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 pegRNA[6]. PE2 harbors five mutations within the M-MLV RT that improve editing efficiency[6]. PE3 uses an additional

69 sgRNA to direct SpCas9^{H840A} to nick the non-edited DNA strand to encourage the edited strand
70 to be utilized as a repair template by DNA repair factors, leading to further increases in editing
71 efficiency[6].
72
73 The ability to precisely install or correct pathogenic mutations regardless of their composition
74 makes prime editing an intriguing approach to perform somatic genome editing in model
75 organisms to study disease processes or to utilize for therapeutic applications. Previous studies
76 have used prime editing to recode loci in cultured cells[6], plants[7], stem cells[8, 9], and mouse
77 zygotes[10, 11]. However, PE delivery in adult animals has not yet been described. In this study, we
78 report a nuclear localization signal sequence optimized PE2 (referred to as PE2*) with higher
79 editing efficiency than PE2. We demonstrate that PE2* enables somatic genome editing in the
80 liver of adult mice, where it can correct a pathogenic disease allele or introduce a directed
81 mutation to drive tumor formation to facilitate cancer modeling. The size of a prime editor
82 precludes its packaging in a single AAV vector. We show that dual AAV-mediated delivery of a
83 split-intein prime editor is functional *in vivo* for gene editing in the mouse liver. These data
84 demonstrate the feasibility of employing PE *in vivo* for the targeted, precise alteration of genomic
85 sequence with potential utility both in model organisms and as a therapeutic modality.
86
87 **Results**
88 **Optimized nuclear localization signal sequence composition improves prime editing**
89
90 PEs have the remarkable ability to introduce a variety of different types of sequence alterations
91 into the genome. Their editing efficiency is influenced by a variety of different parameters (PBS
92 length, position of the RT initiation site relative to the desired sequence alteration, composition of
93 the desired sequence alteration, relative position of the alternate strand nick, etc.). Even under
94 optimal conditions the incorporation rate of the desired edit into the genome is incomplete [6].
95 Previously, we and others have noted that the composition and number of nuclear localization
96 signal (NLS) sequences within a Cas9 effector can influence its efficiency of genome editing[3, 12-
97 16]. The original PE2 contains two SV40 NLS sequences [6], which can be suboptimal for maximizing
98 editing rates by SpCas9 protein-sgRNA complexes[12]. In transient transfection assays of original
99 PE2, we observed incomplete nuclear localization based on immunofluorescence: ~60% of the
100 protein is present in the nucleus in U2OS cells and ~85% is present in the nucleus in HeLa cells;
101 (**Supplementary Fig. S1**). We found that addition of a N-terminal c-Myc NLS[17] and C-terminal

102 variant bipartite SV40 NLS (vBP-SV40)[18] gives rise to nearly complete nuclear localization of the
103 prime editor (PE2*) (**Fig. 1a, Supplementary Fig. S1c,d**).
104
105 In parallel, we constructed PE2* frameworks using the orthogonal Staphylococcus aureus
106 (SaCas9) nickase (N580A)[19] to replace the SpCas9 nickase (H840A). To expand the potential
107 targeting range of these alternate prime editors systems, we utilized both the standard SaCas9
108 backbone that recognizes an NNGRRT PAM (SaPE2*)[19] and the SaCas9^{KKH} variant that broadens
109 the targeting to an NNNRRT PAM (Sa^{KKH}PE2*)[20] (**Fig. 1a**). By immunofluorescence staining using
110 the incorporated 3xHA-tag we observed that SaPE2* or Sa^{KKH}PE2* proteins were localized to the
111 nucleus (**Supplementary Fig. S1a,b**).
112
113 To determine whether the observed improvements in nuclear localization translate into increases
114 in editing efficiency, we evaluated the rate of nucleotide conversion for PE2 and PE2* in HEK293T
115 mCherry reporter cells that contains a premature TAG stop codon that prevents translation of a
116 functional protein (**Supplementary Fig. S2a,d**). PE2 and PE2* were programmed with a pegRNA
117 designed to revert the TAG codon to CAG and delivered without and with different nicking sgRNAs.
118 3 days after transfection, we performed flow cytometry to quantify prime-editing efficiency. PE2*
119 produced a 1.5-1.6 fold increase in editing efficiency (14.3% to 26.4%) relative to PE2 (9.2% to
120 16.5%; **Fig. 1b**). A compatible PAM is also present for Sa^{KKH}PE2* within the reporter to evaluate
121 nucleotide conversion rates, with which we observed more modest editing efficiencies (1.8% to
122 4.7%). All PE systems displayed lower editing activity than an adenine base editor system
123 (ABEmax)[13] for restoration of reporter function (**Fig. 1b**).
124
125 Next, we tested the efficiency for generating a targeted deletion using PE2, PE2* and Sa^{KKH}PE2*
126 in an HEK293T reporter line that can quantify both precise deletions and indel formation. This
127 reporter design shares similarities to the traffic light reporter system [21, 22]. A precise deletion of
128 47bp will remove a sequence insertion disrupting GFP expression, whereas indels that produce
129 a particular reading frame alteration restore mCherry expression (**Supplementary Fig. S2b,d**).
130 PE2* produced a 1.6-1.9 fold increase in the level of precise deletions (5.6%-11.3%) compared
131 to PE2 (3.0%-7.3%; **Fig. 1c**). The relative level of undesired indel formation was roughly
132 proportional to the overall activity levels for PE2 and PE2*. We also observed that Sa^{KKH}PE2*
133 could generate precise 47bp deletion with efficiencies ranging from 1.3% to 4.2% (**Fig. 1c**).
134

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

152 Next, we compared the editing efficiency of PE2 and PE2* for the creation of previously described
153 nucleotide substitutions, deletions and insertions at the *EMX1* locus[6]. HEK293T cells were
154 transfected with different prime editors, pegRNAs and different nicking sgRNAs. Genomic DNA
155 was isolated and editing outcomes at each target site were quantified by high-throughput
156 sequence (HTS). PE2* (3.1%- 6.5%) led to an average 1.9-fold increase in the rate of point
157 mutation introduction compared to PE2 (1.5%- 3.7%) (**Fig. 2a**). Targeted 3bp deletions were
158 generated at 1.4 to 2.1-fold higher rate by PE2* (2.1%- 6.0%) than PE2 (1.5%- 2.9%) (**Fig. 2a**).
159 Targeted 6bp insertions were generated at 1.7 to 2.4-fold higher rate by PE2* (2.2%-3.1%) than
160 PE2 (0.9%-1.8%) (**Fig. 2a**). As observed with the various reporter systems, the level of indel
161 formation was roughly proportional to the activity levels of PE2 and PE2*. Together, these
162 observations suggest that PE2* has broadly improved editing efficiency at endogenous loci.
163
164 We also compared the editing efficiency of SaPE2* and Sa^{KKH}PE2* for the creation of similar
165 nucleotide substitutions, deletions and insertions at the *EMX1* locus. Notably, SaPE2* installed
166 point mutations at two positions in the *EMX1* locus with editing efficiency from 4.7% to 9.3% and
167 a modest indel rate (0.0%-0.5%). Targeted 3-bp deletion and 6-bp insertion were introduced by

168 SaPE2* with an editing efficiency of 4.1% to 9.4% and 2.7% to 5.5%, respectively. Indel induction
169 generated by SaPE2* ranged from 0.0% to 0.6% (**Fig. 2b**). Overall, Sa^{KKH}PE2* exhibited lower
170 editing efficiency at the *EMX1* locus than SaPE2* with the same set of pegRNAs (typically
171 between 1 and 2%; **Fig. 2c**). Notably, at these loci the editing efficiencies for SaPE2* were similar
172 to the rates obtained with PE2*, suggesting that the SaPE2* platform has the potential to broaden
173 the scope of available prime editing systems.
174
175 A homozygous 32-bp deletion in the *CCR5* gene is associated with resistance to human
176 immunodeficiency virus (HIV-1) infection[24-26]. We evaluated the utility of PE2 and PE2* to
177 generate a large, therapeutically relevant 32-bp deletion within *CCR5* that recapitulates the HIV-
178 1 resistance allele (**Fig. 2d**). We utilized linear amplification to incorporate unique molecular
179 identifiers (UMI) prior to sequencing[27] to avoid PCR amplification bias for the assessment of the
180 deletion rate in the population of treated cells. Both PE2 and PE2* were able to generate the
181 desired 32-bp deletion within the *CCR5* locus in HeLa cells, where the PE2* editor displayed
182 higher deletion rates than PE2 (average 1.4-fold across all conditions) with a maximum efficiency
183 of about 6.0% (**Fig. 2e, Supplementary Fig. S3**). Sa^{KKH}PE2* exhibited lower editing efficiency
184 than PE2 for the generation of the 32-bp deletion, with a maximum efficiency of 2.9% (**Fig. 2e,**
185 **Supplementary Fig. S3**). Overall, these results demonstrate that PE can introduce a
186 therapeutically relevant deletion in the *CCR5* gene in human cells.
187
188 **Improved PE increases the correction efficiency of a pathogenic mutation *in vivo***
189
190 Alpha-1 antitrypsin deficiency (AATD) is an inherited disorder that is caused by mutations in the
191 Serpin Peptidase Inhibitor Family A member 1 (*SERPINA1*) gene[28]. The E342K mutation (via
192 G•C-to-A•T) in *SERPINA1* (PiZ allele) is the most frequent mutation and causes severe lung and
193 liver disease[28]. Patients with homozygous mutation in *SERPINA1* (PiZZ) have PiZ protein
194 aggregates in hepatocytes and lack of functional AAT protein in the lung. The PiZ transgenic
195 mouse contains 16 copies of the human *SERPINA1* PiZ allele and is a commonly-used mouse
196 model of human AATD [29]. For the correction of the E342K mutation, there are some challenges
197 for the utilization of an adenine base editor: no optimal NGG PAM is present nearby for SpCas9,
198 and in addition to the target adenine there are several other adenines that may also be susceptible
199 to base conversion ("bystander effect")[30]. Consequently, we investigated the utilization of the PE
200 platform to correct this pathogenic mutation. To test the efficiency of different prime editors at this
201 locus, we first evaluated the generation of the pathogenic E342K mutation (via G•C-to-A•T

