UC San Diego
UC San Diego Previously Published Works

Title

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

Permalink

Journal

ISSN

Authors

Gaudelli, Nicole M
Komor, Alexis C
Rees, Holly A
et al.

Publication Date

DOI

Peer reviewed

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

Nicole M. Gaudelli[1,2,3], **Alexis C. Komor**[1,2,3], **Holly A. Rees**[1,2,3], **Michael S. Packer**[1,2,3], **Ahmed H. Badran**[1,2,3], **David I. Bryson**[1,2,3], and **David R. Liu**[1,2,3,*]

[1]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, 02138

[2]Howard Hughes Medical Institute, Harvard University, Cambridge, MA, 02138

[3]Broad Institute of MIT and Harvard, Cambridge, MA, 02141

Summary

The spontaneous deamination of cytosine is a major source of C•G to T•A transitions, which account for half of known human pathogenic point mutations. The ability to efficiently convert target A•T base pairs to G•C therefore could advance the study and treatment of genetic diseases. While the deamination of adenine yields inosine, which is treated as guanine by polymerases, no enzymes are known to deaminate adenine in DNA. Here we report adenine base editors (ABEs) that mediate conversion of A•T to G•C in genomic DNA. We evolved a tRNA adenosine deaminase to operate on DNA when fused to a catalytically impaired CRISPR-Cas9. Extensive directed evolution and protein engineering resulted in seventh-generation ABEs (*e.g.*, ABE7.10), that convert target A•T to G•C base pairs efficiently (~50% in human cells) with very high product purity (typically ≥99.9%) and very low rates of indels (typically ≤0.1%). ABEs introduce point mutations more efficiently and cleanly than a current Cas9 nuclease-based method, induce less off-target genome modification than Cas9, and can install disease-correcting or disease-suppressing mutations in human cells. Together with our previous base editors, ABEs advance genome editing by enabling the direct, programmable introduction of all four transition mutations without double-stranded DNA cleavage.

---

*Correspondence should be addressed to David R. Liu: drliu@fas.harvard.edu.

Online Content

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

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

The formation of uracil and thymine from the spontaneous hydrolytic deamination of cytosine and 5-methylcytosine, respectively[1,2] occurs an estimated 100–500 times per cell per day in humans[1] and can result in C•G to T•A mutations, accounting for approximately half of all known pathogenic 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 therefore could enable the correction of a substantial fraction of human SNPs associated with disease.

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[3–5]. Compared with standard genome editing methods to introduce point mutations, base editing can proceed more efficiently[3], and with far fewer undesired products such as stochastic insertions or deletions (indels) or translocations[3–8].

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 a ~5-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[5], and (*iv*) nickase activity to nick the non-edited DNA strand, directing cellular DNA repair processes to replace the G-containing DNA strand[3,5]. Together, these components enable efficient and permanent C•G to T•A base pair conversion in bacteria, yeast[4,9], plants[10,11], zebrafish[8,12], mammalian cells[3–8,13,14], mice[8,15,16], and even human embryos[17,18]. Base editing capabilities have expanded through the development of base editors with different protospacer-adjacent motif (PAM) compatibilities[7], narrowed editing windows[7], enhanced DNA specificity[8], and small-molecule dependence[19]. Fourth-generation base editors (BE4 and BE4-Gam) further improve editing efficiency and product purity[5].

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 programmable installation of all four transitions (C to T, A to G, T to C, and G to A) in genomic DNA.

Results

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[20]. 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 occurs on free adenine,

free adenosine, adenosine in RNA, or adenosine in mispaired RNA:DNA heteroduplexes[21], we began by replacing the APOBEC1 component of BE3 with natural adenine deaminases including *E. coli* TadA[22,23], human ADAR2[24], mouse ADA[25], and human ADAT2[26] (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 of untreated cells (Extended Data Fig. E1 and E2b). These results suggest 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 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[3] (APOBEC1 cytidine deaminase fused to dCas9 and UGI), since bacteria lack nick-directed mismatch repair machinery[27] that enables more efficient base editing by BE3. We observed successful rescue of a defective chloramphenicol acetyl transferase (Cam[R]) 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[R] 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.

Our previously described base editors[3,5,7,8] 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[22] that converts adenine to inosine (I) in the single-stranded anticodon loop of tRNA[Arg]. *E. coli* TadA shares homology with the APOBEC enzyme[28] used in our original base editors, and some ABOBECs bind single-stranded DNA in a conformation that resembles tRNA bound to TadA[28]. TadA does not require small-molecule activators (in contrast with ADAR[29]) and acts on polynucleic acid (unlike ADA[25]). Based on these considerations, we chose *E. coli* TadA 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 favorable properties of the Cas9 portion of the editor (Supplementary Table 7). The resulting plasmids were transformed into *E. coli* harboring the Cam[R] H193Y selection (Fig. 2a and Supplementary Table 8). Colonies surviving chloramphenicol challenge were strongly enriched for TadA mutations A106V and D108N (Fig. 2b). Sequence alignment of *E. coli* TadA with *S. aureus* TadA, for which a structure complexed with tRNA[Arg] has been reported[30], predicts that the side-chain of D108

