

# Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage

Nicole M. Gaudelli<sup>1,2,3</sup>, Alexis C. Komor<sup>1,2,3</sup>†, Holly A. Rees<sup>1,2,3</sup>, Michael S. Packer<sup>1,2,3</sup>†, Ahmed H. Badran<sup>1,2,3</sup>, David I. Bryson<sup>1,2,3</sup>† & David R. Liu<sup>1,2,3</sup>

The spontaneous deamination of cytosine is a major source of transitions from C•G to T•A base pairs, which account for half of known pathogenic point mutations in humans. The ability to efficiently convert targeted A•T base pairs to G•C could therefore advance the study and treatment of genetic diseases. The deamination of adenine yields inosine, which is treated as guanine by polymerases, but no enzymes are known to deaminate adenine in DNA. Here we describe adenine base editors (ABEs) that mediate the conversion of A•T to G•C in genomic DNA. We evolved a transfer RNA adenosine deaminase to operate on DNA when fused to a catalytically impaired CRISPR–Cas9 mutant. Extensive directed evolution and protein engineering resulted in seventh-generation ABEs that convert targeted A•T base pairs efficiently to G•C (approximately 50% efficiency in human cells) with high product purity (typically at least 99.9%) and low rates of indels (typically no more than 0.1%). ABEs introduce point mutations more efficiently and cleanly, and with less off-target genome modification, than a current Cas9 nuclease-based method, and can install disease-correcting or disease-suppressing mutations in human cells. Together with previous base editors, ABEs enable the direct, programmable introduction of all four transition mutations without double-stranded DNA cleavage.

The formation of uracil and thymine from the spontaneous hydrolytic deamination of cytosine and 5-methylcytosine, respectively<sup>1,2</sup>, occurs an estimated 100–500 times per cell per day in humans<sup>1</sup> and can result in C•G to T•A mutations, which account for approximately half of all known pathogenic single nucleotide polymorphisms (SNPs) (Fig. 1a). The ability to convert A•T base pairs to G•C base pairs at target loci in the genomic DNA of unmodified cells could therefore make it possible to correct a substantial fraction of human disease-associated SNPs.

Base editing is a form of genome editing that enables direct, irreversible conversion of one base pair to another at a target genomic locus without requiring double-stranded DNA breaks (DSBs), homology-directed repair (HDR) processes, or donor DNA templates<sup>3–5</sup>. Compared with standard genome editing methods to introduce point mutations, base editing can proceed more efficiently<sup>3</sup> and with far fewer undesired products, such as stochastic insertions or deletions (indels) or translocations<sup>3–8</sup>.

The most commonly used base editors are third-generation designs (BE3) comprising (i) a catalytically impaired CRISPR–Cas9 mutant that cannot make DSBs; (ii) a single-strand-specific cytidine deaminase that converts C to uracil (U) within an approximately five-nucleotide window in the single-stranded DNA bubble created by Cas9; (iii) a uracil glycosylase inhibitor (UGI) that impedes uracil excision and downstream processes that decrease base editing efficiency and product purity<sup>5</sup>; and (iv) nickase activity to nick the non-edited DNA strand, directing cellular DNA repair processes to replace the G-containing DNA strand<sup>3,5</sup>. Together, these components enable efficient and permanent C•G to T•A base pair conversion in bacteria, yeast<sup>4,9</sup>, plants<sup>10,11</sup>, zebrafish<sup>8,12</sup>, mammalian cells<sup>3–8,13,14</sup>, mice<sup>8,15,16</sup>, and even human embryos<sup>17,18</sup>. Base editing capabilities have expanded through the development of base editors with different protospacer-adjacent motif (PAM) compatibilities<sup>7</sup>, narrowed editing windows<sup>7</sup>, enhanced DNA specificity<sup>8</sup>, and small-molecule

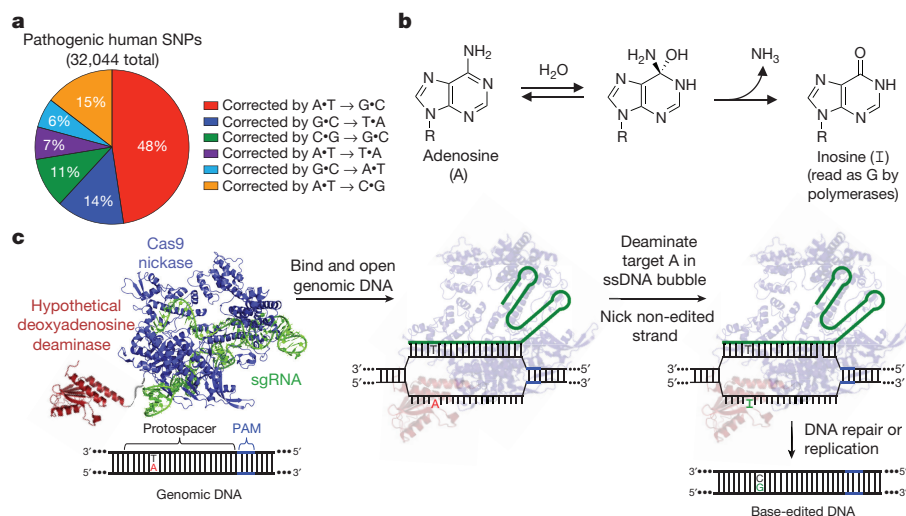
dependence<sup>19</sup>. Fourth-generation base editors (BE4 and BE4-Gam) show further improved editing efficiency and product purity<sup>5</sup>.

To date, all reported base editors mediate C•G to T•A conversion. In this study, we used protein evolution and engineering to develop a new class of adenine base editors (ABEs) that convert A•T to G•C base pairs in DNA in bacteria and human cells. Seventh-generation ABEs efficiently convert A•T to G•C at a wide range of target genomic loci in human cells efficiently and with a very high degree of product purity, exceeding the typical performance characteristics of BE3. ABEs greatly expand the scope of base editing and, together with previously described base editors, enable the programmable installation of all four transitions (C to T, A to G, T to C, and G to A) in genomic DNA.

## Evolution of an adenine deaminase that processes DNA

The hydrolytic deamination of adenosine yields inosine (Fig. 1b). Within the constraints of a polymerase active site, inosine pairs with C and therefore is read or replicated as G<sup>20</sup>. While replacing the cytidine deaminase of an existing base editor with an adenine deaminase could, in theory, provide an ABE (Fig. 1c), no enzymes are known to deaminate adenine in DNA. Although all reported examples of enzymatic adenine deamination occur on free adenine, free adenosine, adenosine in RNA, or adenosine in mispaired RNA–DNA heteroduplexes<sup>21</sup>, we began by replacing the APOBEC1 (apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 1) component of BE3 with natural adenine deaminases such as *Escherichia coli* Tada (ecTada)<sup>22,23</sup>, human ADAR2<sup>24</sup>, mouse ADA<sup>25</sup>, and human ADAT2<sup>26</sup> (Supplementary Sequences 1) to test the possibility that these enzymes might process DNA when present at a high effective molarity. Unfortunately, when plasmids encoding these deaminases fused to Cas9 D10A nickase were transfected into HEK293T cells together with a corresponding single guide RNA (sgRNA), we observed no A•T to G•C editing above that seen in untreated cells (Extended Data Figs 1, 2b). These results suggest

<sup>1</sup>Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>2</sup>Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA. <sup>3</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. †Present addresses: Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093, USA (A.C.K.); Beam Therapeutics, 675 West Kendall Street, Cambridge, Massachusetts 02139, USA (M.S.P., D.I.B.).



**Figure 1 | Scope and overview of base editing by an A•T to G•C base editor.** **a**, Base pair changes required to correct pathogenic human SNPs in the ClinVar database<sup>39</sup>. **b**, The deamination of adenosine (A) forms inosine (I), which is read as guanosine (G) by polymerase enzymes. R = 2'-deoxyribose in DNA, or ribose in RNA. **c**, ABE-mediated A•T to G•C base editing strategy. ABEs contain a hypothetical deoxyadenosine deaminase, which is not known to exist in nature, and a catalytically

impaired Cas9. They bind target DNA in a guide RNA-programmed manner, exposing a small bubble of single-stranded DNA. The hypothetical deoxyadenosine deaminase domain catalyses conversion of adenine to inosine within this bubble. Following DNA repair or replication, the original A•T base pair is replaced with a G•C base pair at the target site.

that the inability of these natural adenine deaminase enzymes to accept DNA precludes their direct use in an ABE.

Given these results, we sought to evolve an adenine deaminase that accepts DNA as a substrate. We developed a bacterial selection method for base editing by creating defective antibiotic resistance genes that contain point mutations at critical positions (Supplementary Table 8 and Supplementary Sequences 2). Reversion of these mutations by base editors restores antibiotic resistance. To validate the selection, we used a bacterial codon-optimized version of BE2<sup>3</sup> (APOBEC1 cytidine deaminase fused to dCas9 and UGI), because bacteria lack the nick-directed mismatch repair machinery<sup>27</sup> that enables more efficient base editing by BE3. We observed successful rescue of a defective chloramphenicol acetyl transferase (Cam<sup>R</sup>) containing an A•T to G•C mutation at a catalytic residue (H193R) by BE2 and an sgRNA programmed to direct base editing to the inactivating mutation.

Next we adapted the selection plasmid for ABE activity by introducing a C•G to T•A mutation in the Cam<sup>R</sup> gene, creating an H193Y substitution that confers minimal chloramphenicol resistance (Supplementary Table 8 and Supplementary Sequences 2). A•T to G•C conversion at the H193Y mutation should restore chloramphenicol resistance, linking ABE activity to bacterial survival.

Previously described base editors<sup>3,5,7,8</sup> exploit the use of cytidine deaminase enzymes that operate on single-stranded DNA but reject double-stranded DNA. This feature is critical to restrict deaminase activity to a small window of nucleotides within the single-stranded bubble created by Cas9. TadA is a tRNA adenine deaminase<sup>22</sup> that converts adenine to inosine (I) in the single-stranded anticodon loop of tRNA<sup>Arg</sup>. *E. coli* TadA shares homology with the APOBEC enzyme<sup>28</sup> used in our original base editors, and some APOBECs bind single-stranded DNA in a conformation that resembles tRNA bound to TadA<sup>28</sup>. TadA does not require small-molecule activators (unlike ADAR<sup>29</sup>) and acts on polynucleic acid (unlike ADA<sup>25</sup>). On the basis of these considerations, we chose ecTadA as the starting point of our efforts to evolve a DNA adenine deaminase.

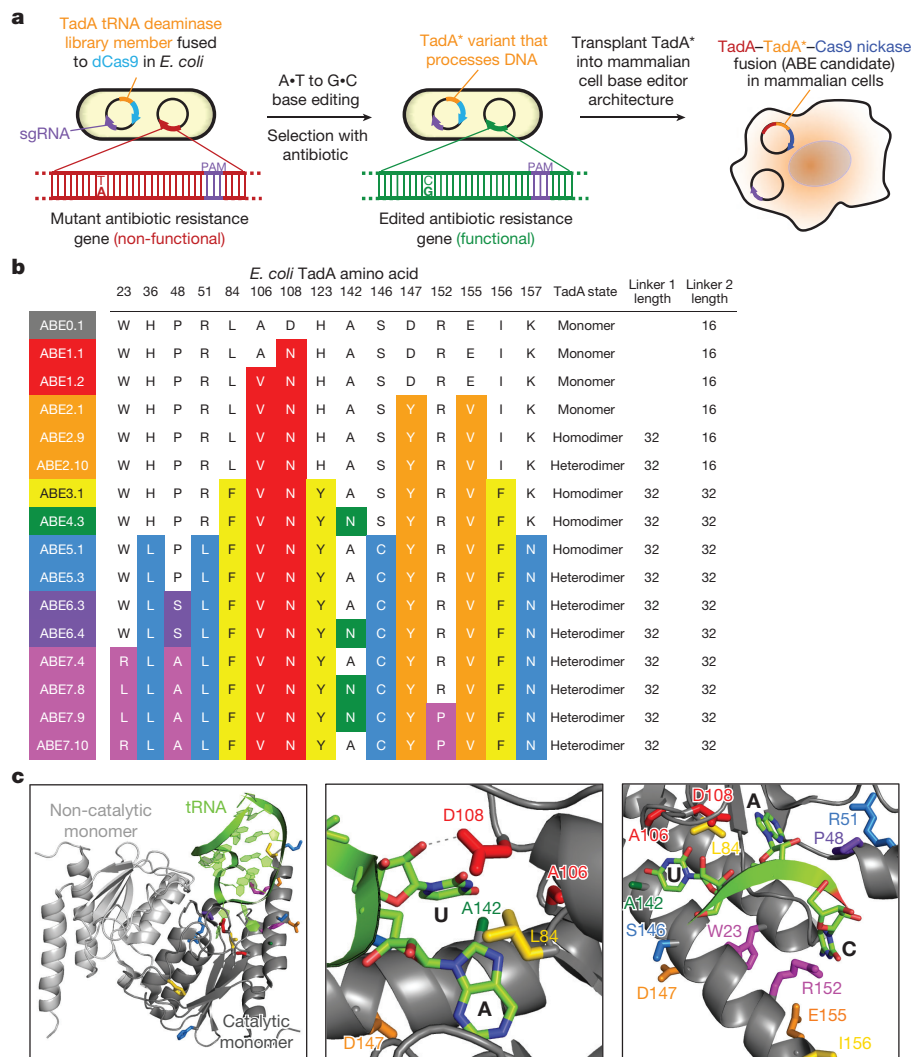
We created unbiased libraries of ecTadA–dCas9 fusions containing mutations only in the adenine deaminase portion of the construct, to avoid altering the favourable properties of the Cas9 portion of the editor (Supplementary Table 7). The resulting plasmids were transformed into *E. coli* harbouring the Cam<sup>R</sup> H193Y selection (Fig. 2a and Supplementary Table 8). Colonies that survived chloramphenicol

challenge were strongly enriched for the TadA mutations A106V and D108N (Fig. 2b). Sequence alignment of ecTadA with *S. aureus* TadA, for which a structure in complex with tRNA<sup>Arg</sup> has been reported<sup>30</sup>, predicts that the side-chain of D108 forms a hydrogen bond with the 2'-OH group of the ribose in the uracil upstream of the substrate adenine (Fig. 2c). Mutations at D108 are likely to abrogate this hydrogen bond, decreasing the energetic opportunity cost of binding DNA. DNA sequencing confirmed that all clones that survived the selection showed A•T to G•C reversion at the targeted site in Cam<sup>R</sup>. Collectively, these results indicate that mutations at or near TadA D108 enable TadA to perform adenine deamination on DNA substrates.

The TadA A106V and D108N mutations were incorporated into a mammalian codon-optimized TadA–Cas9 nickase fusion construct that replaces dCas9 with the Cas9 D10A nickase used in BE3 to manipulate cellular DNA repair to favour the desired base editing outcomes<sup>3</sup>, and adds a C-terminal nuclear localization signal (NLS). We designated the resulting TadA\*–XTEN–nCas9–NLS construct, in which TadA\* represents an evolved TadA variant and XTEN is a 16-amino acid linker used in BE3<sup>3</sup>, as ABE1.2. Transfection of plasmids expressing ABE1.2 and sgRNAs targeting six human genomic sites (Extended Data Fig. 2a) resulted in very low but observable A•T to G•C editing efficiencies (3.2 ± 0.88%; all editing efficiencies are reported as mean ± s.d. of three biological replicates five days after transfection without enrichment for transfected cells unless otherwise noted) at or near protospacer position 5, counting the PAM as positions 21–23 (Fig. 3a). These data confirmed that an ABE capable of catalysing low levels of A•T to G•C conversion emerged from the first round of protein evolution and engineering.

### Improved deaminase variants and ABE architectures

To improve editing efficiencies, we generated an unbiased library of ABE1.2 variants and challenged the resulting TadA\*1.2–dCas9 mutants in bacteria with higher concentrations of chloramphenicol than were used in round 1 (Supplementary Tables 7 and 8). From round 2 we identified two new mutations, D147Y and E155V, that were predicted to lie in a helix adjacent to the TadA tRNA substrate (Fig. 2c). In mammalian cells, ABE2.1 (ABE1.2 + D147Y + E155V) exhibited two-fold to sevenfold higher activity than ABE1.2 at the six genomic sites tested, resulting in an average of 11 ± 2.9% A•T to G•C base editing (Fig. 3a).