202 conversion) in wildtype *SERPINA1* in HEK293T cells. A series of different nicking sgRNAs were
203 also evaluated in conjunction with the PEs (PE3 strategy[6]). We observed 1.6 to 3.4-fold increase
204 for G•C-to-A•T base transition in *SERPINA1* by PE2* (6.4%-15.8%) compared to PE2 (1.9%-
205 9.9%; **Fig. 3a**). The average rate of indel generation with a nicking sgRNA slightly increased with
206 PE2* (0.1% -3.8%) compared to PE2 (0.0%-2.2%). Sa^{KKH}PE2* exhibited lower overall editing
207 efficiency for the installation of the E342K mutation (1.1%-4.4%). Indel generation at the target
208 site by Sa^{KKH}PE2* ranged from 0.0% to 1.4% (**Fig. 3a**).
209
210 Next, we investigated the ability of prime editors to directly correct a pathogenic mutation *in vivo.*
211 Based on the editing results for *SERPINA1* in HEK293T cells, we chose the nicking sgRNA3 for
212 use with the PEs for the *in vivo* experiments. We utilized a different pegRNA designed to revert
213 the E342K mutation (**Fig. 3b)**. A PAM mutation (AGG to AAG) was also included to reduce re-
214 cutting of the locus that introduced a synonymous codon change (**Fig. 3b)**. We introduced PE2
215 or PE2*, pegRNA and nicking sgRNA into the liver of PiZ mice (n=3/group) through hydrodynamic
216 tail vein injection (**Fig. 3c**). Hydrodynamic injection can deliver plasmid DNA to 20-30% of
217 hepatocytes [31]. Using a PE2 plasmid encoding an HA-tag, we observed that 19.98 $\pm$0.88%
218 hepatocytes were HA-tag positive (n=4 mice) (**Supplementary Fig. 4a**). 45 days after injection,
219 livers of PiZ mice were collected and DNA was purified for HTS analysis. Notably, we observed
220 3.1-fold increase for A•T-to-G•C correction in PiZ *SERPINA1* by PE2* (6.7% on average)
221 compared to PE2 (2.1%). The indel rate at the locus was also increased from 0.4% to 2.7% (**Fig.
222 3d**). Interestingly, we observed a low frequency of large deletions between pegRNA and nicking
223 sgRNA (**Supplementary Fig. S4b**). Together, these data demonstrate that prime editors can
224 restore the wild type *SERPINA1* allele thereby enabling pathogenic gene correction in adult mice.
225
226 **Improved PE enhances tumor burden by somatic engineering in the mouse liver**
227
228 *CTNNB1* (β-catenin) is a commonly mutated gene in hepatocellular carcinoma[32]. Overexpression
229 of a mutant *Ctnnb1* and Myc oncogene have been used to generate liver cancer models[14]. To
230 explore the potential of prime editors to drive tumor formation *in vivo*, we delivered PE2 or PE2*,
231 a pegRNA used for installation (via C•G-to-T•A) of the oncogenic S45F mutation in *Ctnnb1* (**Fig.
232 4a**), and a nicking sgRNA to the livers of adult FVB mice (n=4/group) by hydrodynamic tail vein
233 injection (**Fig. 4b**). A MYC transposon and transposase were co-injected to provide a second
234 oncogenic driver necessary for tumor formation in conjunction with the *Ctnnb1* mutation[33]. 25
235 days after injection, livers of adult mice were collected and tumor nodules on the liver were

236 quantified. PE2-treated animals showed an average 5.5 ± 1.1 tumors per mouse, whereas PE2*-
237 treated mice displayed higher rates of tumor formation, with an average 10.0 ± 2.7 tumors on the
238 liver (**Fig. 4c,d**). Consistent with gain of function of the S45F mutation, liver tumors were positive
239 for nuclear β-Catenin[14] (**Supplementary Fig.5**). Sanger sequence of gDNA from the tumor
240 nodules showed precise conversion of S45F in *Ctnnb1* (**Fig. 4e**).
241
242 Prime editors afford the opportunity to install other types of mutations within the genome. To
243 assess the feasibility of generating deletions *in vivo*, we designed a pegRNA to delete the S45
244 codon in *Ctnnb1*, which is a previously described oncogenic mutation at this locus [34](**Fig. 4f**). The
245 prime editor (PE2*), pegRNA for *Ctnnb1* S45 deletion and nicking sgRNA plasmids were delivered
246 by hydrodynamic tail-vein injection along with the MYC transposon and transposase plasmids.
247 pegRNA *Ctnnb1* S45 deletion-treated animals showed extensive tumor formation, whereas
248 pegRNA SERPINA1-treated animals did not induce any tumor formation (**Fig. 4g**). Deep
249 sequencing showed that more than 80% of tumor gDNA contained precise editing removing the
250 S45 codon (**Fig. 4h**). Together, these results demonstrate that prime editors can be used for
251 generating tumor models by somatic cell engineering *in vivo*, and that PE2* provides a platform
252 with improved editing activity.
253
254 **Utilizing a split-intein approach for prime editor delivery**
255
256 The ~6.3-kb coding sequence of PE exceeds the ~4.8-kb packaging size limit of AAV[35]. To deliver
257 prime editors with AAVs, we adapted a split Cas9 dual-AAV strategy[36] in which the original PE2
258 prime editor is divided into an amino-terminal (PE2-N) and carboxy-terminal (PE2-C) segments,
259 which are then reconstituted to full length PE by a *trans*-splicing intein[37] (**Fig. 5a**). To ensure that
260 each prime editor segment is smaller than the AAV packaging size limit, we divided the PE within
261 the SpCas9 amino acid before Ser 714[38]. We generated AAV8 particles encoding the split-intein
262 PE, a nicking sgRNA and a pegRNA (PE3 strategy[6]) to correct the E342K mutation in SERPINA1
263 (**Fig. 5a**). We then characterized the performance of the split-intein AAV prime editor *in vivo*. PiZ
264 mice were treated by tail-vein injection of a low dose dual AAV8-PE (2×10^{11} viral genome total)
265 (**Fig. 5b**). Livers were harvested at 2 weeks (n=2), 6 weeks (n=3) and 10 weeks (n=3) after
266 injection. By targeted deep sequencing, we detected 0.6 ± 0.0 % precise editing at 2 weeks. The
267 precise editing efficiency increased significantly to 2.3 ± 0.4 % at 6 weeks and 3.1 ± 0.6 % at 10
268 weeks (**Fig. 5c**). We observed corresponding increases of indel rates at the target site by split-
269 intein AAVs from 0.1 ± 0.0 % (2 weeks) to 0.4 ± 0.1 % (10 weeks). Utilization of the UDiTaS

270 unidirectional sequencing approach[39] with locus specific primers for library construction affords
271 the opportunity to assess the rate of large deletions or other types of genomic rearrangements in
272 an unbiased manner. UDiTaS analysis at the SERPINA1 transgene locus revealed primarily
273 precise editing among the modified alleles. As observed by the amplicon deep sequencing, a
274 fraction of the modified alleles contained indels and there were a small number of larger deletions
275 between the two nicking sites or extending beyond these sites, although many of the largest
276 deletions (>100 bp) did not meet the level of statistical significance (**Fig. 5d-e, Supplementary**
277 **Fig. 6 and Supplementary table 5**). There was also evidence of a very low rate of AAV insertion
278 at the target site (0.014% of total UMIs). Together, these results demonstrate that delivery of split-
279 intein PE by the dual AAV8 enables low rates of precise editing via the PE3 prime editing strategy
280 *in vivo*.
281
282 **Discussion**
283 Prime editing systems potentially provide a powerful approach for the template-directed
284 incorporation of a variety of types of alterations (nucleotide changes, insertions, deletions) into
285 genomic DNA sequence without relying on homology directed repair [6]. In principle, this provides
286 a strategy for the correction of a variety of different disorders, since prime editing should not be
287 dependent on the cell cycle for efficacy as is HDR[2]. Moreover, unlike HDR and MMEJ[40]
288 approaches for precise sequence insertions, prime editing does not require co-delivery of donor
289 DNA. However, the length of the sequence that can be inserted is limited by the length of the
290 encoded pegRNA. While there are many potential advantages to prime editing, the development
291 of prime editing systems is in its initial stages, and many questions with regard to the utility of this
292 system for genome editing remain to be addressed. This study provides an important
293 demonstration of the *in vivo* utility of prime editors for the template-based modification of genomic
294 sequence with implications for improving the utility of disease model systems and for the eventual
295 translation of this tool to the correction of pathogenic disorders.
296
297 The efficacy of genome editing systems is dependent on a number of factors, one of which is the
298 efficiency of nuclear import. In rapidly proliferating cells, the nuclear envelope provides only a
299 modest barrier to entry for genome editing tools. However, in post-mitotic or quiescent cells the
300 nuclear envelope may provide a greater barrier to the entry of Cas9-based systems, such that the
301 number and composition of the NLS sequences can impact editing efficacy [3, 12-16]. By incorporating
302 additional NLSs, we developed an improved PE (PE2*) that increases the efficiency of genome
303 editing across multiple endogenous sites relative to the original PE2. Importantly, the observed

304 improvements in genome editing for PE2* in cell culture translated to increased rates of genome
305 editing *in vivo*. Thus, demonstrating that NLS sequence composition and architecture is an
306 important parameter to consider in the design of prime editing systems to maximize *in vivo* efficacy,
307 as has been observed for other genome editing systems [3, 36].
308
309 This study provides an important proof-of-concept demonstration of the feasibility of *in vivo* editing
310 of somatic cells in mammalian systems by prime editing. We demonstrated the utility of prime
311 editing systems for two different types of applications: the correction of a pathogenic mutation
312 (AATD) and generation of animal cancer model. The ability to precisely correct a pathogenic
313 mutation or install somatic mutations *in vivo* has important implications for both gene therapy and
314 the development of new mouse models to study cancer and other disorders[41]. Using
315 hydrodynamic injection, PE2* shows higher editing efficiency (~6%, **Fig. 3**) than previously
316 published HDR (~0.5%)[42] in the mouse liver for installing oncogenic point mutations. Future work
317 will address whether PEs can simultaneously generate cooperating oncogenic point mutations (a
318 list of potential mutations that can be studied is in **Supplementary Table 4**).
319
320 Compared with HDR and base editors, prime editing provides a complementary method for
321 generating precise genome alterations. Although a PE is less efficient than base editors for
322 introducing base transitions, a PE can generate precise nucleotide substitutions in target sites
323 where the neighboring sequence composition could potentially be challenging for precise
324 conversion by base editing systems [6]. Both of our *in vivo* target sites have neighboring nucleotides
325 that would potentially be susceptible to conversion when targeted by a base editor (bystander
326 effects). In the case of the *SERPINA1* locus, these bystander changes would introduce missense
327 mutations that could have undesirable effects.
328
329 Prime editing systems can also produce undesired editing outcomes in some instances[11]. The
330 use of the PE2 system in cell culture systems produces primarily precise edits[6]. The rate of precise
331 edits can be increased through the use of an additional nicking sgRNA (PE3 strategy), but this
332 also results in the production of a low rate of indels within the genome[6, 11]. Prime editing in plant
333 protoplasts also produces a fraction of undesired editing outcomes when employing either the
334 PE2 and PE3 strategy[7]. Interestingly, in mouse zygotes the PE3 strategy produces alleles
335 containing the desired edit, but a large fraction also harbor deletions of various sizes between the
336 target site and the nicking site [11]. Reassuringly, for our *in vivo* PE3 editing the majority of modified
337 alleles contain the intended product without additional modifications. Similar to the observation in

338 cell culture experiments, we observe a small fraction of the alleles that contain unintended
339 changes (indels). A small proportion of these (~6.6% modified alleles) contain deletions between
340 the pegRNA and nicking sgRNA sites. Employing a sequential nicking strategy (PE3b) approach
341 instead of PE3 may reduce the indel rates *in vivo*, as observed in other systems[6, 7], when an
342 overlapping nicking RNA can be designed at the prime editor target site.
343
344 We established a dual-AAV system to deliver split-intein prime editors that produces precise
345 editing *in vivo* following a single injection. While we focused on liver editing, the dual AAV-
346 mediated prime editor developed in this study should be applicable to other organ systems.
347 Because AATD requires a high level of gene correction to generate sufficient wildtype protein for
348 disease amelioration[28], our initial AAV PE design is suboptimal for AATD treatment. Future work
349 will be required to increase PE efficiency *in vivo*. The modest editing efficiency of our AAV system
350 may be due in part to the low vector dose ($2x10^{11}$ vg/kg total). In addition, some aspects of our
351 design of split-PE AAV vectors could be improved. We retained the original PE2 architecture in
352 our split-intein design due to its more compact size for vector packaging – substitution of the SV40
353 NLSs with more efficient bipartite SV40[3] or c-myc NLSs[17] may improve *in vivo* activity. Optimizing
354 the position of intein insertion within the PE may also improve activity, as may alteration of the
355 AAV vector layout and promoters that are utilized [36]. The more compact SaPE2* or Sa^{KKH}PE2*
356 systems described herein may also have utility in the context of AAV delivery once they are further
357 optimized. In addition, the development of nanoparticle-mediated delivery or direct RNP delivery
358 could also provide alternate avenues for translation of this tool for *in vivo* therapeutic application
359 without the constraints of AAV cargo capacity. Once the efficiency of prime editing systems *in*
360 *vivo* is improved, it will also be important to explore in greater depth the frequency of off-target
361 editing before these systems are transitioned to gene therapy applications. In summary, our
362 findings demonstrate the broad potential scope of PEs for *in vivo* genome editing, both for disease
363 gene correction and the development of new cancer models in adult animals.
364
365 **Methods**
366 **Generation of plasmids**
367 To generate pegRNA expression plasmids, PCR products including spacer sequences, scaffold
368 sequences and 3' extension sequences were amplified with indicated primers (Supplementary
369 table 2) using Phusion master mix (ThermoFisher Scientifc) or Q5 High-Fidelity enzyme
370 (NewEnglandBioLabs), which were subsequently cloned into a custom vector (Supplementary
371 sequence 2, BfuAI and EcoR I digested) by the Gibson assembly method (NEB). To generate

