



# Hypercompact adenine base editors based on transposase B guided by engineered RNA

Do Yon Kim<sup>1,4</sup>, Yuhee Chung<sup>1,4</sup>, Yujin Lee<sup>2,3</sup>, Dongmin Jeong<sup>2,3</sup>, Kwang-Hyun Park<sup>3</sup>, Hyun Jung Chin<sup>3</sup>, Jeong Mi Lee<sup>3</sup>, Seyeon Park<sup>1</sup>, Sumin Ko<sup>1</sup>, Jeong-Heon Ko<sup>2,3</sup> and Yong-Sam Kim<sup>1,2,3</sup>✉

**Transposon-associated transposase B (TnpB) is deemed an ancestral protein for type V, Cas12 family members, and the closest ancestor to UnCas12f1. Previously, we reported a set of engineered guide RNAs supporting high indel efficiency for Cas12f1 in human cells. Here we suggest a new technology whereby the engineered guide RNAs also manifest high-efficiency programmable endonuclease activity for TnpB. We have termed this technology TaRGET (TnpB-augment RNA-based Genome Editing Technology). Having this feature in mind, we established TnpB-based adenine base editors (ABEs). A Tad-Tad mutant (V106W, D108Q) dimer fused to the C terminus of dTnpB (D354A) showed the highest levels of A-to-G conversion. The limited targetable sites for TaRGET-ABE were expanded with engineered variants of TnpB or optimized deaminases. Delivery of TaRGET-ABE also ensured potent A-to-G conversion rates in mammalian genomes. Collectively, the TaRGET-ABE will contribute to improving precise genome-editing tools that can be delivered by adeno-associated viruses, thereby harnessing the development of clustered regularly interspaced short palindromic repeats (CRISPR)-based gene therapy.**

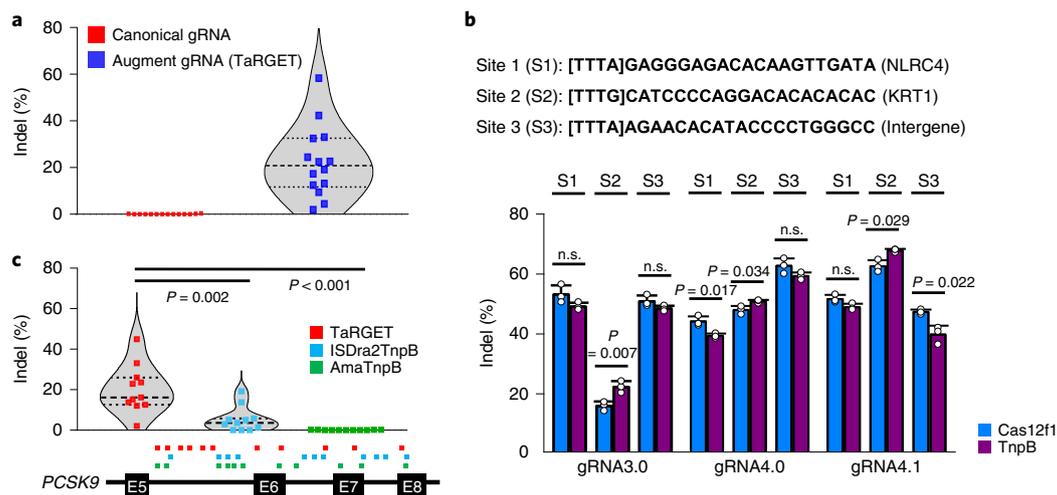
Living organisms are subjected to spontaneous genetic variations, which form the basis for biodiversity and evolution. The random nature of genetic variation as well as non-biological factors are also responsible for a variety of genetic disorders. Of the types of genetic variations identified in humans, single-nucleotide variations (SNVs) account for nearly half of disease-related mutations<sup>1</sup>. This suggests that the development of site-specific, precise genome-editing tools holds promise for the treatment of otherwise intractable genetic disorders. Clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated protein (Cas)-mediated base-editing systems have addressed the unsatisfactory modification efficiency of homology-directed repair by exploiting the high gene-targeting capability of the CRISPR-Cas systems. The catalytically inactive Cas (dCas) or nickase Cas (nCAs) fused to naturally occurring or engineered deaminases led to highly efficient single-nucleotide alterations including the C:G-to-T:A<sup>2</sup> and A:T-to-G:C<sup>3</sup> conversion. The introduction of an *Escherichia coli*-derived uracil DNA N-glycosylase also enabled C:G-to-G:C transversion<sup>4</sup>. Moreover, the base-editing systems guarantee higher levels of safety from a clinical view point because they enable precise genome editing with negligible or low levels of double-strand breaks<sup>2,3</sup>. In addition to such innate high efficiency and safety, base-editing systems have further evolved in various aspects including decreased off-target editing, enhanced conversion specificity, broadened editing windows, increased editing efficiency, and control of unwanted editing<sup>5-7</sup>.

Despite the dramatic improvements of base-editing systems per se, the challenges on the delivery side remain a major hurdle preventing base editors from being widely used in clinical applications<sup>8</sup>. Adeno-associated viruses (AAVs) are considered as a validated delivery platform owing to their relatively high delivery efficiency, low immunogenicity, relatively low level of concern over chromosomal integration, and their other useful and clinically validated features. Moreover, the identification of and engineering efforts for various AAV serotypes has expanded the range of targetable tissues

or organs<sup>9,10</sup>. Nonetheless, the multifaceted merits of AAVs are drastically compromised by their low packaging capacity of 4.7 kb. The limited cargo sizes are even more problematic for base-editing systems owing to heavy Cas proteins and an additional deaminase protein<sup>8,11</sup>. Even the adoption of a light Cas effector such as SaCas9 (3.16 kb) does not alleviate the payload limit issue for AAVs when additional essential components, in this case a promoter, one or more guide RNA (gRNA) cassettes and a poly-adenylation signal sequence are deployed in a vector. Although the oversize issue was, in part, addressed by the split two-vector system<sup>12</sup>, this led to other related challenges for clinical applications; the manufacture of AAV particles on a good manufacturing practice basis is costly, and the need to operate two different manufacturing lines can be burdensome. Furthermore, recent evidence indicates that higher dosages of AAV may induce liver toxicity<sup>13</sup>. Therefore, from a clinical perspective, it is crucial to develop an AAV-packable, but efficient base editor.

Recently, hypercompact programmable RNA-guided DNA nucleases have been suggested as a promising option for more versatile clinical applications<sup>14-16</sup>. In line with this research trend, we transformed the non-functional CRISPR-Cas12f1 system into a highly efficient genome editor through the extensive engineering of gRNA<sup>14</sup>. Here we show that our engineered RNAs can be used as similarly efficient gRNAs for transposon-associated transposase B (TnpB), an evolutionarily direct ancestor for UnCas12f1, suggesting a TaRGET (TnpB/augment RNA-based Genome Editing Technology) system. More importantly, catalytically inactive TnpB (dTnpB) fused to optimally configured deaminases offers a robust platform for an AAV-deliverable adenine base-editing (ABE) system. We also expanded the targetable range by developing protospacer adjacent motif (PAM) variants and a protospacer-protruding mutant of TnpB and using a unique combination of deaminases. This TaRGET-ABE system is expected to contribute to providing wider options for the delivery of base editors.

<sup>1</sup>GenKore, Daejeon, Republic of Korea. <sup>2</sup>KRIBB School of Bioscience, Korea University of Science and Technology, Daejeon, Republic of Korea. <sup>3</sup>Genome Editing Research Center, KRIBB, Daejeon, Republic of Korea. <sup>4</sup>These authors contributed equally: Do Yon Kim, Yuhee Chung. ✉e-mail: [ykim@genkore.com](mailto:ykim@genkore.com)



**Fig. 1 | Augment RNA-guided programmable nuclease using transposase B. a**, Violin plot for indel efficiency of TnpB guided by naturally occurring or engineered gRNA.  $n = 14$  sites. **b**, Sequence information of sites used to compare the indel-forming activity of Cas12f1 and TnpB and comparison of indel efficiency of UnCas12f1 and TnpB in the presence of various engineered gRNA versions (mean  $\pm$  s.d.,  $n = 3$  independent experiments).  $P$  values were derived by a two-tailed Student's  $t$ -test. n.s., not significant. **c**, Violin plots suggesting that the TaRGET system offers a favorable scaffold for adenine base editors owing to the high indel efficiency, as compared to other RNA-guided TnpB systems. Square dots represent the target sites within PCSK9 gene.  $P$  values were derived by a two-tailed Welch's  $t$ -test (TaRGET versus ISDra2TnpB) and by a two-tailed Mann-Whitney Rank Sum test (TaRGET versus AmaTnpB).  $n = 11$  sites.

