

# Human cells based directed evolution of adenine base editors with improved efficiency

**Junhao Fu**

Wenzhou Medical University

**Qing Li**

State Key Laboratory of Cell Biology, Shanghai Key Laboratory of Molecular Andrology, CAS Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell Biology

<https://orcid.org/0000-0002-6392-9699>

**Xiaoyu Liu**

Wenzhou Medical University

**Tianxiang Tu**

Affiliated Eye Hospital of Wenzhou Medical University

**Xiujuan Lv**

Affiliated Eye Hospital of Wenzhou Medical College

**Xidi Yin**

University of Chinese Academy of Sciences

**Zongming Song**

Affiliated Eye Hospital of Wenzhou Medical University <https://orcid.org/0000-0001-5869-7312>

**Jinwei Zhang**

National Institutes of Health <https://orcid.org/0000-0002-2114-173X>

**Jinsong Li**

State Key Laboratory of Cell Biology, Shanghai Institute of Biochemistry and Cell Biology, Center for Excellence in Molecular Cell Science, Chinese Academy of Sciences <https://orcid.org/0000-0003-3456-662X>

**Feng Gu** (✉ [gufenguw@gmail.com](mailto:gufenguw@gmail.com))

Affiliated Eye Hospital of Wenzhou Medical University

---

## Article

**Keywords:** ABE, Variants, Activity, Directed evolution

**Posted Date:** May 6th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-418740/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Nature Communications on October 8th, 2021. See the published version at <https://doi.org/10.1038/s41467-021-26211-0>.

1 **Human cells based directed evolution of adenine base editors**  
2 **with improved efficiency**  
3

4 Junhao Fu<sup>1&</sup>, Qing Li<sup>2&</sup>, Xiaoyu Liu<sup>1&</sup>, Tianxiang Tu<sup>1</sup>, Xiujuan Lv<sup>1</sup>, Xidi Yin<sup>2</sup>, Zongming Song<sup>1,3</sup>,  
5 Jinwei Zhang<sup>4</sup>, Jinsong Li<sup>2\*</sup>, and Feng Gu<sup>1\*</sup>

6 <sup>1</sup>School of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, State Key  
7 Laboratory and Key Laboratory of Vision Science, Ministry of Health and Zhejiang Provincial Key  
8 Laboratory of Ophthalmology and Optometry, Wenzhou, Zhejiang, China

9 <sup>2</sup>State Key Laboratory of Cell Biology, Shanghai Key Laboratory of Molecular Andrology, CAS  
10 Center for Excellence in Molecular Cell Science, Shanghai Institute of Biochemistry and Cell  
11 Biology, Chinese Academy of Sciences, Shanghai, China.

12 <sup>3</sup>Henan Eye Hospital, Henan Eye Institute, Henan Provincial People's Hospital and People's  
13 Hospital of Zhengzhou University, Zhengzhou, Henan 450052, China.

14 <sup>4</sup>Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases,  
15 National Institutes of Health, Bethesda, Maryland, USA.

16  
17 <sup>&</sup>These authors contributed equally to this work.

18 **\* Correspondence:**

19 Feng Gu, **E-mail:** [gufenguw@gmail.com](mailto:gufenguw@gmail.com), Jinsong Li, **E-mail:** [jsli@sibcb.ac.cn](mailto:jsli@sibcb.ac.cn)  
20  
21  
22  
23  
24  
25

## Abstract

Adenine base editors (ABE) are novel genome-editing tools that have been harnessed to introduce precise A•T to G•C conversion in genomic DNA. However, the low activity of ABE remains a major bottleneck that precludes efficacious applications. To address this limitation, we developed a directional screening system in human cells to evolve the deaminase component of the ABE, and identified three high-activity NG-ABEmax variants: NG-ABEmax-SGK (R101S/D139G/E140K), NG-ABEmax-R (Q154R) and NG-ABEmax-K (N127K). With further engineering, we created a new, consolidated variant [NG-ABEmax-KR (N127K/Q154R)] which exhibited superior editing activity both in human cells and in mouse disease models, compared to the original NG-ABEmax. We also found that NG-ABEmax-KR efficiently introduce natural mutations in gamma globin gene promoters with more than four-fold increase in editing activity. This work provides a broadly applicable, rapidly deployable platform to directionally screen and evolve novel, user-specified traits in base editors that extend beyond augmented editing activity.

**Key words:** ABE; Variants; Activity; Directed evolution

## 57 Introduction

58 Recent advances in genome engineering technologies based on CRISPR-Cas9 are enabling  
59 the systematic interrogation of mammalian genome functions and numerous targeted medical  
60 applications<sup>1-3</sup>. Cas9-triggered double-strand DNA break (DSB) at a target sequence in the  
61 genome is subsequently repaired by non-homologous end joining (NHEJ),  
62 microhomology-mediated end joining (MMEJ), or homology-directed repair (HDR)<sup>1,4</sup>.  
63 CRISPR-Cas9 incurs toxicity because it generates DSBs and off-target mutations that are more  
64 prone to occur when a nuclease-induced DSB is present<sup>5</sup>. Base editing is a genome-editing  
65 method that directly generates precise point mutations without creating DSBs<sup>7</sup>. As base editors do  
66 not produce DSBs, they minimize the off-target effect, compared with CRISPR-Cas9 mediated  
67 editing<sup>2</sup>.

68 The current base editing technology can be classified into DNA and RNA base editors. As to  
69 the first one, two classes of DNA base editor have been reported: cytosine base editors (CBE)  
70 which convert a C•G base pair into a T•A pair, and adenine base editors (ABE) which convert an  
71 A•T base pair into a G•C pair<sup>2,3</sup>. CBE and ABE can mediate all four possible transition mutations (C  
72 to T, A to G, T to C and G to A). Base editors have been successfully implemented in various  
73 organisms including human cells, animals and plants<sup>6-12</sup>. Nevertheless, the gene-editing utility of  
74 base editors is hampered by low efficiency and the off-target effect<sup>13,14</sup>. ABE are particularly limited  
75 due to its lower efficiency compared with CBE<sup>15</sup>.

76 To overcome these limitations, several strategies have been adopted to increase the activity of  
77 ABE. These include optimizing the codon usage of ABE, modifying the location and number of  
78 nuclear localization signals (NLS), reconstructing ancestral deaminase component and using  
79 different forms of sgRNA<sup>10,16-19</sup>. These extensive efforts reflect an urgent need to optimize ABE as  
80 an effective gene-editing tool. Additional questions also need to be addressed, including the  
81 control of the editing window for improved target site selection. By introducing F148A mutation in  
82 both the evolved TadA domain (TadA\*) and the wild-type TadA domain (TadA), engineered ABE  
83 possessed the higher fidelity without RNAase editing side effect, also its editing window has been  
84 narrowed<sup>20</sup>. However, for certain sites, such as when the target base exceeds the canonical  
85 editing window relative to the available PAM (protospacer adjacent motif) or the gene is inactivated  
86 by introducing premature termination codons, a wider editing window may be more desired. Here,

with directed evolution in human cells, we sought to establish a general screening platform to engineer NG-ABEmax to identify variants with special traits, i.e., enhanced activity.

## Results

### Generation of a high-throughput *EGFP*-based screening system for adenine base editors

To engineer and augment the existing adenine base editors, we first sought to generate a human cell-based reporter system that allows for rapid and accurate determination of ABE activity and specificity. We previously established an *EGFP*-reporter system in human cells to evaluate the efficiency of CRISPR-Cas9 and CRISPR-Cas12a<sup>21-27</sup>. Herein, when the sequence-specific RNA-guided endonuclease is targeted to the coding region of *EGFP*, it may generate frameshift indel mutations that lead to loss of fluorescence. However, such a screening system is ineffectual to evaluate ABE mutants, because the editing of one or more adenosines to guanosines at different regions of the *EGFP* sequence leads to diverse local protein alterations that can exhibit highly variable effects on *EGFP* folding, structure, and consequently its fluorescence, thus confounding analysis by cytometry.

To create a screening system where base editing completely ablates, rather than unpredictably alters *EGFP* activity, we first tested if ABE editing of a tryptophan TGG codon in *EGFP* (W58) to an amber stop codon TAG inactivates *EGFP* by truncating the protein near its N-terminus<sup>10</sup> (Supplementary Fig. 1a). We observed that the green fluorescence of *EGFP* was abolished by this conversion, suggesting feasibility of this approach (Supplementary Fig. 1b). To avoid any potentially confounding effects of editing neighboring non-target adenosines that flank the W58 site, we then sought to replace all adjacent adenosines with other nucleobases while maintaining *EGFP* activity (Fig. 1a). While the 13 nucleotides upstream of W58 contained no adenosines, the 18 nucleotides downstream of W58 harbor three adenosines, all part of ACC codons encoding Threonine. By mutating the three ACC codons to TCG, we introduced three conservative substitutions: T60S, T63S and T64S, and found that the mutations do not appreciably impact *EGFP* fluorescence (Fig. 1a-c). Lastly, we created a 5'-TG-3' PAM (protospacer adjacent motif) site necessary for Cas9 recognition through a silent mutation (CTC-to-CTG coding for L61). We term this engineered *EGFP* target gene that harbors all four mutations "*EGFP* variant", which exhibited comparable *EGFP* expression levels as the wild-type (Fig. 1a-c).

Base editors such as ABE exhibit variable editing efficiencies at different adenosine sites within the editing window<sup>3</sup>. To measure the positional effects on editing efficiency and to identify an

118 optimal site for subsequent variant screening, we generated 12 partially overlapping single guide  
119 RNAs (sgRNAs) that allow the single editing site to shift gradually from the 5' end (position A1), to  
120 near the 3' end (position A16) of a 20-nucleotide sliding window. To enable single-nucleotide step  
121 scanning in a compact editing window (Fig.1d, e), we used NG-ABEmax to edit either an amber  
122 stop codon (TAG, *dEGFP1*) or an opal stop codon (TGA, *dEGFP2*) back to TGG (encoding W58),  
123 thus restoring full-length *EGFP* transcription and leading to an increase in fluorescence (Fig.1f, g).

124 Consistent with previous reports <sup>3</sup>, our edit site scan analysis revealed a wide editing efficiency  
125 distribution centered around A6 and A7 with efficiencies ranging from 0.05% (A16) to 48% (A6),  
126 suggesting both robust editing efficacy and high site selectivity of the assay system (Fig. 1f-g).

### 127 **Evolution of NG-ABEmax variants with higher editing activity**

128 To demonstrate practical utility of our stop codon reversion-based screening system for  
129 adenine base editors, we used it to screen for NG-ABEmax variants that exhibit augmented editing  
130 efficiencies. To this end, it is appropriate to select an editing site with low-to-intermediate editing  
131 efficiency with the current NG-ABEmax, so that any significant activity enhancement can be  
132 captured and accurately quantified. We selected A12 site with a 3.1% efficiency (Fig. 1f), which  
133 also falls outside the canonical editing window of NG-ABEmax, to screen for variants with either a  
134 generally augmented editing activity, or an ability to edit in a shifted window.

135 We first generated two random variant libraries of NG-ABEmax by error-prone PCR and  
136 named them library 1 and library 2 (Fig. 2a, Supplementary Fig. 2a), which carry mutations in the  
137 evolved TadA domain (TadA\*) or the wild-type TadA domain (TadA), respectively. Library 1  
138 produced about 1428 colonies (numbered L1-1 to L1-1428), and library 2 had 550 colonies  
139 (numbered L2-1 to L2-550). To estimate the frequency of mutagenesis, we randomly picked 20  
140 colonies per library for Sanger sequencing. Sequencing analysis revealed that 19 of the 20  
141 colonies from library 1 harbored unique mutation(s), while the 20th colony had a single-nucleotide  
142 deletion. Very similar results (19 colonies with unique mutation(s); one with deletion) were obtained  
143 with library 2. Further analysis revealed an average rate of ~4.8 substitutions per kilobase in these  
144 two libraries (Supplementary Table 1).