hydrogen bonds with the 2'-OH group of the ribose in the U upstream of the substrate A (Fig. 2c). Mutations at D108 likely abrogate this hydrogen bond, decreasing the energetic opportunity cost of binding DNA. DNA sequencing confirmed that all clones surviving the selection showed A•T to G•C reversion at the targeted site in Cam^R. 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 favor desired base editing outcomes[3], and adds a C-terminal nuclear localization signal (NLS). We designated the resulting TadA*–XTEN–nCas9–NLS construct, where TadA* represents an evolved TadA variant and XTEN is a 16-amino acid linker used in BE3[3], as ABE1.2. Transfection of plasmids expressing ABE1.2 and sgRNAs targeting six diverse human genomic sites (Extended Data Fig. E2a) resulted in very low, but observable A•T to G•C editing efficiencies (3.2±0.88%; all editing efficiencies are reported as mean±SD of three biological replicates 5 days post-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 catalyzing 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, predicted to lie in a helix adjacent to the TadA tRNA substrate (Fig. 2c). In mammalian cells, ABE2.1 (ABE1.2 + D147Y + E155V) exhibited 2- to 7-fold 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).

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. E2c), consistent with our previous findings with BE3[3]. We also varied linker lengths between TadA(2.1)* and Cas9 nickase. An ABE2 variant (ABE2.6) with a linker twice as long (32 amino acids, $(SGGS)_2$-XTEN-$(SGGS)_2$,) as the linker in ABE2.1 offered modestly higher editing efficiencies, now averaging 14±2.4% across the six genomic loci tested (Extended Data Fig. E2c).

Alkyl adenine DNA glycosylase (AAG) catalyzes the cleavage of the glycosidic bond of inosine in DNA[31]. To test if inosine excision impedes 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[31]), or *E. coli* Endo V (inactivated with a D35A mutation[32]). Neither ABE2.1– AAG(E125Q) (ABE2.2) nor ABE2.1–Endo V(D35A) (ABE2.3) exhibited altered A•T to

G•C editing efficiencies in HEK293T cells compared with ABE2.1 (Extended Data Fig. E2d). 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. E2e). Moreover, ABE2.1 induced virtually no indels (≤0.1%) or A•T to non-G•C products (≤0.1%) in HEK293T cells, consistent with inefficient excision of inosine (Extended Data Fig. E3). Taken 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[3,5].

As a final ABE2 engineering study, we explored the role of TadA* dimerization on base editing efficiencies. TadA natively operates as a homodimer, with one monomer catalyzing deamination, and the other monomer acting as a docking station for the tRNA substrate[30]. During selection in *E. coli*, endogenous TadA likely serves as the non-catalytic monomer. In mammalian cells, we hypothesized that tethering an additional wild-type or evolved TadA monomer might improve base editing by minimizing reliance on intermolecular ABE dimerization. Indeed, co-expression with ABE2.1 of either wild-type TadA or TadA*2.1 (ABE2.7 and ABE2.8, respectively), as well as direct fusion of either evolved or wild-type TadA to the N-terminus of ABE2.1 (ABE2.9 and ABE2.10, respectively), substantially improved editing efficiencies (Fig. 3a and Extended Data Fig. E4a). A fused TadA*–ABE2.1 architecture (ABE2.9) was identified to offer the highest editing efficiencies (averaging 20±3.8% across the six genomic loci, a 7.6±2.6-fold average improvement at each site over ABE1.2) and was used in all subsequent experiments (Fig. 2b and 3a).

Finally, we determined which of the two TadA* subunits within the TadA*–ABE2.1 fusion was responsible for A to I catalysis. We introduced an inactivating E59A mutation[22] 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. E4a). These results establish that the internal TadA subunit is responsible for ABE deamination activity.

ABEs That Efficiently Edit a Subset of Targets

Next we performed a third round of bacterial evolution starting with TadA*2.1–dCas9 to further increase editing efficiencies. We increased selection stringency by introducing two early stop codons (Q4stop and W15stop) in the kanamycin resistance gene (Kan[R], aminoglycoside phosphotransferase, Supplementary Table 8 and Supplementary 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±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 efficiencies compared to ABE3.1 (Extended Data Fig. E1 and E4b).