## Results

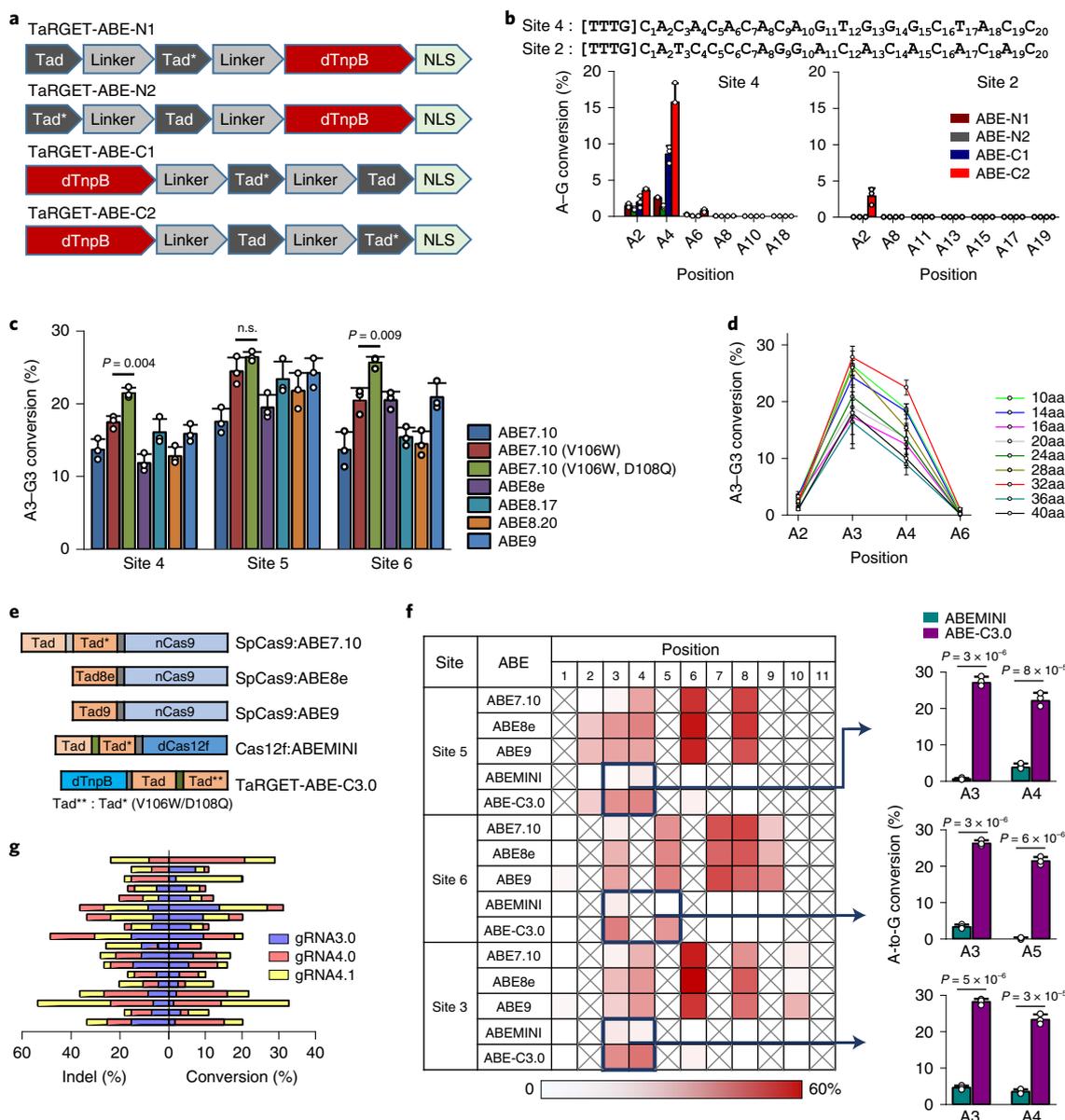
**TaRGET system.** Recently, several hypercompact ancestral genes have been reported to show programmable RNA-guided endonuclease activity. In particular, the IS200/IS605 transposon-encoded TnpB has RuvC-like domains, which can be used as a genome editor when complexed with a compatible gRNA<sup>17</sup>. Karvelis et al. showed that TnpB of *Deinococcus radiodurans* ISDra2 (ISDra2TnpB) is a programmable endonuclease that is guided by right element (RE)-derived RNA (reRNA) to cleave DNA next to the 5'-TTGAT transposon-associated motif (TAM)<sup>18</sup>. TnpB from *Alicyclobacillus macrosporangiidus* also showed an omega gRNA-specific double-stranded DNA (dsDNA) cleavage in vitro with the TCAC TAM preference, though the in vivo indel efficiency was not explored<sup>17</sup>.

Type V Cas proteins, namely Cas12 family members, are likely to have evolved from TnpB, and UnCas12f1 is an early member of the TnpB-origin Cas effectors<sup>19</sup>. Because the IS200/IS605 family element transposase accessory protein TnpB from the *Candidatus Woesearchaeota* archaeon (hereafter TnpB) shares a perfectly matched nucleotide sequence with UnCas12f1 except for 5'-terminal 28 amino acid residues (Supplementary Table 1), the TnpB endonuclease may share certain molecular properties with it, including endonuclease activity and RNA binding. To test this possibility, the RNA-guided programmable nuclease activity was investigated for TnpB. In a previous study<sup>14</sup>, we developed several engineered versions of single guide RNA (sgRNA) for Cas12f, including ge3.0, 4.0, and 4.1. Identical to Cas12f, TnpB showed no indel-formation activity with canonical gRNA in HEK293T cells. However, TnpB exhibited significantly increased indel activity with the engineered (augment) gRNAs (Fig. 1a). Despite the slight target-dependent difference in the indel levels between Cas12f and TnpB, the overall cleavage showed a nearly identical pattern, indicating the orthogonal use of the augment RNAs for TnpB (Fig. 1b). Because the base-editing efficiency usually, but not necessarily, depends on the indel efficiency of wild-type Cas effector proteins, it is required to start with a Cas system that shows sufficiently high indel efficiency. Thus, we compared the indel efficiencies of TaRGET, ISDra2TnpB, and AmaTnpB at the PCSK9 loci in HEK293T cells. The target sites are not exactly shared owing to differences in respective PAM

sequences. Therefore, 11 sites between exon 5 and exon 8 were selected instead in the PCSK9 genome sequence. Interestingly, the TaRGET system showed substantially higher indel efficiency, in comparison to those of ISDra2TnpB and AmaTnpB (Fig. 1c). Therefore, we concluded that the TaRGET system is feasible for use as a platform for the development of a compact base-editing system.

**Feasibility of dTnpB-based adenine base editors.** On the basis of the information on the residues involved in the catalytic activity for Cas12f<sup>20,21</sup>, we constructed four catalytically inactive mutants of TnpB (D354A, E450A, R518A, and D538A). Each mutant was tested as to whether the DNA cleavage activity was completely eradicated while allowing the preservation of the gene-targeting capability. An in vitro DNA digestion assay and an indel assay in HEK293T cells revealed that all of the mutants had null endonuclease activity (Extended Data Fig. 1), and we selected a dTnpB (D354A) mutant from a previous CRISPRa experiment<sup>14</sup>.

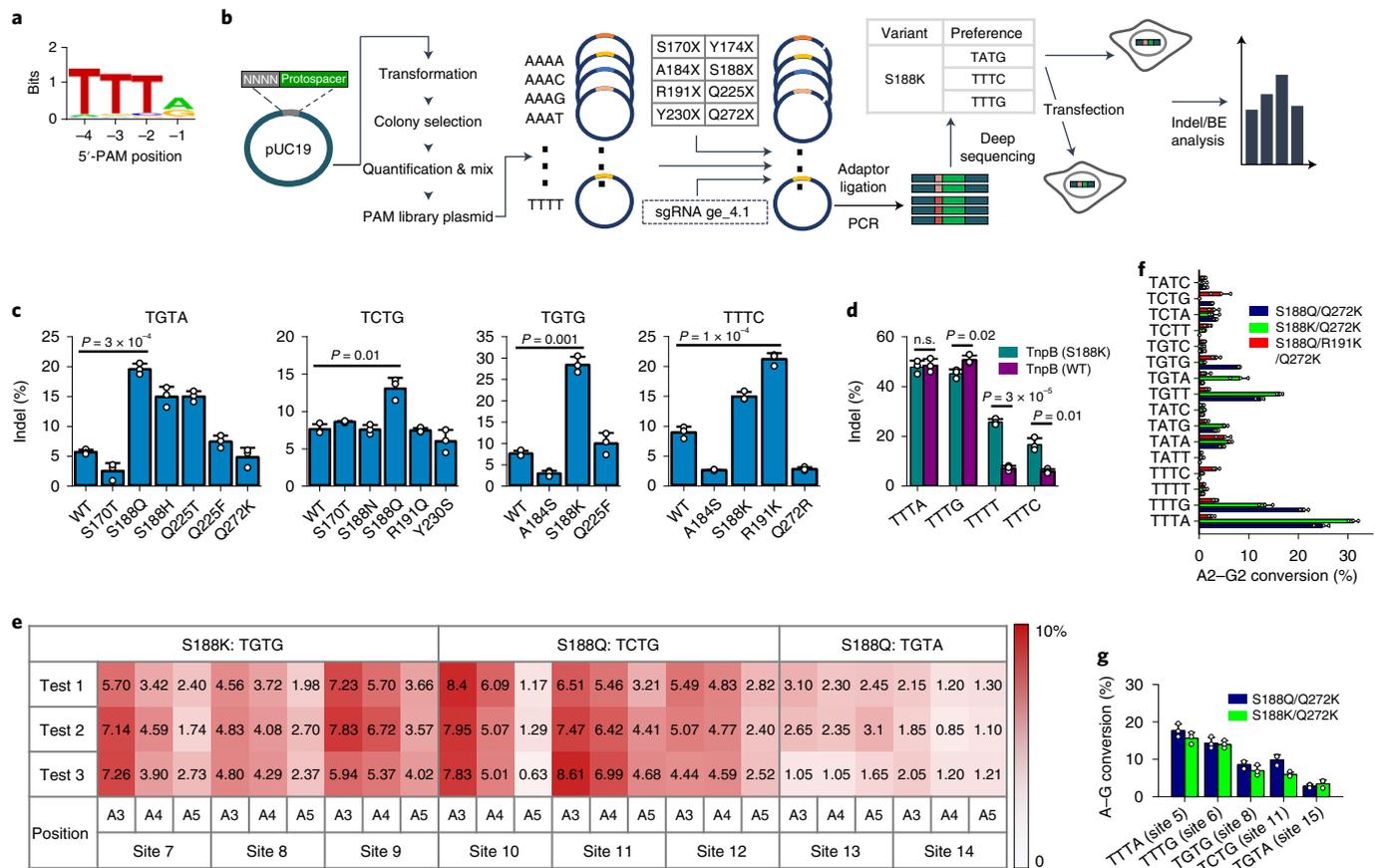
A size-exclusion chromatogram of purified TnpB was used to estimate the molecular mass of sgRNA-bound TnpB as approximately 194 kDa, suggesting that TnpB formed a homo-dimer in the presence of engineered gRNA, similar to UnCas12f1<sup>20,21</sup> (Extended Data Fig. 2). In this case, the orientation of deaminase fusion may affect the base-editing property. To test this possibility, we constructed TnpB-based adenine base editors by fusing the wild-type Tad-mutant Tad (Tad-Tad\*) or Tad\*-Tad to either the N or C terminus of dTnpB<sup>3</sup> (Fig. 2a). These constructs were tested for two validated targets<sup>14</sup>, one showing an A-rich sequence at the PAM-proximal region and the other showing this sequence at the PAM-distal region (Fig. 2b). The deaminase architectures fused to the C-terminal orientation (ABE-C1 and ABE-C2) showed substantial levels of A-to-G conversion activity, whereas the N-terminally fused modules (ABE-N1 and ABE-N2) showed marginal conversion activity. Conversions were only observed at the PAM-proximal regions. Because ABE-C2 showed higher conversion rates in comparison to the other ABEs, it was used to identify a base-editing window. In an experiment where two PAM-proximal A-rich sequences were targeted, a conversion was only observed within the window of A2 to A6, with the most prominent conversion



**Fig. 2 | Feasibility test and optimization of the TaRGET-ABE system.** **a**, The structures and designations of TaRGET-ABE modules according to the orientation and order of the wild-type and mutant Tad (Tad\*). **b**, Investigation of the A-to-G conversion efficiency for two sites in HEK293T cells, each of which carries multiple A sequences at the PAM-proximal or PAM-distal regions (mean ± s.d., n = 3 independent experiments). **c**, A variety of Tad modules as a form of either an engineered monomer or heterodimer fused to the C terminus of dTnpB (D354A). The dTad (D354)-Tad-Tad\* (V106W, D108Q) construct produced the highest substitution efficiency, designated as TaRGET-ABE-3.0 (mean ± s.d., n = 3 independent experiments). Statistical tests were performed by a two-tailed Student's *t*-test. **d**, Determination of the optimized length of a linker used to connect dTnpB and Tad. The length of a linker between Tad and mutant Tad for dimeric Tad modules were fixed at 32 amino acids (mean ± s.d., n = 3 independent experiments). **e**, Architecture of SpCas9-, Cas12f-, and TnpB-based adenine base editors. Tad\*\* denotes the engineered form with V106W and D108Q mutations in Tad\* protein. **f**, Heat map of the A-to-G conversion efficiency at three different sites for SpCas9-, Cas12f-, and TnpB-based adenine base editors. X denotes the non-A sequence. In the panel on the right, the conversion efficiencies were compared between Cas12f-based ABEMINI and TaRGET-ABE-C3.0 (mean ± s.d., n = 3 independent experiments). **g**, Dependence of indel and the A-to-G conversion efficiencies on the types of engineered sgRNAs. n = 18 sites. P values were derived by a two-tailed student's *t*-test.