145 Subsequently, each variant (L1-1 to L1-1428 and L2-1 to L2-550 plasmids) was individually  
146 co-transfected with the corresponding sgRNA (*dEGFP1*-A12) and *dEGFP1* expression cassette  
147 into HEK-293 cells (Supplementary Fig. 2b), respectively. We found that 12 (8 of library 1 and 4 of  
148 library 2) out of the 1978 tested variants exhibited significantly elevated activity at A12, showing at

149 least a 3-fold increase in editing activity (8.9%-11.25%), compared to the wild-type NG-ABEmax  
150 (2.62%-3.4%) (Fig. 2b). These findings suggest that the screening platform is effective at identifying  
151 and enriching high-activity base editor variants.

152 To distinguish whether these 12 high-activity variants possess generally higher activity at all  
153 target sites or shift or expand the editing window to gain activity at the A12 site, we took advantage  
154 of our recently developed method for assessing base editing in human cells, in which editing  
155 efficiency is measured by agarose-gel analysis of editing-mediated inactivation of restriction  
156 enzyme sites<sup>15</sup>. We generated a sgRNA that can simultaneously target a PstI restriction site (A7)  
157 and a Sall restriction site (A12) in the MCS (multiple cloning sites) sequence, which was integrated  
158 into the genome of HEK-293 (termed HEK293-PME) (Fig. 2c). Each of these 12 variants was then  
159 co-transfected with this sgRNA into HEK293-PME cells. At 48 h post-transfection, cells were  
160 collected and isolated genomic DNA. PCR was performed with the genomic DNA as a template  
161 with specific primers. Purified PCR products were treated with PstI (178 bp + 428 bp) or Sall (177  
162 bp + 429 bp), respectively, and then analyzed by agarose gel electrophoresis to estimate editing  
163 efficiency. Remarkably, L1-60, L1-98 and L2-156 showed significantly improved activity (more than  
164 ~3-fold increase) at A7 site (Fig. 2d). While, it may be due to the low sensitivity of this restriction  
165 enzyme digestion-based method, we can't observe the boosting effects at A12 site,  
166 (Supplementary Fig. 3). The enhanced editing efficiencies of L1-60, L1-98 and L2-156 were further  
167 confirmed with above-described *EGFP*-based flow cytometry system (Fig. 2e). Notably, L2-156  
168 possesses the highest activity among all tested variants, with ~2-fold increase in editing efficiencies  
169 at A7 and A9 sites. These results suggest that these three variants possessed across-the-board  
170 higher editing activities both within and outside of the canonical editing window, rather than merely  
171 shifting the editing window.

## 172 **Mapping the key residues underlying enhanced editing activity**

173 To map the exact sequence changes in L1-60, L1-98 and L2-156, we sequenced the TadA\*  
174 domains of L1-60 and L1-98 and the TadA domain of L2-156. We found multiple amino acid  
175 substitutions in L1-60 (R101S/D139G/E140K), L1-98 (E9K/Q154R), and L2-156  
176 (A106S/N127K/E155K), respectively. To deconvolute any combinatorial effects and identify key  
177 residues that drove higher editing activities in L1-60, L1-98 and L2-156, we dissected all possible  
178 single, double, and triple combinations of mutations found in these 3 variants using our  
179 *EGFP*-based system (Fig. 3a and Supplementary Fig. 4).



180 For L1-60, the highest activity occurred when R101S, D139G and E140K were present  
181 simultaneously (hereafter termed NG-ABEmax-SGK, Fig. 3a). For L1-98, the performance of  
182 Q154R alone was comparable to that of E9K/Q154R. Therefore, Q154R is the key mutation for the  
183 enhanced activity of L1-98, and the variant harboring single Q154R mutation termed  
184 NG-ABEmax-R (Fig. 3a). For L2-156, N127K alone accounted for essentially all the enhanced  
185 activity. The activity enhancement was completely abolished when N127K was reverted back to  
186 WT. Thus, N127K substitution was the driving force for the enhanced activity in L2-156 and is  
187 hereafter referred to as NG-ABEmax-K (Fig. 3a). Meanwhile, NG-ABEmax-R and NG-ABEmax-K  
188 have slightly higher activity, compared with that of NG-ABEmax-SGK (Fig. 3a).

189 As these high-efficiency variants were identified and validated on an exogenous *EGFP* gene,  
190 we next verified that they also function similarly at endogenous genomic loci. To this end, we  
191 transfected HEK-293 cells with these two variants NG-ABEmax-R or NG-ABEmax-K to edit a panel  
192 of 19 genomic sites of different sequence context and composition. Upon editing, genomic DNA  
193 was extracted from harvested cells at 48 h post-transfection and next-generation sequencing was  
194 performed.

195 Compared to NG-ABEmax, NG-ABEmax-SGK, NG-ABEmax-R and NG-ABEmax-K exhibited  
196 significantly improved editing efficiency at most genomic sites tested ((Supplementary Fig. 5,6).  
197 The enhancement was particularly pronounced at low-efficiency sites. For NG-ABEmax-R, the  
198 enhanced activity was most obvious at A4 (average: 14.62% of NG-ABEmax-R versus 2.94% of  
199 NG-ABEmax), an increase of 5.0-fold (Supplementary Fig. 5, 6b). For NG-ABEmax-K, the average  
200 activity increased by 2.2-fold at A7 where the enhanced activity was most evident compared to  
201 NG-ABEmax (average: 28.96% of NG-ABEmax-K versus 13.04% of NG-ABEmax) (Fig. 3b,  
202 Supplementary Fig. 6c). Mirroring the findings on the *EGFP* gene, the editing activity of these  
203 variants on genomic sites had been significantly improved both within and outside of the canonical  
204 editing window (Fig. 3, Supplementary Fig. 6). Notably, at the non-canonical editing window (i.e.,  
205 A12), the activity of these variants remains very low, despite significant enhancement by the  
206 mutations (Fig. 3a). We also noticed that the editing activity of NG-ABEmax-K was highest among  
207 these variants at the majority of sites we examined (Fig. 3a. Supplementary Fig. 6). Taken together,  
208 we identified the key editing-enhancing residues of L1-60 (R101S/D139G/E140K), L1-98 (Q154R)  
209 and L2-156 (N127K), and showed that these substitutions led to general enhancement of editing  
210 activity at both exogenous and endogenous sites and both within and outside the canonical editing  
211 window.

## Combinatorial optimization of high-efficiency NG-ABEmax variants

NG-ABEmax-K was selected for further study due to its highest editing activity. We hypothesized that replacing Asn127 in TadA with two other basic amino acids [Histidine (H) or Arginine (R)] may produce a similar effect due to similar charge. We generated NG-ABEmaxH<sup>127</sup> and NG-ABEmaxR<sup>127</sup> and tested the performance (Supplementary Fig. 5). Interestingly, we found that while N127H had little effect on activity, N127R exhibited enhanced activity comparable to N127K within the canonical editing window (Supplementary Fig. 7a). This was most obvious at A7 (ranging from 2.73% to 64.98%, compared to NG-ABEmax [ranging from 1.3% to 48.1%]), with up to 6.02-fold increase on site 3 (Supplementary Fig. 7b). These data suggest that a long-chain basic residue at position 127 in TadA generally boosts the editing efficiency of NG-ABEmax, potentially by enhancing local interactions with the DNA substrate.

We next sought to further enhance the activity by combining mutations from NG-ABEmax-R, and NG-ABEmax-K, because using our EGFP-based system these two have slightly higher activity, compared with that of NG-ABEmax-SGK (L1-60, Fig. 3a). We generated a variant namely NG-ABEmax-KR (N127K mutation in TadA domain, Q154R mutation in TadA\* domain) (Supplementary Fig. 5). We evaluated the activity at the endogenous genomic sites and found that NG-ABEmax-KR exhibited further increased activity (Fig. 3b, Supplementary Fig. 8). Specifically, the average activity of NG-ABEmax-KR was increased by 1.1-fold to 5.4-fold within the canonical editing window, with an activity ranging from 40.87% to 64.22% (7.6% to 60.27% for NG-ABEmax) (Supplementary Fig. 8). Upon additional examination, we found that the editing window has been greatly expanded in both directions in NG-ABEmax-KR (Fig. 3c-e, Supplementary Fig. 9). Specifically, compared with canonical editing window at the A4~A7, the NG-ABEmax-KR has an editing window at A3~A7, with a substantially higher activity at A3, A4, A6 and A7. These findings show that the original mutations acted partially or wholly independently of each other, thus allowing further enhancements by combinatorial optimization.

## NG-ABEmax-KR exhibits superior efficacy in mice disease models and human gene therapy

To test whether NG-ABEmax-KR could be harnessed for efficient generation of mouse disease models, we designed an sgRNA to target c.202 of *Tyr* gene encoding the melanin-producing Tyrosinase, to introduce a missense codon (CAT to CGT, H420R) and generate an albinism mouse model<sup>28</sup> (Fig. 4a). Successful editing would produce *Tyr* mutant mice that have white skin instead of wild-type mice with black skin (Fig. 4b). Remarkably, the efficiency of base editing in the mouse

243 model by NG-ABEmax-KR was drastically higher than the original NG-ABEmax (Fig. 4c). While the  
244 original NG-ABEmax produced only 1 (9%) white and 2 (18%) mosaic (white/black) mice out of 11  
245 in total, the NG-ABEmax-KR produced 15 (75%) white and 2 (10%) mosaic mice out of 20 in total.  
246 Next-generation sequencing of the target site from each mouse showed that the editing efficiency  
247 were greatly boosted at A3 (Fig. 4c, Supplementary Fig. 10), consistent with the phenotype results.  
248 Sanger sequencing further confirmed the successful editing events at the target site (Fig. 4d).

249 Next, we sought to evaluate the performance of NG-ABEmax-KR in introducing  
250 disease-correcting mutations. An effective therapeutic strategy that greatly alleviate the  
251 clinical symptoms of  $\beta$ -globin diseases such as sickle-cell anemia and  $\beta$  thalassaemia is to  
252 recapitulate British-type HPFH and induce fetal hemoglobin production in adults by promoter  
253 manipulation<sup>29</sup>. To this end, we designed a sgRNA to simultaneously mutate the *HBG1* or *HBG2*  
254 (*HBG1/HBG2*) promoters to activate their expression. We found that in HEK-293 cells,  
255 NG-ABEmax-KR efficiently installed the desired T•A to C•G mutations in *HBG1/HBG2* promoters,  
256 with a 4.32-fold increase in editing efficiency over NG-ABEmax (Fig. 4e).