**Figure 2 | Protein evolution and engineering of ABEs. a**, Strategy to evolve a DNA deoxyadenosine deaminase starting from TadA. A library of *E. coli* harbours a plasmid library of mutant *ecTadA* (*TadA\**) genes fused to dCas9 and a selection plasmid requiring targeted A•T to G•C mutations to repair antibiotic resistance genes. Mutations from surviving *TadA\** variants were imported into an ABE architecture for base editing in human cells. **b**, Genotypes of a subset of evolved ABEs. For a list of 57 evolved

*TadA\** genotypes, see Extended Data Fig. 1. The dimerization state (monomer, *TadA\**-*TadA\** homodimer, or wild-type TadA-evolved *TadA\** heterodimer) and linker length (in amino acids) are also shown. **c**, Three views of the *E. coli* TadA deaminase (PDB 1Z3A) aligned with *S. aureus* TadA (not shown) complexed with tRNA<sup>Arg2</sup> (PDB 2B3J). The UAC anticodon loop of the tRNA is the native substrate of wild-type TadA.

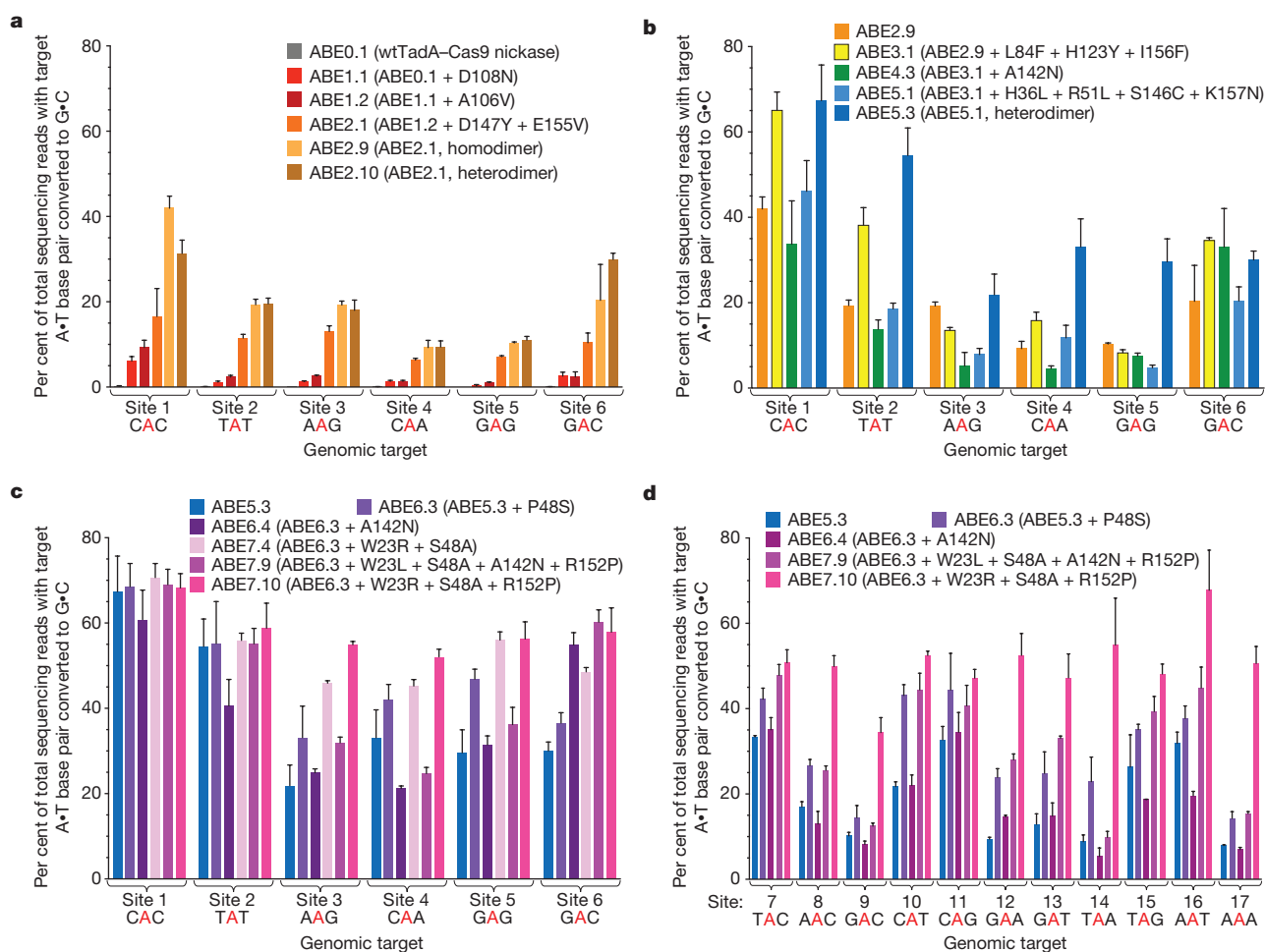
Next we sought to improve ABE2.1 through additional protein engineering. Fusing the *TadA*(2.1)\* domain to the C terminus of Cas9 nickase, instead of the N terminus, abolished editing activity (Extended Data Fig. 2c), as previously shown in BE3<sup>3</sup>. We also varied the length of the linker between *TadA*(2.1)\* and Cas9 nickase. An ABE2 variant (ABE2.6) with a linker twice as long (32 amino acids, (SGGS)<sub>2</sub>-XTEN-(SGGS)<sub>2</sub>) as the linker in ABE2.1 offered modestly higher editing efficiencies, now averaging  $14 \pm 2.4\%$  across the six genomic loci tested (Extended Data Fig. 2c).

Alkyl adenine DNA glycosylase (AAG) catalyses the cleavage of the glycosidic bond of inosine in DNA<sup>31</sup>. To test whether inosine excision impeded ABE performance, we created ABE2 variants designed to minimize potential sources of inosine excision. Given the absence of known protein inhibitors of AAG, we attempted to block endogenous AAG from accessing the inosine intermediate by separately fusing to ABE2.1 catalytically inactivated versions of enzymes involved in inosine binding or removal: human AAG (inactivated with a E125Q mutation<sup>31</sup>) or *E. coli* Endo V (inactivated with a D35A mutation<sup>32</sup>). Neither ABE2.1-AAG(E125Q) (ABE2.2) nor ABE2.1-Endo V(D35A) (ABE2.3) exhibited altered A•T to G•C editing efficiency in HEK293T

cells compared with ABE2.1 (Extended Data Fig. 2d). Indeed, ABE2.1 in Hap1 cells lacking AAG failed to increase base editing efficiency or product purity compared with Hap1 cells containing wild-type AAG (Extended Data Fig. 2e). Moreover, ABE2.1 induced virtually no indels ( $\leq 0.1\%$ ) or A•T to non-G•C products ( $\leq 0.1\%$ ) in HEK293T cells, consistent with inefficient excision of inosine (Extended Data Fig. 3). Together, these observations suggest that cellular repair of inosine intermediates created by ABEs is inefficient, obviating the need to subvert base excision repair. This situation contrasts with that of BE3 and BE4, which are strongly dependent on inhibiting uracil excision to maximize base editing efficiency and product purity<sup>3,5</sup>.

As a final ABE2 engineering study, we investigated the effect of *TadA\** dimerization on base editing efficiency. TadA natively operates as a homodimer, with one monomer catalysing deamination and the other monomer acting as a docking station for the tRNA substrate<sup>30</sup>. During selection in *E. coli*, endogenous TadA probably serves as the non-catalytic monomer. We hypothesized that tethering an additional wild-type or evolved TadA monomer might improve base editing in mammalian cells by minimizing reliance on intermolecular ABE dimerization. Indeed, co-expression with ABE2.1 of either wild-type





**Figure 3 | Evolved ABEs mediate A•T to G•C base editing at human genomic DNA sites. a–c,** A•T to G•C base editing efficiencies in HEK293T cells of round 1 and round 2 ABEs (**a**), round 3, round 4, and round 5 ABEs (**b**), and round 6 and round 7 ABEs (**c**) at six human genomic DNA sites. **d,** Editing efficiencies in HEK293T cells of round 6 and round 7 ABEs at an expanded set of human genomic sites. Values

TadA or TadA\*2.1 (ABE2.7 or ABE2.8, respectively), as well as direct fusion of either evolved or wild-type TadA to the N terminus of ABE2.1 (ABE2.9 or ABE2.10, respectively), substantially improved editing efficiency (Fig. 3a and Extended Data Fig. 4a). A fused TadA\*–ABE2.1 architecture (ABE2.9) offered the highest editing efficiency (averaging  $20 \pm 3.8\%$  across the six genomic loci, a  $7.6 \pm 2.6$ -fold average improvement at each site over ABE1.2) and therefore a dimeric architecture was used in all subsequent experiments (Figs 2b, 3a).

Finally, we determined which of the two TadA\* subunits within the TadA\*–ABE2.1 fusion was responsible for catalysing conversion of adenine to inosine. We introduced an inactivating E59A mutation<sup>22</sup> into either the N-terminal or the internal TadA\* monomer of ABE2.9. The variant with an inactivated N-terminal TadA\* subunit (ABE2.11) demonstrated comparable editing efficiencies to ABE2.9, whereas the variant with an inactivated internal TadA\* subunit (ABE2.12) lost all editing activity (Extended Data Fig. 4a). These results establish that the internal TadA subunit is responsible for catalysing adenine deamination.

### ABEs that efficiently edit a subset of targets

Next we performed a third round of bacterial evolution starting with TadA\*2.1–dCas9 to increase editing efficiency further. We increased selection stringency by introducing two early stop codons (Q4stop and W15stop) into the kanamycin resistance gene (*Kan<sup>R</sup>*, aminoglycoside phosphotransferase, Supplementary Table 8 and Supplementary

and error bars reflect the mean and s.d. of three independent biological replicates performed on different days. Homodimer indicates fused TadA\*–TadA\*–Cas9 nickase architecture; heterodimer indicates fused wtTadA–TadA\*–Cas9 nickase architecture. All ABEs in **b** are homodimers except ABE5.3; ABEs in **c** and **d** are all heterodimers.

Sequences 2). Each of the mutations requires an A•T to G•C reversion to correct the premature stop codon. We subjected a library of TadA\*2.1–dCas9 variants containing mutations in the TadA domain to this higher stringency selection (Supplementary Table 8), resulting in the strong enrichment of three new TadA mutations: L84F, H123Y, and I157F. These mutations were imported into ABE2.9 to generate ABE3.1 (Fig. 2b). In HEK293T cells, ABE3.1 resulted in editing efficiencies averaging  $29 \pm 2.6\%$  across the six tested sites, a 1.6-fold average increase in A•T to G•C conversion at each site over ABE 2.9, and a 11-fold average improvement over ABE1.2 (Fig. 3b). Using longer (64- or 100-amino-acid) linkers between the two TadA monomers, or between TadA\* and Cas9 nickase, did not consistently improve editing efficiency compared to ABE3.1 (Extended Data Figs 1, 4b).

Although ABE3.1 mediated efficient base editing at some targets, such as the CAC in site 1 ( $65 \pm 4.2\%$  conversion), for other sites, such as the GAG in site 5, editing efficiency was much lower ( $8.3 \pm 0.67\%$ ) (Fig. 3b). The results from six genomic loci with different sequence contexts surrounding the target adenine suggested that ABEs from rounds 1–3 strongly preferred target sequence contexts of YAC, where Y is T or C. This preference is likely to have been inherited from the substrate specificity of native *E. coli* TadA, which deaminates the adenine in the UAC anticodon of tRNA<sup>Arg</sup>. The utility of an ABE would be greatly limited, however, by such a target sequence restriction.

To overcome this sequence preference, we initiated a fourth evolution campaign focusing mutagenesis at TadA residues that were predicted



Untreated																						Indel%
Site 2		G <sub>1</sub>	A <sub>2</sub>	G <sub>3</sub>	T <sub>4</sub>	A <sub>5</sub>	T <sub>6</sub>	G <sub>7</sub>	A <sub>8</sub>	G <sub>9</sub>	G <sub>10</sub>	C <sub>11</sub>	A <sub>12</sub>	T <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	0.04
A	0.0	100	0.0	0.0	99.8	0.0	0.1	99.9	0.1	0.0	0.0	99.9	0.0	100	0.0	99.9	0.0	0.0	0.0	0.0	0.0	0.0
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.1	0.0	0.0	0.0	0.0	100	0.0	0.0	99.9	0.0
G	99.9	0.0	100	0.0	0.2	0.0	99.8	0.0	99.9	99.9	0.0	0.0	0.0	0.0	0.0	0.0	100	0.1	0.0	0.0	100	0.1
T	0.0	0.0	0.0	100	0.0	100	0.0	0.1	0.1	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0
ABE7.10																						Indel%
Site 2		G <sub>1</sub>	A <sub>2</sub>	G <sub>3</sub>	T <sub>4</sub>	A <sub>5</sub>	T <sub>6</sub>	G <sub>7</sub>	A <sub>8</sub>	G <sub>9</sub>	G <sub>10</sub>	C <sub>11</sub>	A <sub>12</sub>	T <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	0.1
A	0.1	99.9	0.0	0.0	38.5	0.0	0.2	97.5	0.0	0.0	0.0	0.0	99.8	0.0	100	0.1	100	0.0	0.0	0.0	0.0	0.0
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9
G	99.9	0.1	100	0.0	61.5	0.0	99.7	2.5	99.9	99.9	0.0	0.1	0.0	0.0	0.0	99.9	0.0	0.0	0.0	99.9	0.0	0.0
T	0.1	0.0	0.0	100	0.0	99.9	0.1	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	100	0.1	0.1
ABE7.9																						Indel%
Site 2		G <sub>1</sub>	A <sub>2</sub>	G <sub>3</sub>	T <sub>4</sub>	A <sub>5</sub>	T <sub>6</sub>	G <sub>7</sub>	A <sub>8</sub>	G <sub>9</sub>	G <sub>10</sub>	C <sub>11</sub>	A <sub>12</sub>	T <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	0.04
A	0.1	99.8	0.0	0.0	40.8	0.0	0.2	87.2	0.0	0.0	0.0	0.0	99.9	0.0	100	0.0	100	0.0	0.0	0.0	0.0	0.0
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	99.9
G	99.9	0.2	100	0.0	59.2	0.0	99.8	12.7	100	100	0.0	0.1	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100	0.0
T	0.0	0.0	0.0	100	0.0	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0	100	0.0	0.0
Untreated																						Indel%
Site 6		G <sub>1</sub>	G <sub>2</sub>	A <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	G <sub>6</sub>	A <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>10</sub>	A <sub>11</sub>	G <sub>12</sub>	G <sub>13</sub>	C <sub>14</sub>	C <sub>15</sub>	A <sub>16</sub>	G <sub>17</sub>	G <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	0.02
A	0.0	0.0	100	0.0	0.0	0.1	99.8	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100	100	100	0.0	0.0	0.0	100	99.9	0.0	0.0	0.0	0.0	0.0	100
G	100	100	0.0	0.0	0.0	99.9	0.2	0.0	0.0	0.0	0.0	0.0	100	100	0.0	0.0	0.0	0.0	100	100	100	0.0
T	0.0	0.0	0.0	100	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ABE7.10																						Indel%
Site 6		G <sub>1</sub>	G <sub>2</sub>	A <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	G <sub>6</sub>	A <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>10</sub>	A <sub>11</sub>	G <sub>12</sub>	G <sub>13</sub>	C <sub>14</sub>	C <sub>15</sub>	A <sub>16</sub>	G <sub>17</sub>	G <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	0.05
A	0.0	0.0	98.8	0.0	0.0	0.0	37.1	0.0	0.0	0.0	0.0	97.5	0.0	0.0	0.0	0.0	100	0.1	0.0	0.0	0.0	0.0
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	99.9	99.9	0.0	0.0	0.0	100	100	0.0	0.1	0.0	0.0	0.0	100
G	100	100	7.2	0.0	0.0	100	62.9	0.0	0.0	0.0	2.5	100	100	0.0	0.0	0.0	0.0	99.9	100	100	0.0	0.0
T	0.0	0.0	0.0	100	100	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
ABE7.9																						Indel%
Site 6		G <sub>1</sub>	G <sub>2</sub>	A <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	G <sub>6</sub>	A <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>10</sub>	A <sub>11</sub>	G <sub>12</sub>	G <sub>13</sub>	C <sub>14</sub>	C <sub>15</sub>	A <sub>16</sub>	G <sub>17</sub>	G <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	0.03
A	0.0	0.0	98.6	0.0	0.0	0.0	36.5	0.1	0.0	0.0	0.0	99.8	0.1	0.0	0.0	0.0	100	0.0	0.0	0.0	0.0	0.0
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100	100	0.0	0.0	0.0	100	99.9	0.0	0.0	0.0	0.0	0.0	99.9
G	100	100	1.4	0.0	0.0	100	63.5	0.0	0.0	0.0	0.2	99.9	100	0.0	0.0	0.0	0.0	99.9	100	100	0.0	0.0
T	0.0	0.0	0.0	100	100	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1