372 nicking sgRNA expression plasmids, annealed oligos were cloned into BfuAI-digested vector or
373 pmd264 vector (Supplementary sequence 2). Sequences of all pegRNA and nicking sgRNAs are
374 listed in Supplementary Table 1. PE2* was generated through Gibson assembly, by combining
375 SpyCas9(H840A) and the M-MLV RT from PE2 [6] with additional NLS sequences by PCR and
376 inserting into a NotI/PmeI-digested pCMV-PE2 backbone (Supplementary sequence 3). SaPE2*
377 and Sa^{KKH}PE2* were generated through Gibson assembly, by combining the following three DNA
378 fragments: (i) PCR amplified M-MLV RT with additional NLS sequences from PE2, (ii) a
379 NotI/PmeI-digested PE2 backbone, (iii) a SaCas9 N580A nickase or a Sa^{KKH}-Cas9 nickase
380 (supplementary sequence 3). All plasmids used for *in vitro* experiments were purified using
381 Midiprep kit including endotoxin removal step (Qiagen). pCMV-PE2 was a gift from David Liu
382 (Addgene plasmid # 132775) [6]. AAV-PE-N was generated through Gibson assembly, by
383 combining the following five DNA fragments: (i) gBlock pegRNA driven by U6, (ii) gBlock nicking
384 sgRNA driven by U6, (iii) PCR amplified N-terminal PE2 (amino acid 1-713 of SpCas9 H840A),
385 (iv) gBlock split-intein N terminal, (v) a KpnI/SacI-digested AAV backbone[43]. AAV-PE-C was
386 generated through Gibson assembly, by combining the following four DNA fragments: (i) gBlock
387 split-intein C terminal, (ii) PCR amplified C-terminal PE2 (amino acid 714-1368 of SpCas9 H840A)
388 and M-MLV RT from PE2, (iii) gBlock β-globin poly(A) signal, (iv) a KpnI/NotI-digested AAV
389 backbone.
390
391 **AAV vector production**
392
393 AAV vectors (AAV8 capsids) were packaged at the Viral Vector Core of the Horae Gene Therapy
394 Center at the University of Massachusetts Medical School. Vector titers were determined by gel
395 electrophoresis followed by silver staining and qPCR.
396
397 **Generation reporter cells and Cell culture conditions**
398 HEK293T cells were purchased from ATCC, and cells were maintained in Dulbecco's Modified
399 Eagle's Medium supplemented with 10% FBS. For generation of the mCherry reporter and 47bp-
400 insertion TLR reporter cells, we created single-copy reporter cells using the Invitrogen Flp-In
401 system. Briefly, Flp-In 293T cells were maintained in DMEM, 10% Fetal bovine serum (FBS), 1%
402 pen-strep and 100 µg/ml Zeocin. $1x10^6$ Flp-In 293T cells were plated in a 6 well plate 24 hours
403 before transfection. On the day of transfection, the cells were washed and fresh media without
404 Zeocin was added. The plasmid coding for FLP recombinase and the mCherry reporter or 47bp-
405 insertion TLR reporter plasmid were transfected into the cells at a 9:1 ratio using polyfect
406 (QIAGEN) according to the manufacturer's protocol (900ng mCherry reporter or 47bp-insertion

407 TLR reporter plasmid and 100ng FLP recombinase plasmid to make 1 µg plasmid in total). 48
408 hours following transfection, the cells were washed and split into a 10cm dish with fresh media.
409 100 µg/ml of hygromycin was used to select for cells that contained an integration of the reporter
410 plasmid. Two weeks post selection, hygromycin resistant foci were pooled and propagated for
411 cryopreservation and further experiments. The construction an characterization of the TLR-MCV1
412 reporter cells was described in [23]. All cell types were maintained at 37°C and 5% CO_2 and were
413 tested negative for mycoplasma.
414
415 **Cell culture transfection/electroporation and DNA preparation**
416 For transfection-based editing experiments in HEK293T cells or HEK293T reporter cells, cells
417 were plated 100,000 per well on a 48-well plate. 24 hours later, the cells were co-transfected with
418 540ng of prime editor plasmid, 270ng of pegRNA plasmid and 90ng of Nicking sgRNA plasmid.
419 Lipofectamine 2000 (Invitrogen) was used for the transfection according to the manufacturer's
420 instructions. FACS analysis was performed 3 days after transfection in HEK293T reporter cells.
421 To detect editing efficiency in endogenous genomic loci, HEK293T cells were cultured for 3 days
422 after transfection, and genomic DNA was isolated using QIAamp DNA mini kit (QIAGEN)
423 according to the manufacturer's instructions. For *CCR5* PE editing, the same amount of plasmid
424 were delivered (540ng of prime editor plasmid, 270ng of pegRNA plasmid and 90ng of Nicking
425 sgRNA plasmid), where 2×10^5 HeLa cells were treated per electroporation using the Neon®
426 TransfectionSystem 10 L Kit (Thermo Fisher Scientific) with the recommended electroporation
427 parameters: Pulse voltage (1350 v), Pulse width (10 ms), Pulse number (3).
428
429 **Fluorescent reporter assay**
430 48 h post-transfection cells are trypsinized and harvested into a microcentrifuge tube. Cells are
431 centrifuged at 500×g for 2 min, washed once with 1× PBS, recentrifuged at 500×g for 2 min and
432 resuspended in 1× PBS for flow cytometry (Becton Dickonson FACScan). 10,000 events were
433 counted from each sample for FACS analysis. Experiments were performed in six replicates on
434 different days. The data are reported as mean values with error bars indicating SD.
435
436 **Immunofluorescence and immunohistochemistry**
437 HeLa and U2OS cells are transfected in six-well format via Lipofectamine 2000 (Invitrogen) using
438 the manufacturer's suggested protocol with 300 ng each PE expression plasmid and 150 ng of
439 each pegRNA expression plasmid on a cover slip. 48 h following transfection, transfection media
440 was removed, cells were washed with 1× PBS and fixed with 4% formaldehyde in 1× PBS for 15

441 min at room temperature. Following blocking (blocking solution: 2% BSA, 0.3% Triton X-100,
442 within 1× PBS), samples were stained with mouse antihemagglutinin (Sigma, H9658, 1:500), and
443 Alexa 488 donkey anti-mouse IgG (H+L; Invitrogen, A-21202, 1:2000), sequentially.
444 VECTASHIELD mounting medium with DAPI (Vector Laboratories, H-1200) was used to stain the
445 nuclei and to mount the samples on the slide. Images were taken with ZEISS LSM 710 Confocal
446 Microscope System. We used Fiji[44] and the "CMCI-EMBL" plugin to calculate the signal from the
447 nuclear or cytoplasm comparments as previously described[45]
448 (https://github.com/miura/NucleusRimIntensityMeasurementsV2/). We used 60 to 80 cells for
449 each treatment group to determine the nuclear to cytoplasmic ratio for each PE construct, which
450 was calculated as the intensity value from nucleus divided by the intensity value from cytoplasm
451 and nucleus.
452 IHC staining was performed as previously described[42] using beta-catenin antibody (BD, 610154,
453 1:100) or HA tag (CST, 3724, 1:400).
454
455 **Animal studies**
456 All animal experiments were authorized by the Institutional Animal Care and Use Committee
457 (IACUC) at UMASS medical school. No animals were excluded from the analyses. All prime editor
458 plasmids were prepared by EndoFreeMaxi kit (Qiagen) and were delivered through hydrodynamic
459 tail-vein injection. For PiZ correction, eight-week-old PiZ mice were injected with 2.3ml 0.9%
460 saline containing 30μg PE2 or PE2* (n=3), 15μg pegRNA (*SERPINA1*) and 5μg Nicking sgRNA
461 3. For cancer model generation, FVB/NJ (Strain #001800) was purchased from Jackson
462 Laboratories. Each FVB mouse (n=4) was injected with 2.3ml 0.9% saline containing 30μg PE2
463 or PE2*, 15μg pegRNA (*Ctnnb1*), 15μg Nicking sgRNA 2 (*Ctnnb1*), 5μg pT3 EF1a-MYC (a gift
464 from Xin Chen, Addgene plasmid # 92046)[33] and 1μg CMV-SB10 (a gift from Perry Hackett,
465 Addgene plasmid # 24551).
466
467 **Deep sequencing and data analysis**
468 Library construction for deep sequencing is modified from our previous report[27]. Briefly, 72 h after
469 transfection or electroporation, cells were harvested and genomic DNA was extracted with
470 GenElute Mammalian Genomic DNA Miniprep Kit (Sigma). Genomic loci spanning the target and
471 off-target sites were PCR amplified with locus-specific primers carrying tails complementary to
472 the Truseq adapters (Supplementary Data 3). 50 ng input genomic DNA was PCR amplified with
473 Q5 High-Fidelity DNA Polymerase (New England Biolabs): (98 °C, 15 s; 67 °C 25 s; 72 °C 20 s)
474 ×30 cycles. For the construction of the CCR5 UMI-based library, 50 ng input genomic DNA was

475 first linearly pre-amplified with 10 nM final concentration 5p-CCR5_UMI primer using the Q5 High-
476 Fidelity DNA Polymerase (New England Biolabs): (98 °C, 60 s; 67 °C, 25 s; 72 °C, 20 s) × 10
477 cycles. In the same reaction mix, 500 nM final concentration 5p-DS_constant and 3p-CCR5_DS
478 primers were added for further amplification (98 °C, 60 s; 67 °C, 25 s; 72 °C, 20 s) for 30 cycles.
479 Next, 0.1 µl of each PCR reaction was amplified with index-containing primers to reconstitute the
480 TruSeq adaptors using the Q5 High-Fidelity DNA Polymerase (New England Biolabs): (98 °C, 15
481 s; 67 °C, 25 s; 72 °C, 20 s) x10 cycles. Equal amounts of the PCR products from each
482 experimental condition (identified by different indices) were pooled and gel purified. The purified
483 library was deep sequenced using a paired-end 150 bp Illumina MiniSeq run.
484
485 MiniSeq data analysis was done as previously reported [27]. First, the quality of paired-end
486 sequencing reads (R1 and R2 fastq files) was assessed using FastQC
487 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Raw paired-end reads were
488 combined using paired end read merger (PEAR) (PMID: 24142950) to generate single merged
489 high-quality full-length reads. Reads were then filtered by quality (using Filter FASTQC (PMID:
490 20562416)) to remove those with a mean PHRED quality score under 30 and a minimum per
491 base score under 24. Each group of reads was then aligned to a corresponding reference
492 sequence using BWA (version 0.7.5) and SAMtools (version 0.1.19). To determine indel
493 frequency, size and distribution, all edited reads from each experimental replicate were combined
494 and aligned, as described above. Indel types and frequencies were then cataloged in a text output
495 format at each base using bam-readcount (https://github.com/genome/bam-readcount). For each
496 treatment group, the average background indel frequencies (based on indel type, position and
497 frequency) of the triplicate negative control group were subtracted to obtain the precise editing
498 and indel frequencies for each group. The fraction of precise editing is calculated as sequencing
499 reads with the desired allele editing/ all reads for the target locus. The results were concatenated
500 and loaded into GraphPad Prism 8.4 for data visualization.