257 Enhanced gene-editing activity is generally associated with an elevated off-target activity. To  
258 ask whether NG-ABEmax-KR increases off-target effects, we amplified fragments harboring the  
259 target site by PCR and performed sequencing to detect potential off-target editing. We observed a  
260 slight increase in editing at three off-target sites compared to NG-ABEmax (Supplementary Fig.11).  
261 We also investigated the off-target effects of targeting *Tyr* gene for generation of mice disease  
262 models by the prediction of potential off-target sites, and observed no off-target at the majority of  
263 the predicted sites (Supplementary Fig.12). We did observe off-target editing at certain other sites.  
264 These findings are consistent with the notion that NG-ABEmax-KR enhances the general catalytic  
265 activity of TadA that are manifest at both target and non-target sites. To address whether  
266 NG-ABEmax-KR could increase the on-target:off-target editing ratios, we performed the analysis of  
267 on-target and its off-target. The results revealed a dramatic increase of NG-ABEmax-KR (up to  
268 4037-fold at *HEK 3-A6-OT4* of NG-ABEmax-KR VS 1432-fold of NG-ABEmax) at the off-target  
269 sites analyzed except one (*HEK 2-A7-OT2*, Supplementary Fig.13). Together, these applications in  
270 mouse disease models and human gene therapy show that NG-ABEmax-KR is an effective  
271 gene-editing tool that exhibits superior activity in both human cells and animal models.

## 272 **Molecular underpinnings of the enhanced editing efficiency of NG-ABEmax-KR**

273 To gain a mechanistic understanding of NG-ABEmax-KR's enhanced editing efficiency, we  
274 performed structural modeling. Structural modeling suggests that substitutions of the DNA-facing

side chains of N127 of TadA\* and Q154 of TadA with longer, positively charged side chains may create new, stabilizing interactions with the editing window, potentially augmenting catalysis (Fig.5a, b). Specifically, an N127K substitution is expected to bring its e-ammonium group within hydrogen-bonding distance ( $\sim 3.6$  Å) with the backbone non-bridging oxygen (OP1) of C27, immediately 3' of the deamination site. This notion is consistent with our observation that N127R had a comparable impact on editing activity as N127K while N127H did not. On the opposite side of the DNA substrate, a Q154R substitution of TadA is projected to insert its guanidinium group between the nucleobases of C24 and C25, two nucleosides that immediately precede the deamination site (A26) on the 5' side. Here, the Q154R guanidinium group would be located less than 3.5 Å away from both nucleobases of C24 and C25, and well positioned to make robust cation-p interactions to both, thereby stabilizing the DNA conformation to facilitate catalysis. Thus, *in silico* mutagenesis and structural modeling provide a straightforward explanation for the enhanced activity of NG-ABEmax-KR. Further, our findings provide a proof-of-principle that structure-informed rational redesign of the TadA-DNA interface, in particular, by introducing long, basic side chains at strategic locations, can effectively repurpose an enzyme interface originally evolved to recognize RNA hairpins, to recognize and accommodate topologically constrained, single-stranded DNA substrates.

## Discussion

ABE are particularly promising tools for the studies and applications of gene therapy, because two-thirds of human diseases are caused by single nucleotide mutations, and about half of these can be corrected by A-to-G conversions<sup>30</sup>. However, ABE currently have limited applications due to low editing activity compared with CBEs. Due to the lack of a clear understanding of the artificial TadA-DNA interface within ABE, rational design-based activity optimizations have not borne significant fruit. In this study, we established and applied an *EGFP*-based screening system to efficiently evolve and select NG-ABEmax variants with desired traits, such as those with increased activity. Using this system, we obtained several mutants (NG-ABEmax-SGK, NG-ABEmax-R, and NG-ABEmax-K) that possess higher editing activity (Figs. 2-4). Besides higher editing efficiency, these three variants are able to effectively editing at certain inaccessible positions by the wild-type NG-ABEmax (Fig. 2). Among the three variants NG-ABEmax-K exhibited the best performance at most sites we examined (Fig. 3b, Supplementary Figs. 5-6). To further enhance the activity, NG-ABEmax-KR was created by the combinatorial design using the identified driver mutations (Fig.

306 3b, c, Supplementary Fig. 8). Compared with NG-ABEmax, NG-ABEmax-KR showed superior  
307 editing activity and an expanded editing window (Fig. 3, 4).

308 The Cas9 element in ABE has been extensively studied as a limiting factor for PAM recognition  
309 <sup>31-37</sup>. However, some studies have illustrated that Cas9 elements are also responsible for the  
310 editing window of ABE, which indicates that the Cas9 element may also be selected for the  
311 evolution of ABE <sup>31,33,38,39</sup>. In addition, the sequence or structure of the sgRNA may also modulate  
312 the activity of ABE, suggesting it could be further optimized <sup>10</sup>. As for DNA off-target effects, we  
313 characterized the fidelity of NG-ABEmax and its variants (Supplementary Fig. 11, 12, 13). It is  
314 reported that ABE can also induce RNA off-target effects contributed by the wild-type TadA domain  
315 <sup>20,40-42</sup>. If we deleted the wild-type TadA domain, RNA off-target effects may be ameliorated. In  
316 addition to editing canonical base A by ABE, multiple studies have found that ABE can also edit  
317 base C at a low level <sup>12,40,43,44</sup>. Nonetheless, our study did not observe off-target editing products  
318 involving base C editing. The precise specificity of the NG-ABEmax variants awaits further  
319 comprehensive characterization.

320 While our study was being prepared for publication, Richter and colleagues reported directed  
321 evolution of ABEmax based on phage-assisted non-continuous and continuous evolution (PANCE  
322 and PACE), which produced ABE8e <sup>45</sup>. Based on the synthetic library of TadA sequences that  
323 contains all 20 canonical amino acid substitutions at each position of TadA, Nicole and colleagues  
324 identified ABEmax variants (ABE8s) with higher activity <sup>46</sup>. Notably, we observed that the Q154R  
325 mutation that we selected was also identified in ABE8s but not ABE8e, suggesting that directional  
326 evolution analyses based on different strategies may converge on the same critical sites. Our  
327 high-activity mutations do not overlap with those reported in the ABE8e <sup>45,46</sup>. A comparative  
328 analysis of the editing efficiencies of ABE8e, ABE8s and NG-ABEmax-KR may yield additional  
329 insights into the path leading to improved efficiency and combination of these mutations in the  
330 same enzyme may achieve synergistic effects as we observed in this study. Recently, the structure  
331 of ABE8e has been reported, which provides additional insights for further rational design of ABE  
332 <sup>47</sup>.

333 Taken together, in the present study, we created a high-throughput directional screening  
334 system to evolve, select and assess ABE activity in human cells, and identified NG-ABEmax  
335 variants with significantly increased editing activity. This system can be readily applied to evolve  
336 additional ABE traits besides enhanced activity, such as access to previously inaccessible editing  
337 sites, or modulation of off-target effects. Finally, our analysis provides a mechanistic rationale for

338 the observed activity enhancement, which can guide further optimizations of the TadA-DNA  
339 interface.

## 340 **Material and methods**

### 341 **Plasmid construction**

342 The original plasmid NG-ABEmax expressing the TadA-TadA\*-nSpCas9-NG fusion protein  
343 was obtained from Addgene (Addgene plasmid #124163). The plasmid pSIN-EGFP containing an  
344 *EGFP* gene, *IRES* and *Puromycin* genes were generated as previously described <sup>21</sup>. The sgRNA  
345 expression cassettes were generated by standard protocol. Desired point mutations were  
346 introduced into the coding sequence of *EGFP* by PCR to generate *EGFP*-variant, *dEGFP1*, and  
347 *dEGFP2*. All plasmids were confirmed by Sanger sequencing. The oligonucleotide sequences used  
348 for plasmid construction in this study are listed in Supplementary Tables 2-7.

### 349 **Construction of NG-ABEmax mutant libraries**

350 NG-ABEmax libraries were generated using the protocols which described previously<sup>48</sup>. For  
351 the library 1, the NG-ABEmax plasmids were digested with BseRI, and the digested products  
352 (backbone) was subsequently purified (8277 bp). Then, GeneMorph II Random Mutagenesis Kit  
353 (Agilent) was used to perform error-prone PCR on the TadA\* domain in the NG-ABEmax plasmid  
354 sequence, and the PCR products were purified as fragments (575 bp). The backbone and  
355 fragments were In-fusion assembled (Supplementary Fig 3a). For the library 2, the backbone  
356 plasmids were digested with SacII and BamHI (the final products of 8177 bp). Then, fragments (730  
357 bp) of error-prone PCR on the TadA domain were obtained using the same method  
358 (Supplementary Fig 3a). The resulting library 1 and library 2 were transformed into  
359 electrocompetent *E. coli* DH10B and incubated on LB plates with Ampicillin (0.1 g/mL) at 37°C  
360 overnight. A total of 1978 colonies (1428 of library 1 and 550 of library 2) were obtained. The  
361 plasmids from individual colonies were isolated. The oligonucleotide sequences used for libraries  
362 construction are listed in Supplementary Table 8.

### 363 **Cells and cell culture**

364 HEK-293 cells were obtained from ATCC (CAT#CRL-1573), and HEK293-PME cells were  
365 generated as previously described <sup>21</sup>. HEK-293 cells were grown at 37°C in 5% CO<sub>2</sub> in Dulbecco's  
366 modified Eagle's medium (Life Technologies, Carlsbad, CA) supplemented with 10%  
367 heat-inactivated fetal bovine serum, penicillin/streptomycin. HEK293-PME cells were cultured with  
368 additional puromycin <sup>21</sup>.

## 369 **Transfection protocol, genomic DNA extraction, images and flow cytometry analysis**

370 For NG-ABEmax variants screening experiments, HEK-293 cells were seeded at  $0.9 \times 10^5$  cells  
371 per well on 24-well plates in DMEM medium. 24 h after seeding, cells were co-transfected with 200  
372 ng NG-ABEmax variant plasmids, 100 ng sgRNA expression plasmids, 200 ng dEGFP1 plasmids  
373 and 1.5 ul TurboFect Transfection Reagent (Thermo Fisher Scientific). Medium was changed at 24  
374 h post-transfection and flow cytometry analysis was performed at 48 h post-transfection. For  
375 experiments using HEK293-PME to compare the activity of NG-ABEmax variants, HEK293-PME  
376 cells were seeded at  $0.7 \times 10^5$  cells per well on 24-well plates in the presence of puromycin. 24 h  
377 after seeding, cells were co-transfected with 200 ng NG-ABEmax variant, 100 ng sgRNA  
378 expression plasmids and 1.5 ul TurboFect Transfection Reagent. Genomic DNA extraction of cells  
379 is subsequently performed using genome extraction kit (Vazyme). Images were obtained at 24 h  
380 and 48 h post-transfection, and cells were collected at 48 h post-transfection. Flow cytometry  
381 analysis was performed with FACS Aria II (BD Biosciences).

## 382 **Comparison of the activity of NG-ABEmax variants using HEK293-PME cells**

383 The genomic DNA was used as template for PCR amplification using corresponding primers  
384 and PCR products were subsequently purified and recovered (606 bp). 200 ng purified PCR  
385 products were treated with PstI (178 bp + 428 bp) or Sall (177 bp + 429 bp). Finally, the activity of  
386 NG-ABEmax variants was determined by agarose gel electrophoresis.

## 387 **Sanger sequencing for pJET colonies**

388 Sequence flanking the CRISPR target sites for *Try* was PCR amplified, and products were  
389 inserted into the vector pJET1.2 (CloneJET PCR Cloning Kit, Thermo Fisher Scientific). The ligated  
390 products were transformed into *E. coli*. The corresponding plasmids were isolated and sequenced  
391 on an ABI PRISM 3730 DNA Sequencer.