activity observed at positions A3 and A4 (Extended Data Fig. 3). Base editors evolved with various engineered versions of Tad<sup>22</sup>. Thus, we compared the A3-to-G3 conversion rates as a surrogate for various Tad variants developed thus far, and found that the architectures of the codon-optimized Tad-Tad\* (V106W, D108Q)<sup>23,24</sup> showed the highest conversion rates, as compared to any of the other forms (Fig. 2c). We presented the optimized Tad dimer as

Tad-Tad\*\* and designated this ABE form as TaRGET-ABE-C3.0. The length of the linker connecting dTnpB and a Tad dimer was determined to be 32 amino acids (Fig. 2d and Supplementary Table 1). TaRGET-ABE-C3.0 was compared with several adenine base editors with respect to the editing windows and conversion efficiency (Fig. 2e). As mentioned above, the base-editing window of TaRGET-ABE-C3.0 was formed at a relatively narrow range,



**Fig. 3 | Expanding targetable sites by PAM variants of TnpB engineering.** **a**, Sequence logo analysis of the PAM preference of the wild-type TnpB. **b**, Flow chart describing the development of PAM variants of TnpB and the application of the PAM variants to expanded adenine base editing. A validated target sequence in an *NLRC4* locus (Site 1) was used throughout the experiments including in vitro cleavage and validation in PAM-altered cells. **c**, Investigation of the indel-forming activities of PAM variants for other PAM sequences than the canonical TTTR sequence (mean  $\pm$  s.d.,  $n=3$  independent experiments).  $P$  values were derived by two-tailed Student's  $t$ -test. **d**, Preference of the S188K TnpB variant for expanded TTTN PAM sequences (mean  $\pm$  s.d.,  $n=3$  independent experiments).  $P$  values were derived by two-tailed Student's  $t$ -test. **e**, Heat map showing the A-to-G conversion efficiencies of the PAM variants, S188K and S188Q TnpB PAM variants at non-TTTR PAM sites. Sites with As at positions 3, 4, and 5 were selected to trace the conversion efficiencies. Tests 1–3 indicates triplicate experiments. **f**, Further expansion of targetable sites by stacked PAM mutations (S188Q/Q272K and S188K/Q272K). Efficiency was tested in PAM-altered HEK293T cells (mean  $\pm$  s.d.,  $n=3$  independent experiments). **g**, Multiplexed adenine base editing at sites with various PAM sequences by the S188Q/Q272K and S188K/Q272K PAM variants of TnpB (mean  $\pm$  s.d.,  $n=3$  independent experiments).

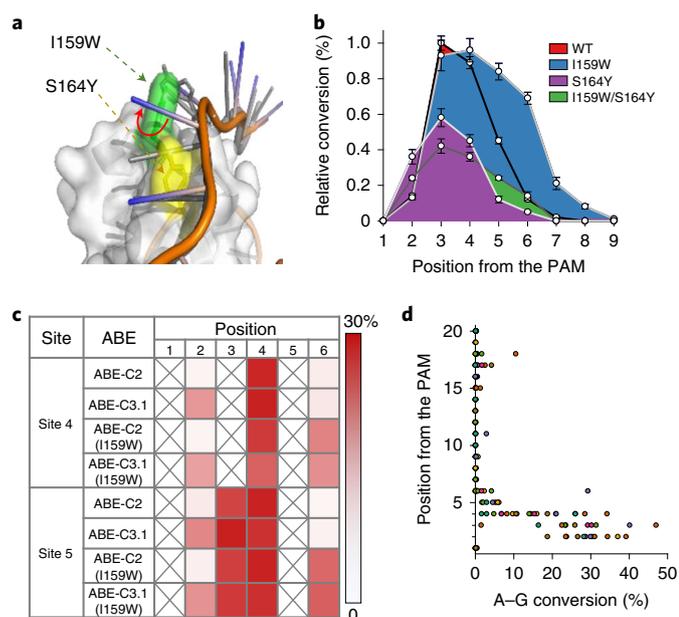
similar to the recently reported ABEMINI<sup>15</sup>. However, the overall conversion efficiency of TaRGET-ABE-C3.0 was significantly higher than that of Cas12f-based ABEMINI, though it was lower than those of SpCas9 nickase-based ABEs, such as ABE7.10 (ref. <sup>3</sup>), ABE8e<sup>25</sup>, and ABE9 (ref. <sup>26</sup>) (Fig. 2f). ABE8e uses monomeric Tad as an engineered form<sup>25</sup> and ABE9 adopts further engineered monomeric Tad protein (V82S/Q154R)<sup>26</sup>.

In a previous study, we presented three different versions of sgRNA for UnCas12f1, ge3.0, ge4.0, and ge4.1, indicating that the selected sgRNA version would affect the base-editing efficiency. To investigate this possibility, we selected 18 targets that show different indel activity outcomes depending on the sgRNA version. Similar to earlier work, 15 out of 18 sites showed the correlation between the gRNA version and indel/conversion efficiency (Fig. 2g). That is, the selection of the most suitable gRNA must come first for the most desirable base-editing outcome. Taken together, TnpB-ABE-C3.0 guided by an optimal gRNA version shows optimal base-editing performance.

**Expanding targetable sites via TnpB and Tad engineering.** Identical to Cas12f1 (ref. <sup>27</sup>), the in vitro cleavage assay indicated that TnpB showed a PAM preference for TTTR (TTTA and TTTG),

which means that targetable sites are quite restricted (Fig. 3a). Thus, we attempted to develop TnpB mutants with preference to non-TTTR PAM and to apply PAM variants to a wider range of sites. To do this, we initially constructed a PAM library vector, which was achieved by securing individual PAM clones ( $4^4=256$  clones) and then mixing them at an equal molar ratio to ensure even distributions of each. The PAM library vectors were digested with sgRNA ge\_4.1 and the TnpB PAM variant proteins. The cleaved vectors were amplified by adaptor ligation and PCR. A deep-sequencing analysis enabled PAM variant–PAM preference matching (Fig. 3b). To select PAM variants with retained dsDNA cleavage activity levels, we prepared different HEK293T clones each carrying different PAM mutants at an *NLRC4* locus via homology-directed repair (Supplementary Fig. 1). This approach would make it possible to derive the relative indel efficiency of each PAM variant.

Because TnpB shows an identical PAM preference and shows sequence conservation in the DNA-binding region with UnCas12f1<sup>27</sup>, we selected candidate amino acids on TnpB on the basis of the structural characterization of UnCas12f1<sup>20,21</sup>, specifically, S170, Y174, A184, S188, R191, Q225, Y230, V271, and Q272. Each candidate site was mutated into all possible 19 amino acids,



**Fig. 4 | Switching a base-editing window by the engineering of Tad and TnpB.** **a**, Modeling of amino acid substitution for shifting or altering an editing window in the PAM-proximal region, indicating Ser164 and Ile159 as the substitution candidate sites. **b**, Base-editing window expanded by the substitution of Ile159 with tryptophan. Relative values were obtained by deriving the conversion efficiencies at each position for three different sites (mean  $\pm$  s.d.,  $n = 3$  independent experiments). **c**, Alteration of a base-editing window through Tad optimization. Values are the means of triplicate experiments. **d**, Multiple-site validation of TaRGET-ABE-C3.1 for A-to-G conversion activity. Values are the means from triplicate experiments. In total, 25 sites were investigated.

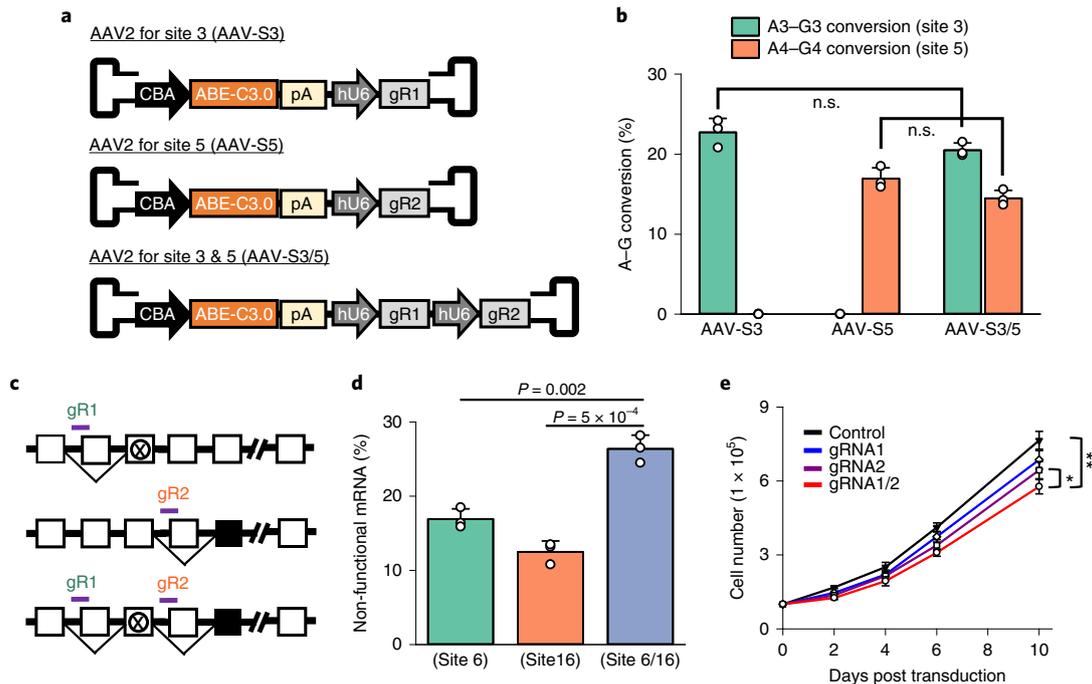
and each PAM variant candidate was tested with respect to the dsDNA cleavage activity in vitro for the altered PAMs as described in the scheme of Fig. 3b. The PAM variant candidates were selected for the criteria of (1) high total sequencing reads and (2) a high sequencing read ratio for a specific PAM. The in vitro cleavage and the deep-sequencing analysis enabled the screening of PAM variant candidates (Supplementary Fig. 2). The results indicate that several variants showed redundantly identical PAM preference. For instance, the S170T, S188Q, S188H, Q225T, Q225F, and Q272K variants showed a high TGTA PAM preference. Among the variants, the S188Q variant showed highest indel frequency for TGTA PAM, when tested in PAM sequence-altered HEK293T cells as described in Supplementary Fig. 1. Likewise, the S188Q, S188K, and R191K mutants showed high indel frequencies for TCTG, TGTG, and TTTC PAMs, respectively (Fig. 3c). The S118K variant showed broader PAM specificity, in this case TTTT and TTTC as well as TTTR, that is, TTTN (Fig. 3d). To test the application of the PAM variants to adenine base editing in a non-TTTR PAM context, the A-to-G conversion activity of the PAM variants were tested for different sites with altered PAMs. As shown in Fig. 2e, each PAM variant showed different levels of A3-to-G3, A4-to-G4, and A5-to-G5 conversion activities. A suitable PAM variant can be selectively used for a specific sequence context. Alternatively, a variant showing a multi-PAM preference can be used for multiplexed base editing. To explore this possibility, we tested the base-editing activity of variants with stacked PAM mutations and found that the S188Q/Q272K and S188K/Q272K variants showed a broader PAM preference including TTTA, TTTG, TTTC, TATG, TGTA, TGTG, TGTC, and TCTG (Fig. 3f). The base-editing activity of the stacked mutants was measured in HEK293T cells carrying the PAM-altered