Although ABE3.1 mediated efficient base editing at some targets, such as the CAC in site 1 (65±4.2% conversion), for other sites, such as the GAG in site 5, editing efficiencies were much lower (8.3±0.67%) (Fig. 3b). The results from six genomic loci with different sequence contexts surrounding the target A suggested that ABEs from rounds 1–3 strongly preferred target sequence contexts of YAC, where Y = T or C. This preference was likely inherited from the substrate specificity of native *E. coli* TadA, which deaminates the A in the UAC anticodon of tRNA[Arg]. 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 predicted to interact with the nucleotides upstream and downstream of the target A[30]. 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 strongly converged 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. E4c), in mammalian cells ABE4.3 exhibited decreased base editing efficiencies (averaging 16±5.8%) compared with ABE3.1 (Fig. 2b and 3b). We hypothesized that the A142N mutation may 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 favor ABE constructs with faster kinetics, we subjected this library to the Cam[R] H193Y selection with higher doses of chloramphenicol after allowing ABE expression for only half the duration (7 h) of the previous rounds of evolution (~14 h) (Supplementary Table 8). Surprisingly, importing a consensus set of mutations from surviving clones (H36L, R51L, S146C, and K157N) into ABE3.1, creating ABE5.1, *decreased* overall editing efficiencies in HEK293T cells by 1.7±0.29-fold (Fig. 2b and 3b).

ABE5.1 included seven mutations since our dimerization state experiments on ABE2.1. We speculated that the accumulation of these mutations may impair 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 efficiencies compared to homodimeric ABE5.1 with two identical evolved TadA* domains.

ABE5.3 editing efficiencies across the six genomic test sites averaged 39±5.9%, with an average improvement at each site of 2.9±0.78-fold relative to ABE5.1 (Fig. 2b and 3b). Importantly, 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 round 5 library to the non-YAC spectinomycin selection used in round 4. Although no highly enriched or beneficial mutations emerged (Extended Data Fig. E5a), mutations from two genotypes emerging 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 Fig. E1 and E5b). Subsequent studies therefore used the ABE5.3 architecture containing heterodimeric wtTadA–TadA*–Cas9 nickase with two 32-residue linkers.

Highly Active ABEs With Broad Sequence Compatibility

A sixth round of evolution aimed to remove any non-beneficial mutations by DNA shuffling and to reexamine 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 through 5 along with wild-type E. coli TadA were shuffled and subjected to the 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. E1). ABE6.3 (ABE5.3+P48S) resulted in 1.3±0.28-fold higher average editing relative to ABE5.3 at the six genomic sites tested, and an average conversion efficiency of 47±5.8% (Fig. 2b and 3c). P48 is predicted to lie ~5 Å from the substrate adenosine 2'-hydroxyl in the TadA crystal structure (Fig. 2c), and we speculate that mutating this residue to Ser may improve compatibility with a deoxyadenosine substrate. While 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±0.13-fold more efficient than editing by ABE6.3, and 1.8±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 (~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 that requires editing T<u>A</u>T, and a new D208N mutation that requires editing T<u>A</u>A (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 (Fig. 2b, 3c, and 3d, and Extended Data Fig. E1, E6a, and E6b). ABE7.10 edited the six genomic test sites with an average efficiency 58±4.0%, an average improvement at each site of 1.3±0.20-fold relative to ABE6.3 (Fig 3c), and 29±7.4-fold compared to ABE1.2. Although mutational dissection revealed that all three of the new mutations contribute to enhanced editing efficiencies (Extended Data Fig. E1, E6a, and E6b), 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 and 2c). We speculate that substitution of Arg for Pro abrogates base-specific enzyme:DNA interactions, broadening target sequence compatibility.

Characterization of Late-Stage ABEs

We characterized in-depth the most promising ABEs from rounds 5–7. 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. E2a). Overall, we observed strong improvement of A•T to G•C editing efficiencies in HEK293T cells during the progression from ABE5 to ABE7 variants (Fig. 3c and 3d). The base editing efficiency of the most active editor overall, ABE7.10, averaged 53±3.7% at the 17 sites tested, exceeded 50% at 11 of these sites, and ranged from 34–68% (Fig. 3c and 3d). These efficiencies compare favorably to the typical C•G to T•A editing efficiencies of BE3[3].

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. E2a). The resulting editing outcomes (Extended Data Fig. E7a), together with an analysis of editing efficiencies at every protospacer position across all 19 sites tested (Extended Data Fig. E7b) suggest that the activity windows of late-stage variants are approximately 4–6 nucleotides wide, from protospacer positions ~4–7 for ABE7.10, and from positions ~4–9 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 ABE7.8-7.10 in U2OS cells at sites 1–6 and observed similar editing results as in HEK293T cells (Extended Data Fig. E6c), demonstrating that ABE activity is not limited to HEK293T cells.

Analysis of individual high-throughput DNA sequencing reads from ABE editing at 6 to 17 genomic sites in HEK293T cells reveals that base editing outcomes at nearby adenines within the editing window are not statistically independent events. The average normalized linkage disequilibrium (LD) between nearby target adenines steadily increased as ABE evolution proceeded (Extended Data Fig. E8), indicating that early-stage ABEs edit nearby adenines more independently, while late-stage ABEs edit nearby adenines more processively. These findings suggest that during the course of evolution, TadA may have evolved kinetic changes that decrease the likelihood of substrate release before additional As within the editing window are converted, resulting in processivity similar to that of BE3[3].

