp
CGGUCCGUCAGCUUGCCGUAGGUGGCAUCGCCCUCGCCUUCGccggacacgcuga 3'
RT template
PBS
d
A to G conversion
Bright
mCherry
Control
ABEmax
PE2*+pegRNA +Nicking sgRNA1
47bp deletion
Bright
GFP
mCherry
Control
PE2*+pegRNA +Nicking sgRNA1
TLR-MCV1 18bp replacement
Bright
GFP
mCherry
Control
PE2*+pegRNA +Nicking sgRNA1

Figure S3

a

CCR5 Delta32 Deletion

PE2

```
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCC      WT        94.82%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCtTAtA--------------------------------TTAAAGATAGTCATCTTGGGGCTGGTCC      Precise    3.86%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCA------------------------------GACATTAAAGATAGTCATCTTGGGGCTGGTCC      -32       0.17%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCtTAtAG-------------------------------TAAAGATAGTCATCTTGGGGCTGGTCC      -32       0.08%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCtTAtA--------------------------------TAAAGATAGTCATCTTGGGGCTGGTCC      -33       0.06%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTC------------------ATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCC      -18       0.04%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGAC---AAAGATAGTCATCTTGGGGCTGGTCC      -3        0.03%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACcATTAAAGATAGTCATCTTGGGGCTGGTCC      +1        0.01%
```

PE2*

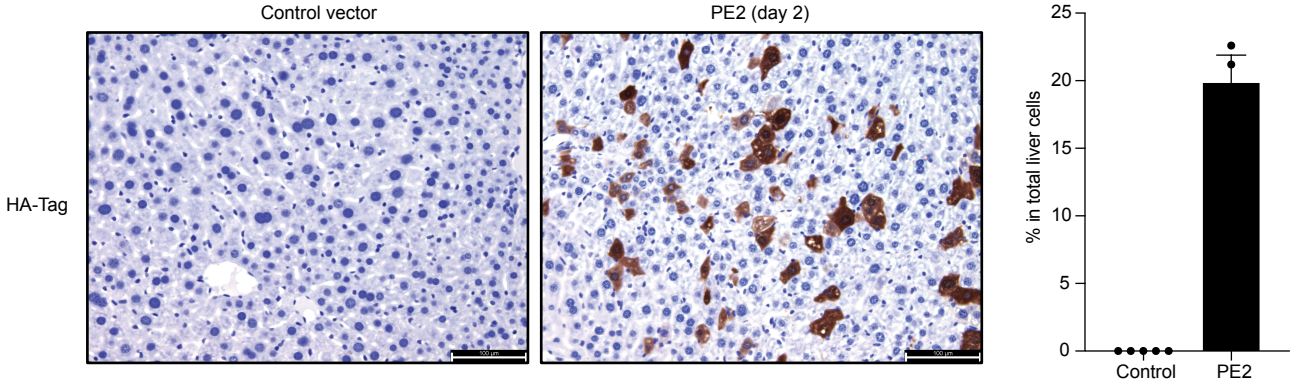
```
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCC      WT        91.25%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCtTAtA--------------------------------TTAAAGATAGTCATCTTGGGGCTGGTCC      Precise    6.08%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCA------------------------------GACATTAAAGATAGTCATCTTGGGGCTGGTCC      -32       0.25%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCtTAtAG-------------------------------TAAAGATAGTCATCTTGGGGCTGGTCC      -32       0.14%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTT-----------------------------------GACATTAAAGATAGTCATCTTGGGGCTGGTCC      -33       0.08%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCtTAtA--------------------------------TAAAGATAGTCATCTTGGGGCTGGTCC      -33       0.06%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTC------------------ATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCC      -18       0.06%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGAC---AAAGATAGTCATCTTGGGGCTGGTCC      -3        0.04%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTT--------GTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCC      -8        0,03%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACcATTAAAGATAGTCATCTTGGGGCTGGTCC      +1        0.03%
```

Sa^{KKH}PE2*

```
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCC      WT        96.36%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACA--------------------------------TTAAAGATAGTCATCTTGGGGCTGGTCC      Precise    2.85%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCA------------------------------GACATTAAAGATAGTCATCTTGGGGCTGGTCC      -32       0.06%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTT-----------------------------------GACATTAAAGATAGTCATCTTGGGGCTGGTCC      -33       0.04%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGAC---AAAGATAGTCATCTTGGGGCTGGTCC      -3        0.02%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGAC-----------TCATCTTGGGGCTGGTCC      -11       0.02%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAA-----------------TAGTCATCTTGGGGCTGGTCC      -17       0.01%
```