*NLRC4* loci as described in Supplementary Fig. 1. The multiplexed base editing was validated in wild-type HEK293T cells by transfecting the stacked PAM mutants of TnpB together with multiple gRNAs. The S188Q/Q272K and S188K/Q272K variants showed the feasibility of base editing at five different endogenous loci, simultaneously (Fig. 3g). Collectively, the engineering of TnpB expanded the occupancy of targetable base-editing sites from 0.78% to 3.12%.

Despite the expansion of targetable sites using PAM variants, the editable incidence was still limited because a prominent editing window formed at positions 3 and 4 (this feature is occasionally favorable for specific editing). Expanding or shifting the window could be an additional option by which to expand the applicability of the TaRGET-ABE system. Structural modeling of the TnpB-gRNA ribonucleoprotein complex identified possible mutation sites at Ile159 and Ser164. The model indicates that the bases at position of 5 and 6 are concealed in a pocket of the WED domain (Supplementary Fig. 3). We speculated that the replacement of Ile159 and Ser164 with a bulky amino acid would make the bases of positions 5 and 6 more protruding (Fig. 4a), which would make deaminases more accessible to those bases. We created the I159W and S164Y mutants and applied them to adenine base editing for several targets, each of which carrying A at a different position. When we compared the editing efficiency of the variants with that of wild-type TaRGET-ABE-C2, the S164Y mutant led to a dramatic compromise in the A-to-G conversion rates at positions 3 and 4 without window expansion. However, the I159W mutant upheld the conversion rates at the positions 5 and 6 with retained A3 and A4 conversions (Fig. 4b). The dead mutant dTnpB (D538A) was used for the I159W Tad mutant. The last approach was related to a divergent architecture of the deaminase module. While constructing various combinations of Tad variants, we fortuitously found that dTnpB-Tad-Tad8e (WQ) modules showed a window expansion at position 2. The eTad (Tad8e) sequence was originally used as a monomeric deaminase for the ABE8e version<sup>25</sup>. The fusion of the Tad-Tad8e (WQ) dimer module to dTnpB (D354A), hereafter referred to as TaRGET-ABE-C3.1, induced dramatically boosted conversion at position 2 with sustained conversion efficiency outcomes at positions 3 and 4 (Fig. 4c).

We attempted to validate the TaRGET-ABE-C3.1 system for 25 endogenous sites (Fig. 4d, Supplementary Table 2 and Supplementary Table 3). The distribution confirmed the most prominent base editing at positions 2–5 without the application of the I159W mutation. Nonetheless, we identified two sites for which positions 17 and/or 18 was edited at a relatively high-efficiency rate. The PAM-proximal editing may arise from the gRNA-independent editing within R-loop<sup>28,29</sup> or from the formation of non-canonical windows. The head-to-head comparison of all versions of TaRGET-ABE in terms of base-editing window and efficiency were made in Extended Data Fig. 4. Taken together, the engineering and reconstruction of TnpB and Tad modules largely broadened the otherwise highly restricted base-editing range by both expanding targetable PAMs and shifting or expanding the base-editing windows.

**Adenine base editing in vitro via AAV delivery.** nSpCas9 (D10A)-based adenine base editors enable highly efficient A-to-G conversions in eukaryotic cells when delivered by plasmid vectors<sup>3,25</sup>. However, the AAV delivery is limited owing to the oversized deaminase-dCas9 modules<sup>8</sup>. This limitation can be overcome by using a split-AAV vector delivery<sup>12</sup> or miniABE8e<sup>30</sup>. While all of the engineered ABEs compromised the full activity of Cas9-based ABE systems, our TaRGET-ABE system is sufficiently compact that it can be delivered in an all-in-one AAV vector. Furthermore, there remains space for additional cargo within a payload size limit of ~4.7 kb. One of the applications utilizing the additional cargo space would be multiplexed base editing. We produced AAV2 particles where the TaRGET-ABE-C3.0 system was charged with one sgRNA



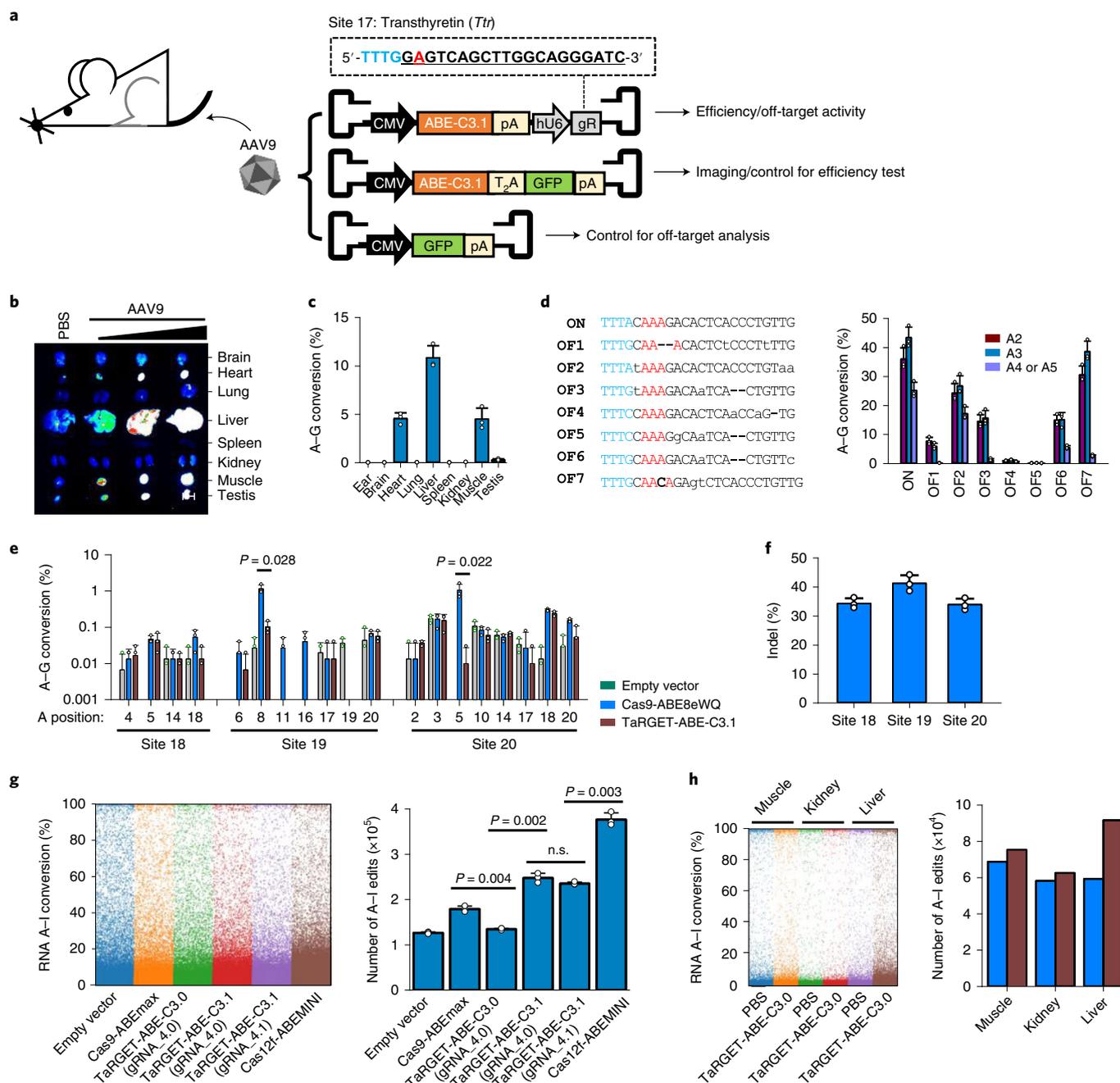
**Fig. 5 | Validation of base-editing activity of the TaRGET-ABE-C3.0 system through AAV delivery in vitro.** **a**, AAVs vector constructs harboring a TaRGET-ABE-C3.0 cassette and single or paired sgRNAs. **b**, Multiplexed adenine base editing by harboring paired sgRNAs in a single AAV particle (mean  $\pm$  s.d.,  $n = 3$  independent experiments). Statistical tests were performed by a two-tailed Student's *t*-test. **c**, Schematic illustration showing the increased generation of non-functional mRNA using paired gRNAs.  $\otimes$ , non-sense mutation leading to an early termination of expression;  $\blacksquare$ , non-functional exon owing to the deprivation of a former one. **d**, Increased levels of non-functional ErbB4 mRNA by paired gRNAs in H661 cells (mean  $\pm$  s.d.,  $n = 3$  independent experiments). *P* values were derived by a two-tailed Student's *t*-test. **e**, Growth inhibition of the TaRGET-ABE-C3.0 system harboring paired gRNA in H661 cells (mean  $\pm$  s.d.,  $n = 7$  independent experiments). The cell number was counted from the day of sub-culturing for 10 days. *P* values were derived by a Fisher's least significant difference post hoc test. \**P* = 0.033, \*\**P* = 0.007.

(sgRNA 1 for site 3 or sgRNA 2 for site 5) or paired sgRNAs simultaneously targeting site3 and site5 (Fig. 5a). HEK293T cells were transduced at a multiplicity of infection (MOI) of 100,000 for vector systems for 10 days, during which the cells were sub-cultured 5 days after initial transduction and the MOI was kept constant through additional treatments of AAV particles with fresh medium. When a single sgRNA was loaded onto the AAV vector, a target-specific base editing was achieved (Fig. 5b). Interestingly, we were able to perform multiplexed A-to-G conversions using paired sgRNAs in a single AAV particle. Moreover, the conversion efficiency at each site obtained using the paired gRNA-AAV particles was not compromised ( $P > 0.05$ ), as compared to those obtained using one sgRNA-charged AAV particle.