**Figure 4 | Product purity of late-stage ABEs.** Product distributions and indel frequencies at two representative human genomic DNA sites in HEK293T cells treated with ABE7.10 or ABE7.9 and the corresponding sgRNA, or in untreated HEK293T cells. At every position, 22,746–111,215 sequencing reads were used.

to interact with the nucleotides upstream and downstream of the target adenine<sup>30</sup>. We subjected TadA\*2.1–dCas9 libraries (Supplementary Table 7) containing randomized amino acids at five such positions (E25, R26, R107, A142, and A143) to a new selection in which A•T to G•C conversion of a non-YAC target (GAT, which causes a T89I mutation in the spectinomycin resistance protein) restores antibiotic resistance (Supplementary Table 8 and Supplementary Sequences 2). Surviving bacteria converted strongly on the TadA mutation A142N. Although apparent A•T to G•C base editing efficiency in bacterial cells with TadA\*4.3–dCas9 (TadA\*3.1 + A142N–dCas9) was higher than with TadA\*3.1–dCas9, as judged by spectinomycin resistance (Extended Data Fig. 4c), in mammalian cells ABE4.3 exhibited decreased base editing efficiency (averaging  $16 \pm 5.8\%$ ) compared with ABE3.1 (Figs 2b, 3b). We hypothesized that the A142N mutation might benefit base editing in a context-dependent manner, and revisited its inclusion in later rounds of evolution (see below).

We performed a fifth round of evolution to increase ABE catalytic performance and broaden target sequence compatibility. We generated a library of TadA\*3.1–dCas9 variants containing unbiased mutations throughout the TadA\* domain as before (Supplementary Table 7). To favour ABE constructs with faster kinetics, we subjected this library to Cam<sup>R</sup> H193Y selection with higher doses of chloramphenicol after allowing ABE to be expressed for only half the duration (7 h) of the previous rounds of evolution (about 14 h) (Supplementary Table 8). Surprisingly, importing a consensus set of mutations from surviving clones (H36L, R51L, S146C, and K157N) into ABE3.1, thereby creating ABE5.1, decreased overall editing efficiencies in HEK293T cells by  $1.7 \pm 0.29$ -fold (Figs 2b, 3b).

ABE5.1 included seven mutations since our dimerization state experiments on ABE2.1. We speculated that the accumulation of these mutations might have impaired the ability of the non-catalytic N-terminal TadA subunit to play its structural role in mammalian cells. In *E. coli*, endogenous wild-type TadA is provided *in trans*, potentially explaining the difference between bacterial selection phenotypes and mammalian cell editing efficiencies. Therefore, we examined the effect of using wild-type TadA instead of evolved TadA\* variants in the N-terminal TadA domain of ABE5 variants. A heterodimeric construct

containing wild-type *E. coli* TadA fused to an internal evolved TadA\* (ABE5.3) exhibited greatly improved editing efficiency compared to homodimeric ABE5.1 with two identical evolved TadA\* domains. ABE5.3 had an average editing efficiency across the six genomic test sites of  $39 \pm 5.9\%$ , with an average improvement at each site of  $2.9 \pm 0.78$ -fold relative to ABE5.1 (Figs 2b, 3b). Notably, ABE5.3 also showed broadened sequence compatibility that now enabled 22–33% editing of non-YAC targets in sites 3–6 (Fig. 3b).

Concurrently, we subjected a library from round 5 to the non-YAC spectinomycin selection used in round 4. Although no highly enriched or beneficial mutations emerged (Extended Data Fig. 5a), mutations from two genotypes that emerged from this selection, N72D + G125A and P48S + S97C, were included in subsequent library generation steps. In addition, eight heterodimeric wild-type TadA–TadA\* ABE5.3 variants (ABE5.5 to ABE5.12) containing 24-, 32-, or 40-residue linkers between the TadA domains or between TadA and Cas9 nickase showed no obvious improvements in base editing efficiency over ABE5.3 (Extended Data Figs 1, 5b). For subsequent studies, we therefore used the ABE5.3 architecture containing heterodimeric wtTadA–TadA\*–Cas9 nickase with two 32-residue linkers.

### Highly active ABEs with broad sequence compatibility

During the sixth round of evolution, we aimed to remove any non-beneficial mutations by DNA shuffling and to re-examine mutations from previous rounds of evolution that may benefit ABE performance once liberated from negative epistasis with other mutations. Evolved TadA\*–dCas9 variants from rounds 1–5 along with wild-type *E. coli* TadA were shuffled and subjected to spectinomycin resistance T89I selection (Supplementary Table 8). Two mutations were strongly enriched from this selection: P48S/T and A142N (first seen in round 4). These mutations were added either separately or together to ABE5.3, forming ABE6.1 to ABE6.6 (Extended Data Fig. 1). ABE6.3 (ABE5.3 + P48S) resulted in  $1.3 \pm 0.28$ -fold higher average editing than ABE5.3 at the six genomic sites tested, and an average conversion efficiency of  $47 \pm 5.8\%$  (Figs 2b, 3c). P48 is predicted to lie approximately 5 Å from the substrate adenosine 2'-hydroxyl in the TadA crystal structure (Fig. 2c), and we speculated that mutating this

residue to Ser might improve compatibility with a deoxyadenosine substrate. Although at most sites ABE6 variants that contained the A142N mutation were less active than ABEs that lack this mutation, editing by ABE6.4 (ABE6.3 + A142N) at site 6, which contains a target A at position 7 in the protospacer, was  $1.5 \pm 0.13$ -fold more efficient than editing by ABE6.3, and  $1.8 \pm 0.16$ -fold more efficient than editing by ABE5.3 (Fig. 3c). These results suggest that ABEs containing A142N may offer improved editing of adenines closer to the PAM than position 5.

Although six rounds of evolution and engineering yielded substantial improvements, ABE6 editors still suffered from reduced editing efficiencies (about 20–40%) at target sequences containing multiple adenines near the targeted A (Fig. 3c). To address this challenge, we performed a seventh round of evolution in which new unbiased libraries of TadA\*6–dCas9 variants were targeted to two separate sites in the kanamycin resistance gene: the Q4stop mutation used in round 3, which requires editing of TAT, and a new D208N mutation that requires editing of TAA (Supplementary Table 7 and 8, Supplementary Sequences 2). Surviving clones contained three enriched sets of mutations: W23L/R, P48A, and R152H/P.

Introducing these mutations separately or in combinations into mammalian cell ABEs (ABE7.1 to ABE7.10) substantially improved editing efficiencies, especially at targets that contain multiple A residues (Figs 2b, 3c, d, and Extended Data Figs 1, 6a, b). ABE7.10 edited the six genomic test sites with an average efficiency of  $58 \pm 4.0\%$ , an average improvement at each site of  $1.3 \pm 0.20$ -fold relative to ABE6.3 (Fig. 3c), and  $29 \pm 7.4$ -fold compared to ABE1.2. Although mutational dissection revealed that all three of the new mutations contributed to the enhanced editing efficiency (Extended Data Figs 1, 6a, b), the R152P substitution is particularly noteworthy, as this residue is predicted to contact the C in the UAC anticodon loop of the tRNA substrate (Fig. 2b, c). We speculate that substitution of Arg for Pro abrogates base-specific enzyme–DNA interactions and thereby broadens target sequence compatibility.

### Characterization of late-stage ABEs

We characterized the most promising ABEs from rounds 5–7 in depth. We chose an expanded set of 17 human genomic targets that place a target A at position 5 or 7 of the protospacer and collectively include all possible NAN sequence contexts (Extended Data Fig. 2a). Overall, we observed strong improvement of A•T to G•C editing efficiency in HEK293T cells during the progression from ABE5 to ABE7 variants (Fig. 3c, d). The base editing efficiency of the most active editor overall, ABE7.10, averaged  $53 \pm 3.7\%$  at the 17 sites tested, exceeded 50% at 11 of these sites, and ranged from 34–68% (Figs 3c, d, 4). These results compare favourably to the typical C•G to T•A editing efficiency of BE3<sup>3</sup>.

Next, we further characterized the base editing activity window of late-stage ABEs. We chose a human genomic site containing an alternating 5'-A-N-A-N-A-N-3' sequence that could be targeted with either of two sgRNAs such that an A would be located either at every odd position (site 18) or at every even position (site 19) from 2 to 9 in the protospacer (Extended Data Fig. 2a). The resulting editing outcomes (Extended Data Fig. 7a), together with an analysis of editing efficiencies at every protospacer position across all 19 sites tested (Extended Data Fig. 7b), suggest that the activity windows of late-stage variants are approximately 4–6 nucleotides wide, from approximately protospacer positions 4 to 7 for ABE7.10, and positions 4 to 9 for ABE6.3, ABE7.8, and ABE7.9, counting the PAM as positions 21–23 (Fig. 5). We note that the precise editing window boundaries can vary in a target-dependent manner (Supplementary Table 1), as is the case with BE3 and BE4. We also tested ABE7.8–ABE7.10 in U2OS cells at sites 1–6 and observed similar editing results to those obtained in HEK293T cells (Extended Data Fig. 6c), demonstrating that ABE activity is not limited to HEK293T cells.

Analysis of individual high-throughput DNA sequencing reads from ABE editing at 6–17 genomic sites in HEK293T cells revealed that base

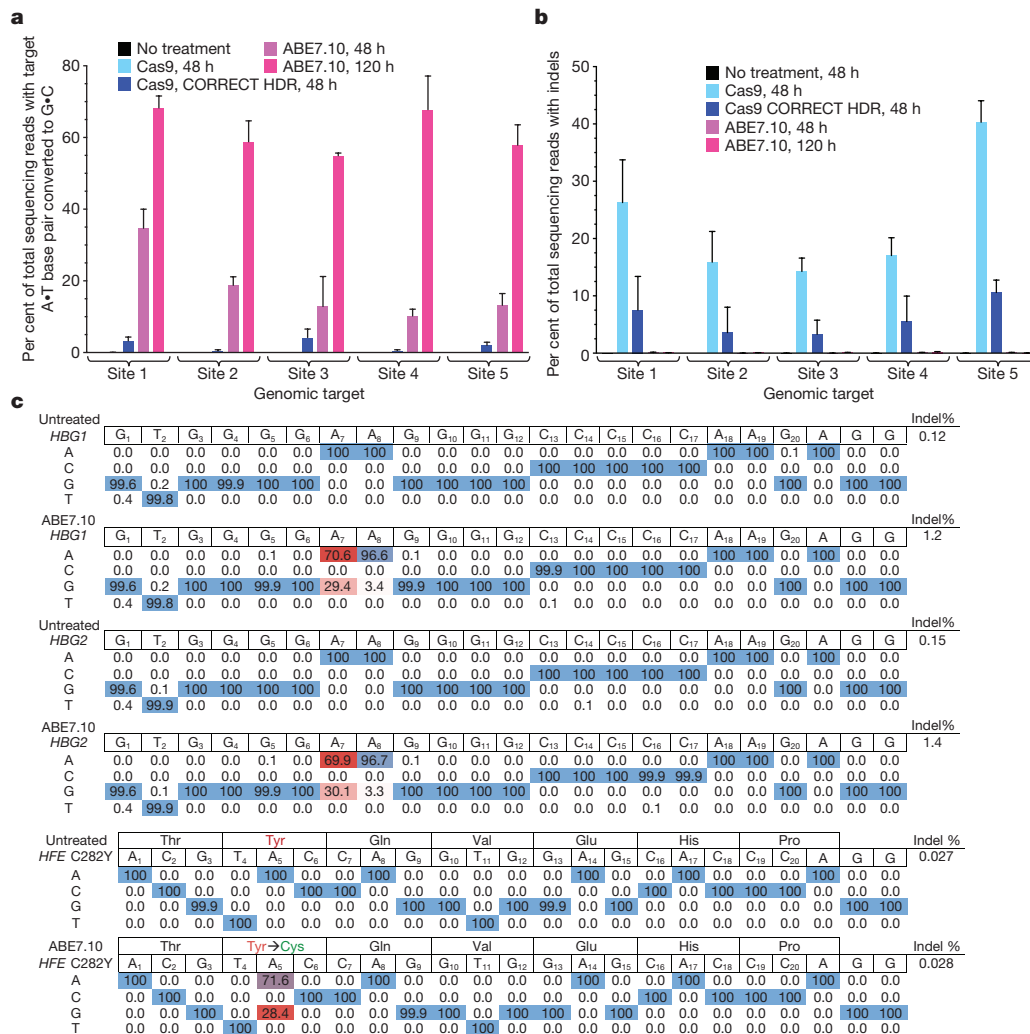
editing outcomes at nearby adenines within the editing window are not statistically independent events. The average normalized linkage disequilibrium between nearby target adenines increased steadily as ABE evolution proceeded (Extended Data Fig. 8), indicating that early-stage ABEs edit nearby adenines more independently, whereas late-stage ABEs edit nearby adenines more processively. These findings suggest that TadA might have evolved kinetic changes that decrease the likelihood of substrate release before additional adenines within the editing window are converted, resulting in processivity similar to that of BE3<sup>3</sup>.

In contrast to the formation of C to non-T edits and indels that can arise from BE3-mediated base editing of cytidines, ABEs convert A•T to G•C very cleanly in HEK293T and U2OS cells, with indel frequencies and A to non-G editing similar to those of untreated cells (typically 0.1% or less) among the 17 genomic NAN sites tested (Fig. 4 and Supplementary Table 1). The undesired products of BE3 arise from uracil excision and downstream repair processes<sup>5</sup>. The remarkable product purity of all tested ABE variants suggests that the activity or abundance of enzymes that remove inosine from DNA may be low compared to those of uracil N-glycosylase (UNG), resulting in minimal base excision repair following ABE editing.

We compared the efficiency of ABE7.10-catalysed A•T to G•C editing to that of a current Cas9 nuclease-mediated HDR method, CORRECT<sup>33</sup>. At five genomic loci in HEK293T cells, we observed average target point mutation frequencies ranging from 0.47% to 4.2% with 3.3% to 10.6% indels using the CORRECT HDR method under optimized 48-h conditions in HEK293T cells (Fig. 5a). At the same five genomic loci, ABE7.10 resulted in average target mutation frequencies of 10–35% after 48 h, and 55–68% after 120 h (Fig. 5a), with fewer than 0.1% indels (Fig. 5b). The target mutation/indel ratio averaged 0.43 for CORRECT HDR, and more than 500 for ABE7.10, representing an improvement of over 1,000-fold in product selectivity for ABE7.10. Although HDR is well-suited to introduce insertions and deletions into genomic DNA, these results demonstrate that ABE7.10 can introduce A•T to G•C point mutations with much higher efficiency and far fewer undesired products than a current Cas9 nuclease-mediated HDR method.