501

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

503

504 For tagmentation, transposome was assembled as previously described[39] using purified Tn5
505 protein and oligonucleotides purchased from IDT. 200ng of genomic DNA was incubated with 2ul
506 of assembled transposome at 55 degree for 7 mins, and the product was cleaned up (20ul) with
507 a Zymo column (Zymo Research, #D4013). Tagmented DNA was used for the 1st PCR using
508 PlatinumTM SuperFi DNA polymerase (Thermo) with i5 primer and gene specific primers

509 (Supplementary table 3). Two different libraries were prepared for gDNA from each mouse with
510 different combinations of primers (i5+Locus_F [UDiTaS], i5+Locus_R [UDiTaS]). The i7 index was
511 added in the 2nd PCR and the PCR product was cleaned up with Ampure XP SPRI beads
512 (Agencourt, 0.9X reaction volume). Completed libraries were quantified by Tapestation and Qubit
513 (Agilent), pooled with equal mole and sequenced with 150 bp paired-end reads on an Illumina
514 MiniSeq instrument.
515
516 **UDiTaS data analysis**
517 The analysis pipeline was built using python code. Briefly, the analysis steps are as follows:
518 i) Demultiplexing. Raw BCL files were converted and demultiplexed using the appropriate
519 sequencing barcodes, allowing up to one mismatch in each barcode. Unique molecular
520 identifiers (UMIs) for each read were extracted for further downstream analysis.
521 ii) Trimming. Remove 3′ adapters using cutadapt, version 3.0
522 http://journal.embnet.org/index.php/embnetjournal/article/view/200/479
523 iii) Create reference sequence based the UDiTaS locus-specific primer position and AAV plasmid
524 map separately. Build index files for the reference using bowtie2-index[46], version 2.4.0.
525 iv) Alignment analysis. Paired reads were then globally aligned (end-to-end mode) to all the
526 reference amplicons using bowtie2's very sensitive parameter. Finally, samtools[47] (version
527 0.1.19) was used to create and index sorted bam files. Paired-end reads covering a window
528 between pegRNA targeting site and nicking sgRNA targeting sites were extracted and the
529 total number of unique UMIs were counted. Precise editing or small indels were analyzed as
530 previously described[27]. Pindel[48] (version 0.2.5b8) was used to detect breakpoints of large
531 deletions. Raw sequencing reads that align to the reference sequence were collapse to a
532 single read by common UMI and categorized as an exemplar for each UMI to a specific
533 category—for example, Wild Type, precise editing, small indel/substitution and Large
534 Deletions. Then the number of UMIs assigned per category is determined to define the ratio
535 of each event.
536 v) AAV integration. Extract the unmapped reads that did not locally align to the AAV/plasmid in
537 steps 3 and 4 using bedtools bamtofastq. With bowtie2, index the AAV plasmid sequence and
538 then do a local alignment of the reads. Of the reads that locally align to the AAV plasmid, first
539 filter out those reads which are directly adjacent to the UDiTaS primer (on read 2) and do not
540 contain any target locus sequence. This removes reads that are due to false priming. Of the
541 remaining reads, collapse these by UMI and count the UMIs. Classify the exemplar read for
542 each UMI as 'AAV/Plasmid Integrations'.

543
544 The software used for the above analysis is available at:
545 https://github.com/locusliu/GUIDESeq-Preprocess_from_Demultiplexing_to_Analysis
546 https://github.com/editasmedicine/uditas. https://github.com/ericdanner/REPlacE_Analysis.
547 and https://github.com/locusliu/PCR_Amplicon_target_deep_seq/blob/master/CRESA-lpp.py.

548 **Statistical analysis**

549 The fold changes of editing (precise editing or indels) are calculated between the corresponding
550 groups: pegRNA_only between PE2 and PE2*, or with specific Nicking sgRNA between PE2 and
551 PE2*. All raw data is listed in Supplementary Table 6 and statistical analyses were performed
552 using GraphPad Prism 8.4. Sample size was not pre-determined by statistical methods, but
553 rather, based on preliminary data. Group allocation was performed randomly. In all studies, data
554 represent biological replicates (n) and are depicted as mean ± s.d. as indicated in the figure
555 legends. Comparison of mean values was conducted with one-way ANOVA with Tukey's multiple
556 comparisons test, as indicated in the figure legends. In all analyses, P values < 0.05 were
557 considered statistically significant.
558

559 **Data availability**

560 Illumina Sequencing data have been submitted to the Sequence Read Archive. These datasets
561 are available under BioProject Accession number PRJNA496245.The authors declare that all
562 other data supporting the findings of this study are available within the paper and its
563 Supplementary Information files or upon reasonable request. Backbone plasmids used for
564 pegRNA and sgRNA cloning are available from addgene.
565

566 **Acknowledgements**

567 We thank C. Mello and P. Zamore for helpful discussions. We thank Y. Liu, J. Xie, and Q. Su in
568 the UMass Morphology and Viral Vector Cores for support. W.X. was supported by grants from
569 the National Institutes of Health (DP2HL137167, P01HL131471 and UG3HL147367), American
570 Cancer Society (129056-RSG-16-093), the Lung Cancer Research Foundation, and the Cystic
571 Fibrosis Foundation. P.Z., E.M., K.P. and S.A.W. were supported in part by the National Institutes
572 of Health (R01GM115911, and UG3TR002668) and the Rett Syndrome Research Trust. A.M. and
573 E.J.S. were supported by UG3TR002668.
574

575 **Author contributions**

576 PL and SQL performed the experiments, analyzed data and wrote the manuscript. CZ, EM, YGZ,
577 KP, and AM performed experiments. GG, EJS, TRF, SAW, and WX supervised the study and
578 wrote the manuscript with all co-authors.
579
580 **Competing interests**
581 The authors declare that they have no competing interests.
582
583 **Supplementary data accompanies this paper**
584

585 **Figure Legends:**

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

609

610 **Fig. 2 Improved PE2* increases editing efficiency at endogenous loci**. (a) Comparison of
611 editing efficiency for nucleotide substitution, targeted 3-bp deletion, and 6-bp insertion with PE2
612 and PE2* at *EMX1* locus in HEK293T cells. Indels broadly indicates mutations to the endogenous
613 sequence that do not result in the desired sequence alteration. (b-c) Editing efficiency for
614 nucleotide substitution, targeted 3-bp deletion, and 6-bp insertion with SaPE2* (b) and Sa^{KKH}PE2*
615 (c) at *EMX1* locus in HEK293T cells. (d) Sequence of CCR5 locus and pegRNA used for the 32bp
616 deletion. Two mutations in red were included to demonstrate that sequence collapse was not a
617 function of nuclease-induced microhomology mediated deletion and to reduce re-cutting of
618 deletion allele. Bottom panel shows the alignment of pegRNA with the *CCR5* sense strand. (e)

619 Comparison of efficiency for generating a targeted 32-bp deletion with PE2, PE2*, and Sa^{KKH}PE2*
620 within *CCR5* in HeLa cells. All expression vectors were delivered by transient transfection. The
621 presence of sgRNAs to promote nicking of the complementary strand is indicated in each figure
622 panel. Results were obtained from three independent experiments and presented as mean ± SD.
623 *P<0.05, **P<0.01, ***P<0.001, ****P<0.0001 by one-way ANOVA with Tukey's multiple
624 comparisons test. ns, not significant.
625
626 **Fig. 3 Improved PE2* increases the correction efficiency of a pathogenic mutation *in vivo*.**
627 (a) Installation (via G•C-to-A•T) of the pathogenic *SERPINA1* E342K mutation in HEK293T cells
628 using PE2, PE2*, and Sa^{KKH}PE2*. Editing efficiencies reflect sequencing reads which contain the
629 desired edit. The presence of sgRNAs to promote nicking of the complementary strand is
630 indicated on the x-axis. Results were obtained from three independent experiments and are
631 presented as mean ± SD. (b) pegRNA used for correction (via A•T-to-G•C) of the E342K mutation
632 includes a spacer sequence, a sgRNA scaffold, a RT template including edited bases (red) and
633 a primer-binding site (PBS). A PAM mutation (AGG to AAG) was introduced to reduce re-cutting
634 of the locus that results in a synonymous codon change. (c) Schematic overview of correction
635 strategy of the *SERPINA1* E342K mutation in PiZ transgenic mouse model of AATD. Prime editor,
636 pegRNA and nicking sgRNA plasmid were delivered by hydrodynamic tail-vein injection. (d)
637 Comparison of the efficiency of K342E correction and indels in mouse livers in PE2 or PE2*
638 treatment groups. Precise editing is defined as the fraction of sequencing reads with both A to G
639 prime editing and synonymous PAM modification. Results were obtained from three mice and
640 presented as mean ± SD. **P<0.01, ***P<0.001, ****P<0.0001 by one-way ANOVA with Tukey's
641 multiple comparisons test.
642
643 **Fig. 4 Generating mouse cancer models using improved PE2***. (a) pegRNA used for
644 installation (via C•G-to-T•A) of the oncogenic S45F in *Ctnnb1* in mouse liver. (b) Schematic
645 overview of the somatic cell editing strategy to drive tumor formation. Prime editor (PE2 or PE2*),
646 pegRNA for *Ctnnb1* S45F and nicking sgRNA plasmids were delivered by hydrodynamic tail-vein
647 injection along with the MYC transposon and transposase plasmids. (c) Representative images
648 of tumor burden in mouse liver with PE2 or PE2*. (d) Tumor numbers in the livers of mice 25 days
649 after injection with PE2 or PE2*. Control group was pegRNA only. (e) Sanger sequencing from
650 normal liver and representative tumors. The dashed box denotes C to T editing in tumors.
651 *P<0.05 by one-way ANOVA with Tukey's multiple comparisons test. (f) Schematic of *Ctnnb1* S45
652 deletion strategy using PE2* (S45del). pegRNA used for 3bp deletion (TCC) is shown. (g) PE2*

653 treatment leads to oncogenic activation of *Ctnnb1*. Prime editor (PE2*), pegRNA (*Ctnnb1* S45del
654 or SERPINA1) and nicking sgRNA plasmids were delivered by hydrodynamic tail-vein injection
655 along with the MYC transposon and transposase plasmids. Mice treated with the peg*Ctnnb1*
656 S45del (n=4) displayed a large number of liver tumors whereas mice treated with pegSERPINA1
657 as a control displayed no noticeable oncogenic lesions. beta-Catenin (CTNNB1) IHC staining was
658 performed. Scale bars: 100 μm (20X lens). (h) Prime editing efficiency and indels determined by
659 targeted deep sequencing in control liver and representative tumors.