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%) among the 17 genomic NAN sites tested (Fig. 4 and Supplementary Table 1). We recently showed that undesired products of BE3 arise from uracil excision and downstream repair processes[5]. 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 the those of UNG, resulting in minimal base excision repair following ABE editing.

We compared ABE7.10-catalyzed A•T to G•C editing efficiencies to those of a current Cas9 nuclease-mediated HDR method, CORRECT[33]. 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 these 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 < 0.1% indels (Fig. 5b). The target mutation:indel ratio averaged 0.43 for CORRECT HDR, and > 500 for ABE7.10, representing a > 1,000-fold improvement in product selectivity favoring ABE7.10. While we note that 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. Since no method yet exists to comprehensively profile off-target activity of ABEs, we assumed that off-target ABE editing primarily occurs at the off-target sites that are edited when Cas9 nuclease is complexed with the same guide RNA, as we and others observed to be the case with BE3[3,8,34]. We treated HEK293T cells with three well-characterized guide RNAs[35] 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[35]. The efficiency of on-target indels by Cas9 and the efficiency of on-target base editing by ABE7.10 both averaged 54% (Supplementary Table 2–4). We observed detectable modification (≥0.2% indels) by Cas9 nuclease at nine of the 12 (75%) known off-target loci (Fig. 5c and Supplementary Table 2–4). In contrast, when complexed with the same sgRNAs, ABE7.10, ABE7.9, or ABE 7.8 led to ≥0.2% off-target base editing at only four of the 12 (33%) known Cas9 off-target sites. Moreover, the nine confirmed Cas9 off-target loci were modified with an average efficiency of 14% indels, while the four confirmed ABE off-target loci were modified with an average of only 1.3% A•T to G•C mutation (Supplementary Table 2–4). Although seven of the nine confirmed Cas9 off-target loci contained at least one A 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 of 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 in abundant mRNAs or in RNAs containing regions of homology to the native tRNA substrate of *E. coli* TadA in ABE7.10-treated HEK293T cells compared to untreated cells (Extended Data Fig. E9), nor any apparent ABE toxicity in bacterial or human cells under the conditions used in this study. 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 β-globin genes cause a variety of blood diseases. Humans with the rare benign condition HPFH (hereditary persistence of fetal hemoglobin) are resistant to some β-globin diseases including sickle-cell anemia. In certain patients, this phenotype is mediated by mutations in the promoters of the γ-globin genes *HBG1* and *HBG2* that enable sustained expression of fetal hemoglobin, which is normally silenced in humans around birth[36]. We designed an sgRNA that programs ABE to simultaneously mutate -198T to C in the promoter driving *HBG1* expression, and -198T to C in the promoter driving *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[37]. 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. E10).

The iron storage disorder hereditary hemochromatosis (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[38]. We transfected DNA encoding ABE7.10 and a guide RNA that places the target A at protospacer position 5 into an immortalized lymphoblastoid cell line (LCL) harboring the *HFE* C282Y genomic mutation. Due to the extreme resistance of LCL cells to transfection, we isolated transfected cells and measured editing efficiency by 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 future clinical therapies for diseases with a genetic component, including the development of ABEs that accept a wide variety of PAMs[7], 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 A 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[3] and BE4[5], 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.

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 generated in this work were assembled using the USER cloning method as previously described[39] 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 a 1-piece blunt-end ligation of a PCR product containing a variable 20-nt sequence corresponding to the desired sgRNA targeted site. Primers and templates used in the synthesis of all sgRNA plasmids used in this work are listed in Supplementary Table 5. All mammalian ABE constructs, sgRNA plasmids and bacterial constructs were transformed and stored as glycerol stocks at −80°C in Mach1 T1^R Competent Cells (Thermo Fisher Scientific), which are *recA⁻*. 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 ZympPURE Plasmid Midiprep (Zymo Research Corportion), 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)

Briefly, 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 × 25 μL PCR reactions containing 75 ng-1.2 μg 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 Dpn1 (NEB) at 37°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 μg/mL ethidium bromide. The PCR product was extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen) and eluted with 30 μL of H$_2$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 × 50 μL PCR reactions, 66 °C annealing, 20-s extension) in order to install the appropriate USER junction sequences onto the 5' and 3' end 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 μL per well in a 98-well PCR plate, 5–6 plates total, Tm 66 °C, 4.5-min extension) following the manufacturer's protocol. Each PCR reaction was combined with 300 mL of 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 5 column volumes of PE wash buffer (Qiagen) and the DNA fragment was eluted with 800 μL of H_2O per column. Both DNA fragments were quantified using a NanoDrop 1000 Spectrophotometer (Themo Fisher Scientific).