Figure S4

a



b

K342E correction (via A•T-to-G•C in mice)

PAM modification

Correction

Nicking sgRNA PAM

PE2

```
CTCTCCCCTCCAGGCCGTGCATAAGGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCGA   WT        93.8%
CTCTCCCCTCCAGGCCGTGCATAAaGCTGTGCTGACCATCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   Precise   2.87%
CTCTCCCCTCCAGGCCGTGC---------TGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   -12       0.14%
CTCTCCCCTCCAGGC--------GTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   -12       0.11%
CTCTCCCCTCCAGGCCGTGCATAAaGCTGTGCTGACCATCGACgAGA-------------------GCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   -20       0.05%
CTCTCCCCTCCAGG--------------------------------GACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   -37       0.05%
CTCTCCCCTCCAGGCCGTG----------TGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   -10       0.03%
CTCTCCCCTCCAGGCCGTGC-------------------------------------------------------------------------------CCCCCCGA   -80       0.03%
```

PE2*

```
CTCTCCCCTCCAGGCCGTGCATAAGGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   WT        84.7%
CTCTCCCCTCCAGGCCGTGCATAAaGCTGTGCTGACCATCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   Precise   8.68%
CTCTCCCCTCCAGGCCGTGC---------TGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   -12       0.81%
CTCTCCCCTCCAGGC--------GTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   -12       0.72%
CTCTCCCCTCCAGGCCGTGC-------------------------------------------------------------------------------CCCCCCGA   -80       0.67%
CTCTCCCCTCCAGGCCGTGCATAAaGCTGTGCTGACCATCGACgAGA-------------------GCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   -20       0.43%
CTCTCCCCTCCAGGCC-------------------------------------------------------------------ATACCCATGTCTATCCCCCCCGA   -69       0.26%
CTCTCCCCTCCAGGCCGTGCgATAAaGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   +1        0.11%
CTCTCCCCTCCAGGCCGTGCATAAGGCTGTGCTGAGCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   sub       0.07%
CTCTCCCCTCCAGGCCGTGCATAAGGCTGTGCTGACCATCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   sub       0.06%
CTCTCCCCTCCAGGCCGTGCATAAaGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   sub       0.04%
```

Figure S5

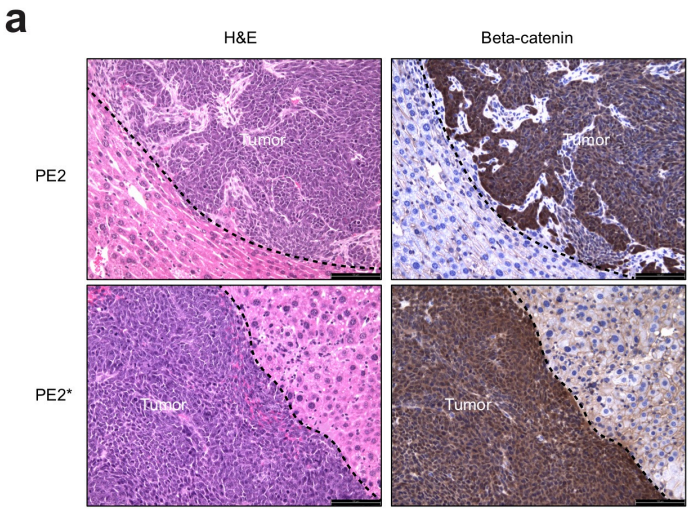


Figure S6

a

K342E correction (via dual AAV)

PAM modification | Correction | Nicking sgRNA PAM | Amplicon-seq

```
GGCCGTGCATAAGGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   WT        93.04%
GGCCGTGCATAAaGCTGTGCTGACCATCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   Precise   3.39%
GGCCGTGCATAAaGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   sub       0.51%
GGCCGTGCATAAaGCTGTGCTGACCATCGACtAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   sub       0.25%
GGCCGTGCATAAaGCTGTGCTGACCATtGAtGAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   sub       0.21%
GGCCGTGCATAAaGCTGTGCTGACCATCGACAAGAAAGtGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   sub       0.12%
GGCCGTGC---------TGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   -12       0.05%
GGCCGTGC---------------------------------------------------------------------------CCCCCCGA   -80       0.04%
GGCCGTG----------------------------------------------------------------------------GATTGAACAAA   -119      0.02%
GGCCGTGC-TAAGGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGA   -1        0.02%
```

b

K342E correction (via dual AAV)