## 392 **Mice**

393 All animal procedures were carried out in accordance with the current guidelines of the  
394 Institutional Animal Care and Use Committee (IACUC) at the Center for Excellence in Molecular  
395 Cell Science, Shanghai Institute of Biochemistry and Cell Biology, Shanghai, China. B6D2F1  
396 (C57BL/6×DBA2) female mice were used as oocyte donors. The males (C57BL/6) were used to  
397 mate with B6D2F1 females to obtain zygotes. ICR females were used as pseudo-pregnant foster  
398 mothers.

## ***In vitro* transcription**

The mRNA transcriptional templates of NG-ABEmax and NG-ABEmax-KR were amplified by PCR using Phanta Max Super-Fidelity DNA Polymerase (Vazyme), purified by the Universal DNA Purification Kit (TIANGEN) and then transcribed using the mMACHINE T7 ULTRA transcription Kit (Invitrogen) following the manufacturer's instruction. The transcriptional template of *Tyr* sgRNA was amplified from Px330-mCherry plasmids <sup>49</sup>(Addgene#98750) and transcribed *in vitro* using the MEGAscript T7 kit (Invitrogen) following the manufacturer's instructions. mRNA and sgRNA were subsequently purified with the MEGAclean Transcription Clean-Up Kit (Invitrogen), resuspended in RNase-free water and then stored at  $-80^{\circ}\text{C}$ .

## **Microinjection and embryo transfer**

The mixture of ABE mRNA (100 ng/ $\mu\text{l}$ ) and sgRNA (100 ng/ $\mu\text{l}$ ) was diluted in RNAase-free water, centrifuging at  $4^{\circ}\text{C}$ , 13,400 g for 10 min and then injected into the cytoplasm of zygotes harvested from B6D2F1 females (mated with C57BL/6 males) using a micromanipulator (Olympus) and a FemtoJet microinjector (Eppendorf). The injected embryos were cultured in EmbryoMax KSOM Medium (Sigma-Aldrich) until the two-cell stage and then transferred into oviducts of recipients at 0.5 days post-coitum (dpc). Recipient mothers delivered pups at 19.5 dpc and we analyzed phenotypes of offspring at day 10 after birth.

## **Next-generation sequencing (NGS)**

Extracted genomic DNA from transfected cells or mice tails were performed by standard protocol. NGS library was constructed using genomic DNA as template through two rounds of PCR. First-step PCR amplification of 100-220 bp sequences from on/off-target sites were performed using specific primers. For the second-step PCR amplification, we fixed the barcodes, index and adaptor sequences to the first-step PCR amplification products. The second-step PCR amplification products were purified and pooled, and subsequently subjected to paired-end read sequencing using the Hiseq-PE150 strategy at Novogene (Nanjing, China). Finally, open-sourced "CRISPResso" software (version 1.0.10) was used to analyze the status of base editing. The oligonucleotide sequences used for NGS are listed in Supplementary Tables 9, 10.

## **Molecular modeling**

Single amino-acid substitutions are generated *in silico* using Coot <sup>50</sup>. Side chain orientations are then geometrically optimized in Coot. Molecular graphics are prepared using MacPyMOL (Schrödinger, Inc).



431 **Statistics**

432 All data were expressed as mean  $\pm$  SD Differences were determined by 2-tailed Student's  
433 t-test or Mann-Whitney test between two groups. The criterion for statistical significance was  $*P <$   
434  $0.05$ ,  $**P < 0.01$ ,  $***P < 0.001$ ,  $****P < 0.0001$ .

436 **Acknowledgments**

437 We thank our group member for technical assistance. This work was supported by grants from  
438 National Natural Science Foundation of China (81201181), National Key R&D Program of China  
439 [2018YFA0107304], Science Technology project of Zhejiang Province (2017C37176), and Project  
440 of State Key Laboratory of Ophthalmology, Optometry and Visual Science, Wenzhou Medical  
441 University (J02-20190201), Shanghai Super Postdoctoral Incentive Program, China National  
442 Postdoctoral Program for Innovative Talents (BX20200348) and by the Intramural Research  
443 Program of the NIH, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)  
444 (ZIADK075136).

445 **Conflict of interest**

446 The authors declare that they have a patent (pending) for the screening platform and the  
447 mutants in this study.

448 **Data availability**

449 The data that support the findings of this study are available in the Supplementary Materials.  
450 The deep sequencing data are available at the NCBI Sequence Read Archive (SRA)  
451 (<https://www.ncbi.nlm.nih.gov/sra/>) under BioProject PRJNA 714183 (SRA: SRR13950186-SRR  
452 13950197; SRR13948600-SRR13948607; SRR13950619-SRR13950626; SRR13958607-  
453 SRR13958611; sample accession number, SAMN 18291703); BioProject PRJNA 713866 (SRA:  
454 SRR13945386- SRR13945395; sample accession number, SAMN 18273806). Pre-processed data  
455 is available upon request.

456 **Author contributions**

457 F.G and J.L designed research, J.F, Q. L, X.L, T.T, X.L and X.Y performed research, J.F, Z.S,  
458 J.Z, J.L and F.G. performed data analyses, and J.Z, Q.L, and F.G. wrote the manuscript. All  
459 authors have read and approved the final manuscript.

460

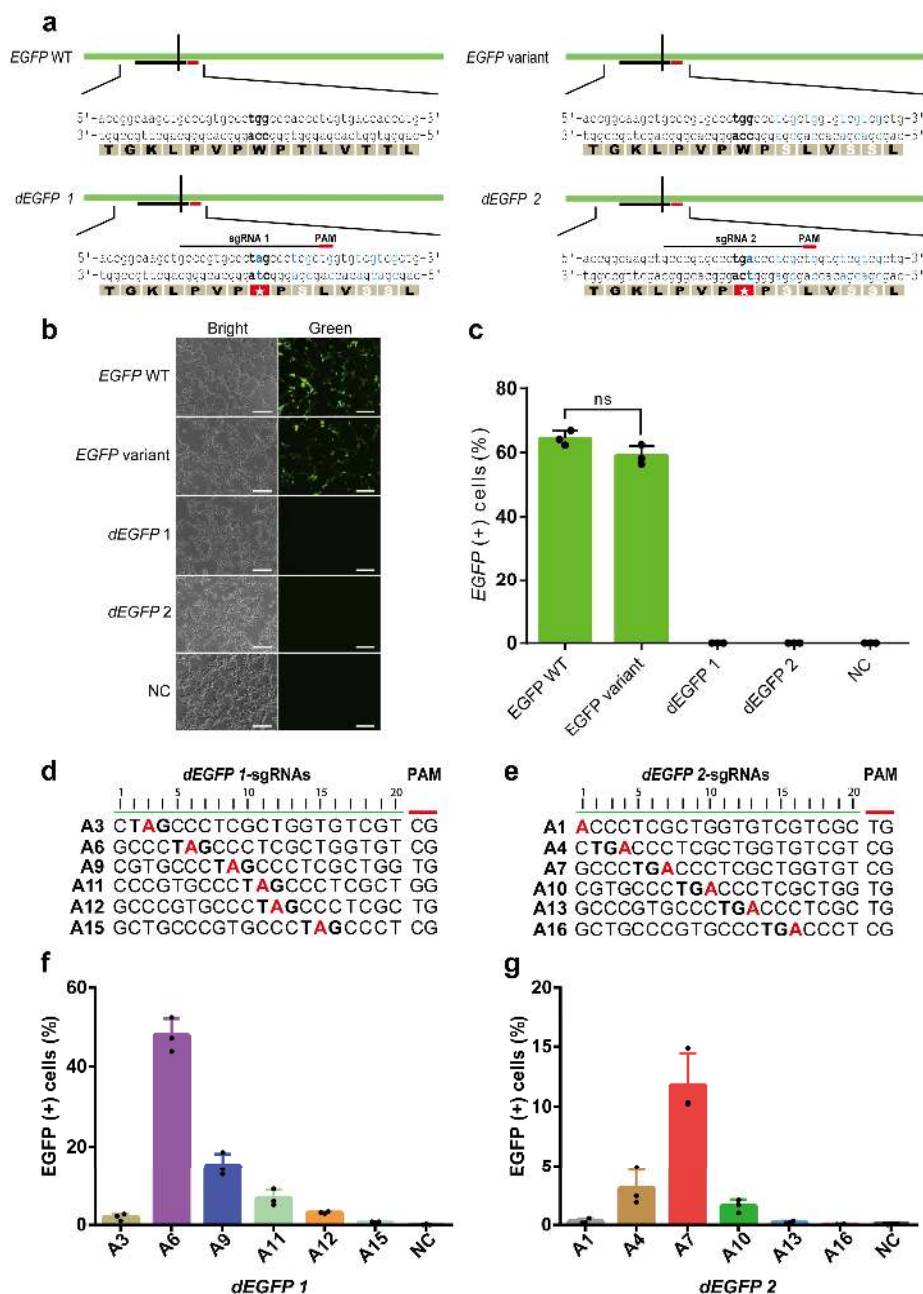
## References

1. Hsu, P.D. et al. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014).
2. Komor, A.C. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-4 (2016).
3. Gaudelli, N.M. et al. Programmable base editing of A\*T to G\*C in genomic DNA without DNA cleavage. *Nature* **551**, 464-471 (2017).
4. Iyer, S. et al. Precise therapeutic gene correction by a simple nuclease-induced double-stranded break. *Nature* **568**, 561-565 (2019).
5. Hsu, P.D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-32 (2013).
6. Hess, G.T. et al. Methods and Applications of CRISPR-Mediated Base Editing in Eukaryotic Genomes. *Mol Cell* **68**, 26-43 (2017).
7. Li, G. et al. Highly efficient and precise base editing in discarded human tripronuclear embryos. *Protein Cell* **8**, 776-779 (2017).
8. Ryu, S.M. et al. Adenine base editing in mouse embryos and an adult mouse model of Duchenne muscular dystrophy. *Nat Biotechnol* **36**, 536-539 (2018).
9. Villiger, L. et al. Treatment of a metabolic liver disease by in vivo genome base editing in adult mice. *Nat Med* **24**, 1519-1525 (2018).
10. Li, C. et al. Expanded base editing in rice and wheat using a Cas9-adenosine deaminase fusion. *Genome Biol* **19**, 59 (2018).
11. Kang, B.C. et al. Precision genome engineering through adenine base editing in plants. *Nat Plants* **4**, 427-431 (2018).
12. Liu, Z. et al. Efficient generation of mouse models of human diseases via ABE- and BE-mediated base editing. *Nat Commun* **9**, 2338 (2018).
13. Jin, S. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292-295 (2019).
14. Zuo, E. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289-292 (2019).
15. Lv, X. et al. Development of a Simple and Quick Method to Assess Base Editing in Human Cells. *Mol Ther Nucleic Acids* **20**, 580-588 (2020).