The capability of charging paired sgRNAs in an all-in-one AAV vector system can make an additive effect on substitution-based therapeutic strategy of certain diseases. We illustrate this concept for a possible treatment strategy for cancer. Epidermal growth factor receptor 4 (ErbB4; HER4) is a kinase that stimulates oncogenesis and cancer progression in many cancer types, and chemical or biological inhibitors are used for the treatment of cancer<sup>31</sup>. We loaded two sets of sgRNAs together with TaRGET-ABE-C3.0 in an all-in-one AAV vector. sgRNA 1 aims to induce exon skipping and a frame-shift mutation by substituting a splicing acceptor consensus sequence (–AG–) with a splicing-skipping sequence (–GG–)<sup>32,33</sup>. sgRNA 2 induces the skipping of the exon involved in the binding of growth factors. The use of either one of the two sgRNAs can produce non-functional receptors, but the concomitant charge of the two sgRNAs can further increase the frequencies of the occurrence of non-functional ErbB receptors (Fig. 5c). We screened two intron-exon interface targets that meet the requirements of PAM and the

reading frame: one (sgRNA 1) is at the intron I–exon 2 and the other (sgRNA 2) at intron III–exon 4. AAV2 particles carrying either one of the two sgRNAs or both were produced in HEK293T cells and treated in H661 cells at an MOI of  $10^5$ . TaRGET-ABE-C3.0 carrying sgRNA 1 or sgRNA 2 produced non-functional mRNAs with at the frequency of  $17.3 \pm 2.5\%$  and  $13.2 \pm 2.3\%$ , respectively. However, the percentage of non-functional mRNAs increased to  $26.8 \pm 3.7\%$  for AAV particles carrying both sgRNAs (Fig. 5d and Supplementary Fig. 4). This additive effect was manifested in the stalling of the growth of cancer cells, where the two sgRNAs collaborated to retard the growth of ErbB4-positive H661 cells (Fig. 5e). Taken together, these results suggest that the hypercompact TnpB-based adenine base editors provide a useful and precise genome-editing tool delivered by AAV. It is important to note that a more universal PAM variant would expand the application of the exon-skipping strategy for gene knockout, particularly genes consisting of a few exons, including the transthyretin (TTR)<sup>34</sup> and proprotein convertase subtilisin/kexin type 9 (PCSK9)<sup>35,36</sup>.

**Specificity and A-to-G conversion in vivo via AAV delivery.** Next, we attempted to explore the feasibility of in vivo base editing using AAV-delivered TaRGET-ABE modules and to gauge the specificity of TaRGET-ABE system. We targeted a locus in the murine transthyretin (*Ttr*) gene, where editable A is in position 2 from the PAM. AAV9 harboring the TaRGET-ABE-C3.1 module was administered by tail vein injections. AAV9 particles carrying TaRGET-ABE-C3.1 and T2A-linked GFP construct without gRNA cassette was used for imaging and as a control for efficiency tests. GFP-carrying AAV9 particles were used as a control for off-target analysis (Fig. 6a). Intensive GFP expression by AAV9 delivery was observed in the



**Fig. 6 | Assessment of base-editing activity in vivo via AAV delivery and off-target property of TaRGET-ABE system.** **a**, AAV9 vector constructs harboring a TaRGET-ABE-C3.1 and an sgRNA targeting a transthyretin loci. **b**, Fluorescence images of tissues from mice transduced by AAV9 harboring GFP gene. Scale bar, 1 cm. The image shows representative tissues obtained from three transduced mice. **c**, A-to-G conversion efficiency at a *Ttr* loci for various tissues obtained from AAV9-transduced mice (mean  $\pm$  s.d.,  $n=3$  independent experiments). **d**, gRNA-dependent off-target levels at the seven potential off-target sites (mean  $\pm$  s.d.,  $n=3$  independent experiments). **e**, gRNA-independent off-target activity of TaRGET-ABE-3.1 at R-loops formed by CRISPR-dAsCas12a in comparison with that of Cas9-ABE8e (WQ) system (mean  $\pm$  s.d.,  $n=3$  independent experiments).  $P$  values were derived by a two-tailed Student's  $t$ -test. **f**, Indel-forming efficiency of CRISPR-AsCas12a at sites 18–20 used for R-loop off-target activity (mean  $\pm$  s.d.,  $n=3$  independent experiments). **g**, gRNA-independent RNA editing by various ABEs in HEK293T cells. The left panel indicates a representative Jitter plot of triplicated experiments. The right panel shows the number of A-to-I RNA reads in HEK293T cells transduced by AAV2 particles harboring ABEmax, TaRGET-ABE-3.0, or ABEmini (mean  $\pm$  s.d.,  $n=3$  independent experiments).  $P$  values were derived by a two-tailed Student's  $t$ -test. **h**, gRNA-independent RNA editing by TaRGET-ABE-3.0 in the muscle, kidney, and liver of AAV9-transduced mice.

heart, liver, muscle, and testis (Fig. 6b). Deep-sequencing analysis indicated that A2-to-G2 conversions were observed in the tissues with high GFP expressions, among which liver shows the highest conversion efficiency of  $10.9 \pm 1.2\%$  four weeks after transduction.

Moderate conversion efficiencies were also observed in the heart and muscles (Fig. 6c).

Biased gRNA-dependent off-target activity was investigated at seven potential off-target loci that were screened using

Cas-OFFinder<sup>37</sup>, which have TTTR PAM sequence for five sites and TTTC for two sites (OF4 and OF5). The potential off-target sites were high sequence similarities in the PAM-proximal regions to on-target sites, and all of them carry 3 As in the position 2–5 from the PAM. For off-target sites with TTTR PAM sequence, non-trivial levels of off-target conversions were observed for all sites (Fig. 6d). In particular, OF7 showed almost comparable A-to-G conversion efficiency to that of the on-target site, indicating that efforts to improve specificity of TaRGET-ABE system need to be made. One more off-target issue is related to a gRNA-independent DNA and RNA editing by deaminases. First, the gRNA-independent conversion of DNA was investigated in the R-loop regions formed by CRISPR-dCas12a system<sup>28,29</sup>. Out of three sites tested, two sites showed the unguided conversions over 1% by both Cas9-based ABE8e (WQ). However, the gRNA-independent off-target base-editing activity was significantly lower for TaRGET-ABE-C3.1, as compared to the Cas9-based ABE8e (WQ) for those sites (Fig. 6e). Nonetheless, the overall R-loop off-target activity at those sites was not high for both base-editing editors in spite of high indel-forming activity of the catalytically active Cas12a (Fig. 6f). Finally, the gRNA-independent RNA editing was monitored using RNA sequencing analysis *in vitro* as well as *in vivo* following the forced overexpression of various base editors<sup>23</sup> (Fig. 6g,h). Previous results indicated that ABE8e (WQ) showed reduced RNA editing frequency<sup>23,24</sup>. When compared to ABE8e (WQ) and ABEMINI, the TaRGET-ABE-3.0 system showed a lower level of A-to-I conversion rates<sup>15</sup>. However, it is noteworthy that TaRGET-ABE-C3.1 has higher levels of RNA A-to-I editing, as compared to TaRGET-ABE-C3.0 (Fig. 6g). As expected, there were no differences in the RNA-editing rates by the use of different version of gRNA (that is, gRNA\_4.0 versus gRNA\_4.1). The RNA A-to-I conversion level was also observed *in vivo* in proportion to infectivity of AAV particles (Fig. 6h). These results indicate that efforts to improve TaRGET-ABE system particularly with respect to gRNA-dependent off-target activity need to be made, including the previous ones to mitigate these side effects<sup>5</sup>.

## Discussion

Precise genome editing would provide a radical treatment option for various genetic disorders<sup>38,39</sup>, most of which rely on symptomatic therapies without a curative regimen. The base-editing strategy is assumed to bring significant clinical benefits for potential patients with a single substitution mutation. Recently, an adenine base editor was proposed for safer gene knockout without dsDNA breaks through an exon-skipping strategy<sup>35,36</sup>. Despite these wider applications to genetic manipulations *in vivo*, the clinical utility of adenine base editors has been largely restricted owing to delivery failure when using AAV particles. Nguyen et al., made an arduous effort to deliver an nCas9-based adenine base editor in the all-in-one AAV vector<sup>30</sup>, but the conversion efficiency was quite low (less than 1%). Lipid nanoparticles are considered as a delivery system for these heavyweight CRISPR editors, but clinically approved ones are confined to the targeting of the liver tissue<sup>34</sup>. Thus, the development of hypercompact TaRGET-ABE would expand the spectrum of precise genome-editing-based treatments. The TaRGET-ABE system would also make the development of gene therapy applicable to a wider range of genetic diseases.

The property of persistency related to AAV-assisted gene delivery provokes safety concerns over *in vivo* gene therapy. Because the TaRGET-ABE system adopts catalytically dead TnpB instead of a nickase, the issue of a residual level of indel mutations that are occasionally observed for nickase Cas proteins can be excluded. However, the issue of gRNA-independent as well as gRNA-dependent off-target activity remains<sup>5</sup>. Our previous study indicated less tolerance to mismatches and a low incidence of off-targets for the CRISPR-Cas12f1 system, but the TaRGET system showed the ‘higher-than-expected’ levels of gRNA-dependent off-target conversion activity. Furthermore, gRNA-independent off-target concerns must

be further addressed and, TaRGET-specific strategies to mitigate those concerns need to be developed.

We attempted to develop a cytosine base editor using dTnpB by employing various cytidine deaminases validated beforehand. However, we found that the C-to-T conversion efficiency was not satisfactory (in fact, <5% for tested targets) in HEK293T cells. This low efficiency may arise from the accessibility of fused deaminases to nucleotides within editing windows, which is possibly due to incompatible orientation or any structural hindrance. A scrutinized design of deaminase modules on the basis of structural modeling would enable one to render a functional TaRGET-CBE system. Furthermore, the development of a nickase TnpB, which could be obtained possibly by an alternative strategy to the previous sequence manipulations inside the RuvC domain, might allow one to establish a prime editing system that can be delivered in an all-in-one AAV vector system.

## Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41589-022-01077-5>.