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

681

682 **Supplementary Fig. 1 Optimized prime editor increases the degree of nuclear import.** (a)
683 Representative immunofluorescence images of prime editors transfected into U2OS cells
684 immunostained with a HA tag antibody to visualize subcellular localization. PE2 contains two
685 SV40 NLSs, whereas the PE2* variant include an extra N-terminal c-Myc NLS and a C-terminal

686 variant bipartite SV40 NLS (vBPSV40). DNA was stained with DAPI. Scale bar, 100μm. (b) The
687 nuclear/cytoplasm ratio of different prime editors as determined by confocal microscopy. (c-d)
688 Representative immunofluorescence images and quantification in HeLa cells. ***P<0.001 by one-
689 way ANOVA with Tukey's multiple comparisons test.
690
691 **Supplementary Fig. 2 Prime editing in reporter cells by PE2 and PE2*.** (a) Sequence of
692 reporter locus and pegRNA used for repair (via A•T-to-G•C transition) of mCherry reporter in
693 HEK293T cells. Bar above the cDNA indicates the stop codon with the target "t" for conversion
694 indicated in red. Two additional silent mutations are included to discourage recutting of the
695 repaired DNA sequence (b) Sequence of reporter and pegRNA used for generation of a 47bp
696 deletion to restore function of the GFP reporter in HEK293T cells. The bars above the cDNA
697 indicates three nucleotide blocks that correspond to codons in the GFP reporter. (c) Sequence of
698 reporter and pegRNA used for replacement of a 18bp element to restore function of the GFP
699 reporter. The bars above the cDNA indicates three nucleotide blocks that correspond to codons
700 in the GFP reporter. (d) Representative images of HEK293T reporter cells transfected with control,
701 ABEmax or PE2*. Scale bar, 400μm.
702
703 **Supplementary Fig.3 Sequencing analysis of the *CCR5* prime editing by PE2, PE2*, or**
704 **Sa^{KKH}PE2*.**
705 The percentage of most common sequences in PE2, PE2* and Sa^{KKH}PE2* transfected cells is
706 shown on the right (representative of n=3, determined by Illumina sequencing). The PE target site
707 is underlined. The PAM sequences are indicated in light blue for SpCas9 and green for $SaCas9^{KKH}$.
708 The box denotes the 32bp sequence deleted in CCR5delta32. Two "t" mutations in red were
709 included in the PE2 and PE2* RT template (Fig. 2d) to demonstrate that sequence collapse was
710 not a function of nuclease-induced microhomology mediated deletion and to reduce re-cutting of
711 deletion allele. Deleted bases are indicated by dashes and inserted bases are in blue.
712
713 **Supplementary Fig.4 Sequencing analysis of *SERPINA1* editing in the liver of PiZ mouse.**
714 (a) Evaluating prime editor expression in mouse liver. Left panel: FVB mice were injected with
715 30μg of control vector or PE2 plasmid with HA-tag. Livers were harvested at day 2 and IHC
716 staining were performed with an HA-tag antibody. Representative IHC images are shown. Scale
717 bars: 100 μm (20X lens). Right panel: Quantification of HA^+ cells. Numbers are mean ± sem (n=4
718 mice).

719 (b) The percentage of most common sequences in the liver of PE2 and PE2*-treated mice is
720 shown on the right (representative liver of n=3, determined by Illumina sequencing). The PE target
721 site is underlined. The PAM sequences are in light blue. Nucleotide substitutions are labeled in
722 red. Deleted bases are indicated by dashes. Inserted bases are shown in blue/lower case.
723
724 **Supplementary Fig.5 Liver tumors are positive for nuclear beta-Catenin**. (a), Representative
725 H&E and beta-catenin IHC staining in PE2 or PE2*-induced S45F tumors in Fig. 4. Scale bars:
726 100 μm (20X lens).
727
728 **Supplementary Fig. 6 Sequencing analysis of the *Ctnnb1* and *SERPINA1* prime editing by**
729 **PE2 or PE2*.**
730 (a), The percentage of most common sequences in the liver of dual AAV-treated PiZ mice
731 determined by locus amplification is shown on the right (representative liver of n=3). A portion of
732 the PE target site is underlined. The PAM sequences are in light blue. Nucleotide substitutions
733 are labeled in red. Deleted bases are indicated by dashes. (b), Length distribution of precise
734 editing and other indels at *SERPINA1 (K342E correction)* and *Ctnnb1* target sites by UDiTaS.
735 Also included are sequence modifications that may be associated with pegRNA scaffold
736 insertions[7]. A representative liver is shown (n=3 mice).
737
738 **References:**

739 1. Porteus, M.H. A New Class of Medicines through DNA Editing. *N Engl J Med* **380**, 947-
740 959 (2019).
741 2. Panier, S. & Boulton, S.J. Double-strand break repair: 53BP1 comes into focus. *Nat Rev*
742 *Mol Cell Biol* **15**, 7-18 (2014).
743 3. Suzuki, K. et al. In vivo genome editing via CRISPR/Cas9 mediated homology-
744 independent targeted integration. *Nature* **540**, 144-149 (2016).
745 4. Pickar-Oliver, A. & Gersbach, C.A. The next generation of CRISPR-Cas technologies
746 and applications. *Nat Rev Mol Cell Biol* **20**, 490-507 (2019).
747 5. Rees, H.A. & Liu, D.R. Base editing: precision chemistry on the genome and
748 transcriptome of living cells. *Nature Reviews Genetics* **19**, 770-788 (2018).
749 6. Anzalone, A.V. et al. Search-and-replace genome editing without double-strand breaks or
750 donor DNA. *Nature* **576**, 149–157 (2019).
751 7. Lin, Q. et al. Prime genome editing in rice and wheat. *Nat Biotechnol* **38**, 582-585
752 (2020).
753 8. Schene, I.F. et al. Prime editing for functional repair in patient-derived disease models.
754 *Nature Communications* **11**, 5352 (2020).
755 9. Geurts, M.H. et al. Evaluating CRISPR-based Prime Editing for cancer modeling and
756 CFTR repair in intestinal organoids. *bioRxiv*, 2020.2010.2005.325837 (2020).

757 10. Liu, Y. et al. Efficient generation of mouse models with the prime editing system. *Cell*
758 *Discov* **6**, 27 (2020).
759 11. Aida, T. et al. Prime editing primarily induces undesired outcomes in mice. *bioRxiv*,
760 2020.2008.2006.239723 (2020).
761 12. Wu, Y. et al. Highly efficient therapeutic gene editing of human hematopoietic stem
762 cells. *Nat Med* **25**, 776-783 (2019).
763 13. Koblan, L.W. et al. Improving cytidine and adenine base editors by expression
764 optimization and ancestral reconstruction. *Nat Biotechnol* (2018).
765 14. Zafra, M.P. et al. Optimized base editors enable efficient editing in cells, organoids and
766 mice. *Nat Biotechnol* (2018).
767 15. Song, C.-Q. et al. Adenine base editing in an adult mouse model of tyrosinaemia. *Nature*
768 *Biomedical Engineering* **4**, 125-130 (2020).
769 16. Walton, R.T., Christie, K.A., Whittaker, M.N. & Kleinstiver, B.P. Unconstrained genome
770 targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* (2020).
771 17. Ray, M., Tang, R., Jiang, Z. & Rotello, V.M. Quantitative tracking of protein trafficking
772 to the nucleus using cytosolic protein delivery by nanoparticle-stabilized nanocapsules.
773 *Bioconjugate chemistry* **26**, 1004-1007 (2015).
774 18. Makkerh, J.P., Dingwall, C. & Laskey, R.A. Comparative mutagenesis of nuclear
775 localization signals reveals the importance of neutral and acidic amino acids. *Curr Biol* **6**,
776 1025-1027 (1996).
777 19. Ran, F.A. et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**,
778 186-191 (2015).
779 20. Kleinstiver, B.P. et al. Broadening the targeting range of Staphylococcus aureus
780 CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* **33**, 1293-1298 (2015).
781 21. Certo, M.T. et al. Tracking genome engineering outcome at individual DNA breakpoints.
782 *Nature methods* **8**, 671-676 (2011).
783 22. Mir, A. et al. Heavily and fully modified RNAs guide efficient SpyCas9-mediated
784 genome editing. *Nat Commun* **9**, 2641 (2018).
785 23. Iyer, S. et al. Efficient Homology-directed Repair with Circular ssDNA Donors. *bioRxiv*,
786 864199 (2019).
787 24. Dean, M. et al. Genetic restriction of HIV-1 infection and progression to AIDS by a
788 deletion allele of the CKR5 structural gene. Hemophilia Growth and Development Study,
789 Multicenter AIDS Cohort Study, Multicenter Hemophilia Cohort Study, San Francisco
790 City Cohort, ALIVE Study. *Science* **273**, 1856-1862 (1996).
791 25. Liu, R. et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some
792 multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367-377 (1996).
793 26. Samson, M. et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant
794 alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722-725 (1996).
795 27. Bolukbasi, M.F. et al. Orthogonal Cas9-Cas9 chimeras provide a versatile platform for
796 genome editing. *Nat Commun* **9**, 4856 (2018).
797 28. Loring, H.S. & Flotte, T.R. Current status of gene therapy for alpha-1 antitrypsin
798 deficiency. *Expert Opin Biol Ther* **15**, 329-336 (2015).
799 29. Carlson, J.A. et al. Accumulation of PiZ alpha 1-antitrypsin causes liver damage in
800 transgenic mice. *J Clin Invest* **83**, 1183-1190 (1989).
801 30. Gaudelli, N.M. et al. Programmable base editing of A*T to G*C in genomic DNA
802 without DNA cleavage. *Nature* **551**, 464-471 (2017).

803 31. Liu, F., Song, Y. & Liu, D. Hydrodynamics-based transfection in animals by systemic
804 administration of plasmid DNA. *Gene Ther* **6**, 1258-1266 (1999).
805 32. Zucman-Rossi, J., Villanueva, A., Nault, J.C. & Llovet, J.M. Genetic landscape and
806 biomarkers of hepatocellular carcinoma. *Gastroenterology* **149**, 1226-1239 (2015).
807 33. Liu, P. et al. A functional mammalian target of rapamycin complex 1 signaling is
808 indispensable for c-Myc-driven hepatocarcinogenesis. *Hepatology* **66**, 167-181 (2017).
809 34. Marquardt, J.U., Andersen, J.B. & Thorgeirsson, S.S. Functional and genetic
810 deconstruction of the cellular origin in liver cancer. *Nat Rev Cancer* **15**, 653-667 (2015).
811 35. Wang, D.A.-O., Tai, P.A.-O. & Gao, G.A.-O. Adeno-associated virus vector as a
812 platform for gene therapy delivery.
813 36. Levy, J.M. et al. Cytosine and adenine base editing of the brain, liver, retina, heart and
814 skeletal muscle of mice via adeno-associated viruses. *Nat Biomed Eng* **4**, 97-110 (2020).
815 37. Zettler, J., Schütz, V. & Mootz, H.D. The naturally split Npu DnaE intein exhibits an
816 extraordinarily high rate in the protein trans-splicing reaction. *FEBS Lett* **583**, 909-914
817 (2009).
818 38. Wright, A.V. et al. Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci*
819 *U S A* **112**, 2984-2989 (2015).
820 39. Giannoukos, G. et al. UDiTaS™, a genome editing detection method for indels and
821 genome rearrangements. *BMC genomics* **19**, 212 (2018).
822 40. Yao, X. et al. CRISPR/Cas9 - Mediated Precise Targeted Integration In Vivo Using a
823 Double Cut Donor with Short Homology Arms. *EBioMedicine* **20**, 19-26 (2017).
824 41. Maddalo, D. et al. In vivo engineering of oncogenic chromosomal rearrangements with
825 the CRISPR/Cas9 system. *Nature* **516** (2014).
826 42. Xue, W. et al. CRISPR-mediated direct mutation of cancer genes in the mouse liver.
827 *Nature* **514**, 380-385 (2014).
828 43. Yin, H. et al. Therapeutic genome editing by combined viral and non-viral delivery of
829 CRISPR system components in vivo. *Nat Biotechnol* **34**, 328-333 (2016).
830 44. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat*
831 *Methods* **9**, 676-682 (2012).
832 45. Miura, K. Measurements of Intensity Dynamics at the Periphery of the Nucleus. In:
833 Miura K., Sladoje N. (eds) Bioimage Data Analysis Workflows. Learning Materials in
834 Biosciences. Springer, Cham. (2020).
835 46. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods*
836 **9**, 357-359 (2012).
837 47. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
838 2078-2079 (2009).
839 48. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth
840 approach to detect break points of large deletions and medium sized insertions from
841 paired-end short reads. *Bioinformatics* **25**, 2865-2871 (2009).
842