- 492 16. Koblan, L.W. et al. Improving cytidine and adenine base editors by expression optimization  
493 and ancestral reconstruction. *Nat Biotechnol* **36**, 843-846 (2018).
- 494 17. Qin, W. et al. Precise A\*T to G\*C base editing in the zebrafish genome. *BMC Biol* **16**, 139  
495 (2018).
- 496 18. Wang, M. et al. Optimizing base editors for improved efficiency and expanded editing scope  
497 in rice. *Plant Biotechnol J* **17**, 1697-1699 (2019).
- 498 19. Hua, K. et al. Simplified adenine base editors improve adenine base editing efficiency in rice.  
499 *Plant Biotechnol J* **18**, 770-778 (2020).
- 500 20. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by  
501 mutagenesis. *Nature* **571**, 275-278 (2019).
- 502 21. Zhang, Y. et al. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA  
503 cleavage in human cells. *Sci Rep* **4**, 5405 (2014).
- 504 22. Tu, M. et al. A 'new lease of life': FnCpf1 possesses DNA cleavage activity for genome  
505 editing in human cells. *Nucleic Acids Res* **45**, 11295-11304 (2017).
- 506 23. Sun, H. et al. A Single Multiplex crRNA Array for FnCpf1-Mediated Human Genome Editing.  
507 *Mol Ther* **26**, 2070-2076 (2018).
- 508 24. Lin, L. et al. Engineering the Direct Repeat Sequence of crRNA for Optimization of  
509 FnCpf1-Mediated Genome Editing in Human Cells. *Mol Ther* **26**, 2650-2657 (2018).
- 510 25. He, X. et al. Boosting activity of high-fidelity CRISPR/Cas9 variants using a  
511 tRNA(Gln)-processing system in human cells. *J Biol Chem* **294**, 9308-9315 (2019).
- 512 26. Xie, H. et al. High-fidelity SaCas9 identified by directional screening in human cells. *PLoS*  
513 *Biol* **18**, e3000747 (2020).
- 514 27. Liu, X. et al. Lb2Cas12a and its engineered variants mediate genome editing in human cells.  
515 *FASEB J* **35**, e21270 (2021).
- 516 28. Kwon, B.S. et al. Molecular basis of mouse Himalayan mutation. *Biochem Biophys Res*  
517 *Commun* **161**, 252-60 (1989).
- 518 29. Frangoul, H. et al. CRISPR-Cas9 Gene Editing for Sickle Cell Disease and  $\beta$ -Thalassemia.  
519 *N Engl J Med* **384**, 252-260 (2021).
- 520 30. Landrum, M.J. et al. ClinVar: public archive of interpretations of clinically relevant variants.  
521 *Nucleic Acids Res* **44**, D862-8 (2016).
- 522 31. Hua, K. et al. Precise A.T to G.C Base Editing in the Rice Genome. *Mol Plant* **11**, 627-630  
523 (2018).

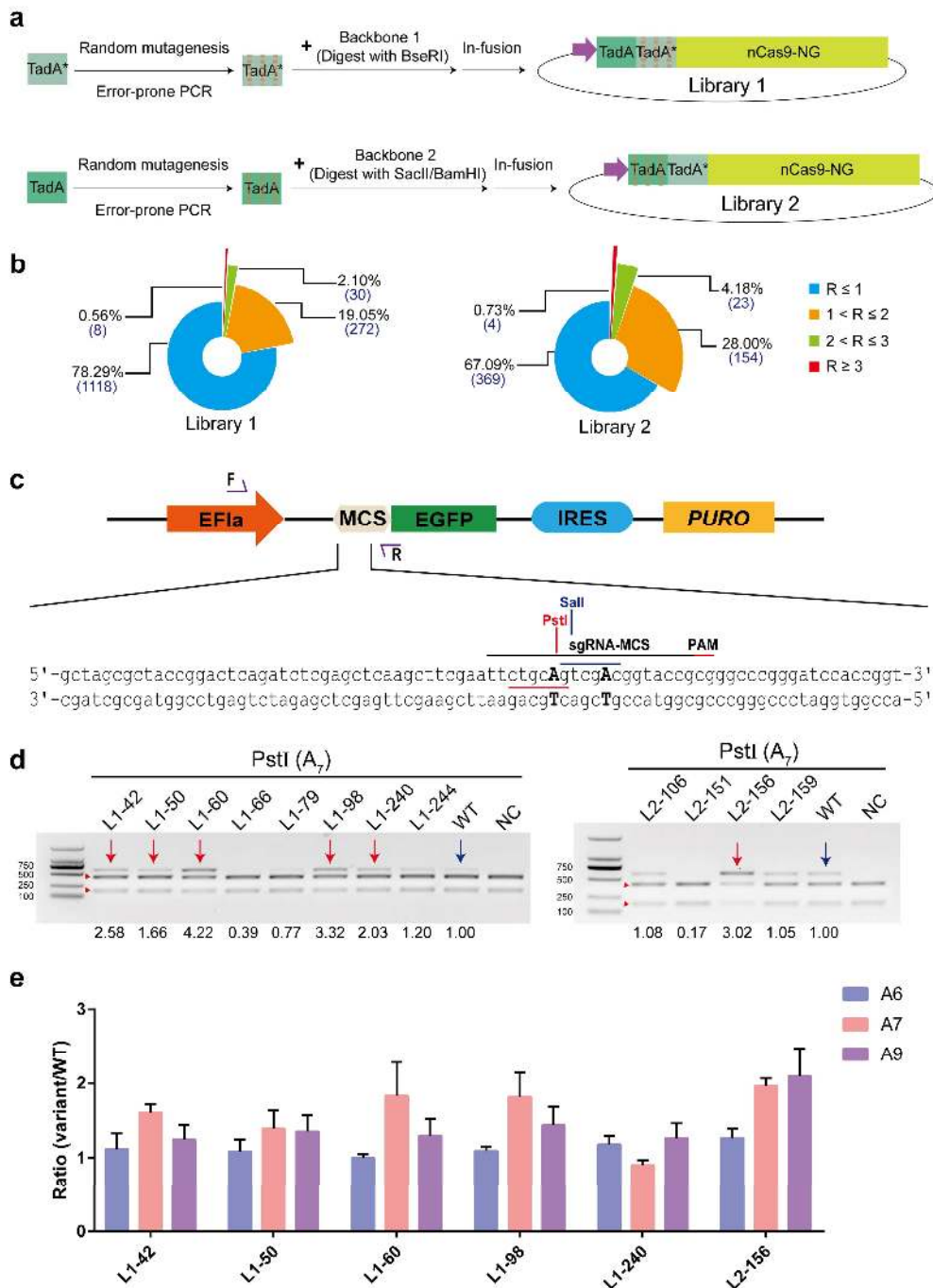
- 524 32. Yang, L. et al. Increasing targeting scope of adenosine base editors in mouse and rat  
525 embryos through fusion of TadA deaminase with Cas9 variants. *Protein Cell* **9**, 814-819  
526 (2018).
- 527 33. Huang, T.P. et al. Author Correction: Circularly permuted and PAM-modified Cas9 variants  
528 broaden the targeting scope of base editors. *Nat Biotechnol* **37**, 820 (2019).
- 529 34. Qin, R. et al. Developing a highly efficient and widely adaptive CRISPR-SaCas9 toolset for  
530 plant genome editing. *Plant Biotechnol J* **17**, 706-708 (2019).
- 531 35. Ren, B. et al. Cas9-NG Greatly Expands the Targeting Scope of the Genome-Editing Toolkit  
532 by Recognizing NG and Other Atypical PAMs in Rice. *Mol Plant* **12**, 1015-1026 (2019).
- 533 36. Huang, S. et al. Developing ABEmax-NG with Precise Targeting and Expanded Editing  
534 Scope to Model Pathogenic Splice Site Mutations In Vivo. *iScience* **15**, 640-648 (2019).
- 535 37. Walton, R.T. et al. Unconstrained genome targeting with near-PAMless engineered  
536 CRISPR-Cas9 variants. *Science* **368**, 290-296 (2020).
- 537 38. Yang, L. et al. Correction to: Increasing targeting scope of adenosine base editors in mouse  
538 and rat embryos through fusion of TadA deaminase with Cas9 variants. *Protein Cell* **10**, 700  
539 (2019).
- 540 39. Nguyen Tran, M.T. et al. Engineering domain-inlaid SaCas9 adenine base editors with  
541 reduced RNA off-targets and increased on-target DNA editing. *Nat Commun* **11**, 4871  
542 (2020).
- 543 40. Grunewald, J. et al. CRISPR DNA base editors with reduced RNA off-target and self-editing  
544 activities. *Nat Biotechnol* **37**, 1041-1048 (2019).
- 545 41. Grunewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided  
546 DNA base editors. *Nature* **569**, 433-437 (2019).
- 547 42. Rees, H.A. et al. Analysis and minimization of cellular RNA editing by DNA adenine base  
548 editors. *Sci Adv* **5**, eaax5717 (2019).
- 549 43. Lee, H.K. et al. Targeting fidelity of adenine and cytosine base editors in mouse embryos.  
550 *Nat Commun* **9**, 4804 (2018).
- 551 44. Kim, H.S. et al. Adenine base editors catalyze cytosine conversions in human cells. *Nat*  
552 *Biotechnol* **37**, 1145-1148 (2019).
- 553 45. Richter, M.F. et al. Phage-assisted evolution of an adenine base editor with improved Cas  
554 domain compatibility and activity. *Nature Biotechnology* **38**, 883-891 (2020).

- 555 46. Gaudelli, N.M. et al. Directed evolution of adenine base editors with increased activity and  
556 therapeutic application. *Nat Biotechnol* **38**, 892-900 (2020).
- 557 47. Lapinaite, A. et al. DNA capture by a CRISPR-Cas9-guided adenine base editor. *Science*  
558 **369**, 566-571 (2020).
- 559 48. Liu, X. et al. Engineered FnCas12a with enhanced activity through directional evolution in  
560 human cells. *J Biol Chem*, 100394 (2021).
- 561 49. Wu, Y. et al. Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem*  
562 *Cell* **13**, 659-62 (2013).
- 563 50. Emsley, P. et al. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**,  
564 486-501 (2010).
- 565
- 566
- 567
- 568
- 569
- 570
- 571
- 572
- 573
- 574
- 575
- 576
- 577
- 578
- 579
- 580
- 581
- 582
- 583
- 584
- 585
- 586



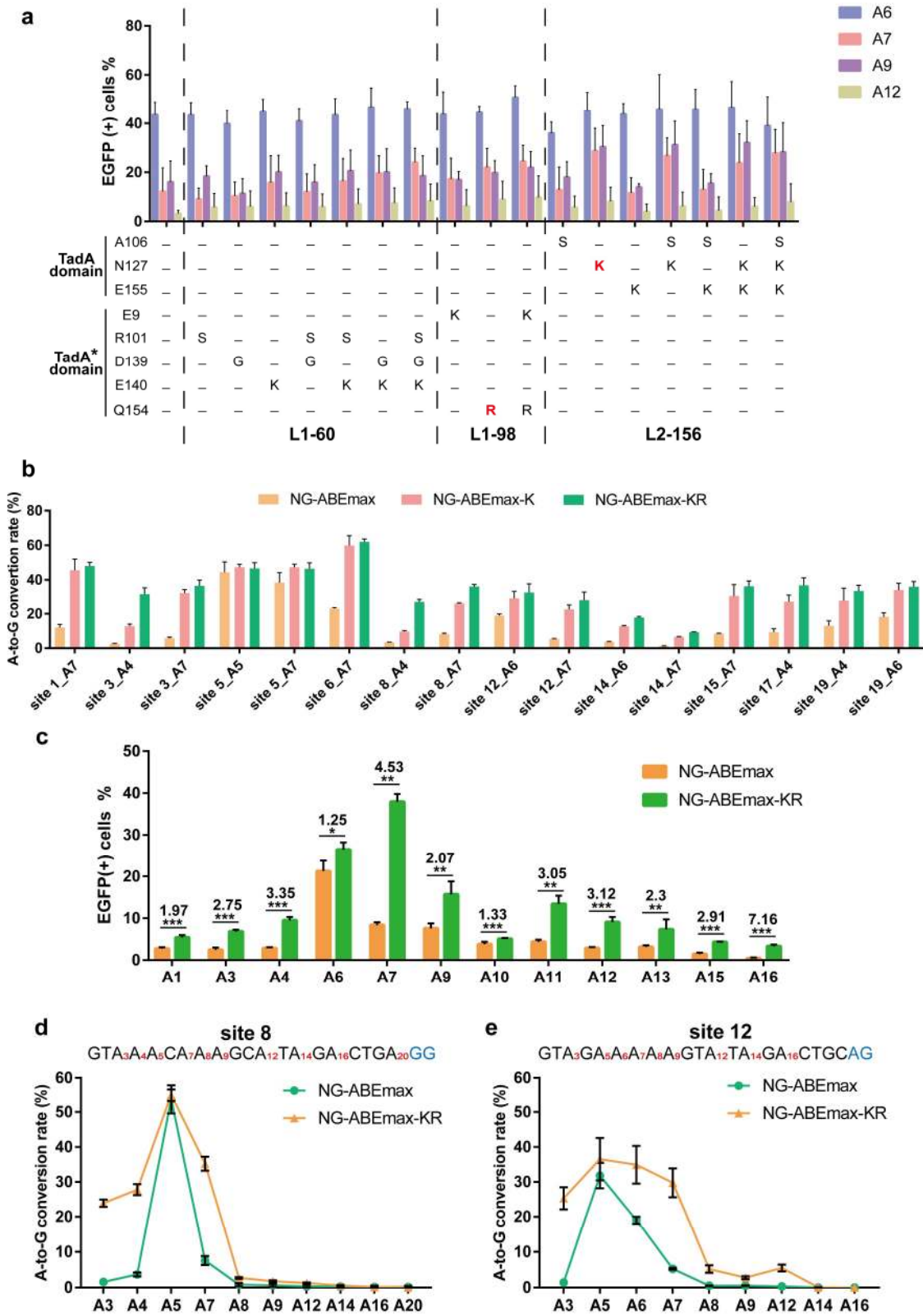
**Fig. 1 Design of the stop-codon reversion-based screening system.**