Next we examined off-target editing by ABE7 variants. As no method yet exists to comprehensively profile the off-target activity of ABEs, we assumed that off-target ABE editing occurred primarily at the off-target sites that are edited when Cas9 nuclease is complexed with the same guide RNA, as is the case with BE3<sup>3,8,34</sup>. We treated HEK293T cells with three well-characterized guide RNAs<sup>35</sup> and either Cas9 nuclease or ABE7 variants, and sequenced the on-target loci and the 12 most active off-target human genomic loci associated with these guide RNAs as identified by the genome-wide GUIDE-Seq method<sup>35</sup>. The efficiency of on-target indels by Cas9 and the efficiency of on-target base editing by ABE7.10 both averaged 54% (Supplementary Tables 2–4). We observed detectable modification (0.2% indels or more) by Cas9 nuclease at nine of the 12 (75%) known off-target loci (Fig. 5c and Supplementary Tables 2–4). In contrast, when complexed with the same sgRNAs, ABE7.10, ABE7.9, or ABE 7.8 produced 0.2% or more off-target base editing at only four of the twelve (33%) known Cas9 off-target sites. Moreover, the nine confirmed Cas9 off-target loci were modified with an average efficiency of 14% indels, whereas the four confirmed ABE off-target loci were modified with an average of only 1.3% A•T to G•C mutations (Supplementary Tables 2–4). Although seven of the nine confirmed Cas9 off-target loci contained at least one adenine within the ABE activity window, three of these seven off-target loci were not detectably edited by ABE7.8, 7.9, or 7.10. Together, these data strongly suggest that ABE7 variants may be less prone to off-target genome modification than Cas9 nuclease, although a comprehensive, unbiased method of profiling the DNA specificity of ABEs is needed. In addition, we did not detect any apparent ABE-induced A•T to G•C DNA editing outside on-target or off-target protospacers following ABE treatment.

Although additional studies are needed to examine possible RNA editing by ABEs, we observed no elevated adenine mutation rate among



**Figure 5 | Comparison of ABE7.10-mediated base editing and Cas9-mediated HDR, and application of ABE7.10 to two disease-relevant SNPs. a, A•T to G•C base editing efficiencies in HEK293T cells treated either with ABE7.10 or with Cas9 nuclease and an ssDNA donor template (following the CORRECT HDR method<sup>33</sup>) targeted to five human genomic DNA sites. b, Indel formation in HEK293T cells treated as described in a. c, Application of ABE to install a disease-suppressing SNP, or to correct a**

disease-inducing SNP. Top, ABE7.10-mediated –198T→C mutation (on the strand complementary to the one shown) in the promoter region of *HBG1* and *HBG2* genes in HEK293T cells. The target adenine is at protospacer position 7. Bottom, ABE7.10-mediated reversion of the C282Y mutation in the *HFE* gene in LCL cells. The target adenine is at protospacer position 5.

four abundant mRNAs in ABE7.10-treated HEK293T cells compared to untreated cells (Extended Data Table 1), nor any apparent ABE toxicity in bacterial or human cells under the conditions used here. We speculate that the evolved mutations at TadA residues known to interact with the ribose 2'-hydroxyl (Fig. 2c), the fused Cas9 nickase, or ABE nuclear localization may impede RNA editing.

**Installation of disease-relevant mutations with ABE**

Finally, we tested the potential of ABEs to introduce disease-suppressing mutations and to correct pathogenic mutations in human cells. Mutations in  $\beta$ -globin genes cause a variety of blood diseases. Humans with the rare benign condition HPFH (hereditary persistence of fetal hemoglobin) are resistant to some  $\beta$ -globin diseases, including sickle-cell anaemia. In certain patients, this phenotype is mediated by mutations in the promoters of the  $\gamma$ -globin genes *HBG1* and *HBG2* that enable sustained expression of fetal hemoglobin, which is normally silenced in humans around birth<sup>36</sup>. We designed an sgRNA that programs ABE to simultaneously mutate –198T to C in the promoter that drives *HBG1* expression, and –198T to C in the promoter that drives *HBG2* expression, by placing the target A•T base pair at protospacer position 7. These mutations are known to confer British-type

HPFH and enable fetal hemoglobin production in adults<sup>37</sup>. ABE7.10 installed the desired T•A to C•G mutations in the *HBG1* and *HBG2* promoters with 29% and 30% efficiency, respectively, in HEK293T cells (Fig. 5c, Extended Data Fig. 9).

The iron storage disorder hereditary haemochromatosis (HHC) is an autosomal recessive genetic disorder commonly caused by a G to A mutation at nucleotide 845 in the human *HFE* gene, resulting in a C282Y substitution in the HFE protein that leads to excessive iron absorption and potentially life-threatening elevation of serum ferritin<sup>38</sup>. We transfected DNA encoding ABE7.10 and a guide RNA that places the target adenine at protospacer position 5 into an immortalized lymphoblastoid cell line (LCL) harbouring the *HFE* C282Y genomic mutation. Owing to the extreme resistance of LCL cells to transfection, we isolated transfected cells and measured editing efficiency by high-throughput DNA sequencing (HTS) of their genomic DNA. We observed the clean conversion of the Tyr282 codon to Cys282 in 28% of sequencing reads from transfected cells, with no evidence of undesired editing or indels at the on-target locus (Fig. 5c). Although much additional research is needed to develop these and other ABE editing strategies into potential clinical therapies for diseases with a genetic component, including the development of ABEs that accept



a wide variety of PAMs<sup>7</sup>, these examples demonstrate the potential of ABEs to correct disease-driving mutations, and to install mutations known to suppress genetic disease phenotypes, in human cells.

In summary, seven rounds of evolution and engineering transformed a protein with no ability to deaminate adenine at target loci in DNA (wild-type TadA-dCas9) into forms that edit DNA weakly (ABE1s and ABE2s), variants that edit limited subsets of sites efficiently (ABE3s, ABE4s, and ABE5s), and, ultimately, highly active ABEs with broad sequence compatibility (ABE6s and ABE7s). We recommend ABE7.10 for general A•T to G•C base editing. When the target adenine is at protospacer positions 8–10, ABE7.9, ABE7.8, or ABE6.3 may offer higher editing efficiencies than ABE7.10, although conversion efficiencies at these positions are typically lower than at protospacer positions 4–7. The development of ABEs greatly expands the capabilities of base editing and the fraction of pathogenic SNPs that can be addressed by genome editing without introducing DSBs (Fig. 1a). Together with BE3<sup>3</sup> and BE4<sup>5</sup>, these ABEs advance the field of genome editing by enabling the direct installation of all four transition mutations at target loci in living cells with a minimum of undesired byproducts.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

**Received 5 October; accepted 17 October 2017.**

**Published online 25 October 2017.**

- Krokan, H. E., Drabløs, F. & Slupphaug, G. Uracil in DNA—occurrence, consequences and repair. *Oncogene* **21**, 8935–8948 (2002).
- Lewis, C. A. Jr, Crayle, J., Zhou, S., Swanstrom, R. & Wolfenden, R. Cytosine deamination and the precipitous decline of spontaneous mutation during Earth's history. *Proc. Natl Acad. Sci. USA* **113**, 8194–8199 (2016).
- 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).
- Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729 (2016).
- Komor, A. C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci. Adv.* **3**, eaao4774 (2017).
- Komor, A. C., Badran, A. H. & Liu, D. R. Editing the genome without double-stranded DNA breaks. *ACS Chem. Biol.* (2017).
- Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
- Rees, H. A. et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).
- Satomura, A. et al. Precise genome-wide base editing by the CRISPR Nickase system in yeast. *Sci. Rep.* **7**, 2095 (2017).
- Lu, Y. & Zhu, J.-K. Precise editing of a target base in the rice genome using a modified CRISPR/Cas9 system. *Mol. Plant* **10**, 523–525 (2017).
- Zong, Y. et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
- Zhang, Y. et al. Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. *Nat. Commun.* **8**, 118 (2017).
- Billon, P. et al. CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. *Mol. Cell* **67**, 1068–1079 (2017).
- Kuscu, C. et al. CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. *Nat. Methods* **14**, 710–712 (2017).
- Kim, K. et al. Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
- Chadwick, A. C., Wang, X. & Musunuru, K. *In vivo* base editing of PCSK9 (proprotein convertase subtilisin/kexin Type 9) as a therapeutic alternative to genome editing. *Arterioscler. Thromb. Vasc. Biol.* **37**, 1741–1747 (2017).
- Liang, P. et al. Correction of  $\beta$ -thalassaemia mutant by base editor in human embryos. *Protein Cell* <https://doi.org/10.1007/s13238-017-0475-6> (2017).
- Li, G. et al. Highly efficient and precise base editing in discarded human triprenuclear embryos. *Protein Cell* **8**, 776–779 (2017).
- Tang, W., Hu, J. H. & Liu, D. R. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. *Nat. Commun.* **8**, 15939 (2017).
- Yasui, M. et al. Miscoding properties of 2'-deoxyinosine, a nitric oxide-derived DNA Adduct, during translesion synthesis catalyzed by human DNA polymerases. *J. Mol. Biol.* **377**, 1015–1023 (2008).
- Zheng, Y., Lorenzo, C. & Beal, P. A. DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. *Nucleic Acids Res.* **45**, 3369–3377 (2017).
- Kim, J. et al. Structural and kinetic characterization of *Escherichia coli* TadA, the wobble-specific tRNA deaminase. *Biochemistry* **45**, 6407–6416 (2006).
- Wolf, J., Gerber, A. P. & Keller, W. tadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. *EMBO J.* **21**, 3841–3851 (2002).
- Matthews, M. M. et al. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat. Struct. Mol. Biol.* **23**, 426–433 (2016).
- Grunebaum, E., Cohen, A. & Roifman, C. M. Recent advances in understanding and managing adenosine deaminase and purine nucleoside phosphorylase deficiencies. *Curr. Opin. Allergy Clin. Immunol.* **13**, 630–638 (2013).
- Gerber, A. P. & Keller, W. An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* **286**, 1146–1149 (1999).
- Fukui, K. DNA mismatch repair in eukaryotes and bacteria. *J. Nucleic Acids* **2010**, 260512 (2010).
- Shi, K. et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. *Nat. Struct. Mol. Biol.* **24**, 131–139 (2017).
- Macbeth, M. R. et al. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* **309**, 1534–1539 (2005).
- Loosey, H. C., Ruthenburg, A. J. & Verdine, G. L. Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. *Nat. Struct. Mol. Biol.* **13**, 153–159 (2006).
- Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D. & Ellenberger, T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl Acad. Sci. USA* **97**, 13573–13578 (2000).
- Vik, E. S. et al. Endonuclease V cleaves at inosines in RNA. *Nat. Commun.* **4**, 2271 (2013).
- Paquet, D. et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125–129 (2016).
- Kim, D. et al. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017).
- Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
- Traxler, E. A. et al. A genome-editing strategy to treat  $\beta$ -hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. *Nat. Med.* **22**, 987–990 (2016).
- Wienert, B. et al. KLF1 drives the expression of fetal hemoglobin in British HPFH. *Blood* **130**, 803–807 (2017).
- Alexander, J. & Kowdley, K. V. HFE-associated hereditary hemochromatosis. *Genet. Med.* **11**, 307–313 (2009).
- Landrum, M. J. et al. ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–D868 (2016).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** This work was supported by DARPA HR0011-17-2-0049, US NIH RM1 HG009490, R01 EB022376, and R35 GM118062, and HHMI. A.C.K. and D.I.B. were Ruth L. Kirchstein National Research Service Awards Postdoctoral Fellows (F32 GM 112366 and F32 GM106621, respectively). M.S.P. was an NSF Graduate Research Fellow and was supported by training grant T32 GM008313. We thank Z. Niziolek for technical assistance. N.M.G. thanks A. E. Martin for his encouragement.

**Author Contributions** N.M.G. designed the research, performed all evolution experiments, conducted human cell experiments, analysed data, and wrote the manuscript. A.C.K. assisted with experimental design and human cell experiments and analysed data. H.A.R. performed HDR and off-target experiments. M.S.P. performed computational data analyses and developed HTS processing scripts. A.H.B. contributed to selection design and evolution strategy. D.I.B. assisted with cloning of late-stage ABEs. D.R.L. designed and supervised the research and wrote the manuscript. All of the authors contributed to editing the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to D.R.L. ([driiu@fas.harvard.edu](mailto:driiu@fas.harvard.edu)).

## METHODS

**General methods.** DNA amplification was conducted by PCR using Phusion U Green Multiplex PCR Master Mix (ThermoFisher Scientific) or Q5 Hot Start High-Fidelity 2× Master Mix (New England BioLabs) unless otherwise noted. All mammalian cell and bacterial plasmids were assembled using the USER cloning method as previously described<sup>40</sup> and starting material gene templates were synthetically accessed as either bacterial or mammalian codon-optimized gBlock Gene Fragments (Integrated DNA Technologies). All sgRNA expression plasmids were constructed by one-piece blunt-end ligation of a PCR product containing a variable 20-nucleotide sequence corresponding to the desired sgRNA targeted site. Primers and templates used in the synthesis of all sgRNA plasmids are listed in Supplementary Table 5. All mammalian ABE constructs, sgRNA plasmids and bacterial constructs were transformed and stored as glycerol stocks at  $-80^{\circ}\text{C}$  in Mach1 T1<sup>R</sup> Competent Cells (Thermo Fisher Scientific), which are *recA*<sup>-</sup>. Molecular biology grade, Hyclone water (GE Healthcare Life Sciences) was used in all assays and PCR reactions. All vectors used in evolution experiments and mammalian cell assays were purified using ZymoPURE Plasmid Midiprep (Zymo Research Corporation), which includes endotoxin removal. Antibiotics used for either plasmid maintenance or selection during evolution were purchased from Gold Biotechnology.

**Generation of bacterial *TadA*\* libraries (evolution rounds 1–3, 5 and 7).** In brief, libraries of bacterial ABE constructs were generated by two-piece USER assembly of a PCR product containing a mutagenized *E. coli TadA* gene and a PCR product containing the remaining portion of the editor plasmid (including the XTEN linker, dCas9, sgRNA, selectable marker, origin of replication, and promoter). Specifically, mutations were introduced into the starting template (Supplementary Table 7) in  $8 \times 25\ \mu\text{l}$  PCR reactions containing 75 ng–1.2  $\mu\text{g}$  of template using Mutazyme II (Agilent Technologies) following the manufacturer's protocol and primers NMG-823 and 824 (Supplementary Table 6). After amplification, the resulting PCR products were pooled and purified from polymerase and reaction buffer using a MinElute PCR Purification Kit (Qiagen). The PCR product was treated with DpnI (NEB) at  $37^{\circ}\text{C}$  for 2 h to digest any residual template plasmid. The desired PCR product was subsequently purified by gel electrophoresis using a 1% agarose gel containing 0.5  $\mu\text{g}/\text{ml}$  ethidium bromide. The PCR product was extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen) and eluted with  $30\ \mu\text{l}$  H<sub>2</sub>O. Following gel purification, the mutagenized *ecTadA* DNA fragment was amplified with primers NMG-825 and NMG-826 (Supplementary Table 6) using Phusion U Green Multiplex PCR Master Mix ( $8 \times 50\ \mu\text{l}$  PCR reactions,  $66^{\circ}\text{C}$  annealing, 20-s extension) in order to install the appropriate USER junction sequences onto the 5' and 3' ends of the fragment. The resulting PCR product was purified by gel electrophoresis. Next, the backbone of the bacterial base editor plasmid template (Supplementary Table 7), was amplified with primers NMG-799 and NMG-824 (Supplementary Table 6) and Phusion U Green Multiplex PCR Master Mix ( $100\ \mu\text{l}$  per well in a 96-well PCR plate, 5–6 plates total,  $66^{\circ}\text{C}$ , 4.5-min extension) following the manufacturer's protocol. Each PCR reaction was combined with 300  $\mu\text{l}$  PB DNA binding buffer (Qiagen) and 25 ml of the solution was loaded onto a HiBind DNA Midi column (Omega Bio-Tek). Bound DNA was washed with five column volumes of PE wash buffer (Qiagen) and the DNA fragment was eluted with  $800\ \mu\text{l}$  H<sub>2</sub>O per column. Both DNA fragments were quantified using a NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific).