TadA* libraries were assembled following a previously reported USER assembly procedure[39] with the following conditions: 0.22 pmol of ecTadA mutagenized DNA fragment 1, 0.22 pmol of 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 μL of USER assembly mixture were combined in 50 mM potassium acetate, 20 mM Tris-acetate, 10 mM magnesium acetate, 100 μg/mL BSA at pH7.9 (1× CutSmart Buffer, New England Biolabs). Generally, each round of evolution required ~1 mL of USER assembly mixture (22 nmol of each DNA assembly fragment) which was distributed into 10-μL aliquots across multiple 8-well PCR strips. The reactions were warmed to 37 °C for 60 min, then heated to 80 °C for 3 min to denature the two enzymes. The assembly mixture was slowly cooled to 12 °C at 0.1 °C/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 of PB buffer (Qiagen) to the assembly reaction mixture and binding the material onto a MinElute column (480 μL per column). ABE hybridized library constructs were eluted in 30 μL of H_2O per column and 2 μL of this eluted material was added to 20 μL of NEB 10-beta electrocompotent *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 ~300 electroporations to generate 5–10 million colony forming units (cfu). Freshly electroporated *E. coli* were recovered in 200 mL pre-warmed Davis Rich Media (DRM) at 37 °C, and incubated with shaking at 200 rpm in a 500-mL vented baffled flask for 15 min before carbenicillin (for plasmid maintenance) was added to 30 μg/mL. The culture was incubated at 37 °C with shaking at 200 rpm for 18 h. The plasmid library was isolated with a ZympPURE Plasmid Midiprep kit following manufacturer's procedure (50 mL culture per DNA column), except the plasmid library was eluted in 200 μL pre-warmed water per column. Evolution rounds 1–3, 5 and 7 followed this procedure in order to generate the corresponding library 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 using ecTadA*(2.1)-dCas9 as a template and amplifying with appropriately designed degenerate NNK-containing primers (Supplementary Table 6). Briefly, ecTadA*(2.1)-dCas9 template was amplified separately with two sets of primers: NMG-1197 + NMG-1200, and NMG-1199 + NMG-1200, using Phusion U Green Multiplex PCR Master Mix, forming 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 μL of H_2O per 200 μL of PCR reaction. In a third PCR reaction, 1 μL of PCR product 1 and 1 μL PCR product 2 were combined with 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[40]. Solutions of 10 mM each of dATP, dCTP, dGTP and dTTP/dUTP (3 parts dUTP: 7 parts dTTP) 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 μM each of dATP, dCTP, dGTP, and dUTP/dTTP (3:7) in 1× ThermoPol Reaction Buffer (Tm 63 °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 μL of H_2O per extraction column. The purified DNA product was digested with 2 U of USER enzyme per 40 μL in 1× CutSmart Buffer at 37 °C and monitored by analytical agarose gel electrophoresis until digestion was complete. The reaction was quenched with 10 vol of PN1 binding buffer (Qiagen) when the starting material was no longer observed (typically 3–4 h at 37 °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 μL of pre-warmed H_2O per column.

The purified shuffled TadA* fragment was reassembled into full-length TadA*-XTEN-dCas9 product by an internal primer extension procedure. The eluted digested DNA fragments (25 μL) were combined with 4 U of Vent Polymerase (NEB), 800 μM each of dATP, dCTP, dGTP, and dTTP, 1 U of *Taq* DNA polymerase in 1× ThermoPol Buffer supplemented with 0.5 mM $MgSO_4$. The thermocycler program for the reassembly procedure was the following: 94 °C for 3 min, 40 cycles of denaturation at 92 °C for 30 s,

annealing over 60 s at increasing temperatures starting at 30 °C and adding 1 °C per cycle (cooling ramp = 1 °C/s), and extension at 72 °C for 60 s with an additional 4 s per cycle, ending with one final cycle of 72 °C for 10 min. The reassembled product was amplified by PCR with the following conditions: 15 μL of unpurified internal assembly was combined with 1 μM each of USER primers NMG-825 and NMG-826, 100 μL of Phusion U Green Multiplex PCR Master Mix and H_2O to a final volume of 200 μL, 63 °C annealing, extension time of 30 s. The PCR product was purified by gel electrophoresis and assembled using thhe 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. The library of evolution 6 constructs was isolated using a ZymoPURE Plasmid Midiprep kit following the manufacturer's procedure following transformation of the hybridized library into NEB 10-beta electrocompotent *E. coli*.

Bacterial evolution of TadA variants

The previously described strain S1030[41] was used in all evolution experiments and an electrocompotent version of the bacteria was prepared as previously described[39] harboring the appropriate selection plasmid specific to each round of evolution (Supplementary Table 7). Briefly, 2 μL of freshly generated TadA* library (300–600 ng/μL) prepared as described above was added to 22 μL of freshly prepared electrocompotent 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 $5–10 \times 10^6$ cfu. After electroporation, freshly transformed S1030 cells were recovered in a total of 250 mL of pre-warmed DRM media at 37 °C shaking at 200 rpm 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 following 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_{BAD} promoter. The culture was grown to saturation at 37 °C with shaking at 200 rpm for 18 h, except 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 500-cm^2 square culture dishes containing 1.8% agar-2xYT, 30 μg/mL of plasmid maintenance antibiotics, and a concentration of the selection antibiotic pre-determined to be above the MIC of the S1030 strain harboring the antibiotic 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) electrocompotent cells, incubated with maintenance antibiotic and L-Ara and re-challenged with the selection condition. After 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 (ATTC 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*-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_2. 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 post-seeding, cells were transfected at approximately 70% confluency with 1.5 μL of Lipofectamine 2000 (Thermo Fisher Scientific) according to the manufacturer's protocols and 750 ng of ABE plasmid, 250ng of sgRNA expression, and 10 ng of a GFP expression plasmid (Lonza). Unless otherwise stated, cells were cultured for 5 days, with a media change on day 3. Media 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⁻* cells and genomic DNA extraction