**a**, Different *EGFP* variants. Single nucleotide conversions are indicated by blue and residue substitutions are indicated by white. The stop codon was highlighted with red pentagram. **b, c**, Fluorescence image (**b**) and editing efficiencies (**c**) of HEK-293 cells co-transfect with *EGFP* WT, variant, *dEGFP* 1 and *dEGFP* 2, respectively. *EGFP* variant maintains the expression, while *dEGFP* 1 and *dEGFP* 2 abolishes the expression. (**c**). *EGFP* positive cells were quantified by flow cytometry. Scale bar, 10  $\mu$ m. Data are mean  $\pm$  s.d. of three technical replicates. **d, e**, The sequence of different sgRNAs (5'-NG-3' PAM) which were suitable for *dEGFP* 1 and *dEGFP* 2, respectively. The stop codon and target A base have been highlighted in bold and red. **f, g**, Editing efficiencies of A to G mutations at different positions of *dEGFP* 1 and *dEGFP* 2. Error bars indicate mean  $\pm$  s.d. of three technical replicates.



**Fig. 2 Evolution of NG-ABEmax variants with improved A•T to G•C base editing activity.**

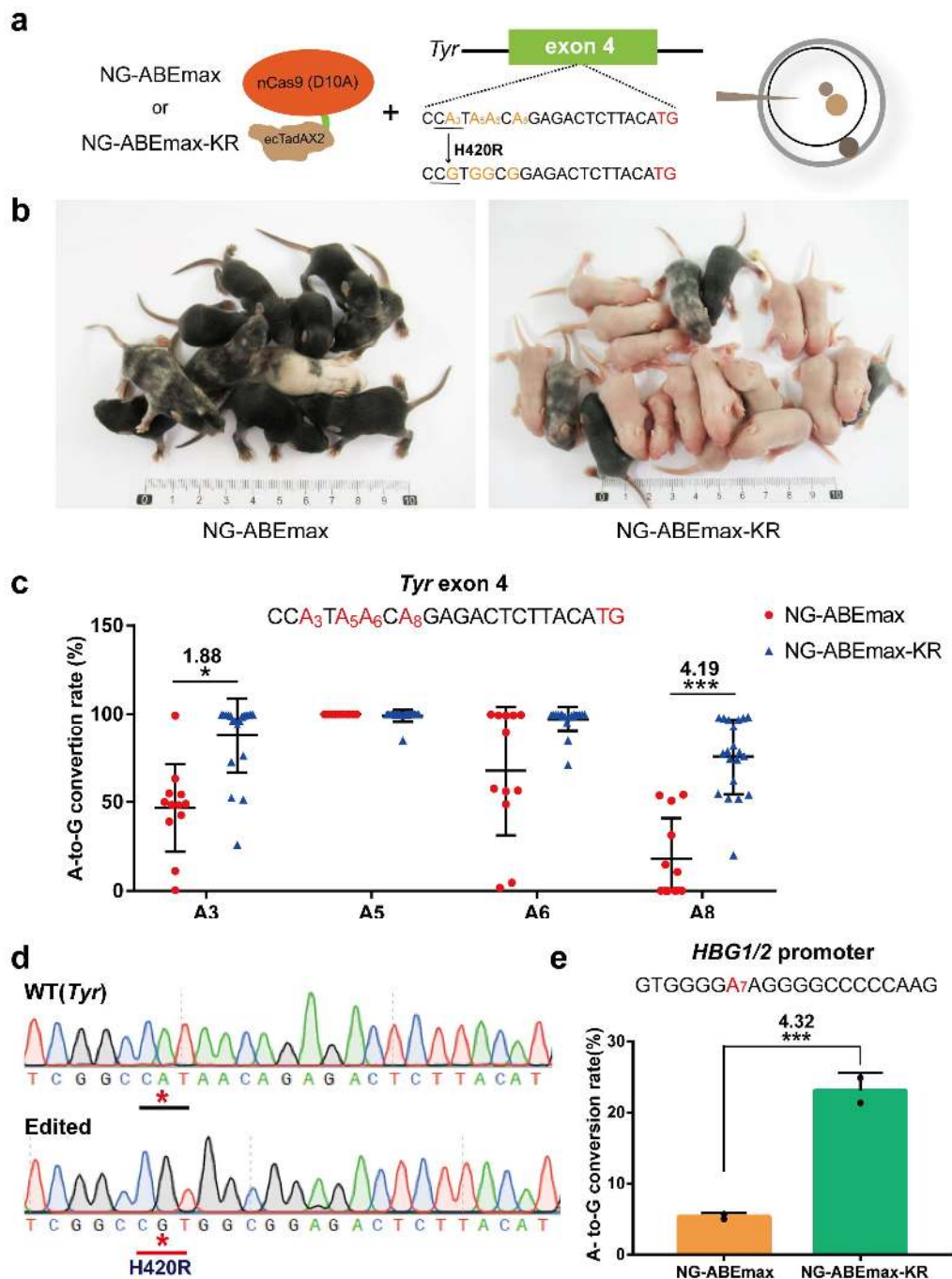
**a**, Scheme of libraries generation design. The library 1 contain TadA\* mutations and the library 2 contain TadA mutations. Red lines represent the point mutations. **b**, Summary of the library 1 and library 2 (A•T to G•C base editing efficiencies at A12). R represents the ratio of editing efficiency of mutants to NG-ABEmax. **c**, Schematic for assessing A•T to G•C base editing efficiencies of NG-ABEmax variants using the HEK293-PME cell line. Restriction sites are highlighted in red and blue. **d**, Agarose gel electrophoresis results of testing editing efficiency of NG-ABEmax variants with PstI. Cleaved bands from PstI are labeled with red triangle. The amplicon is 606-bp. The variants with high-activities are highlighted with red arrows. WT represents NG-ABEmax and highlighted with blue arrows. **e**, Boosted editing efficiencies of selected variants via EGFP-based reporter system. Error bars indicate mean  $\pm$  s.d. of three technical replicates.



**Fig. 3 Engineering of the NG-ABEmax variants.**

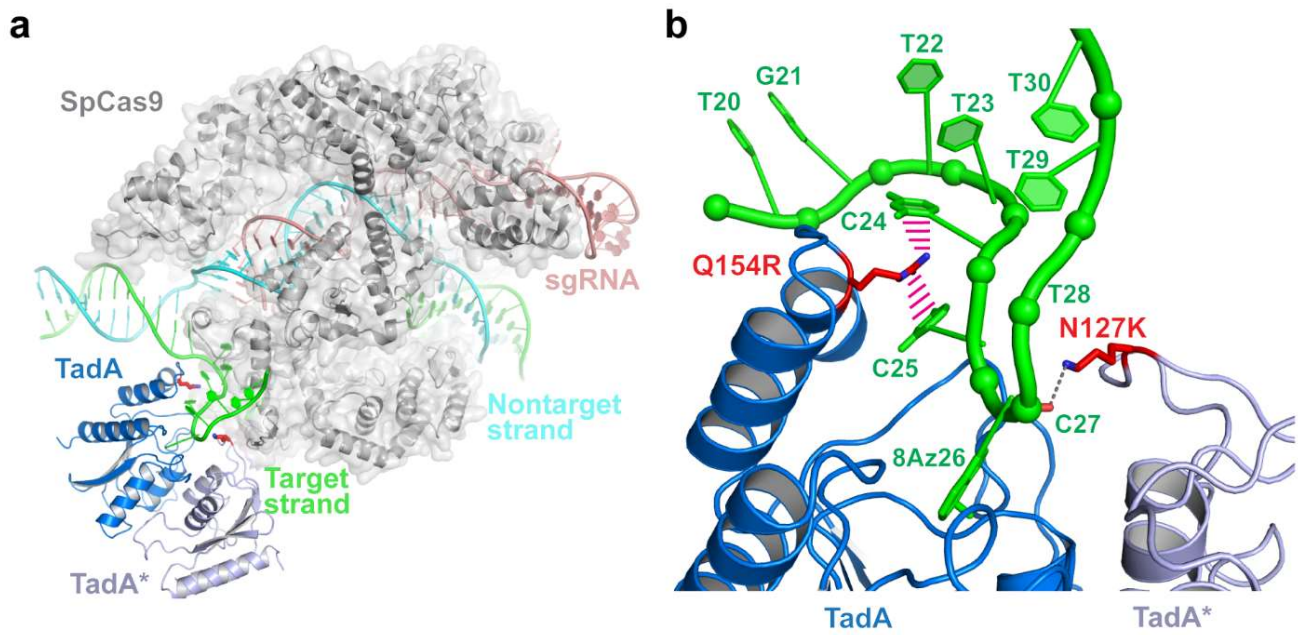
**a**, Mapping the key residues for L1-60, L1-98 and L2-156 using *EGFP*-based reporter system. Editing efficiencies were quantified by flow cytometry. **b**, Boosted editing efficiencies of NG-ABEmax-KR at endogenous genomic sites. **c**, Summary of boosted editing efficiencies of NG-ABEmax-KR at different A base at the *EGFP* sites. Error bars indicate mean  $\pm$  s.d. of three technical replicates. **d**, **e**, Increased editing window of NG-ABEmax-KR in site 8 and site 10. Each A base was highlighted in red. Error bars indicate mean  $\pm$  s.d. of three technical replicates.