Figure 1

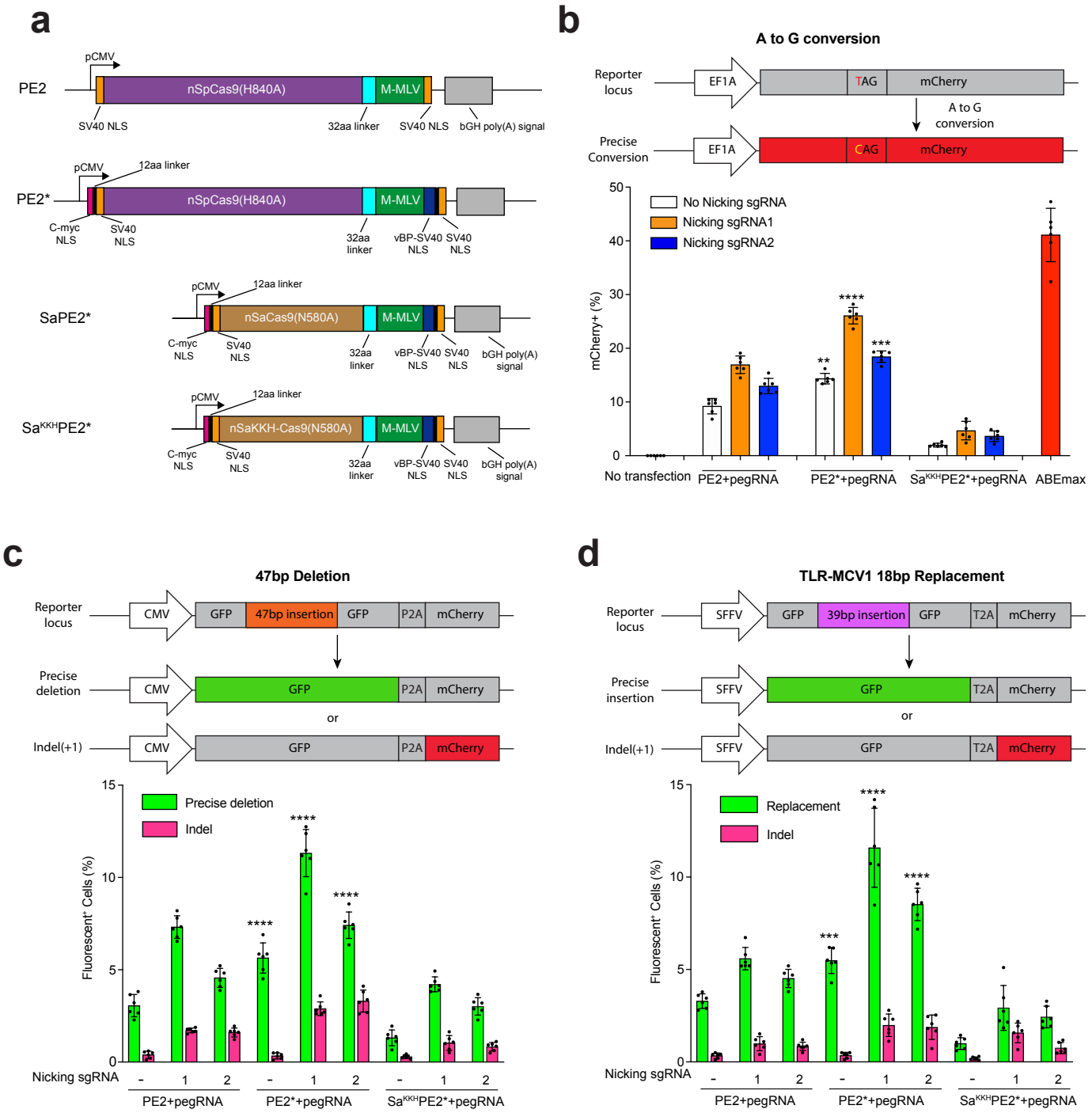

a
PE2
pCMV
nSpCas9(H840A)
M-MLV
SV40 NLS
32aa linker
SV40 NLS
bGH poly(A) signal
PE2*
12aa linker
C-myc NLS
SV40 NLS
vBP-SV40 NLS
SaPE2*
nSaCas9(N580A)
SaKKHPE2*
nSaKKH-Cas9(N580A)
b
A to G conversion
Reporter locus
EF1A
TAG
mCherry
A to G conversion
Precise Conversion
CAG
No Nicking sgRNA
Nicking sgRNA1
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

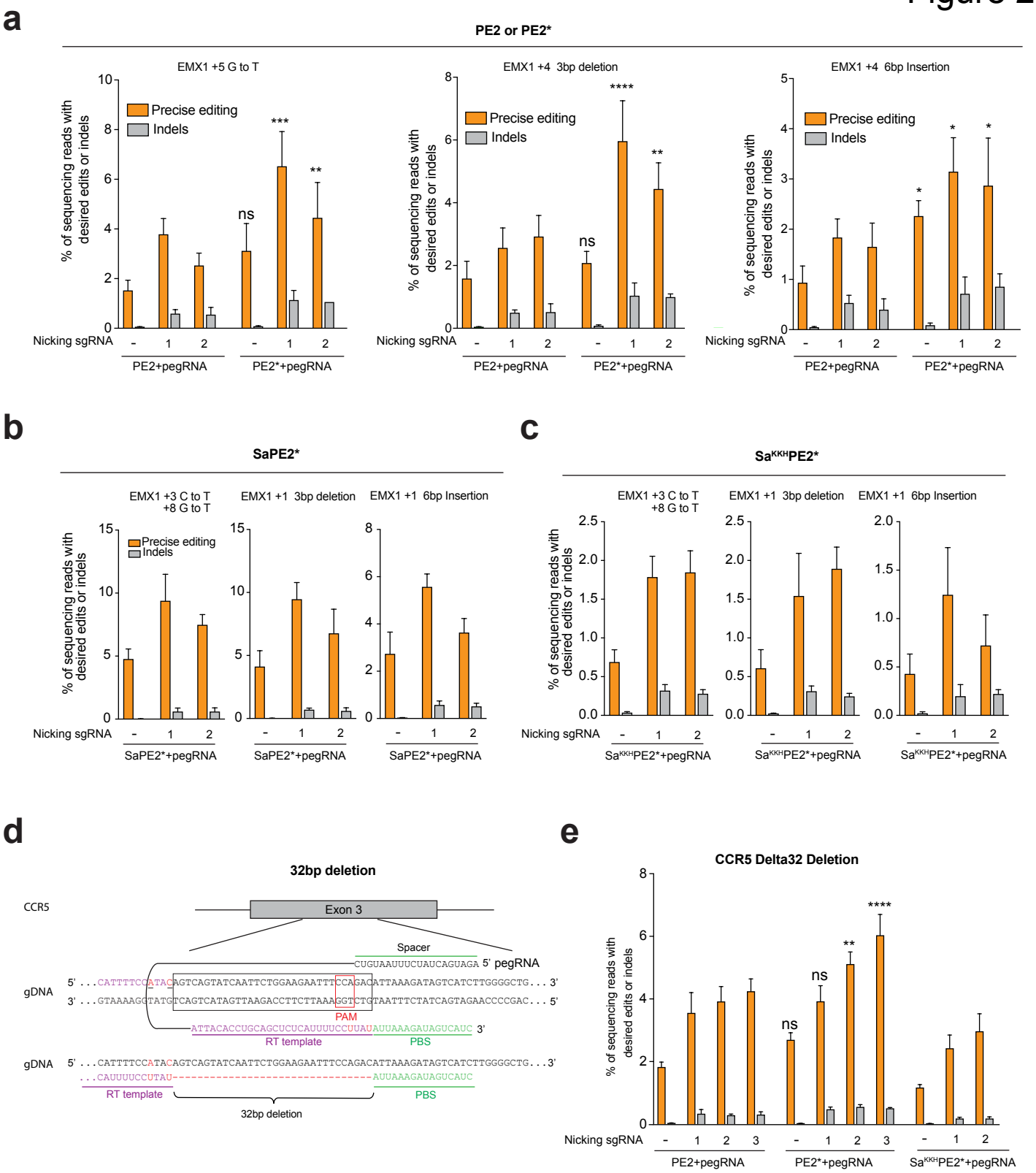

a
PE2 or PE2*
EMX1 +5 G to T
EMX1 +4 3bp deletion
EMX1 +4 6bp Insertion
Precise editing
Indels
% of sequencing reads with desired edits or indels
Nicking sgRNA
PE2+pegRNA
PE2*+pegRNA
b
SaPE2*
EMX1 +3 C to T +8 G to T
EMX1 +1 3bp deletion
EMX1 +1 6bp Insertion
SaPE2*+pegRNA
c
SaKKHPE2*
SaKKHPE2*+pegRNA
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