*TadA*\* libraries were assembled following a previously reported USER assembly procedure<sup>40</sup> with the following conditions: 0.22 pmol *ecTadA* mutagenized DNA fragment 1, 0.22 pmol plasmid backbone fragment 2, 1 U of USER (Uracil-Specific Excision Reagent, New England Biolabs) enzyme, and 1 U of DpnI enzyme (New England Biolabs) per  $10\ \mu\text{l}$  USER assembly mixture were combined in 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100  $\mu\text{g}/\text{ml}$  BSA at pH 7.9 (1× CutSmart Buffer, New England Biolabs). Generally, each round of evolution required  $\sim 1\ \text{ml}$  USER assembly mixture (22 pmol of each DNA assembly fragment) which was distributed into  $10\text{-}\mu\text{l}$  aliquots across multiple 8-well PCR strips. The reactions were warmed to  $37^{\circ}\text{C}$  for 60 min, then heated to  $80^{\circ}\text{C}$  for 3 min to denature the two enzymes. The assembly mixture was slowly cooled to  $12^{\circ}\text{C}$  at  $0.1^{\circ}\text{C}/\text{s}$  in a thermocycler to promote annealing of the freshly generated ends of the two USER junctions.

With a library of constructs in hand, we removed denatured enzymes and reaction buffer from the assembly mixture by adding 5 vol PB buffer (Qiagen) to the assembly reaction mixture and binding the material onto a MinElute column ( $480\ \mu\text{l}$  per column). ABE hybridized library constructs were eluted in  $30\ \mu\text{l}$  H<sub>2</sub>O per column and  $2\ \mu\text{l}$  of this eluted material was added to  $20\ \mu\text{l}$  NEB 10-beta electrocompetent *E. coli* and electroporated with a Lonza 4D-Nucleofector System using bacterial program 5 in a 16-well Nucleocuvette strip. A typical round of evolution used  $\sim 300$  electroporations to generate 5–10 million colony forming units (c.f.u.). Freshly electroporated *E. coli* were recovered in 200 ml pre-warmed Davis Rich Medium (DRM) at  $37^{\circ}\text{C}$ , and incubated with shaking at 200 r.p.m. in

a 500-ml vented baffled flask for 15 min before the addition of 30  $\mu\text{g}/\text{ml}$  carbenicillin (for plasmid maintenance). The culture was incubated at  $37^{\circ}\text{C}$  with shaking at 200 r.p.m. for 18 h. The plasmid library was isolated with a ZymoPURE Plasmid Midiprep kit following the manufacturer's procedure (50 ml culture per DNA column), except that the plasmid library was eluted in  $200\ \mu\text{l}$  pre-warmed water per column. Evolution rounds 1–3, 5 and 7 followed this procedure in order to generate the corresponding libraries with minor variations (Supplementary Table 7).

**Generation of site-saturated bacterial *TadA*\* library (evolution round 4).** Mutagenesis at Arg24, Glu25, Arg107, Ala142, and Ala143 of *ecTadA* was achieved by using *ecTadA*\*(2.1)-dCas9 as a template and amplifying with appropriately designed degenerate NNK-containing primers (Supplementary Table 6). Briefly, the *ecTadA*\*(2.1)-dCas9 template was amplified separately with two sets of primers: NMG-1197 + NMG-1200, and NMG-1199 + NMG-1202, using Phusion U Green Multiplex PCR Master Mix, to form PCR product 1 and PCR product 2, respectively. Both PCR products were purified individually using PB binding buffer and a MiniElute column and eluted with  $20\ \mu\text{l}$  H<sub>2</sub>O per  $200\ \mu\text{l}$  PCR reaction. In a third PCR reaction,  $1\ \mu\text{l}$  PCR product 1 and  $1\ \mu\text{l}$  PCR product 2 were combined with the exterior, uracil-containing primers NMG-1202 and NMG-1197, and amplified by Phusion U Green Multiplex PCR Master Mix to form the desired extension-overlap PCR product with flanking uracil-containing USER junctions. In a fourth PCR reaction, *ecTadA*\*(2.1)-dCas9 was amplified with NMG-1201 and NMG-1198 to generate the backbone DNA fragment for USER assembly. After DpnI digestion and gel purification of both USER assembly fragments, the extension-overlap PCR product (containing the desired NNK mutations in *ecTadA*) was incorporated into the *ecTadA*\*(2.1)-dCas9 backbone by USER assembly as described above. The freshly generated NNK library was transformed into NEB 10-beta electrocompetent *E. coli* and the DNA was harvested as described above.

**Generation of DNA-shuffled bacterial *TadA*\* library (evolution round 6).** DNA shuffling was achieved by a modified version of the nucleotide exchange and excision technology (NExT) DNA shuffling method<sup>41</sup>. Solutions of 10 mM each of dATP, dCTP, dGTP, and dTTP/dUTP (7/3) were freshly prepared. Next, the *TadA*\* fragment was amplified from 20 fmol of a pool of *TadA*\*-XTEN-dCas9 bacterial constructs isolated from evolution rounds 1–5 in equimolar concentrations using *Taq* DNA Polymerase (NEB), primers NMG-822 and NMG-823 (Supplementary Table 6), and 400  $\mu\text{M}$  each of dATP, dCTP, dGTP, and dUTP/dTTP (3/7) in 1× ThermoPol Reaction Buffer ( $63^{\circ}\text{C}$ , 1.5-min extension time). The freshly generated uracil-containing DNA library fragment was purified by gel electrophoresis and extracted with QIAquick Gel Extraction Kit (Qiagen), eluting with  $20\ \mu\text{l}$  H<sub>2</sub>O per extraction column. The purified DNA product was digested with 2 U of USER enzyme per  $40\ \mu\text{l}$  in 1× CutSmart Buffer at  $37^{\circ}\text{C}$  and monitored by analytical agarose gel electrophoresis until digestion was complete. The reaction was quenched with 10 vol PN1 binding buffer (Qiagen) when the starting material was no longer observed (typically 3–4 h at  $37^{\circ}\text{C}$ ). Additional USER enzyme was added to the reaction if needed. The digested material was purified with QiaexII kit (Qiagen) using the manufacturer's protocol and the DNA fragments were eluted in  $50\ \mu\text{l}$  pre-warmed H<sub>2</sub>O per column.

The purified shuffled *TadA*\* fragments were reassembled into full-length *TadA*\* amplicons by an internal primer extension procedure as follows. The eluted digested DNA fragments ( $25\ \mu\text{l}$ ) were combined with 4 U Vent Polymerase (NEB), 800  $\mu\text{M}$  each of dATP, dCTP, dGTP, and dTTP, and 1 U *Taq* DNA polymerase in 1× ThermoPol Buffer supplemented with 0.5 mM MgSO<sub>4</sub>. The thermocycler program for the reassembly procedure was as follows:  $94^{\circ}\text{C}$  for 3 min, 40 cycles of denaturation at  $92^{\circ}\text{C}$  for 30 s, annealing over 60 s at increasing temperatures starting at  $30^{\circ}\text{C}$  and adding  $1^{\circ}\text{C}$  per cycle (cooling ramp  $1^{\circ}\text{C}/\text{s}$ ), and extension at  $72^{\circ}\text{C}$  for 60 s with an additional 4 s per cycle, ending with one final cycle of  $72^{\circ}\text{C}$  for 10 min. The full-length reassembled product was amplified by PCR with the following conditions:  $15\ \mu\text{l}$  unpurified internal assembly was combined with  $1\ \mu\text{M}$  each of USER primers NMG-825 and NMG-826,  $100\ \mu\text{l}$  Phusion U Green Multiplex PCR Master Mix and H<sub>2</sub>O to a final volume of  $200\ \mu\text{l}$ ,  $63^{\circ}\text{C}$  annealing, extension time 30 s. The PCR product was purified by gel electrophoresis and assembled using the USER method into the corresponding *ecTadA*\*-XTEN-dCas9 backbone with corresponding flanking USER junctions generated from amplification of the backbone with USER primers NMG-799 and NMG-824 as before. Following transformation of the USER assembly products into NEB 10-beta electrocompetent *E. coli*, the library of evolution round 6 constructs was isolated using a ZymoPURE Plasmid Midiprep kit following the manufacturer's instructions.

**Bacterial evolution of *TadA* variants.** The previously described strain S1030<sup>42</sup> was used in all evolution experiments and an electrocompetent version of the bacteria was prepared as previously described<sup>40</sup> harbouring the appropriate selection plasmid specific to each round of evolution (Supplementary Table 7). In brief,  $2\ \mu\text{l}$  freshly generated *TadA*\* library (300–600 ng/ $\mu\text{l}$ ) prepared as described above was added to  $22\ \mu\text{l}$  freshly prepared electrocompetent S1030 cells containing the



target selection plasmid and electroporated with a Lonza 4D-Nucleofector System using bacterial program 5 in a 16-well Nucleocuvette strip. A typical selection used  $5\text{--}10 \times 10^6$  c.f.u. After electroporation, freshly transformed S1030 cells were recovered in a total of 250 ml pre-warmed DRM at 37 °C with shaking at 200 r.p.m. for 15 min. Following this brief recovery incubation, carbenicillin was added to a final concentration of 30 µg/ml to maintain the library plasmid, along with the appropriate antibiotic to maintain the selection plasmid; see Supplementary Table 7 for the list of selection conditions, including the antibiotics used for each round. Immediately after the addition of the plasmid maintenance antibiotics, 100 mM of L-arabinose was added to the culture to induce translation of *TadA*\*-dCas9 fusion library members, which were expressed from the *P*<sub>BAD</sub> promoter. The culture was grown to saturation at 37 °C with shaking at 200 r.p.m. for 18 h, except that the incubation time for evolution round 5 was only 7 h.

Library members were challenged by plating 10 ml of the saturated culture onto each of four 245 mm<sup>2</sup> square bioassay dishes containing 1.8% agar-2×YT, 30 µg/ml plasmid maintenance antibiotics, and a concentration of the selection antibiotic pre-determined to be above the MIC of the S1030 strain harbouring the selection plasmid alone (Supplementary Table 8). Plates were incubated at 37 °C for 2 days and ~500 surviving colonies were isolated. The *TadA*\* genes from these colonies were amplified by PCR with primers NMG-822 and NMG-823 (Supplementary Table 6) and submitted for DNA sequencing. Concurrently, the colonies were inoculated separately into 1-ml DRM cultures in a 96-deep well plate and grown overnight at 37 °C, 200 r.p.m. Aliquots (100 µl) of each overnight culture were pooled, the plasmid DNA was isolated, and the *TadA*\* genes were amplified with USER primers NMG-825 and NMG-826 (Supplementary Table 6). The *TadA*\* genes were subcloned back into the plasmid backbone (containing the XTEN linker-dCas9, and appropriate guide RNAs) with the USER assembly protocol described above. This enriched library was transformed into the appropriate S1030 (+ selection plasmid) electrocompetent cells, incubated with maintenance antibiotic and L-Ara and re-challenged with the selection condition. After a 2-day incubation, 300–400 surviving clones were isolated as described above and their *TadA*\* genes were sequenced. Mutations arising from each selection round were imported into mammalian ABE constructs and tested in mammalian cells as described below.

**General mammalian cell culture conditions.** HEK293T (ATCC CRL-3216) and U2OS (ATCC HTB-96) were purchased from ATCC and cultured and passaged in Dulbecco's modified Eagle's medium (DMEM) plus GlutaMax (ThermoFisher Scientific) supplemented with 10% (v/v) fetal bovine serum (FBS). Hap1 (Horizon Discovery, C631) and Hap1 AAG<sup>-</sup> cells (Horizon Discovery, HZGHC001537c002) were maintained in Iscove's modified Dulbecco's medium (IMDM) plus GlutaMax (ThermoFisher Scientific) supplemented with 10% (v/v) FBS. Lymphoblastoid cell lines (LCL) containing a C282Y mutation in the *HFE* gene (Coriell Biorepository, GM14620) were maintained in Roswell Park Memorial Institute medium 1640 (RPMI-1640) plus GlutaMax (ThermoFisher Scientific) supplemented with 20% FBS. All cell types were incubated, maintained, and cultured at 37 °C with 5% CO<sub>2</sub>. Cell lines were authenticated by the suppliers and tested negative for mycoplasma.

**HEK293T tissue culture transfection protocol and genomic DNA preparation.** HEK293T cells grown in the absence of antibiotic were seeded on 48-well poly-D-lysine coated plates (Corning). 12–14 h after seeding, cells were transfected at approximately 70% confluency with 1.5 µl Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's protocols and 750 ng ABE plasmid, 250 ng sgRNA expression plasmid, and 10 ng GFP expression plasmid (Lonza). Unless otherwise stated, cells were cultured for 5 days, with a medium change on day 3. Medium was removed, cells were washed with 1 × PBS solution (Thermo Fisher Scientific), and genomic DNA was extracted by addition of 100 µl freshly prepared lysis buffer (10 mM Tris-HCl, pH 7.0, 0.05% SDS, 25 µg/ml Proteinase K (ThermoFisher Scientific)) directly into each well of the tissue culture plate. The genomic DNA mixture was transferred to a 96-well PCR plate and incubated at 37 °C for 1 h, followed by an 80 °C enzyme denaturation step for 30 min. Primers used for mammalian cell genomic DNA amplification are listed in Supplementary Table 9.

**Nucleofection of HAP1 and HAP1 AAG<sup>-</sup> cells and genomic DNA extraction.** HAP1 and HAP1 AAG<sup>-</sup> cells were nucleofected using the s.e. Cell Line 4D-Nucleofector X Kit S according to the manufacturer's protocol. In brief,  $4 \times 10^5$  cells were nucleofected with 300 ng of ABE plasmid and 100 ng of sgRNA expression plasmid using the 4D-Nucleofector program DZ-113 and cultured in 250 µl medium in a 48-well poly-D-lysine coated culture plate for 3 days. DNA was extracted as described above.

**Nucleofection of U2OS cells and genomic DNA extraction.** U2OS cells were nucleofected using the SG Cell Line 4D-Nucleofector X Kit (Lonza) according to the manufacturer's protocol. In brief,  $1.25 \times 10^5$  cells were nucleofected in 20 µl SG buffer along with 500 ng ABE plasmid and 100 ng sgRNA expression plasmid using the 4D-Nucleofector program EH-100 in a 16-well Nucleocuvette strip (20 µl cells per well). Freshly nucleofected cells were transferred into 250 µl medium in

a 48-well poly-D-lysine coated culture plate. Cells were incubated for 5 days and medium was changed every day. DNA was extracted as described above.

**Electroporation of LCL HFE C282Y cells.** LCL cells were electroporated using a Gene Pulser Xcell Electroporator (BioRad) and 0.4-cm gap Gene Pulser electroporation cuvettes (BioRad). In brief,  $1 \times 10^7$  LCL cells were resuspended in 250 µl RPMI-160 plus GlutaMax. To this medium was added 65 µg plasmid expressing ABE7.10, GFP, and the corresponding sgRNA targeting the C282Y mutation in the *HFE* gene. The mixture was added to a pre-chilled 0.4-cm gap electroporation cuvette and the cell-DNA mixture was incubated in the cuvette on ice for 10 min. Cells were pulsed at 250 V and 950 µF for 3 ms. Cells were transferred back onto ice for 10 min, then transferred to 15 ml pre-warmed RPMI-160 supplemented with 20% FBS in a T-75 flask. The next day, an additional 5 ml medium was added to the flask and cells were left to incubate for a total of 5 days. After incubation, cells were isolated by centrifugation, resuspended in 400 µl medium, filtered through a 40-µm strainer (Thermo Fisher Scientific), and sorted for GFP fluorescence using an FACSARIA III Flow Cytometer (Becton Dickinson Biosciences). GFP-positive cells were collected in a 1.5-ml tube containing 500 µl medium. After centrifugation, the medium was removed and cells were washed twice with 600 µl 1 × PBS (Thermo Fisher Scientific). Genomic DNA was extracted as described above.

**Comparison between ABE 7.10 and HDR using the CORRECT method.** HEK293T cells grown in the absence of antibiotic were seeded on 48-well poly-D-lysine coated plates (Corning). After 12–14 h, cells were transfected at ~70% confluency with 750 ng Cas9 or base editor plasmid, 250 ng sgRNA expression plasmid, 1.5 µl Lipofectamine 3000 (Thermo Fisher Scientific), and for HDR assays 0.7 µg single-stranded donor DNA template (100 nt, PAGE-purified from IDT) according to the manufacturer's instructions. Single-stranded 100-mer oligonucleotide donor templates are listed in Supplementary Table 10.