PAM modification | Correction | Nicking sgRNA PAM | UDiTaS UMI/Total UMI | Editing (%) | p

```
GGCCGTGCATAAGGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGG        WT      (72173/78458)  91.9
GGCCGTGCATAAaGCTGTGCTGACCATCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGG        Precise(2115/78458)  2.70   0.034
GGCCGTGCATAAaGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGG        sub     (137/78458)  0.17   0.036
GGCCGTGCATAAaGCTGTGCTGACCATCGACtAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGG        sub     (115/78458)  0.15   0.045
GGCCGTGCATAAaGCTGTGCTGACCATtGAtgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGG        sub     (89/78458)   0.11   0.049
GGCCGTGCATAAaGCTGTGCTGACCATCGAtAAGAAAGtGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGG        sub     (51/78458)   0.07   0.036
GGCCGTGC-TAAGGCTGTGCTGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGG        -1      (29/78458)   0.04   0.034
GGCCGTGC---------TGACCATCGACAAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGG        -12     (19/78458)   0.02   0.036

GGCCGTGCATAAGGCTGTGCTGACCATCGACAAGAAAGGGACTGAAG-------------------------------------------[5bp]   -56    (8/78458)    0.01   0.148
GGCC-------------------------------------------------------------------ATACCCATGTCTATCCCCCCCGAGG        -69    (8/78458)    0.01   0.045
GGCCGTGCATAAaGCTGTGCTGACCATCGACgAGAAAGGGACTGAAGCTGCTGGGGCCATGTTTTT-------------------------[50bp]   -72    (12/78458)   0.02   0.034
[12bp]-----------------------------------------------------------GAGGCCATACCCATGTCTATCCCCCCCGAGG        -79    (13/78458)   0.02   0.036
GGCCGTGC-----------------------------------------------------------------------------CCCCCCCGAGG        -80    (4/78458)    0.01   0.071
G----------------------------------------------------------------------------------CGAGG        -92    (3/78458)    0.00   0.144
[12bp]-----------------------------------------------------------------------------TCTATCCCCCCCGAGG        -95    (6/78458)    0.01   0.057
GGCCGTG------------------------------------------------------------------------------[7bp]   -98    (73/78458)   0.09   0.034
Deletions between pegRNA and nicking sgRNA (<100bp)

[13bp]-----------------------------------------------------------------------------------[50bp]   -162   (18/78458)   0.02   0.049
GGCCGTG----------------------------------------------------------------------------------[31bp]   -119   (15/78458)   0.02   0.036
[8bp]------------------------------------------------------------------------------------[80bp]   -186   (9/78458)    0.01   0.085
[12bp]-----------------------------------------------------------------------------------[868bp]  -978   (3/78458)    0.00   0.176
[2316bp]---------------------------------------------------------------------------------[28bp]   -2442  (2/78458)    0.00   0.144
Deletions spaning pegRNA and nicking sgRNA (>100bp)

GGCCGTGCATAAaGCTGTGCTGACCATCGACgAGAgAGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGG        Precise(12/78458)    0.02
GGCCGTGCATAAaGCTGTGCTGACCATCGACgAGAggGGGACTGAAGCTGCTGGGGCCATGTTTTTAGAGGCCATACCCATGTCTATCCCCCCCGAGG        Precise(7/78458)     0.01
Scaffold insertion
```

S45 deletion (via TCC deletion)

Nicking sgRNA PAM | TCC deletion | Peg sgRNA PAM | UDiTaS UMI/Total UMI

```
GAAAAGCTGCTGTCAGCCACTGGCAGCAGCAGTCTTACTTGGATTCTGGAATCCATTCTGGTGCCACCACCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCT        WT      2026/45539
GAAAAGCTGCTGTCAGCCACTGGCAGCAGCAGTCTTACTTGGATTCTGGAATCCATTCTGGTGCCACCACCACAGCTCCT---CTGAGTGGCAAGGGCAACCCT        Precise 42153/45539
[63bp]-----------------------------------ACTTGGATTCTGGAATCCATTCTGGTGCCACCACCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCT        -99    12/45539
GAAAAGCTGCTGTCAGCCACTGGCAGCAGCAGT-------------------------------------------------------------------[7bp]   -78    14/45539
GAAAAGCTGCTGTCAGCCACTGGCAGCAGCAGTCTTACTTGGATTCTGGAATCCATTCTGGTGCCACCA----------------------------------[39bp]  -74    16/45539
GAAAAGCTGCTGTCAGCCACTGGCAGCAGCAGTCTTACTTGGATTCTGGAATCCATTCTGGTGCCACCACCACAGCTC--------------------------[15bp]  -41    12/45539
[506bp]-----------------------------------------------------------------------------------------------[71bp]  -681   8/45539
```