Received: 4 February 2022; Accepted: 2 June 2022;

Published online: 01 August 2022

## References

- Bamshad, M. J. et al. Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12**, 745–755 (2011).
- Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
- 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).
- Kurt, I. C. et al. CRISPR C-to-G base editors for inducing targeted DNA transversions in human cells. *Nat. Biotechnol.* **39**, 41–46 (2021).
- Porto, E. M. et al. Base editing: advances and therapeutic opportunities. *Nat. Rev. Drug Discov.* **19**, 839–859 (2020).
- Naeem, M., Majeed, S., Hoque, M. Z. & Ahmad, I. Latest developed strategies to minimize the off-target effects in CRISPR–Cas-mediated genome editing. *Cells* **9**, 1608 (2020).
- Molla, K. A. & Yang, Y. CRISPR/Cas-mediated base editing: technical considerations and practical applications. *Trends Biotechnol.* **37**, 1121–1142 (2019).
- Stevanovic, M., Piotter, E., McClements, M. E. & MacLaren, R. E. CRISPR systems suitable for single AAV vector delivery. *Curr. Gene Ther.* **22**, 1–14 (2022).
- Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).
- Li, C. & Samulski, R. J. Engineering adeno-associated virus vectors for gene therapy. *Nat. Rev. Genet.* **21**, 255–272 (2020).
- Wu, Z., Yang, H. & Colosi, P. Effect of genome size on AAV vector packaging. *Mol. Ther.* **18**, 80–86 (2010).
- Levy, J. M. et al. Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses. *Nat. Biomed. Eng.* **4**, 97–110 (2020).
- Chand, D. et al. Hepatotoxicity following administration of onasemnogene abeparvovec (AVXS-101) for the treatment of spinal muscular atrophy. *J. Hepatol.* **74**, 560–566 (2021).
- Kim, D. Y. et al. Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat. Biotechnol.* **40**, 94–102 (2021).
- Xu, X. et al. Engineered miniature CRISPR–Cas system for mammalian genome regulation and editing. *Mol. Cell* **81**, 4333–4345 (2021).
- Wu, Z. et al. Programmed genome editing by a miniature CRISPR–Cas12f nuclease. *Nat. Chem. Biol.* **17**, 1132–1138 (2021).
- Altae-Tran, H. et al. The widespread IS200/IS605 transposon family encodes diverse programmable RNA-guided endonucleases. *Science* **374**, 57–65 (2021).
- Karvelis, T. et al. Transposon-associated TnpB is a programmable RNA-guided DNA endonuclease. *Nature* **599**, 692–696 (2021).
- Shmakov, S. et al. Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.* **15**, 169–182 (2017).

20. Takeda, S. N. et al. Structure of the miniature type V-F CRISPR–Cas effector enzyme. *Mol. Cell* **81**, 558–570 (2021).
21. Xiao, R., Li, Z., Wang, S., Han, R. & Chang, L. Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR–Cas12f nuclease. *Nucleic Acids Res.* **49**, 4120–4128 (2021).
22. Molla, K. A. et al. Precise plant genome editing using base editors and prime editors. *Nat. Plants* **7**, 1166–1187 (2021).
23. Rees, H. A., Wilson, C., Dorman, J. L. & Liu, D. R. Analysis and minimization of cellular RNA editing by DNA adenine base editors. *Sci. Adv.* **8**, eaax5717 (2019).
24. Jeong, Y. K. et al. Adenine base editor engineering reduces editing of bystander cytosines. *Nat. Biotechnol.* **39**, 1426–1433 (2021).
25. Richter, M. F. et al. Phage-assisted evolution of an adenine base editor with improved Cas domain compatibility and activity. *Nat. Biotechnol.* **38**, 883–891 (2020).
26. Yan, D. et al. High-efficiency and multiplex adenine base editing in plants using new TadA variants. *Mol. Plant* **14**, 722–731 (2021).
27. Harrington, L. B. et al. Programmed DNA destruction by miniature CRISPR–Cas14 enzymes. *Science* **362**, 839 (2018).
28. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat. Biotechnol.* **38**, 620–628 (2020).
29. Yu, Y. et al. Cytosine base editors with minimized unguided DNA and RNA off-target events and high on-target activity. *Nat. Commun.* **11**, 2052 (2020).
30. Nguyen Tran, M. T. et al. Engineering domain-inlaid SaCas9 adenine base editors with reduced RNA off-targets and increased on-target DNA editing. *Nat. Commun.* **11**, 4871 (2020).
31. Sharma, B., Singh, V. J. & Chawla, P. A. Epidermal growth factor receptor inhibitors as potential anticancer agents: an update of recent progress. *Bioorg. Chem.* **116**, 105393 (2021).
32. Huang, S. et al. Developing ABEmax-NG with precise targeting and expanded editing scope to model pathogenic splice site mutations in vivo. *iScience* **15**, 640–648 (2019).
33. Yuan, J. et al. Genetic modulation of RNA splicing with a CRISPR-guided cytidine deaminase. *Mol. Cell* **72**, 380–394 (2018).
34. Gillmore, J. D. et al. CRISPR–Cas9 in vivo gene editing for transthyretin amyloidosis. *N. Engl. J. Med.* **385**, 463–502 (2021).
35. Rothgangl, T. et al. In vivo adenine base editing of PCSK9 in macaques reduces LDL cholesterol levels. *Nat. Biotechnol.* **39**, 949–957 (2021).
36. Musunuru, K. et al. In vivo CRISPR base editing of PCSK9 durably lowers cholesterol in primates. *Nature* **593**, 429–434 (2021).
37. Bae, S., Park, J. & Kim, J.-S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473–1475 (2014).
38. Anzalone, A. V. et al. Search-and-replace genome editing without double-strand breaks or donor DNA. *Nature* **576**, 149–157 (2019).
39. Doudna, J. A. The promise and challenge of therapeutic genome editing. *Nature* **578**, 229–236 (2020).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2022

## Methods

**Plasmid vector construction.** Human codon-optimized TnpB gene (Supplementary Table 1) was synthesized and cloned into a pCas12f-2A-EGFP vector (Addgene) by replacing the TnpB-coding sequence with Cas12f1 sequence. Various versions of Tad sequences were then fused at either 5'- or 3'-region of TnpB with 10–40 amino-acid-long linkers using NEBuilder HiFi DNA assembly master mix (New England BioLabs). gRNA sequences were positioned under U6 promoter 5'-upstream of cytomegalovirus promoter using the MluI restriction enzyme. Spacer sequences were cloned into by digesting the vectors with BbsI restriction enzyme for 1 h at 37°C. The catalytically dead TnpB were generated by site-directed mutagenesis by primers incorporating the intended base changes. Plasmid vectors for cell transfection were prepared using Nucleobond Xtra midi (MACHEREY-NEGEL). All vector constructs were sequence verified using Sanger sequencing.

**TnpB engineering.** The construction of TnpB PAM variants and mutants (I159W and S164Y) was performed by site-directed mutagenesis. PCR amplifications were performed using Q5 Hot Start high-fidelity DNA polymerase (NEB) and the PCR products were ligated using KLD Enzyme Mix (NEB). The ligated products were transformed into DH5 $\alpha$  *E. coli* cells. Mutagenesis was confirmed by Sanger sequencing analysis. The modified plasmid vectors were purified using a NucleoBond Xtra Midi EF kit (MN).

**PAM library construction and PAM preference determination.** Oligonucleotides harboring a protospacer (5'-CACACACACAGTGGGCTACC-3') and PAM library sequence (NNNN) were synthesized (Bionics) and cloned into a PUC19 vector using an All in One PCR cloning kit (Biofact). Each cloned vector was used to transform DH5 $\alpha$  *E. coli* cells using an electroporator (Bio-Rad). Each transformant colony was grown at 37°C in LB broth until the culture reached an optical density of 0.6. Cells were collected by centrifugation at 3,500g for 15 min. Plasmid vectors were prepared using a plasmid preparation kit (Biofact). Sequence was verified using Sanger sequencing analysis. Each vector was spectrophotometrically quantified at 265 nm and mixed at an equal molar ratio to prepare 256 PAM library vectors. Plasmid vectors encoding TnpB PAM variants were used to transform BL21(DE3) *E. coli* cells. Each transformant colony was grown at 37°C in LB broth until the culture reached an optical density of 0.6. Cells were incubated at 18°C overnight in the presence of 0.1 mM isopropylthio- $\beta$ -D-galactoside and then collected by centrifugation at 3,500g for 15 min. Cells were resuspended in 20 mM Tris-HCl (pH 7.6), 500 mM NaCl, 5 mM  $\beta$ -mercaptoethanol, 5% glycerol. Cell lysates were prepared by sonication followed by centrifugation at 15,000g for 15 min. TnpB proteins were purified on a Ni<sup>2+</sup>-affinity column and a Heparin column. gRNA was synthesized using T7 RNA polymerase (NEB) in the presence of 1  $\mu$ g purified plasmid and 4 mM NTPs (Jena Bioscience), purified using a Monarch RNA cleanup kit (NEB) and aliquoted into cryogenic vials before storage in liquid nitrogen. Purified TnpB (5  $\mu$ g), gRNA (1  $\mu$ g), and PAM library plasmid vector (1  $\mu$ g) were mixed at a final mixture 100  $\mu$ l in 5 mM Tris-HCl, pH 7.5, 25 mM NaCl, 5 mM MgCl<sub>2</sub>, 1 mM dithiothreitol buffer and incubated at 37°C for 2 h. The incubated samples were end-repaired using a NEBNext Ultra II End Repair/DA-Tailing Module (NEB, E7546) at 20°C for 30 min, and the reaction was terminated by incubating at 65°C for 30 min. After treated with RNase A at 100 mg ml<sup>-1</sup> at room temperature for 15 min, plasmid DNA was purified using a HiGene Gel&PCR purification kit (Biofact). The purified DNAs (200 ng per 10  $\mu$ l) were ligated with a 200 ng adaptor DNA (5'-AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3') using a LigaFast Rapid DNA ligation kit (Promega). DNAs were PCR-amplified using a forward primer (5'-GTAAAACGACGGCCAGT-3') and a reverse primer (5'-GTGACTGGAGTTC-3') using a KOD One PCR Master Mix (TOYOBO). The resulting PCR amplicons were labeled with Illumina TruSeq HT dual indexes. The final PCR products were subjected to 150-bp paired-end sequencing using an Illumina iSeq 100.

**PAM-mutant cell lines.** HEK293T cells (LentX-293T, Takara) were maintained in Dulbecco's modified Eagle medium (Corning) supplemented with 10% heat-inactivated fetal bovine serum (VWR) and 1% penicillin–streptomycin (WELGENE) in an incubator (37°C, 5% CO<sub>2</sub> atmosphere). PAM-varied oligonucleotides (90-mer) were synthesized as donor DNA. For transfection, 4  $\mu$ g SpCas9 plasmid vector was transfected with 4  $\mu$ g donor DNAs into 4  $\times$  10<sup>5</sup> HEK293T cells using a Neon transfection system (Invitrogen). The electroporation conditions were as follows: 1,300 V, 20 mA, 2 pulses. Three days after transfection, a single cell was placed in each well of 24-well plates (Corning) and grown for 3 weeks. Genomic DNA was prepared from each colony cells using a PureHelix genomic DNA preparation kit (NanoHelix). PAM-containing region was amplified using KOD One PCR Master Mix (TOYOBO) according to the manufacturer's instructions. The PAM sequence was verified by deep sequencing using an Illumina iSeq 100.