HAP1 and HAP1 AA*G⁻* cells were nucleofected using the SE Cell Line 4D-Nucleofector X Kit S according to the manufacturer's protocol. Briefly, 4×10^5 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 of media 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 manufacture's protocol. Briefly, 1.25×10^5 cells were nucleofected in 20 μL of SG buffer along with 500 ng of ABE plasmid and 100 ng of sgRNA expression plasmid using the 4D-Nucleofector program EH-100 in a 16-well Nucleocuvette strip (20 μL of cells per well). Freshly nucleofected cells were transferred into 250 μL of media in a 48-well poly-D-lysine coated culture plate. Cells were incubated for 5 days and media was changed every day. DNA was extracted as described above.

Electroporation of LCL *HFE* C828Y cells

LCL cells were electroporated using a Gene Pulser Xcell Electroporater (BioRad) and 0.4 cm gap Gene Pulser electroporation cuvettes (BioRad). Briefly, 1×10^7 LCL cells were resuspended in 250 μL RPMI-160 plus GlutaMax. To this media was added 65 μg of 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 on ice for 10 min, then transferred to 15 mL of pre-warmed RPMI-160 supplemented with 20% FBS in a T-75 flask. The next day, an additional 5 mL of media 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 of media, filtered through a 40 μm strainer (Thermo Fisher Scientific), and sorted for GFP fluorescence using an FACSAria III Flow Cytometer (Becton Dickenson Biosciences). GFP-positive cells were collected in a 1.5-mL tube containing 500 μL of media. After centrifugation, the media was removed and cells were washed twice with 600 μL of 1× PBS (Thermo Fisher Scientific). Genomic DNA was extracted as described above.

Comparison between ABE 7.10 and homology directed repair using the 'CORRECT' method[42]

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 of Cas9 or base editor plasmid, 250 ng of sgRNA expression plasmid, 1.5 μL of Lipofectamine 3000 (Thermo Fisher Scientific), and for HDR assays 0.7 μg of single-stranded donor DNA template (100 nt, PAGE-purified from IDT) according to the manufacturer's instructions. 100-mer single-stranded oligonucleotide donor templates are listed in Supplementary Table 10.

Genomic DNA was harvested 48 h post-transfection (as described by Tessier-Lavigne *et. al.* during the development of the CORRECT method[42]) 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 there was minimal risk of amplifying donor oligo template.

High-throughput DNA sequencing (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 discussed in this work 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 of genomic DNA extract and 12.5 μL of 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 of unpurified PCR 1 reaction mixture, and 12.5 μL of 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 of H_2O. 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 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.

Due 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). Briefly, 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 centered on the protospacer sequence. Indels were counted as reads in which this base calling window was ≥1 bp different in length. This Python script yields output

with identical quality (estimated base calling error rate of < 1 in 1,000), but in far less time due 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 were 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 using qPCR as described above. The number of cycles for each amplicon was determined empirically for each primer pair: for B-actin, B-catenin, GAPDH, Reticulocalbin: 29 cycles; for MN1 and RSLD1: 25 cycles. 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 disequlilbrium analysis