**Fig. 4 Application of NG-ABEmax-KR for the generation of mice disease models and human gene therapy.**

**a**, Schematic of comparing editing activity of NG-ABEmax and NG-ABEmax-KR in mouse embryos via zygote intracytoplasmic injection. **b**, The newborn pups (days 10) produced by intracytoplasmic injection of NG-ABEmax or NG-ABEmax-KR mRNA and *Tyr* sgRNA. The *Tyr* mutant mice (H420R) are in white and the wild-type are in black, respectively. **c**, Statistical analysis of on-target A-to-G base conversions induced by NG-ABEmax ( $n = 11$ ), NG-ABEmaxKR ( $n = 20$ ) in all pups. Data are mean  $\pm$  s.d for the indicated numbers of mice. Each A base was highlighted in red. \* $P < 0.05$ , \*\*\* $P < 0.001$  by Student's unpaired two-sided t-test. **d**, Sanger sequence chromatograms confirmed the editing events. The desired mutation was highlighted with red star. **e**, Boosted editing efficiency (A7) of NG-ABEmax-KR at *HBG1/2*.



627

628

**Fig. 5 Molecular basis of enhanced editing efficiency of NG-ABEmax-KR.**

629

**a**, Overall structural model of NG-ABEmax-KR showing SpCas9 (gray), target-strand DNA (TS, green),

630

nontarget-strand DNA (NTS, cyan), single-target RNA (sgRNA, red), TadA (blue) and TadA\* (light blue). The side

631

chains of N127K of TadA\* and Q154R of TadA are shown as red sticks. **b**, Modelled Interactions between key residues

632

of NG-ABEmax-KR with TS DNA, colored as in (a).

# Figures

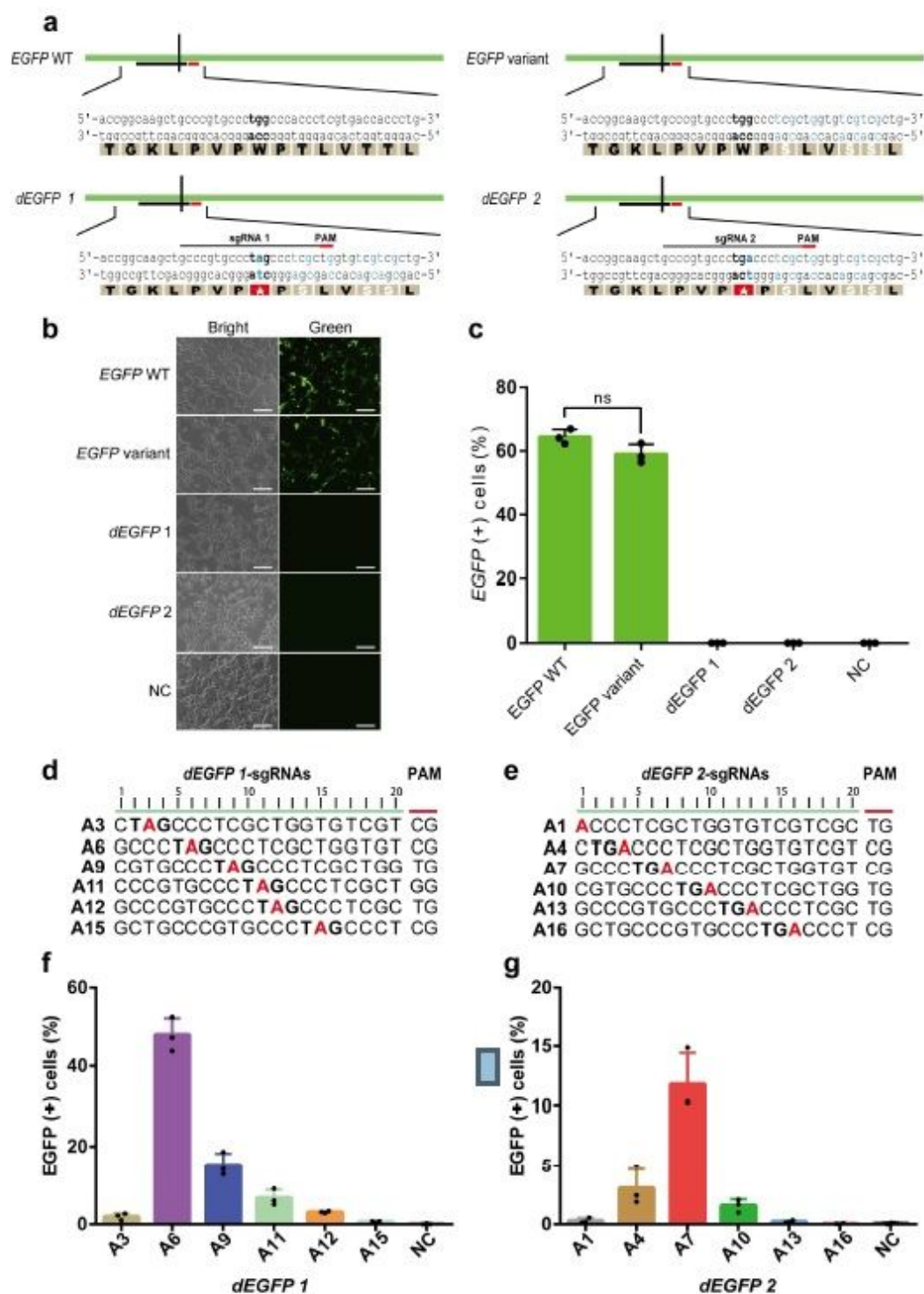
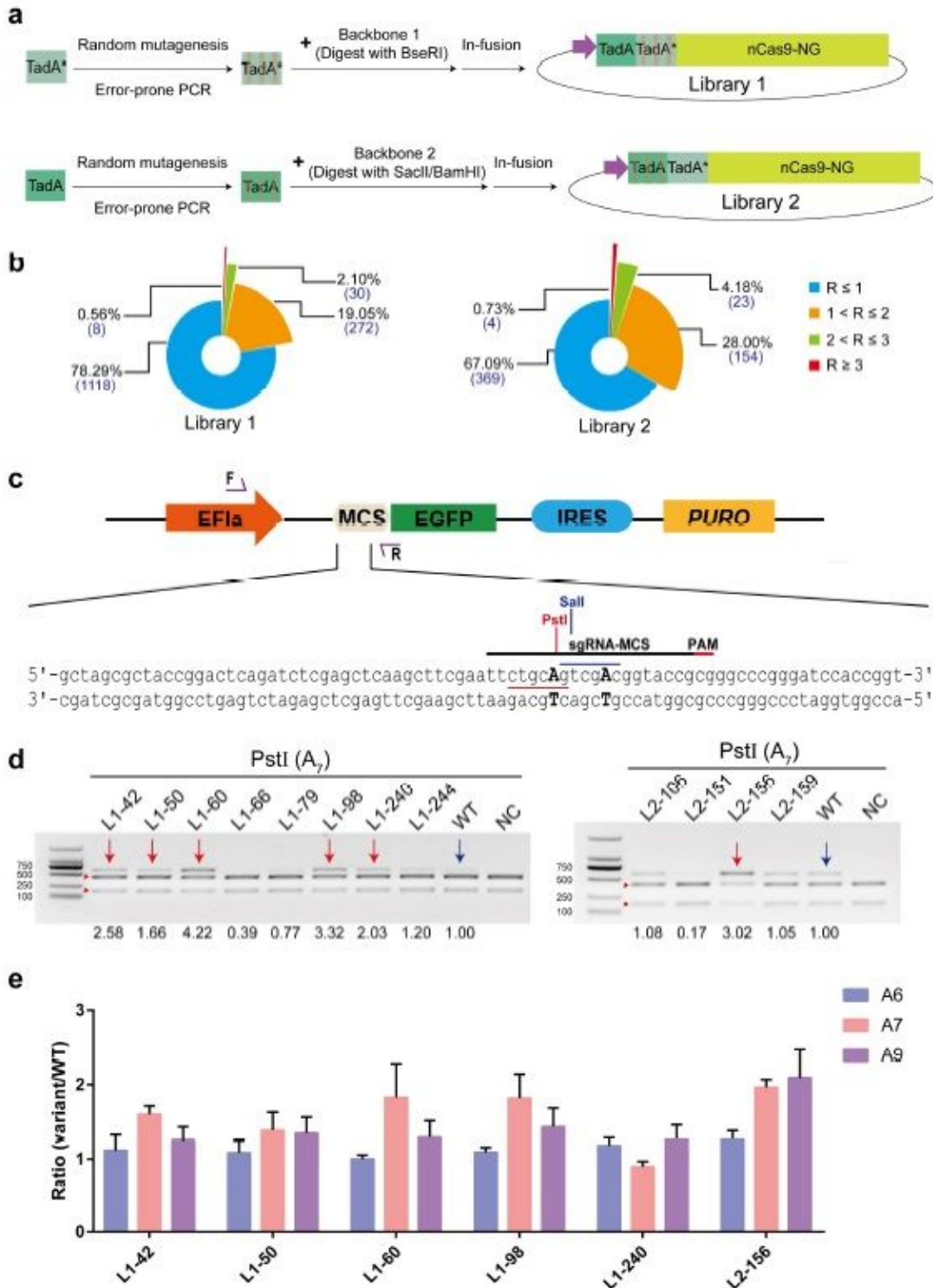


Figure 1

Design of the stop-codon reversion-based screening system. a, Different EGFP variants. Single nucleotide conversions are indicated by blue and residue substitutions are indicated by white. The stop codon was highlighted with red pentagram. b, c, Fluorescence image(b) and editing efficiencies(c) of HEK-293 cells