Genomic DNA was harvested 48 h after transfection (as described<sup>43</sup>) using the Agencourt DNAdvance Genomic DNA isolation Kit (Beckman Coulter) according to the manufacturer's instructions. A size-selective DNA isolation step ensured that there was no risk of contamination by the single-stranded donor DNA template in subsequent PCR amplification and sequencing steps. We re-designed amplification primers to ensure that there was minimal risk of amplifying the donor oligo template.

**HTS of genomic DNA samples.** Genomic sites of interest were amplified by PCR with primers containing homology to the region of interest and the appropriate Illumina forward and reverse adapters (Supplementary Table 9). Primer pairs used in this first round of PCR (PCR 1) for all genomic sites can be found in Supplementary Table 9. Specifically, 25 µl of a given PCR 1 reaction was assembled containing 0.5 µM of each forward and reverse primer, 1 µl genomic DNA extract and 12.5 µl Phusion U Green Multiplex PCR Master Mix. PCR reactions were carried out as follows: 95 °C for 2 min, then 30 cycles of (95 °C for 15 s, 62 °C for 20 s, and 72 °C for 20 s), followed by a final 72 °C extension for 2 min. PCR products were verified by comparison with DNA standards (Quick-Load 100 bp DNA ladder) on a 2% agarose gel supplemented with ethidium bromide. Unique Illumina barcoding primer pairs were added to each sample in a secondary PCR reaction (PCR 2). Specifically, 25 µl of a given PCR 2 reaction was assembled containing 0.5 µM of each unique forward and reverse Illumina barcoding primer pair, 2 µl unpurified PCR 1 reaction mixture, and 12.5 µl Q5 Hot Start High-Fidelity 2 × Master Mix. The barcoding PCR 2 reactions were carried out as follows: 95 °C for 2 min, then 15 cycles of (95 °C for 15 s, 61 °C for 20 s, and 72 °C for 20 s), followed by a final 72 °C extension for 2 min. PCR products were purified by electrophoresis with a 2% agarose gel using a QIAquick Gel Extraction Kit, eluting with 30 µl H<sub>2</sub>O. DNA concentration was quantified with the KAPA Library Quantification Kit-Illumina (KAPA Biosystems) and sequenced on an Illumina MiSeq instrument according to the manufacturer's protocols.

**General HTS data analysis.** Sequencing reads were demultiplexed in MiSeq Reporter (Illumina). Alignment of amplicon sequences to a reference sequence was performed as previously described using a Matlab script with improved output format (Supplementary Note 1). In brief, the Smith–Waterman algorithm was used to align sequences without indels to a reference sequence; bases with a quality score of less than 30 were converted to 'N' to prevent base miscalling as a result of sequencing error. Indels were quantified separately using a modified version of a previously described Matlab script in which sequencing reads with more than half the base calls below a quality score of Q30 were filtered out (Supplementary Note 2). Indels were counted as reads which contained insertions or deletions of greater than or equal to 1 bp within a 30-bp window surrounding the predicted Cas9 cleavage site.

Owing to homology in the *HBG1* and *HBG2* loci, primers were designed that would amplify both loci within a single PCR reaction. In order to computationally separate sequences of these two genomic sites, sequencing experiments involving this amplicon were processed using a separate Python script (Supplementary Note 3). In brief, reads were disregarded if more than half of the base calls were below Q30, and base calls with a quality score below Q30 were converted to N. *HBG1* or *HBG2*



reads were identified as having an exact match to a 37-bp sequence containing two SNPs that differ between the sites. A base calling and indel window were defined by exact matches to 10-bp flanking sequences on both sides of a 43-bp window centred on the protospacer sequence. Indels were counted as reads in which this base calling window differed in length by more than 1 bp. This Python script yields output with identical quality to the aforementioned Matlab script (estimated base calling error rate of <1 in 1,000), but in far less time owing to the absence of an alignment step.

To calculate the total number of edited reads as a proportion of the total number of successfully sequenced reads, the fraction of edited reads as measured by the alignment algorithm were multiplied by (1 – fraction of reads containing an indel).

**RNA isolation from HEK293T cells and analysis.** HEK293T cells were plated and a subset was transfected with ABE 7.10 as described above and incubated for five days before being removed from the plate using TrypLE Express (Thermo Fisher Scientific) and pelleted. RNA was extracted using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. cDNA was generated from the isolated RNA using the ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs) according to the manufacturer's instructions with a mixture of random primers and Oligo-dT primers. Amplification of the cDNA for high-throughput sequencing was performed to the top of the linear range (29 cycles for all four amplicons) using qPCR as described above. High-throughput sequencing of the amplicons was performed as described above. Sequences were aligned to the reference sequence for each RNA, obtained from the NCBI.

**Linkage disequilibrium analysis.** A custom Python script (Supplementary Note 4) was used to assess editing probabilities at the primary target adenine ( $P_1$ ), at the secondary target adenine ( $P_2$ ), and at both the primary and secondary target adenines ( $P_{1,2}$ ). Linkage disequilibrium was then evaluated as  $P_{1,2} - (P_1 \times P_2)$ . Linkage disequilibrium values were normalized with a normalization factor of  $\text{Min}(P_1(1 - P_2), (1 - P_1)P_2)$ . This normalization controls for allele frequency and yields a normalized linkage disequilibrium value from 0 to 1.

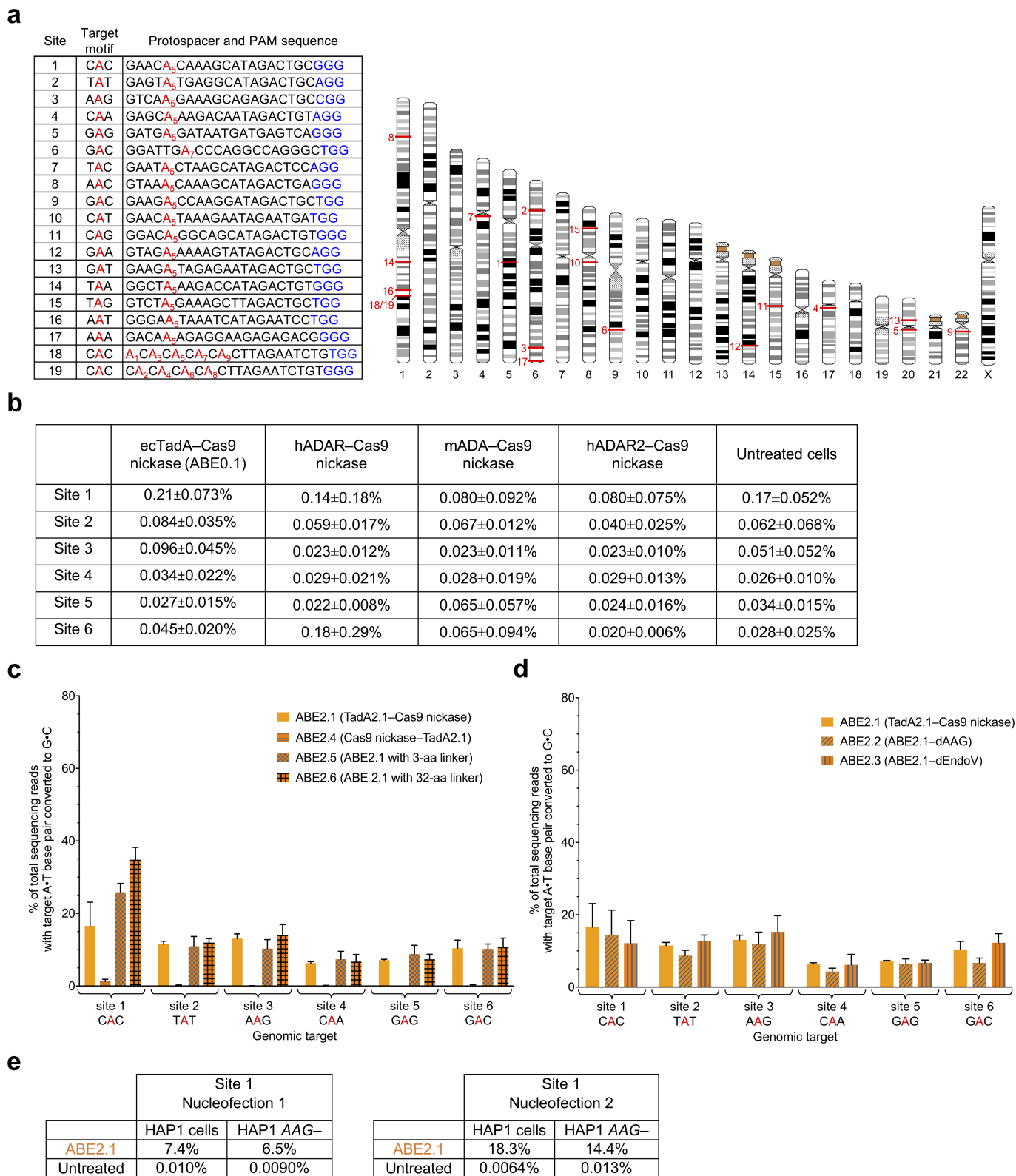
**Code availability.** Scripts used in this work can be found in Supplementary Information, Supplementary Notes 1–4.

**Data availability.** High-throughput sequencing data have been deposited in the NCBI Sequence Read Archive database under accession code SRP119577. Plasmids encoding ABE6.3, ABE7.8, ABE7.9, and ABE7.10 are available from Addgene.

40. Badran, A. H. *et al.* Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* **533**, 58–63 (2016).
41. Müller, K. M. *et al.* Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution. *Nucleic Acids Res.* **33**, e117 (2005).
42. Esvelt, K. M., Carlson, J. C. & Liu, D. R. A system for the continuous directed evolution of biomolecules. *Nature* **472**, 499–503 (2011).
43. Kwart, D., Paquet, D., Teo, S. & Tessier-Lavigne, M. Precise and efficient scarless genome editing in stem cells using CORRECT. *Nat. Protocols* **12**, 329–354 (2017).

	<i>E. coli</i> TadA amino acid																TadA state	Linker 1 length	Linker 2 length	Other modifications relative to TadA*-Cas9 nickase						
	23	26	36	37	48	49	51	72	84	97	106	108	123	125	142	146					147	152	155	156	157	161
ABE0.1	W	R	H	N	P	I	R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K	monomer		16	wtTadA-Cas9 nickase
ABE0.2	W	R	H	N	P	I	R	N	L	S	A	D	H	G	A	S	D	R	E	I	K	K	homodimer	32	32	wtTadA-wtTadA-Cas9 nickase
ABE1.1	W	R	H	N	P	I	R	N	L	S	A	N	H	G	A	S	D	R	E	I	K	K	monomer		16	
ABE1.2	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	D	R	E	I	K	K	monomer		16	
ABE2.1	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	monomer		16	
ABE2.2	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	monomer		16	C-terminal dAAG(E125Q)
ABE2.3	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	monomer		16	C-terminal dEndoV(D35A)
ABE2.4	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	monomer		16	TadA2.1 at C-terminus of Cas9 nickase
ABE2.5	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	monomer		3	
ABE2.6	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	monomer		32	
ABE2.7	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	heterodimer		16	overexpression of wt TadA
ABE2.8	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	homodimer		16	overexpression of TadA2.1
ABE2.9	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	homodimer	32	16	TadA2.1-TadA2.1-Cas9 nickase
ABE2.10	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	heterodimer	32	16	wtTadA-TadA2.1-Cas9 nickase
ABE2.11	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	heterodimer	32	16	dTadA2.1(E59A)-TadA2.1-Cas9 nickase
ABE2.12	W	R	H	N	P	I	R	N	L	S	V	N	H	G	A	S	Y	R	V	I	K	K	heterodimer	32	16	TadA2.1-dTadA2.1(E59A)-Cas9 nickase
ABE3.1	W	R	H	N	P	I	R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K	homodimer	32	32	
ABE3.2	W	R	H	N	P	I	R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K	heterodimer	32	32	
ABE3.3	W	R	H	N	P	I	R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K	heterodimer	100	32	
ABE3.4	W	R	H	N	P	I	R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K	homodimer	100	32	
ABE3.5	W	R	H	N	P	I	R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K	heterodimer	100	64	
ABE3.6	W	R	H	N	P	I	R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K	homodimer	100	64	
ABE3.7	W	R	H	N	P	I	R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K	heterodimer	32	64	
ABE3.8	W	R	H	N	P	I	R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	K	homodimer	32	64	
ABE4.1	W	R	H	N	P	I	R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K	homodimer	32	32	
ABE4.2	W	G	H	N	P	I	R	N	L	S	V	N	H	G	N	S	Y	R	V	I	K	K	homodimer	32	32	
ABE4.3	W	R	H	N	P	I	R	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K	homodimer	32	32	
ABE5.1	W	R	L	N	P	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	homodimer	32	32	
ABE5.2	W	R	H	S	P	I	R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T	homodimer	32	32	
ABE5.3	W	R	L	N	P	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	32	32	
ABE5.4	W	R	H	S	P	I	R	N	F	S	V	N	Y	G	A	S	Y	R	V	F	K	T	heterodimer	32	32	
ABE5.5	W	R	L	N	P	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	24	24	
ABE5.6	W	R	L	N	P	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	24	32	
ABE5.7	W	R	L	N	P	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	24	40	
ABE5.8	W	R	L	N	P	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	32	24	
ABE5.9	W	R	L	N	P	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	32	40	
ABE5.10	W	R	L	N	P	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	40	24	
ABE5.11	W	R	L	N	P	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	40	32	
ABE5.12	W	R	L	N	P	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	40	40	
ABE5.13	W	R	H	N	P	I	L	D	F	S	V	N	Y	A	A	S	Y	R	V	F	K	K	homodimer	32	32	
ABE5.14	W	R	H	N	S	I	L	N	F	C	V	N	Y	G	A	S	Y	R	V	F	K	K	homodimer	32	32	
ABE6.1	W	R	H	N	S	I	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	K	K	heterodimer	32	32	
ABE6.2	W	R	H	N	T	V	L	N	F	S	V	N	Y	G	N	S	Y	R	V	F	N	K	heterodimer	32	32	
ABE6.3	W	R	L	N	S	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	32	32	
ABE6.4	W	R	L	N	S	I	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K	heterodimer	32	32	
ABE6.5	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	32	32	
ABE6.6	W	R	L	N	T	V	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K	heterodimer	32	32	
ABE7.1	W	R	L	N	A	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	32	32	
ABE7.2	W	R	L	N	A	I	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K	heterodimer	32	32	
ABE7.3	L	R	L	N	A	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	32	32	
ABE7.4	R	R	L	N	A	I	L	N	F	S	V	N	Y	G	A	C	Y	R	V	F	N	K	heterodimer	32	32	
ABE7.5	W	R	L	N	A	I	L	N	F	S	V	N	Y	G	A	C	Y	H	V	F	N	K	heterodimer	32	32	
ABE7.6	W	R	L	N	A	I	L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K	heterodimer	32	32	
ABE7.7	L	R	L	N	A	I	L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K	heterodimer	32	32	
ABE7.8	L	R	L	N	A	I	L	N	F	S	V	N	Y	G	N	C	Y	R	V	F	N	K	heterodimer	32	32	
ABE7.9	L	R	L	N	A	I	L	N	F	S	V	N	Y	G	N	C	Y	P	V	F	N	K	heterodimer	32	32	
ABE7.10	R	R	L	N	A	I	L	N	F	S	V	N	Y	G	A	C	Y	P	V	F	N	K	heterodimer	32	32	

Extended Data Figure 1 | Genotypes of 57 ABEs described in this work. Mutations are coloured according to the round of evolution in which they were identified.