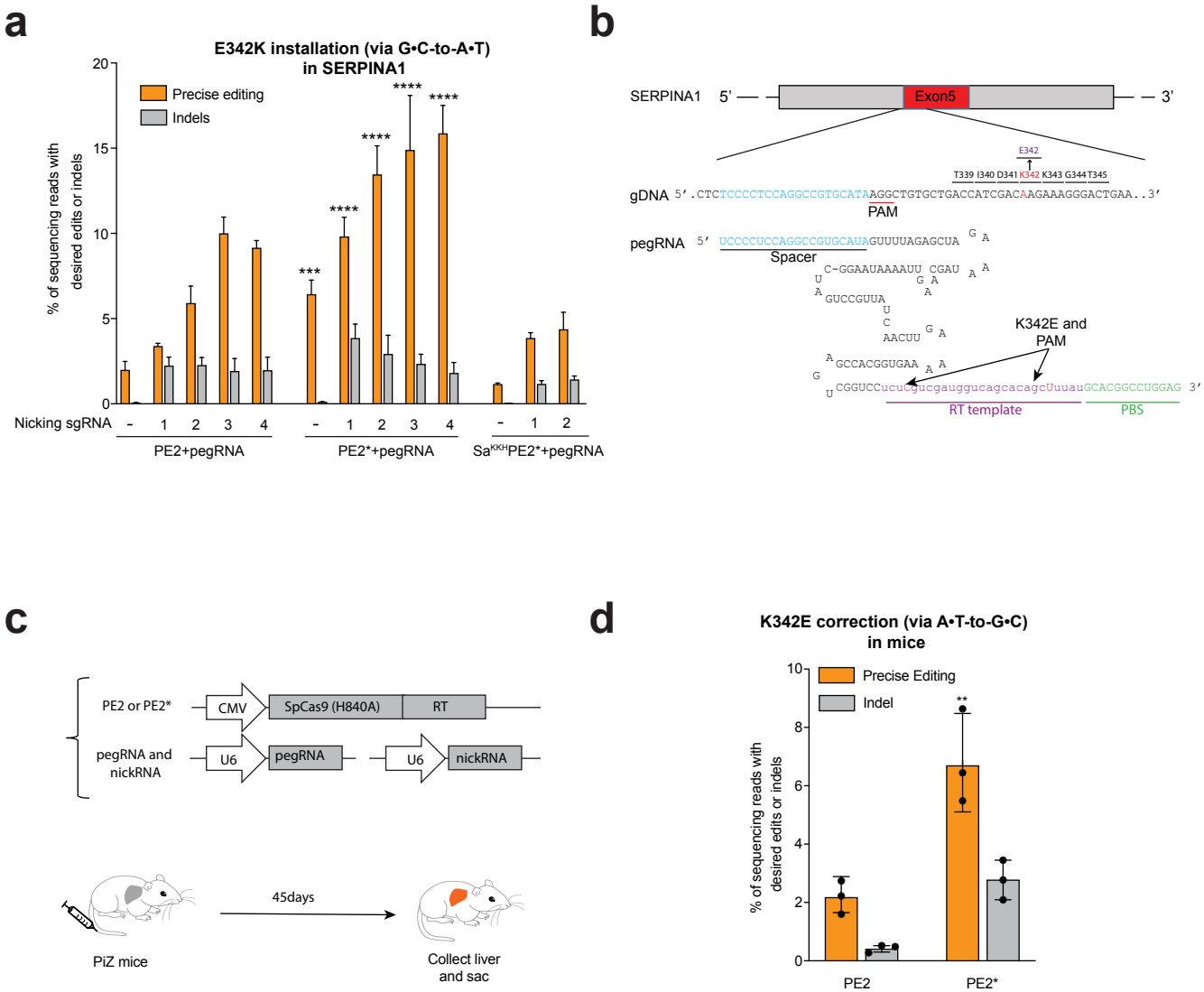

a
E342K installation (via G•C-to-A•T) in SERPINA1
Precise editing
Indels
% of sequencing reads with desired edits or indels
Nicking sgRNA
PE2+pegRNA
PE2*+pegRNA
Sa^KKH PE2*+pegRNA
b
SERPINA1
Exon5
gDNA
PAM
pegRNA
Spacer
K342E and PAM
RT template
PBS
c
PE2 or PE2*
CMV
SpCas9 (H840A)
RT
pegRNA and nickRNA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

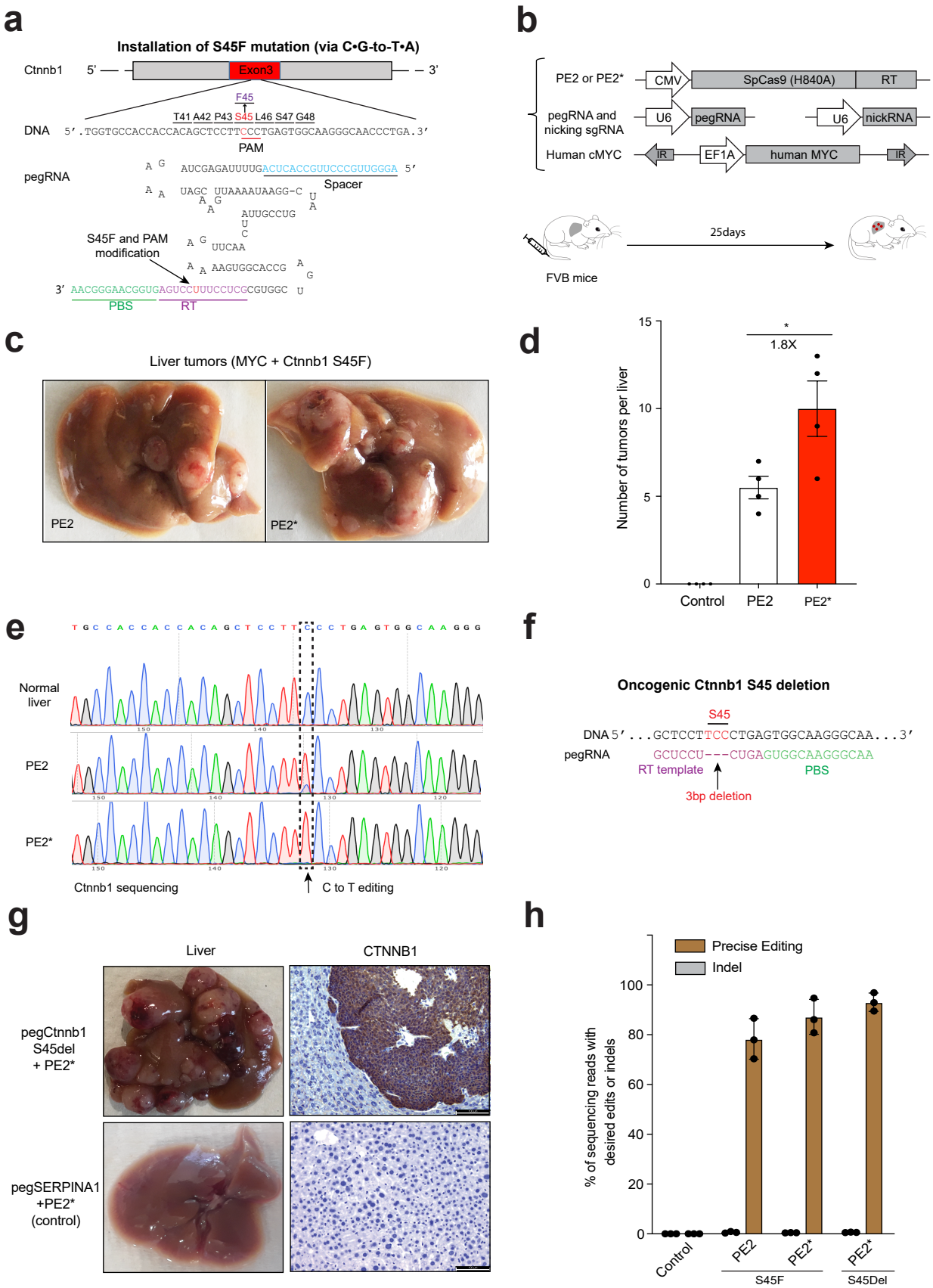

a
Installation of S45F mutation (via C•G-to-T•A)
Ctnnb1
5'
Exon3
3'
F45
T41 A42 P43 S45 L46 S47 G48
DNA
5'.TGGTGCCACCACCACAGCTCCTTCCCTGAGTGGCAAGGGCAACCCTGA.3'
PAM
pegRNA
AUCGAGAUUUUGACUCACCGUUCCCGUUGGGA 5'
Spacer
UAGC UUAAAAUAAGG-C
AUUGCCUG
UUCAA
AAGUGGCACCG
S45F and PAM modification
3' AACGGGAACGGUGAGUCCUUUCCUCGCGUGGC
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

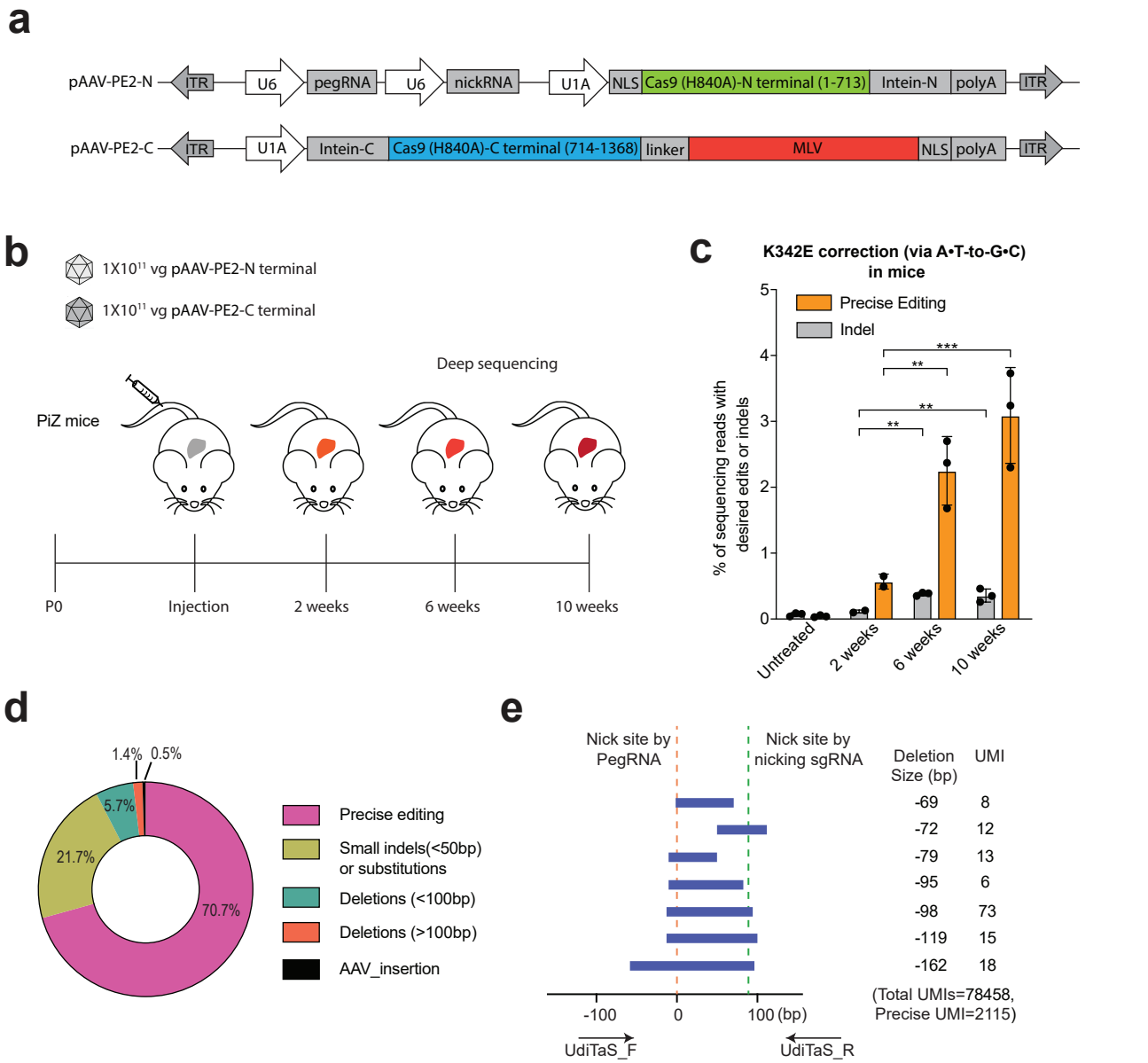

a
pAAV-PE2-N
ITR
U6
pegRNA
U6
nickRNA
U1A
NLS
Cas9 (H840A)-N terminal (1-713)
Intein-N
polyA
ITR
pAAV-PE2-C
ITR
U1A
Intein-C
Cas9 (H840A)-C terminal (714-1368)
linker
MLV
NLS
polyA
ITR
b
1X10¹¹ vg pAAV-PE2-N terminal
1X10¹¹ vg pAAV-PE2-C terminal
Deep sequencing
PiZ mice
P0
Injection
2 weeks
6 weeks
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