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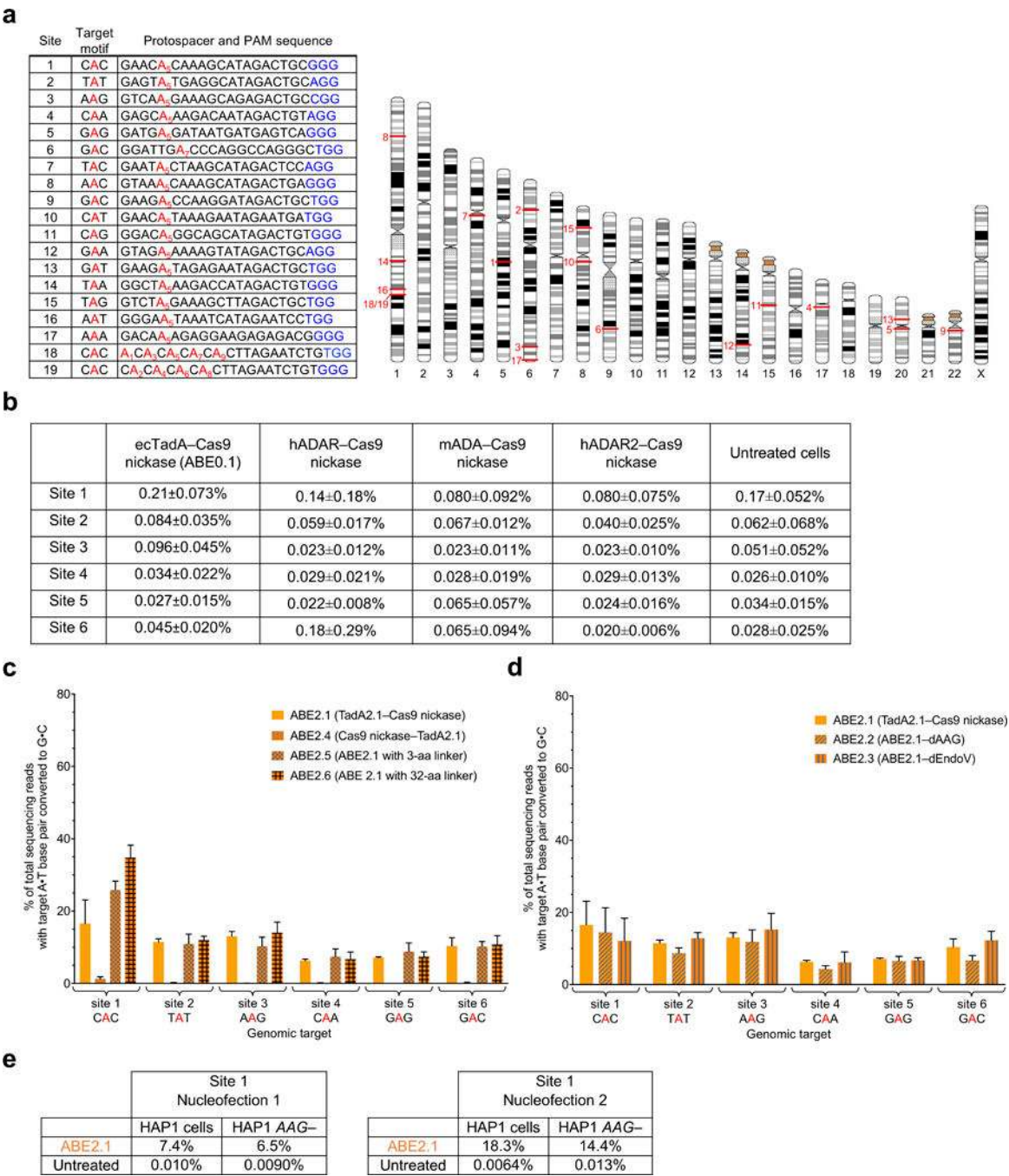
Data availability

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.

Extended Data

E. coli TadA amino acid

	23	26	36	37	48	49	51	72	84	97	106	108	123	125	142	146	147	152	155	156	157	161	TadA state	Linker 1 length	Linker 2 length	Other modifications relative to TadA*–Cas9 nickase
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-teminus 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	R	D	F	S	V	N	Y	G	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		R	V	F	K	K	monomer	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 E1. Genotypes of 57 ABEs described in this work
Mutations are colored based on the round of evolution in which they were identified.

a

Site	Target motif	Protospacer and PAM sequence
1	CAC	GAACA$_5$CAAAGCATAGACTGCGGG
2	TAT	GAGTA$_5$TGAGGCATAGACTGCAGG
3	AAG	GTCAA$_5$GAAAGCAGAGACTGCCGG
4	CAA	GAGCA$_4$AAGACAATAGACTGTAGG
5	GAG	GATGA$_5$GATAATGATGAGTCAGGG
6	GAC	GGATTGA$_7$CCCAGGCCAGGGCTGG
7	TAC	GAATA$_5$CTAAGCATAGACTCCAGG
8	AAC	GTAAA$_4$CAAAGCATAGACTGAGGG
9	GAC	GAAGA$_5$CCAAGGATAGACTGCTGG
10	CAT	GAACA$_5$TAAAGAATAGAATGATGG
11	CAG	GGACA$_5$GGCAGCATAGACTGTGGG
12	GAA	GTAGA$_5$AAAAGTATAGACTGCAGG
13	GAT	GAAGA$_5$TAGAGAATAGACTGCTGG
14	TAA	GGCTA$_5$AAGACCATAGACTGTGGG
15	TAG	GTCTA$_5$GAAAGCTTAGACTGCTGG
16	AAT	GGGAA$_5$TAAATCATAGAATCCTGG
17	AAA	GACAA$_5$AGAGGAAGAGAGACGGGG
18	CAC	A$_1$CA$_3$CA$_5$CA$_7$CA$_9$CTTAGAATCTGTGG
19	CAC	CA$_2$CA$_4$CA$_6$CA$_8$CTTAGAATCTGTGGG