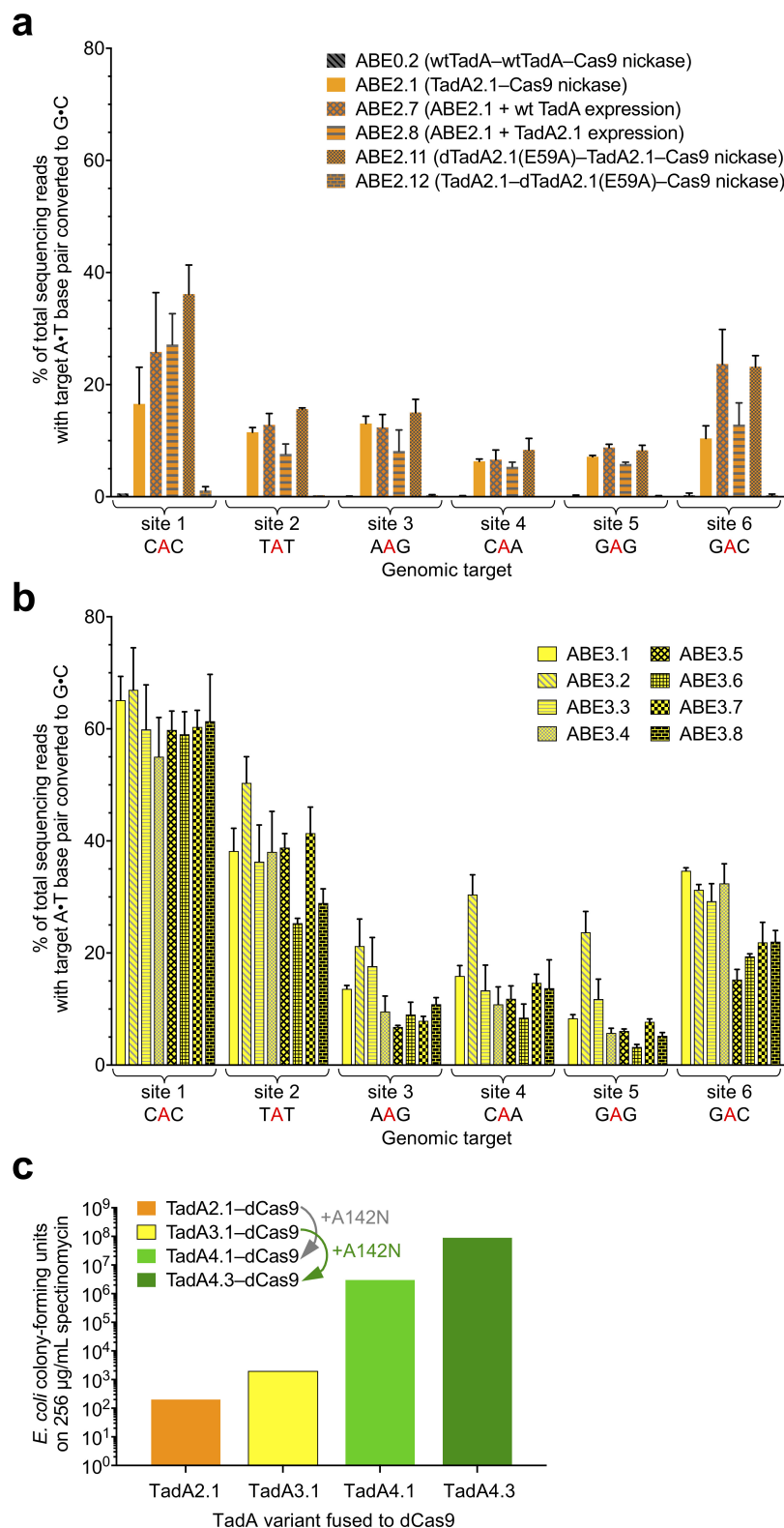


**Extended Data Figure 2 | Base editing efficiencies of additional early-stage ABE variants.** **a**, Table of 19 human genomic DNA test sites (left) with corresponding locations on human chromosomes (right). The sequence context (target motif) of the edited A in red is shown for each site. PAM sequences are shown in blue. **b**, A•T to G•C base editing efficiencies in HEK293T cells of various wild-type RNA adenine deaminases fused to Cas9 nickase at six human genomic target DNA sites. Values reflect the mean and s.d. of three biological replicates performed on different days. **c**, A•T to G•C base editing efficiencies in HEK293T cells

of ABE2 editors with altered fusion orientations and linker lengths at six human genomic target DNA sites. **d**, A•T to G•C base editing efficiencies in HEK293T cells at six human genomic target DNA sites of ABE2 editors fused to catalytically inactivated alkyl-adenosine glycosylase (AAG) or endonuclease V (EndoV), two proteins that bind inosine in DNA. **e**, A•T to G•C base editing efficiencies of ABE2.1 in HAP1 cells at site 1 with or without AAG. Values and error bars in **b** and **c** are the mean and s.d. of three independent biological replicates performed on different days.

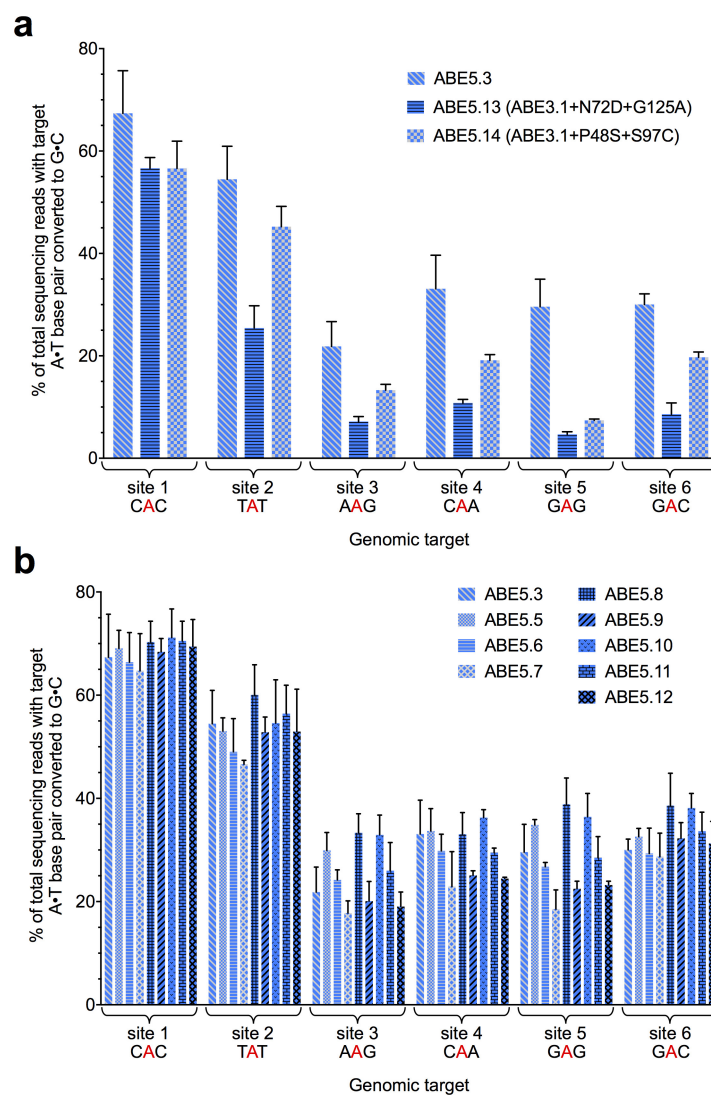


Untreated																							Indel%		
Site 1	G <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	C <sub>4</sub>	A <sub>5</sub>	C <sub>6</sub>	A <sub>7</sub>	A <sub>8</sub>	A <sub>9</sub>	G <sub>10</sub>	C <sub>11</sub>	A <sub>12</sub>	T <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	G	G	G	0.059	
A	0.0	100.0	100.0	0.0	99.8	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0		
C	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0		
G	100.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	99.8	100.0	100.0		
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0		
ABE2.1																							Indel%		
Site 1	G <sub>1</sub>	A <sub>2</sub>	A <sub>3</sub>	C <sub>4</sub>	A <sub>5</sub>	C <sub>6</sub>	A <sub>7</sub>	A <sub>8</sub>	A <sub>9</sub>	G <sub>10</sub>	C <sub>11</sub>	A <sub>12</sub>	T <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	G	G	G	0.048	
A	0.0	99.9	99.7	0.0	78.0	0.0	99.9	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
C	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0		
G	100.0	0.0	0.3	0.0	22.0	0.1	0.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	99.9	100.0	
T	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1	0.0	0.0	0.0		
Untreated																							Indel%		
Site 2	G <sub>1</sub>	A <sub>2</sub>	G <sub>3</sub>	T <sub>4</sub>	A <sub>5</sub>	T <sub>6</sub>	G <sub>7</sub>	A <sub>8</sub>	G <sub>9</sub>	G <sub>10</sub>	C <sub>11</sub>	A <sub>12</sub>	T <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	A	G	G	0.058	
A	0.0	100.0	0.0	0.0	99.9	0.0	0.1	99.9	0.0	0.1	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1		
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0		
G	99.9	0.0	100.0	0.0	0.0	0.0	99.8	99.9	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	99.9	0.0	0.0	100.0	99.9		
T	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0		
ABE2.1																							Indel%		
Site 2	G <sub>1</sub>	A <sub>2</sub>	G <sub>3</sub>	T <sub>4</sub>	A <sub>5</sub>	T <sub>6</sub>	G <sub>7</sub>	A <sub>8</sub>	G <sub>9</sub>	G <sub>10</sub>	C <sub>11</sub>	A <sub>12</sub>	T <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	A	G	G	0.047	
A	0.0	100.0	0.0	0.0	87.7	0.0	0.1	99.9	0.0	0.1	0.0	100.0	0.0	99.9	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.1		
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	0.0	0.0	0.0		
G	99.9	0.0	99.9	0.0	12.3	0.0	99.9	0.1	100.0	99.9	0.0	0.0	0.0	0.1	99.9	0.0	0.0	0.0	100.0	0.0	0.0	100.0	99.9		
T	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0		
Untreated																							Indel%		
Site 3	G <sub>1</sub>	T <sub>2</sub>	C <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	G <sub>6</sub>	A <sub>7</sub>	A <sub>8</sub>	A <sub>9</sub>	G <sub>10</sub>	C <sub>11</sub>	A <sub>12</sub>	G <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	C	G	G	0.075	
A	0.0	0.0	0.0	99.9	100.0	0.0	100.0	100.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1		
C	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9	100.0	0.0	0.0		
G	100.0	0.0	0.0	0.1	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	99.9		
T	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0		
ABE2.1																							Indel%		
Site 3	G <sub>1</sub>	T <sub>2</sub>	C <sub>3</sub>	A <sub>4</sub>	A <sub>5</sub>	G <sub>6</sub>	A <sub>7</sub>	A <sub>8</sub>	A <sub>9</sub>	G <sub>10</sub>	C <sub>11</sub>	A <sub>12</sub>	G <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	C	G	G	0.085	
A	0.0	0.0	0.0	99.9	85.9	0.0	99.4	99.9	100.0	0.0	0.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0		
C	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0	0.0	0.0		
G	100.0	0.0	0.0	0.1	14.1	100.0	0.6	0.1	0.0	100.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	100.0		
T	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0		
Untreated																							Indel%		
Site 4	G <sub>1</sub>	A <sub>2</sub>	G <sub>3</sub>	C <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	A <sub>7</sub>	G <sub>8</sub>	A <sub>9</sub>	G <sub>10</sub>	A <sub>11</sub>	A <sub>12</sub>	T <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	T <sub>20</sub>	A	G	G	0.042	
A	0.0	100.0	0.0	0.0	100.0	100.0	100.0	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.1	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
C	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0		
G	100.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	0.0	100.0	100.0		
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	99.9	0.0	0.0	0.0		
ABE2.1																							Indel%		
Site 4	G <sub>1</sub>	A <sub>2</sub>	G <sub>3</sub>	C <sub>4</sub>	A <sub>5</sub>	A <sub>6</sub>	A <sub>7</sub>	G <sub>8</sub>	A <sub>9</sub>	G <sub>10</sub>	A <sub>11</sub>	A <sub>12</sub>	T <sub>13</sub>	A <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	C <sub>17</sub>	T <sub>18</sub>	G <sub>19</sub>	T <sub>20</sub>	A	G	G	0.024	
A	0.0	100.0	0.0	0.0	93.9	96.7	99.5	0.0	100.0	0.0	100.0	100.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	
C	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0		
G	100.0	0.0	100.0	0.0	6.1	3.3	0.5	100.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0		
T	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.0	0.0		
Untreated																							Indel%		
Site 5	G <sub>1</sub>	A <sub>2</sub>	T <sub>3</sub>	G <sub>4</sub>	A <sub>5</sub>	G <sub>6</sub>	A <sub>7</sub>	T <sub>8</sub>	A <sub>9</sub>	A <sub>10</sub>	T <sub>11</sub>	G <sub>12</sub>	A <sub>13</sub>	T <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	G <sub>17</sub>	T <sub>18</sub>	C <sub>19</sub>	A <sub>20</sub>	G	G	G	0.036	
A	0.0	99.9	0.0	0.1	99.9	0.0	99.9	0.0	100.0	100.0	0.0	0.1	100.0	0.0	0.1	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0		
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0		
G	100.0	0.0	0.0	99.9	0.1	100.0	0.0	0.0	0.0	0.0	0.0	99.9	0.0	0.0	99.9	0.0	100.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	
T	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0		
ABE2.1																							Indel%		
Site 5	G <sub>1</sub>	A <sub>2</sub>	T <sub>3</sub>	G <sub>4</sub>	A <sub>5</sub>	G <sub>6</sub>	A <sub>7</sub>	T <sub>8</sub>	A <sub>9</sub>	A <sub>10</sub>	T <sub>11</sub>	G <sub>12</sub>	A <sub>13</sub>	T <sub>14</sub>	G <sub>15</sub>	A <sub>16</sub>	G <sub>17</sub>	T <sub>18</sub>	C <sub>19</sub>	A <sub>20</sub>	G	G	G	0.060	
A	0.0	100.0	0.0	0.1	93.0	0.0	94.8	0.0	99.9	100.0	0.0	0.1	100.0	0.0	0.1	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0		
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0		
G	99.9	0.0	0.0	99.9	7.0	100.0	5.2	0.0	0.1	0.0	0.0	99.8	0.0	0.0	99.8	0.0	100.0	0.0	0.0	0.0	100.0	100.0	100.0		
T	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	100.0	0.1	0.0	100.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0		
Untreated																							Indel%		
Site 6	G <sub>1</sub>	G <sub>2</sub>	A <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	G <sub>6</sub>	A <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>10</sub>	A <sub>11</sub>	G <sub>12</sub>	G <sub>13</sub>	C <sub>14</sub>	C <sub>15</sub>	A <sub>16</sub>	G <sub>17</sub>	G <sub>18</sub>	G <sub>19</sub>	C <sub>20</sub>	T	G	G	0.014	
A	0.0	0.1	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0		
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	99.9	100.0	100.0	0.0	0.0	0.0	1											



**Extended Data Figure 4 | Base editing efficiencies of additional ABE2 and ABE3 variants, and the effect of adding A142N to TadA\*–dCas9 on antibiotic selection survival in *E. coli*.** **a**, A•T to G•C base editing efficiencies in HEK293T cells at six human genomic target DNA sites of ABE2 variants with different engineered dimeric states. A control ABE variant containing two wild-type TadA domains and no evolved TadA\* domains (ABE0.2) did not result in A•T to G•C editing at the six genomic sites tested, confirming that dimerization alone is insufficient to mediate ABE activity. **b**, A•T to G•C base editing efficiencies in HEK293T cells at six human genomic target DNA sites of ABE3.1 variants differing in their dimeric state (homodimer of TadA\*–TadA\*–Cas9 nickase, or heterodimer