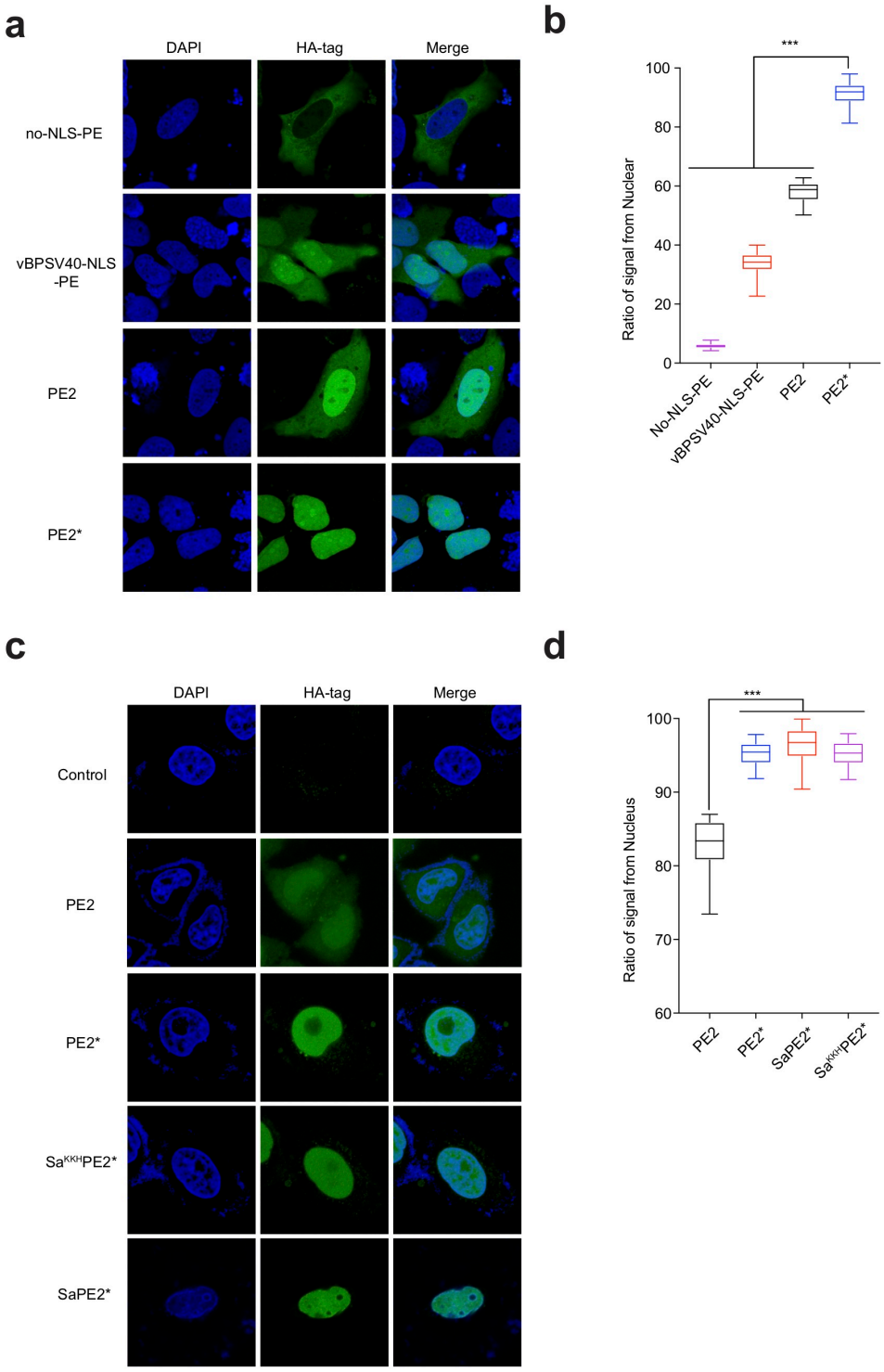

a
DAPI
HA-tag
Merge
no-NLS-PE
vBPSV40-NLS
-PE
PE2
PE2*
b
***
Ratio of signal from Nuclear
100
80
60
40
20
0
No-NLS-PE
vBPSV40-NLS-PE
PE2
PE2*
c
DAPI
HA-tag
Merge
Control
PE2
PE2*
SaKKHPE2*
SaPE2*
d
***
Ratio of signal from Nucleus
100
90
80
70
60
PE2
PE2*
SaPE2*
SaKKHPE2*

Figure S2

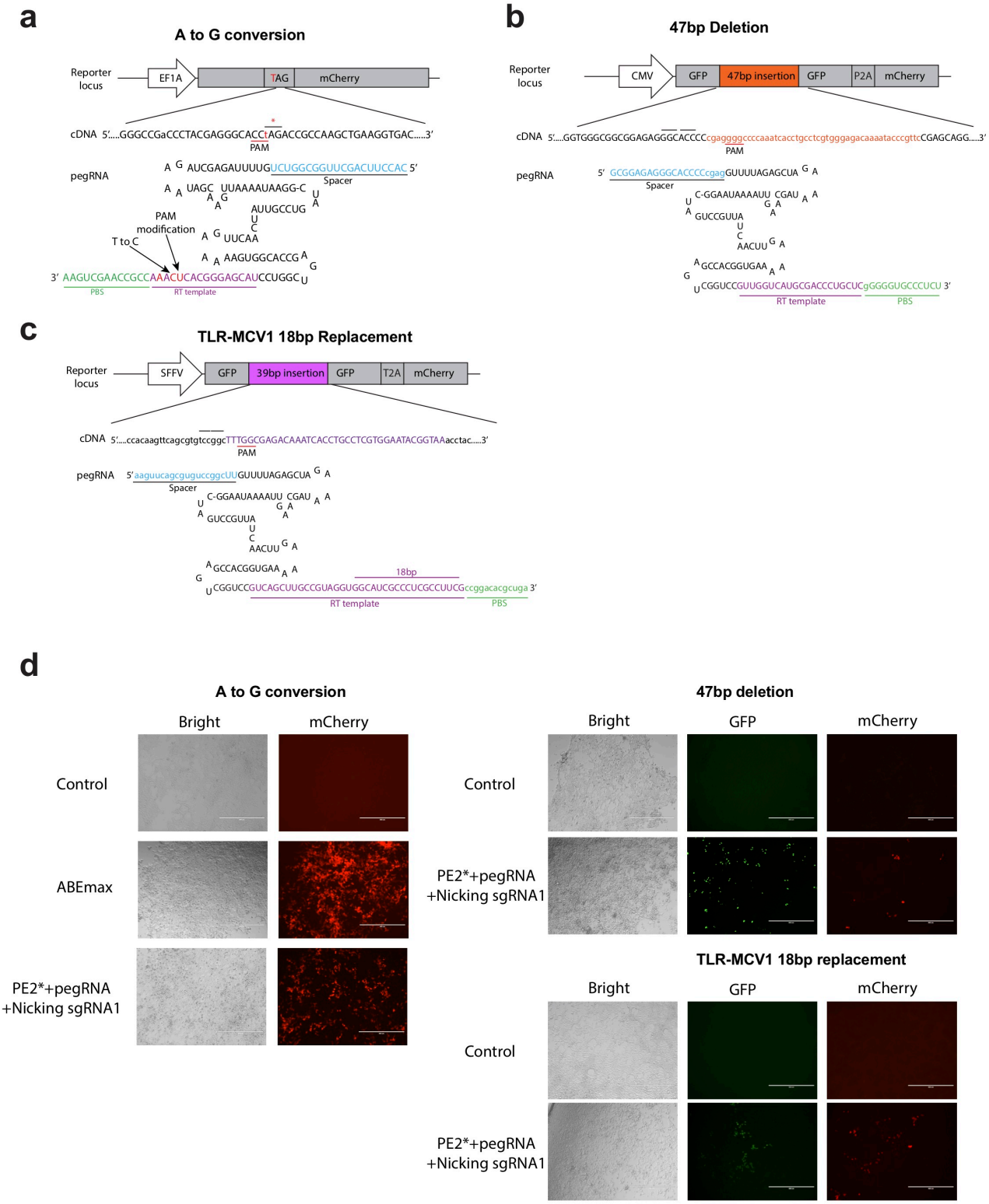

a
A to G conversion
Reporter locus
EF1A
TAG
mCherry
cDNA 5'....GGGCCGaCCCTACGAGGGCACCtAGACCGCCAAGCTGAAGGTGAC.....3'
PAM
pegRNA
AUCGAGAUUUUGUCUGGCGGUUCGACUUCCAC 5'
Spacer
UAGC UUAAAAUAAGG-C
AUUGCCUG
UUCAA
AAGUGGCACCG
PAM modification
T to C
3' AAGUCGAACCGCCAAACUCACGGGAGCAUCCUGGC
PBS
RT template
b
47bp 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*

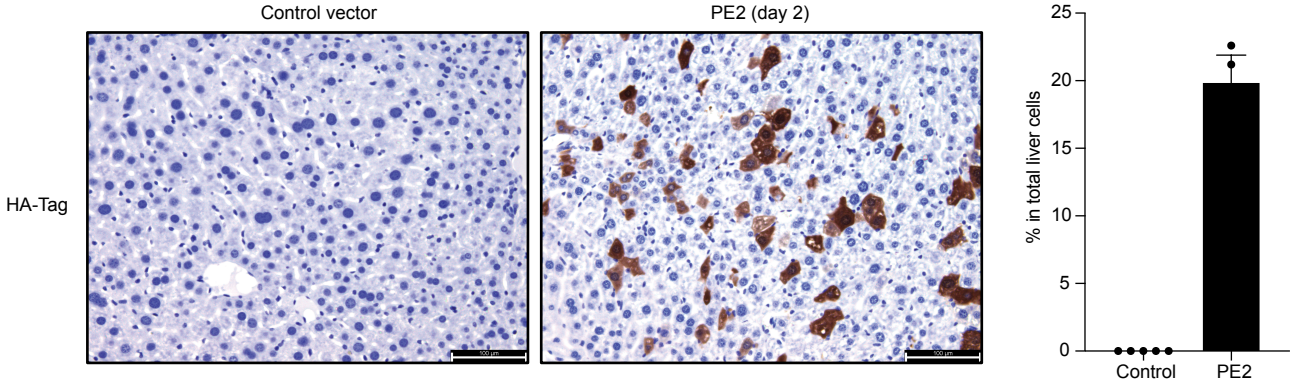
```
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCATACAGTCAGTATCAATTCTGGAAGAATTTCCAGACATTAAAGATAGTCATCTTGGGGCTGGTCC      WT        91.25%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCtTAtA--------------------------------TTAAAGATAGTCATCTTGGGGCTGGTCC      Precise    6.08%
CAAAAAGAAGGTCTTCATTACACCTGCAGCTCTCATTTTCCA------------------------------GACATTAAAGATAGTCATCTTGGGGCTGGTCC      -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

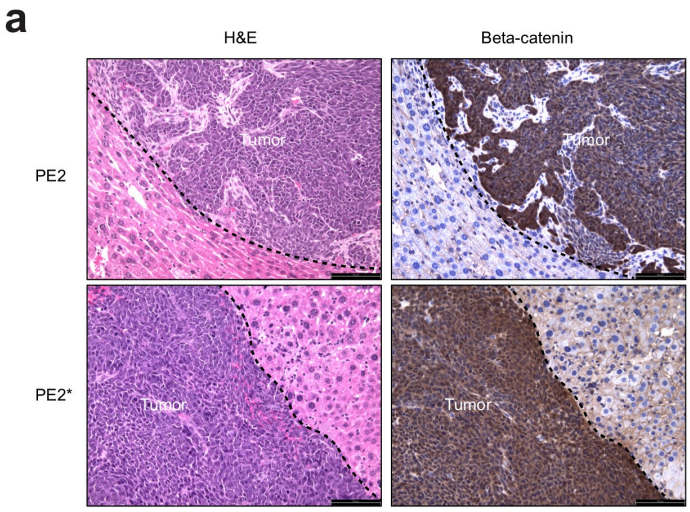


Figure S6

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
```