**Measurement of substitution efficiency.** HEK293T cells were transfected with vectors using a lipofection method. HEK293T or H661 cells were seeded into 24-well plates at a density of 1.0  $\times$  10<sup>5</sup> per well 1 days before transfection. Six microliters FuGene reagents (Promega) were mixed with 1.5  $\mu$ g TnpB-ABE vector plus 500 ng gRNA-encoding PCR amplicon in 300  $\mu$ l Opti-MEM and incubated

at room temperature for 15 min. The mixtures were added to each well, and cells were grown for 3 to 5 days at 37°C and 5% CO<sub>2</sub>. Genomic DNA was extracted by cell lysis with Martin's solution (50 mM Tris-HCl, pH 8.5, 1 mM EDTA, 0.005% SDS, proteinase K). Samples for deep-sequencing analysis were prepared by three rounds of PCR amplifications. For the primary PCR, 1  $\mu$ l cell lysis was amplified by target-specific primers designed to amplify targeted locus in 10  $\mu$ l total volume. One microliter primary PCR products were amplified by primers with Illumina adapter sequence to produce 150-bp-long amplicons. Finally, Illumina TruSeq HT dual indexes were labeled on the PCR amplicons by PCR reactions. All PCR reactions were performed using KOD one PCR master mix (Toyobo) according to the manufacturer's instructions. Pooled amplicons were column-purified using a PCR purification kit (BioFact). The final PCR products were subjected to 150-bp paired-end sequencing using an iSeq Control software (v.1.4.1.1700) installed in an Illumina iSeq 100. Indel frequencies were calculated by MAUND, which is available at <https://github.com/ibs-cge/maund>. The tested sites were compiled in Supplementary Table 4.

**RNA sequencing analysis.** Libraries were prepared for 151-bp paired-end sequencing using TruSeq stranded mRNA Sample Preparation Kit (Illumina). In brief, total mRNA molecules were purified and fragmented from 1  $\mu$ g total RNA using oligo (dT) magnetic beads. The fragmented mRNAs were synthesized as single-stranded complementary DNAs (cDNAs) through random hexamer priming. The synthesized cDNAs were used as a template for the preparation of dsDNAs. After sequential process of end repair, A-tailing and adapter ligation, cDNA libraries were PCR-amplified using using KOD One PCR Master Mix (TOYOBO). Quality of these cDNA libraries was evaluated with the Agilent 2100 BioAnalyzer (Agilent). They were quantified with the KAPA library quantification kit (Kapa Biosystems) according to the manufacturer's library quantification protocol. Following cluster amplification of denatured templates, sequencing was conducted as paired-end (2  $\times$  151 bp) using Illumina NovaSeq6000 (Illumina). The paired-end RNA sequencing results were aligned to the reference genome (GRCh38) by using bwa mem with default options<sup>40</sup>. The alignment results were sorted using samtools sort and single-nucleotide substitution was analyzed by REDIttools (<https://github.com/BioinfoUNIBA/REDIttools2>). The adapter sequences and the ends of the reads less than Phred quality score 20 were trimmed and simultaneously the reads shorter than 50 bp were removed by using cutadapt v.2.8 (ref. <sup>41</sup>). We screened the positions with adenosines that were partly read as inosine and then calculated the A-to-I frequencies at all positions.

**Production of adeno-associated virus.** AAV particles were produced as described previously with procedural modifications<sup>44,42</sup>. AAVpro 293T cells (Takara) were seeded at a density of 2  $\times$  10<sup>7</sup> onto a triple flask with a dimension of 500 cm<sup>2</sup> (Thermo) before transfection. Vectors encoding base editor components, pHelper, and pAAV-RC2/2 (or pAAV-RC2/9) were mixed at 1:1:1 molar ratio, and the vector mixture (150  $\mu$ g) was used to transfect AAVpro 293T cells at a confluency of ~70% using a Polythyleneimine (Polysciences) transfection reagent according to the manufacturer's instruction. Three days after transfection, cells were harvested and lysed by three cycles of freezing and thawing. Cell lysates were treated with DNase I (Enzygnomics) at the final concentration of 10 units per microliter at 37°C for 30 min. A gradient solution was prepared by adding 4 ml 60% iodixanol solution, 4 ml 40% iodixanol, 5 ml 25% iodixanol, and 7 ml 15% iodixanol sequentially into a tube. Lysed samples were poured with care on top of the gradient solution. The gradient mixture was centrifuged in a 50.2 Ti ultracentrifuge rotor (Beckman) at 50,000 rpm at 10°C for 2.5 h. The viral fraction was retrieved using an 18-gauge needle at the 40% iodixanol layer. After washed with PBS, the viral fraction was stored at -80°C until the use for transfection.

**Mouse experiment.** Mice were maintained on a 12-h light/dark cycle at room temperature (20–25°C) with constant humidity (50  $\pm$  5%). Six-week-old female Balb/c mice (Jackson Laboratory) were intravenously injected with AAV9 particles at 1  $\times$  10<sup>12</sup> vg using 4-mm, 32-gauge needles. Mice were sacrificed 1 or 2 weeks after systemic injections and tissues were resected immediately after sacrifice. Tissues were subjected to either fluorescence imaging using a LuminoGraph III image analyzer (Atto) or preparations of genomic DNAs. All mouse studies were performed in accordance with the KRIBB Institutional Animal Care and Use Committee, under animal use protocol number KRIBB-AEC-21201.

**Statistical analysis.** Statistical significance tests were performed using SigmaPlot software (v.14.0) through a two-tailed Student's *t*-test or Welch's *t*-test. In cases where normality fails, a Mann–Whitney Rank Sum test was employed. *P* values < 0.05 were considered significant. Data points in violin plots represent the full range of values with the interquartile range (25th–75th percentile), and average values are indicated by horizontal lines. The error bars in all dot and bar plots show the standard deviation and were plotted with SigmaPlot (v.14.0). We did not predetermine sample sizes on the basis of statistical methods. For all experimental results, the *n* is reported in the accompanying figure legend.

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

### Data availability

Data that support the findings of this study are available within the Article and its Supplementary Information. Deep-sequencing data for large-scale validation and RNA-seq data were deposited at the NCBI Sequence Read Archive database (<http://www.ncbi.nlm.nih.gov/sra>) under accession number PRJNA823884. All other data that support the findings of the present study and plasmid vectors are available from the corresponding author upon request. Source data are provided with this paper.

### Code availability

Reditools is available at <https://github.com/BioinfoUNIBA/REDIttools2>. MAUND is available at <https://github.com/ibs-cge/maund>.

### References

- Jung, Y. & Han, D. BWA-MEME: BWA-MEM emulated with a machine learning approach. *Bioinformatics* **38**, 2404–2413 (2022).
- Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* **17**, 10 (2011).
- Verdera, H. C., Kuranda, K. & Mingozzi, F. AAV vector immunogenicity in humans: a long journey to successful gene transfer. *Mol. Ther.* **28**, 723–746 (2020).

### Acknowledgements

This work was supported by a grant through the KRIBB Research Initiative Program (KGM5382221 to Y.L., D.J., K.-H.P, H.J.C, J.M.L, J.-H.K and Y.-S.K.), the National Research Foundation of Korea (INNOPOLIS) grant (2021-DD-RD-0178-01 to D.Y.K., Y.C., S.P. and Y.-S.K.), and the National Research Foundation of Korea grant (2022R1C1C1013085 to D.Y.K and S.K.) funded by the Ministry of Science and ICT,

the 'Alchemist Project' funded by the Ministry of Trade, Industry and Energy (20012445 to D.Y.K. and Y.-S.K.), and Cooperative Research Program for Agriculture Science and Technology Development (PJ0165422022 to D.Y.K., Y.C., S.P. and Y.-S.K.) funded by Rural Development Administration.

### Author contributions

Y.-S.K. conceived the study and designed the experiments. D.Y.K. and Y.C. performed overall experiments. D.J. and Y.L. constructed the PAM library vectors and PAM variants of TnpB. H.J.C. and J.M.L. performed base-editing assays in cells. S.P. and S.K. derived PAM-mutant HEK293T cells. K.-H.P. performed the structural analysis for TnpB engineering. J.-H.K. and Y.-S.K. interpreted data. Y.-S.K. wrote the manuscript.

### Competing interests

D.Y.K., Y.C. and Y.-S.K. have filed patent applications on the TaRGET-ABE and PAM variants of TnpB through GenKOre. Y.-S.K. and D.Y.K. are co-founders of GenKOre. The remaining authors declare no other competing interests.

### Additional information

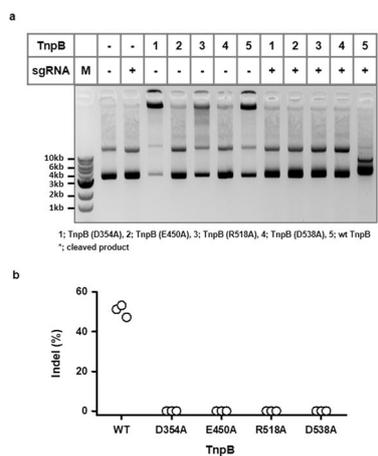
**Extended data** is available for this paper at <https://doi.org/10.1038/s41589-022-01077-5>.

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41589-022-01077-5>.

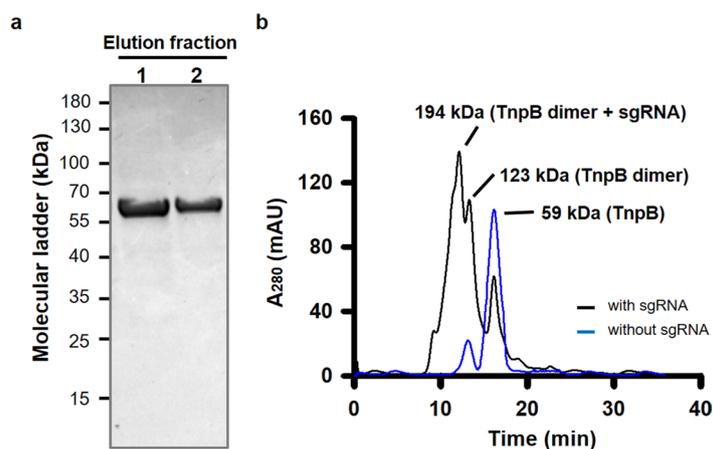
**Correspondence and requests for materials** should be addressed to Yong-Sam Kim.

**Peer review information** *Nature Chemical Biology* thanks Sangsu Bae, Rahul Kohli and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).