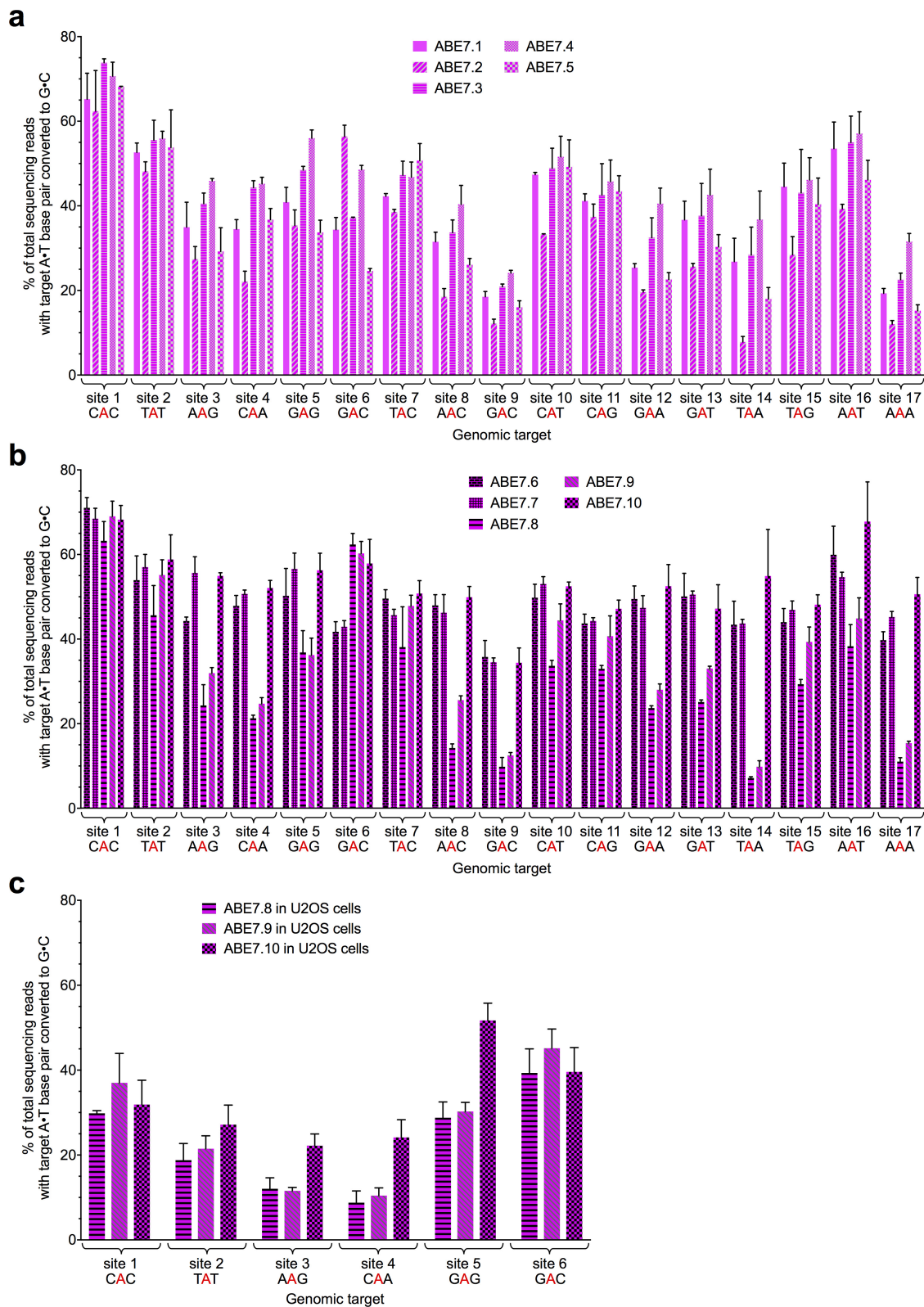
of wild-type TadA–TadA\*–Cas9 nickase), in the length of the TadA–TadA linker, and in the length of the TadA–Cas9 nickase linker. See Extended Data Fig. 1 for ABE genotypes and architectures. **c**, Colony-forming units on  $2 \times \text{YT}$  agar with  $256 \mu\text{g ml}^{-1}$  of spectinomycin of *E. coli* cells expressing an sgRNA targeting the I89T defect in the spectinomycin resistance gene and a TadA\*–dCas9 editor lacking or containing the A142N mutation identified in evolution round 4. Successful A•T to G•C base editing at the target site restores spectinomycin resistance. Values and error bars in **a** and **b** show the mean and s.d. of three independent biological replicates performed on different days.



**Extended Data Figure 5 | Base editing efficiencies of additional ABE5 variants.** **a**, A•T to G•C base editing efficiencies in HEK293T cells at six human genomic target DNA sites of two ABE3.1 variants with two pairs of mutations isolated from spectinomycin selection of the round 5 library. **b**, A•T to G•C base editing efficiencies in HEK293T cells at six human

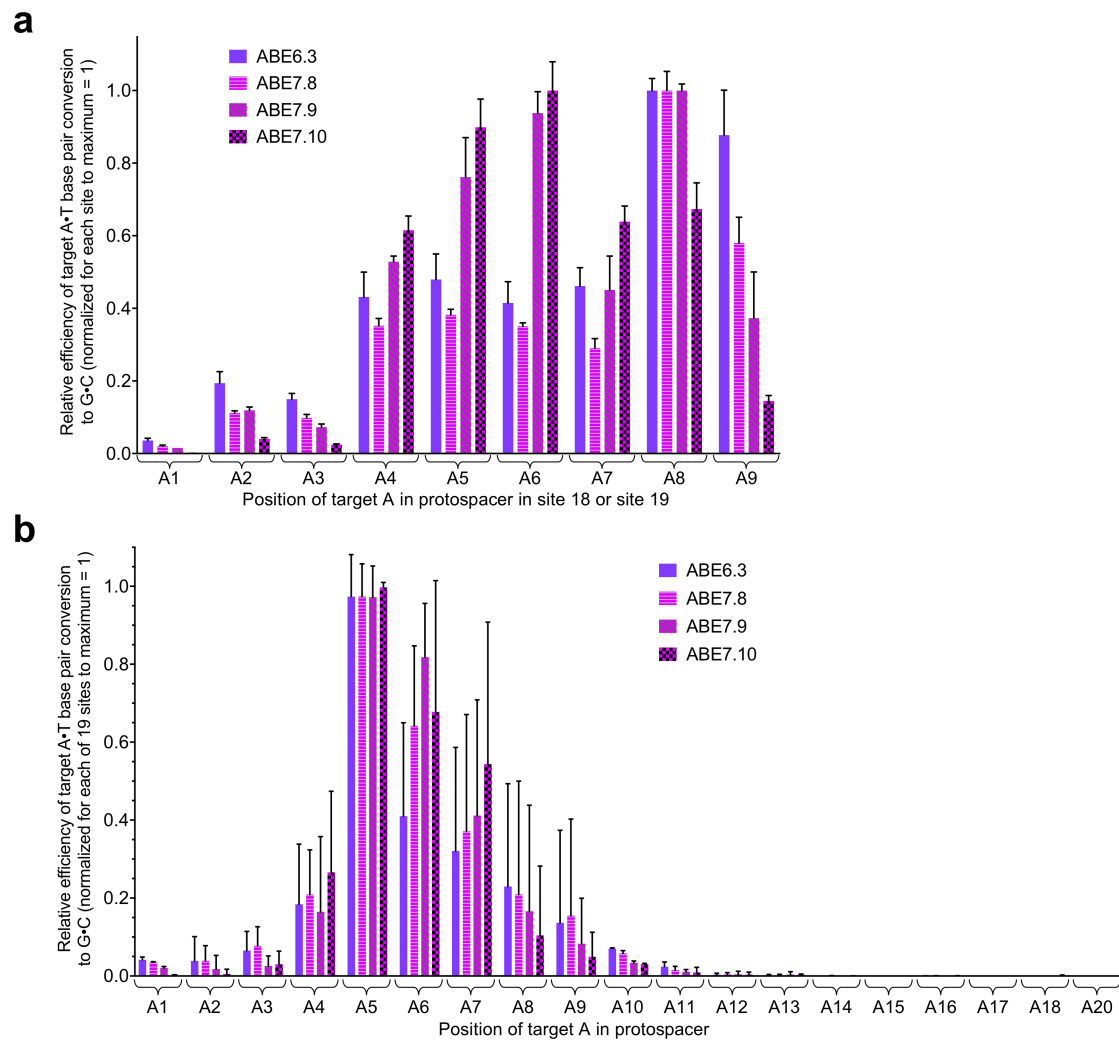
genomic target DNA sites of ABE5 variants with different linker lengths. See Extended Data Fig. 1 for ABE genotypes and architectures. Values and error bars show the mean and s.d. of three independent biological replicates performed on different days.





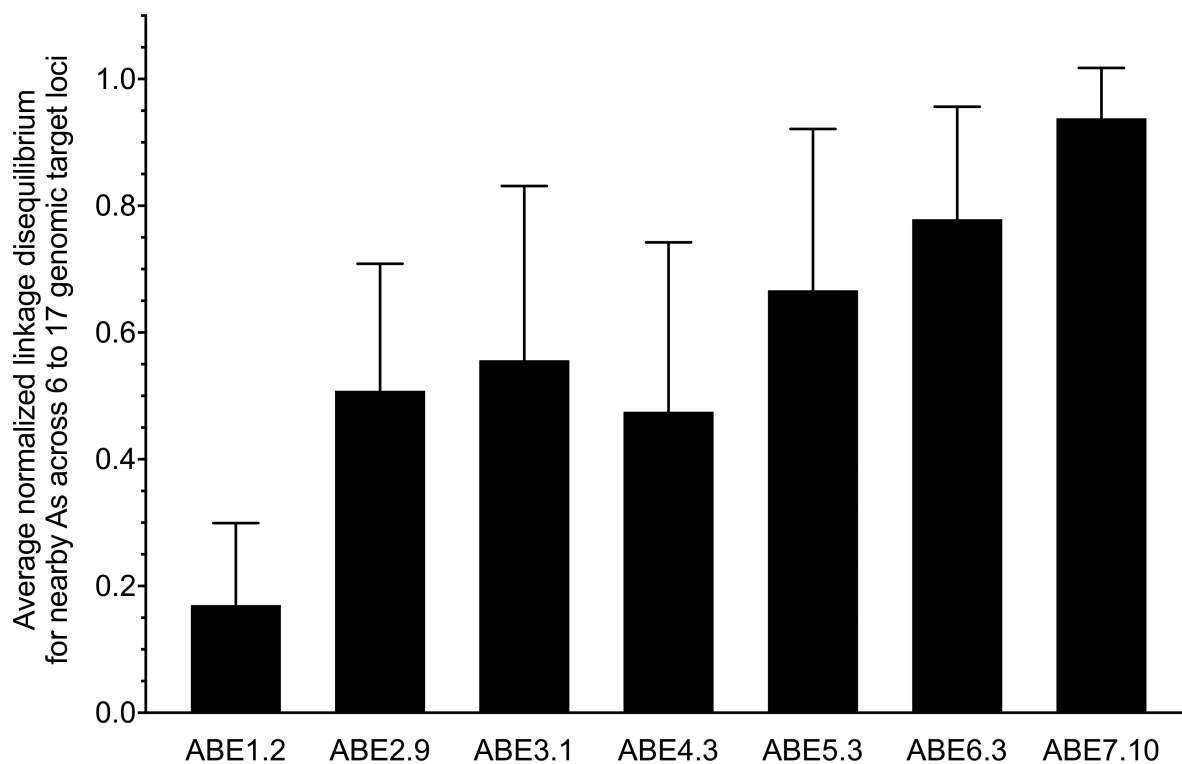
**Extended Data Figure 6 | Base editing efficiencies of ABE7 variants at 17 genomic sites. a, b,** A•T to G•C base editing efficiencies in HEK293T cells at 17 human genomic target DNA sites of ABE7.1–ABE7.5 (a), and ABE7.6–7.10 (b). See Extended Data Fig. 1 for ABE genotypes and architectures. **c,** A•T to G•C base editing efficiencies in U2OS cells at six human genomic target DNA sites of ABE7.8–ABE7.10. The lower editing

efficiencies observed in U2OS cells compared with HEK293T cells are consistent with differences in transfection efficiency between the two cell lines; we observed transfection efficiencies of about 40–55% in U2OS cells under the conditions used in this study, compared to about 65–80% in HEK293T cells. Values and error bars show the mean and s.d. of three independent biological replicates performed on different days.



**Extended Data Figure 7 | Activity window of late-stage ABEs.** **a**, Relative A•T to G•C base editing efficiencies in HEK293T cells of late-stage ABEs at protospacer positions 1–9 in two human genomic DNA sites that together place an adenine at each of these positions. Values are normalized so that the maximum observed efficiency at each of the two sites for each ABE is 1. **b**, Relative A•T to G•C base editing efficiencies in HEK293T

cells of late-stage ABEs at protospacer positions 1–18 and 20 across all 19 human genomic DNA sites tested. Values are normalized so that the maximum observed efficiency at each of the 19 sites for each ABE is 1. Values and error bars show the mean and s.d. across 19 sites, each with three independent biological replicates performed on different days.



**Extended Data Figure 8 | Rounds of evolution and engineering increased ABE processivity**. The calculated mean normalized linkage disequilibrium between nearby target adenines at 6–17 human genomic target DNA sites for the most active ABEs emerging from each round of evolution and engineering. Higher linkage disequilibrium values indicate that an ABE is more likely to edit an adenine if a nearby adenine

in the same DNA strand (the same sequencing read) is also edited. Linkage disequilibrium values are normalized from 0 to 1 in order to be independent of editing efficiency. Values and error bars show the mean and s.d. of normalized linkage disequilibrium values across 6–17 sites, each with three independent biological replicates performed on different days.





Extended Data Table 1 | Analysis of cellular RNAs in ABE7.10-treated cells compared with untreated cells

RNA	# of As > 0.2% mutated in any sample	ABE-treated 1	ABE-treated 2	Untreated 1	Untreated 2
beta-actin	11	0.41±0.38%	0.65±0.67%	0.17±0.30%	0.37±0.93%
beta-catenin	11	0.38±0.23%	0.59±0.38%	0.11±0.34%	0.23±0.73%
GAPDH	21	0.51±0.95%	0.94±1.5%	0.64±0.97%	0.74±1.4%
RB1	20	0.41±0.29%	0.66±0.47%	0.22±0.39%	0.30±0.61%

beta-actin amplicon (position 1-30): GACAAAACCTAACTTGCGCAGAAAACAAGA...

beta-catenin amplicon (position 1-30): TTTGATGGAGTTGGACATGGCCATGGAACC...

GAPDH amplicon (position 1-30): GGCTACAGCAACAGGGTGGTGGACCTCATG...

RB1 amplicon (position 1-30): GGAAGGATTATGATAGGGACAAGGATGATA...

RNA from HEK293T cells treated with ABE7.10 and the sgRNA targeting site 1, or from untreated HEK293T cells, was isolated and reverse transcribed into cDNA. The cDNAs corresponding to four abundant cellular RNAs (encoding  $\beta$ -actin,  $\beta$ -catenin, GAPDH and RB1) were amplified and analysed by HTS. Within each amplicon, the A to G mutation frequency at each adenine position for which the mutation rate was  $\geq 0.2\%$  in any sample is shown for two ABE7.10-treated biological replicates (ABE-treated 1 and ABE-treated 2) and for two untreated biological replicates (untreated 1 and untreated 2). The start of each amplicon is shown.

## Life Sciences Reporting Summary

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

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

### ► Experimental design

#### 1. Sample size

Describe how sample size was determined.

Experiments were done in biological triplicate, n=3 on different days. In previous studies we determined this sample size to be sufficient to ensure reproducibility.

#### 2. Data exclusions

Describe any data exclusions.

No data was excluded.

#### 3. Replication

Describe whether the experimental findings were reliably reproduced.

All attempts at replication were successful, and standard deviations were within expected ranges.

#### 4. Randomization

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

Different cell passages were used for each biological replicate.

#### 5. Blinding

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

Not applicable

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

#### 6. Statistical parameters

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

n/a Confirmed

- The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g.  $P$  values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.



## ► Software

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

### 7. Software

Describe the software used to analyze the data in this study.

Matlab, Python - scripts provided in SI

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

## ► Materials and reagents

Policy information about [availability of materials](#)

### 8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

No restrictions

### 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Not applicable.

### 10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

U2OS/HEK293T (ATCC) LCL (Coriell Biorepository) Hap1 cells (Horizon Discovery)

b. Describe the method of cell line authentication used.

Cells were authenticated by the supplier.

c. Report whether the cell lines were tested for mycoplasma contamination.

HEK293T cells tested negative for mycoplasma.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

No commonly misidentified cell lines were used.

## ► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

### 11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Not applicable

Policy information about [studies involving human research participants](#)

### 12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Not applicable