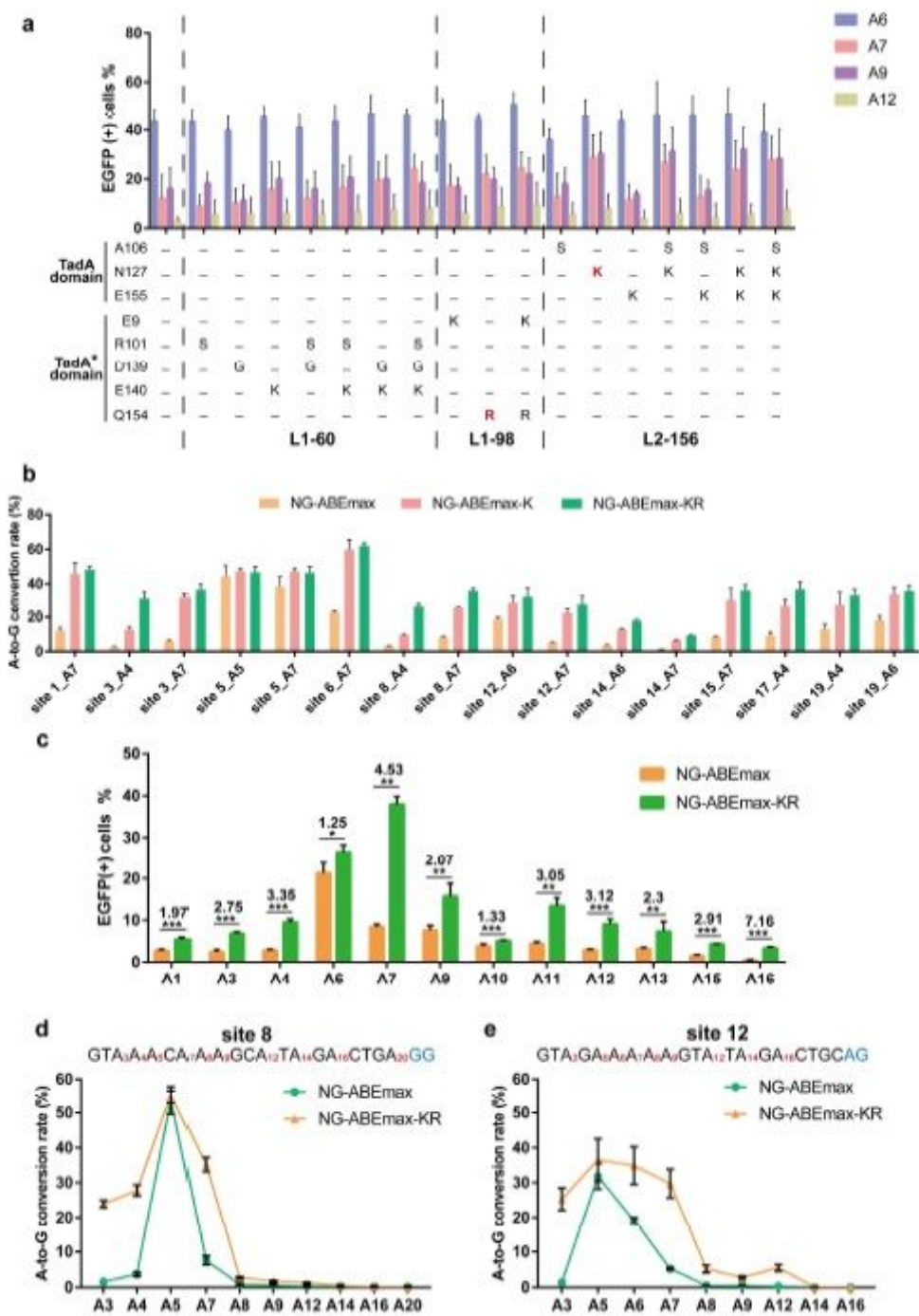
co-transfect with EGFP WT, variant, dEGFP 1 and dEGFP 2, respectively. EGFP variant maintains the expression, while dEGFP 1 and dEGFP 2 abolishes the expression. (c). EGFP positive cells were quantified by flow cytometry. Scale bar, 10  $\mu$ m. Data are mean  $\pm$  s.d. of three technical replicates. d, e, The sequence of different sgRNAs (5'-NG-3' PAM) which were suitable for dEGFP 1 and dEGFP 2, respectively. The stop codon and target A base have been highlighted in bold and red. f, g, Editing efficiencies of A to G mutations at different positions of dEGFP 1 and dEGFP 2. Error bars indicate mean  $\pm$  s.d. of three technical replicates.



## Figure 2

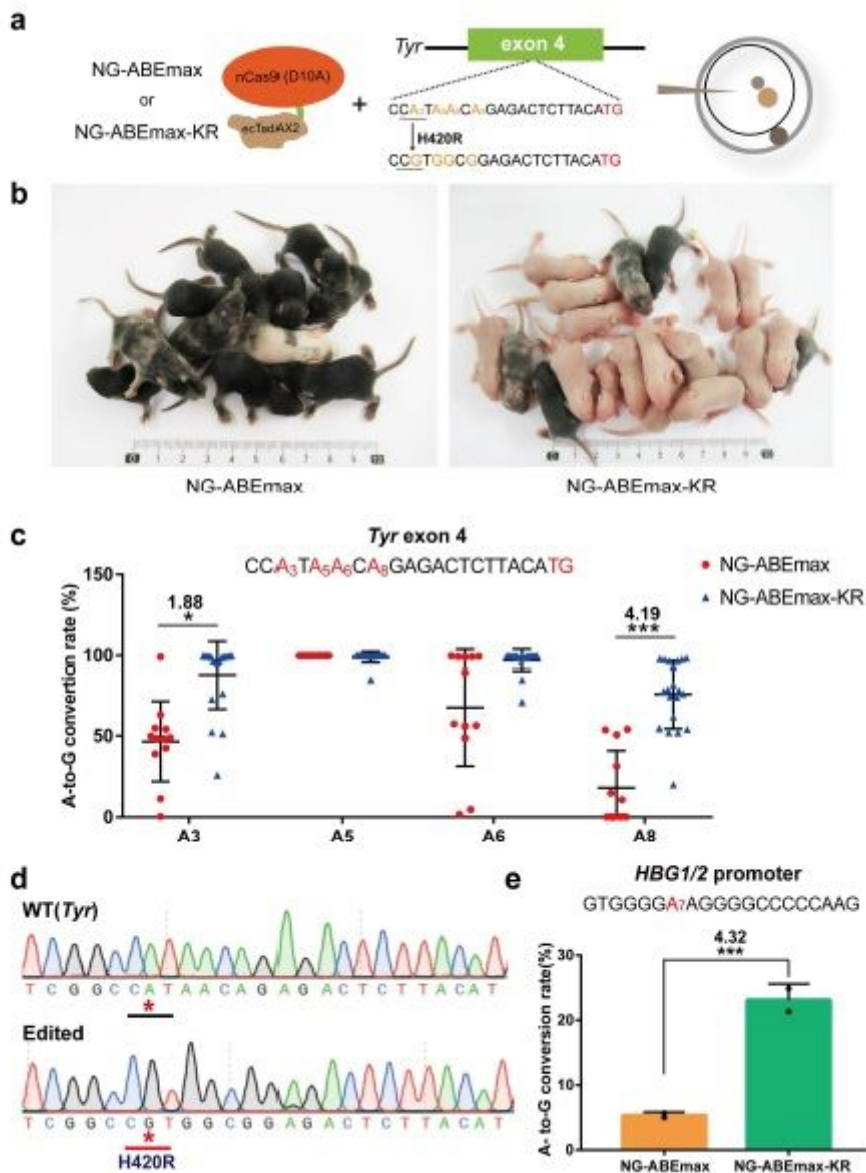
Evolution of NG-ABEmax variants with improved A•T to G•C base editing activity. a, Scheme of libraries generation design. The library 1 contain TadA\* mutations and the library 2 contain TadA mutations. Red lines represent the point mutations. b, Summary of the library 1 and library 2 (A•T to G•C base editing efficiencies at A12). R represents the ratio of editing efficiency of mutants to NG-ABEmax. c, Schematic for assessing A•T to G•C base editing efficiencies of NG-ABEmax variants using the HEK293-PME cell line. Restriction sites are highlighted in red and blue. d, Agarose gel electrophoresis results of testing editing efficiency of NG-ABEmax variants with PstI. Cleaved bands from PstI are labeled with red triangle. The amplicon is 606-bp. The variants with high-activities are highlighted with red arrows. WT represents NG-ABEmax and highlighted with blue arrows. e, Boosted editing efficiencies of selected variants via EGFP-based reporter system. Error bars indicate mean  $\pm$  s.d. of three technical replicates.





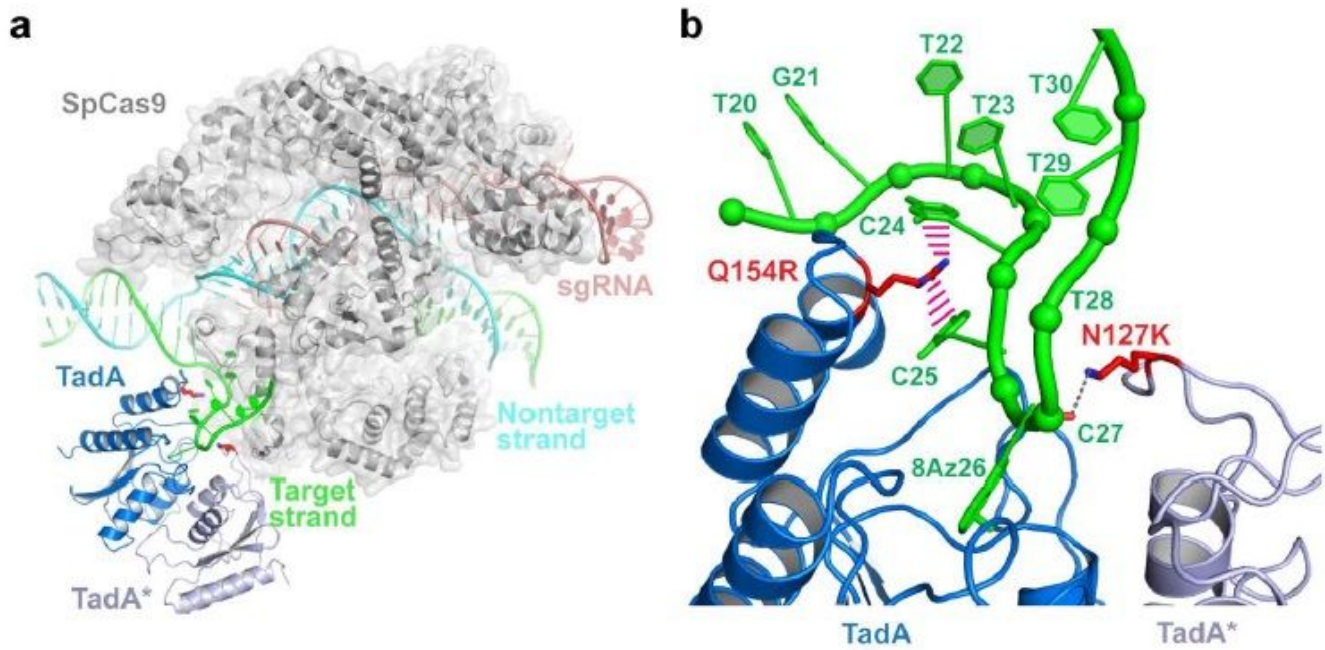
**Figure 3**

engineering of the NG-ABEmax variants. a, Mapping the key residues for L1-60, L1-98 and L2-156 using EGFP-based reporter system. Editing efficiencies were quantified by flow cytometry. b, Boosted editing efficiencies of NG-ABEmax-KR at endogenous genomic sites. c, Summary of boosted editing efficiencies of NG-ABEmax-KR at different A base at the EGFP sites. Error bars indicate mean  $\pm$  s.d. of three technical replicates. d, e, Increased editing window of NG-ABEmax-KR in site 8 and site 10. Each A base was highlighted in red. Error bars indicate mean  $\pm$  s.d. of three technical replicates.



**Figure 4**

Application of NG-ABEmax-KR for the generation of mice disease models and human gene therapy. a, Schematic of comparing editing activity of NG-ABEmax and NG-ABEmax-KR in mouse embryos via zygote intracytoplasmic injection. b, The newborn pups (days 10) produced by intracytoplasmic injection of NG-ABEmax or NG-ABEmax-KR mRNA and Tyr sgRNA. The Tyr mutant mice (H420R) are in white and the wild-type are in black, respectively. c, Statistical analysis of on-target A-to-G base conversions induced by NG-ABEmax (n = 11), NG-ABEmax-KR (n = 20) in all pups. Data are mean  $\pm$  s.d for the indicated numbers of mice. Each A base was highlighted in red. \*P < 0.05, \*\*\*P < 0.001 by Student's unpaired two-sided t-test. d, Sanger sequence chromatograms confirmed the editing events. The desired mutation was highlighted with red star. e, Boosted editing efficiency (A7) of NG-ABEmax-KR at HBG1/2.



**Figure 5**

Molecular basis of enhanced editing efficiency of NG-ABEmax-KR. a, Overall structural model of NG-ABEmax-KR showing SpCas9 (gray), target-strand DNA (TS, green), nontarget-strand DNA (NTS, cyan), single-target RNA (sgRNA, red), TadA (blue) and TadA\* (light blue). The side chains of N127K of TadA\* and Q154R of TadA are shown as red sticks. b, Modelled Interactions between key residues of NG-ABEmax-KR with TS DNA, colored as in (a).

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarydata.pdf](#)