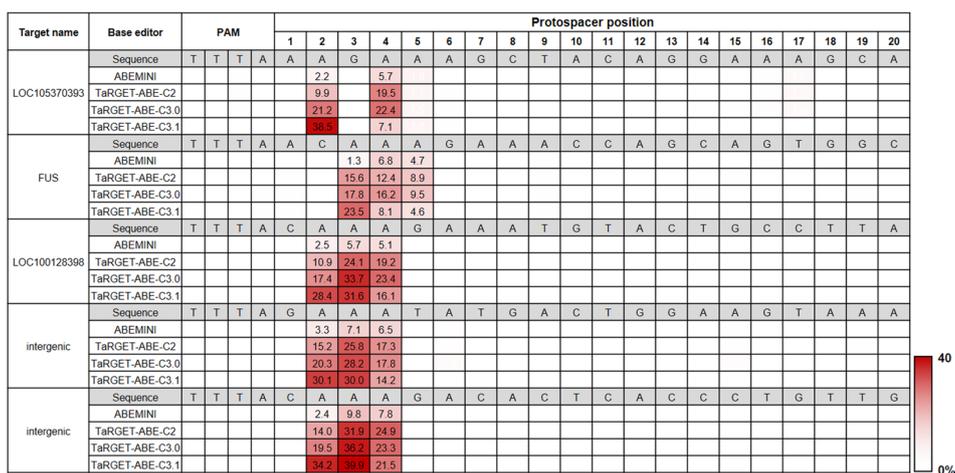
**Extended Data Fig. 1 | Construction of catalytically inactive TnpB (dTnpB) and confirmation of indel-null activity in HEK293T cells. a.** An agarose gel image showing cleavage pattern of plasmid vector by wild-type or several catalytically dead mutants of TnpB. Plasmid vectors were constructed to harbor protospacer and PAM sequence for sgRNA. Five microgram of plasmid vectors were incubated with 1 microgram of gRNA and 2 microgram of TnpB or dTnpB at 37 °C for 1h. The cleaved vector samples were resolved on a 0.8% agarose gel. M denotes molecular ladders. The image corresponds to a representative experiment for three independent experiments. **b.** Indel efficiencies of TnpB or dead mutants of TnpB at an *NLRC4* locus (5'-TTTAGAGGGAGACACAAGTTGATA-3') in HEK293T cells. n = 3 independent experiments.



**Extended Data Fig. 2 | Dimerization of TnpB in the presence of single-stranded guide RNA.** **a.** An SDS-PAGE gel image for purified TnpB in elution fractions. The gel image corresponds to a representative experiment for six independent experiments. **b.** Size-exclusion chromatography profiles of TnpB proteins in the presence or absence of sgRNA. TnpB proteins were incubated with gRNA and the RNP complex was resolved on a Superdex 200 column. The molecular mass was estimated from a standard curve derived from the mobility of molecular ladder proteins.

	PAM										Protospacer															
<b>Site a</b>	T	T	T	A	C	A	A	A	G	A	C	A	C	T	C	A	C	C	C	T	G	T	T	G		
A	0	0	0	99.93	0	97.54	75.62	82.73	0.04	98.85	0	99.96	0	0	0	99.94	0	0	0	0	0.11	0	0	0.04		
C	0	0.09	0.07	0	100	0	0	0	0	0	99.91	0	99.78	0.05	99.86	0	99.88	99.73	100	0	0	0.08	0.03	0		
G	0	0	0	0.07	0	2.46	24.38	17.27	99.96	1.15	0	0.04	0	0	0	0.06	0	0	0	0	99.89	0	0	99.96		
T	100	99.91	99.93	0	0	0	0	0	0	0.09	0	0.22	99.95	0.14	0	0.12	0.27	0	100	0	99.92	99.97	0	0		
<b>Site b</b>	T	T	T	G	A	T	A	C	A	G	A	A	A	T	C	C	T	A	A	A	T	G	G	T		
A	0	0	0	100	99.64	0	78.89	0	88.65	0	99.24	99.98	99.64	0	0	0	0	99.77	99.64	100	0	0.07	0	0		
C	0.03	0.05	0	0	0	0.13	0	99.68	0	0	0	0	0	0.10	99.82	99.95	0.13	0	0	0	0	0	0	0.03		
G	0	0	0	0	0.36	0	21.11	0	11.35	100	0.76	0.02	0.36	0	0	0	0	0.23	0.36	0	0	99.93	100	0		
T	99.97	99.95	100	0	0	99.87	0	0.32	0	0	0	0	0	99.90	0.18	0.05	99.87	0	0	0	100	0	0	99.97		
<b>Site c</b>	T	T	T	A	A	A	G	A	A	A	G	C	T	A	C	A	G	G	A	A	A	G	C	A		
A	0	0	0	99.97	98.65	96.34	0.03	80.14	90.07	98.47	0.15	0	0	99.71	0	99.24	1.04	0.29	99.81	99.66	99.47	0.02	0	99.18		
C	0	0.11	0.06	0	0	0	0	0	0	0	0	0	99.81	0	0	99.83	0	0	0	0	0	0	100	0		
G	0	0	0	0.03	1.35	3.66	99.97	19.86	9.93	1.53	99.85	0	0	0.29	0	0.76	98.96	99.71	0.19	0.34	0.53	99.98	0	0.82		
T	100	99.89	99.94	0	0	0	0	0	0	0	0	0.19	100	0	0.17	0	0	0	0	0	0	0	0	0		
<b>Site d</b>	T	T	T	A	A	C	A	A	A	G	A	A	A	C	C	A	G	C	A	G	T	G	G	C		
A	0	0	0	99.93	99.99	0	82.22	86.77	96.04	0.64	99.90	99.81	99.70	0	0	99.48	0.09	0	99.64	0.07	0	0.44	0.11	0		
C	0.09	0.03	0.07	0	0	99.85	0	0	0	0	0	0	0	0	99.92	100	0	0	99.93	0	0	0	0	99.87		
G	0	0	0	0.07	0.04	0	17.76	13.23	3.96	99.36	0.10	0.19	0.30	0	0	0.52	99.91	0	0.36	99.93	0	99.56	99.89	0		
T	99.91	99.97	99.93	0	0	0.15	0	0	0	0	0	0	0	0	0.08	0	0	0	0.07	0	0	100	0	0.13		
<b>Site e</b>	T	T	T	A	C	A	A	A	G	A	A	A	T	G	T	A	C	T	G	C	C	T	T	A		
A	0	0	0	99.90	0	97.62	88.29	88.02	0.02	98.85	99.66	98.14	0	0.27	0	99.86	0	0	0	0	0	0	0	99.69		
C	0	0.04	0.14	0	99.92	0	0	0	0	0	0	0	0	0.10	0	0.03	0	99.89	0	0	98.60	100	0.05	0.03	0	
G	0	0	0	0.10	0	2.38	21.71	11.88	99.98	1.15	0.34	1.86	0	99.73	0	0.14	0	0	100	1.08	0	0	0	0.31		
T	100	99.96	99.86	0	0.08	0	0	0	0	0	0	0	0	99.90	0	99.97	0	0.11	100	0	0.32	0	99.95	99.97	0	

**Extended Data Fig. 3 | Identification of the base-editing window of TaRGET-ABE-C2 system.** The substitution profile of TaRGET-ABE-C2 was explored for five endogenous sites in HEK293T cells to define a base editing window. The endogenous sites were selected from the validated sites showing substantial A-to-G conversion activities and also carrying multiple adenine sequences in the PAM-proximal region. The TaRGET-ABE system elicits A-to-G conversions in the range of positions 2–6, but most predominantly at the positions 3–4. The intensity of colors is not proportional to the editing efficiency, but highlights the qualitatively high efficiency positions.



**Extended Data Fig. 4 | Head-to-head comparison of TaRGET-based adenine base editors and Cas12f-based ABEMINI.** The efficiency of TaRGET-ABEs was compared with that of ABEMINI at five endogenous *loci* in HEK293T cells. Specifically, TaRGET-ABE-C2 was compared with ABEMINI to rule out the contribution of Tad engineering to the editing efficiency. The results show significant difference in A-to-G conversion activities in the base-editing window between TaRGET-ABE-C2 and ABEMINI, indicating TnpB as a preferred nuclease for hypercompact adenine base editors. The intensity of colors represents of the conversion efficiency. The values represent the means of three independent experiments.

## Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

### Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided  
*Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g.  $F$ ,  $t$ ,  $r$ ) with confidence intervals, effect sizes, degrees of freedom and  $P$  value noted  
*Give  $P$  values as exact values whenever suitable.*
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's  $d$ , Pearson's  $r$ ), indicating how they were calculated

*Our web collection on [statistics for biologists](#) contains articles on many of the points above.*

### Software and code

Policy information about [availability of computer code](#)

Data collection Deep sequencing data were collected using an iSeq Control software (v1.4.1.1700) installed in an iSeq 100 sequencer (Illumina). Quantitative RT-PCR data were collected using a QuantStudio Design & Analysis Software (v1.5.1).

Data analysis SigmaPlot (v14.0), MAUND (v0.4.0.0, available at <https://github.com/ibs-cge/maund>), Excel (2016), REDIttools (<https://github.com/BioinfoUNIBA/REDIttools2>), Cas-OFFinder (<http://www.rgenome.net/cas-offinder>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

### Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Deep sequencing data for large-scale validation and RNA-seq data were deposited at the NCBI Sequence Read Archive database (<http://www.ncbi.nlm.nih.gov/sra>) under accession number PRJNA\_823884. Raw data and related information will be provided in the Source Data file, Supplementary Figures and Supplementary Table.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences       Behavioural & social sciences       Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical analyses were performed to predetermine sample sizes. Most of other papers deal with less than 10 sites to gauge the base-editing efficiency. Because we investigated the efficiency at 25 sites, our sample size is larger than those in any other publications. Most of experiments were performed in, at least, triplicate.
Data exclusions	Data obtained by wild-type TnpB were not excluded. There were several targets that showed zero efficiency with TnpB PAM variants and those targets were excluded.
Replication	All experiments were repeated at least three times. Critical experiments were confirmed for repeatability by 3 different testers.
Randomization	Mammalian cells used in this study were grown under identical conditions; no randomization was used because all experiments were controlled so that other covariates can minimally affect the results.
Blinding	Comparative study for Cas9-, Cas12f-, and TnpB-based base editors were performed by 3 different testers blinded to information with respect to the Cas type. For other experiments, blinding was not necessary because 3 different testers performed independent experiments and their results were averaged to derive mean values.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input type="checkbox"/>	<input checked="" type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Human research participants
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern

### Methods

n/a	Involvement in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Eukaryotic cell lines

Policy information about [cell lines](#)

Cell line source(s)	HEK293T (Lenti-X 293T, Takara), H661 (ATCC)
Authentication	Cells were authenticated by the supplier using STR analysis.
Mycoplasma contamination	Mycoplasma contamination was regularly tested and non-contaminated cells were only used in this study.
Commonly misidentified lines (See <a href="#">ICLAC</a> register)	Not used.

## Animals and other organisms

Policy information about [studies involving animals](#); [ARRIVE guidelines](#) recommended for reporting animal research

Laboratory animals	The maintenance conditions for mice were described in Methods Six-week-old female Balb/c mice (The Jackson Laboratory) were used for AAV injection.
--------------------	---

Wild animals

No wild animals were used in this study.

Field-collected samples

No field-collected samples were used in this study.

Ethics oversight

All mouse studies were performed in accordance with the KRIBB Institutional Animal Care and Use Committee, under animal use protocol number KRIBB-AEC-21201.

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