b

	ecTadA–Cas9 nickase (ABE0.1)	hADAR–Cas9 nickase	mADA–Cas9 nickase	hADAR2–Cas9 nickase	Untreated cells
Site 1	0.21±0.073%	0.14±0.18%	0.080±0.092%	0.080±0.075%	0.17±0.052%
Site 2	0.084±0.035%	0.059±0.017%	0.067±0.012%	0.040±0.025%	0.062±0.068%
Site 3	0.096±0.045%	0.023±0.012%	0.023±0.011%	0.023±0.010%	0.051±0.052%
Site 4	0.034±0.022%	0.029±0.021%	0.028±0.019%	0.029±0.013%	0.026±0.010%
Site 5	0.027±0.015%	0.022±0.008%	0.065±0.057%	0.024±0.016%	0.034±0.015%
Site 6	0.045±0.020%	0.18±0.29%	0.065±0.094%	0.020±0.006%	0.028±0.025%

c **d**

e

	Site 1 Nucleofection 1			Site 1 Nucleofection 2	
	HAP1 cells	HAP1 AAG–		HAP1 cells	HAP1 AAG–
ABE2.1	7.4%	6.5%	ABE2.1	18.3%	14.4%
Untreated	0.010%	0.0090%	Untreated	0.0064%	0.013%

Extended Data Figure E2. 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 standard deviation 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) reflect the mean and s.d. of three independent biological replicates performed on different days.

Untreated — Site 1 (Indel% 0.059)

	G1	A2	A3	C4	A5	G6	A7	A8	A9	G10	C11	A12	T13	A14	G15	A16	C17	T18	G19	C20	G	G	G
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 — Site 1 (Indel% 0.048)

	G1	A2	A3	C4	A5	G6	A7	A8	A9	G10	C11	A12	T13	A14	G15	A16	C17	T18	G19	C20	G	G	G
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	100.0	0.0	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	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 — Site 2 (Indel% 0.058)

	G1	A2	G3	T4	A5	T6	G7	A8	G9	G10	C11	A12	T13	A14	G15	A16	C17	T18	G19	C20	A	G	G
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	0.0	99.9	99.9	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 — Site 2 (Indel% 0.047)

	G1	A2	G3	T4	A5	T6	G7	A8	G9	G10	C11	A12	T13	A14	G15	A16	C17	T18	G19	C20	A	G	G
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 — Site 3 (Indel% 0.075)

	G1	T2	C3	A4	A5	G6	A7	A8	A9	G10	C11	A12	G13	A14	G15	A16	C17	T18	G19	C20	C	G	G
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 — Site 3 (Indel% 0.085)

	G1	T2	C3	A4	A5	G6	A7	A8	A9	G10	C11	A12	G13	A14	G15	A16	C17	T18	G19	C20	C	G	G
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	100.0	100.0	0.0	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 — Site 4 (Indel% 0.042)

	G1	A2	G3	C4	A5	A6	A7	G8	A9	G10	A11	A12	T13	A14	G15	A16	C17	T18	G19	T20	A	G	G
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	0.1	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	99.9	0.0	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 — Site 4 (Indel% 0.024)

	G1	A2	G3	C4	A5	A6	A7	G8	A9	G10	A11	A12	T13	A14	G15	A16	C17	T18	G19	T20	A	G	G
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	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	100.0	0.0	0.0	0.0	100.0	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	100.0	0.0	0.0	0.0

Untreated — Site 5 (Indel% 0.036)

	G1	A2	T3	G4	A5	G6	A7	T8	A9	A10	T11	G12	A13	T14	G15	A16	G17	T18	C19	A20	G	G	G
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	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 — Site 5 (Indel% 0.060)

	G1	A2	T3	G4	A5	G6	A7	T8	A9	A10	T11	G12	A13	T14	G15	A16	G17	T18	C19	A20	G	G	G
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 — Site 6 (Indel% 0.014)

	G1	G2	A3	T4	T5	G6	A7	C8	C9	C10	A11	G12	G13	C14	G15	A16	C17	G18	G19	C20	T	G	G
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	100.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0
G	100.0	99.8	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	99.9	100.0
T	0.0	0.0	0.0	100.0	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

ABE2.1 — Site 6 (Indel% 0.014)

	G1	G2	A3	T4	T5	G6	A7	C8	C9	C10	A11	G12	G13	C14	G15	A16	C17	G18	G19	C20	T	G	G
A	0.0	0.0	99.9	0.0	0.0	0.0	87.1	0.1	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.1	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	99.9	99.9	99.9	0.0	0.0	0.0	99.9	99.9	0.0	0.0	0.0	99.9	0.0	0.0	0.0	0.0
G	100.0	99.9	0.1	0.0	0.0	100.0	12.9	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	99.9	99.9	99.9	0.0	0.0	99.9	100.0
T	0.0	0.1	0.0	100.0	100.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.1	0.1	0.0	0.1	0.0	0.1	0.0	100.0	0.0	0.0	0.0

Extended Data Figure E3. High-throughput DNA sequencing analysis of HEK293T cells treated with ABE2.1 and sgRNAs targeting each of six human genomic sites

One representative replicate is shown. Data from untreated HEK293T cells are shown for comparison.
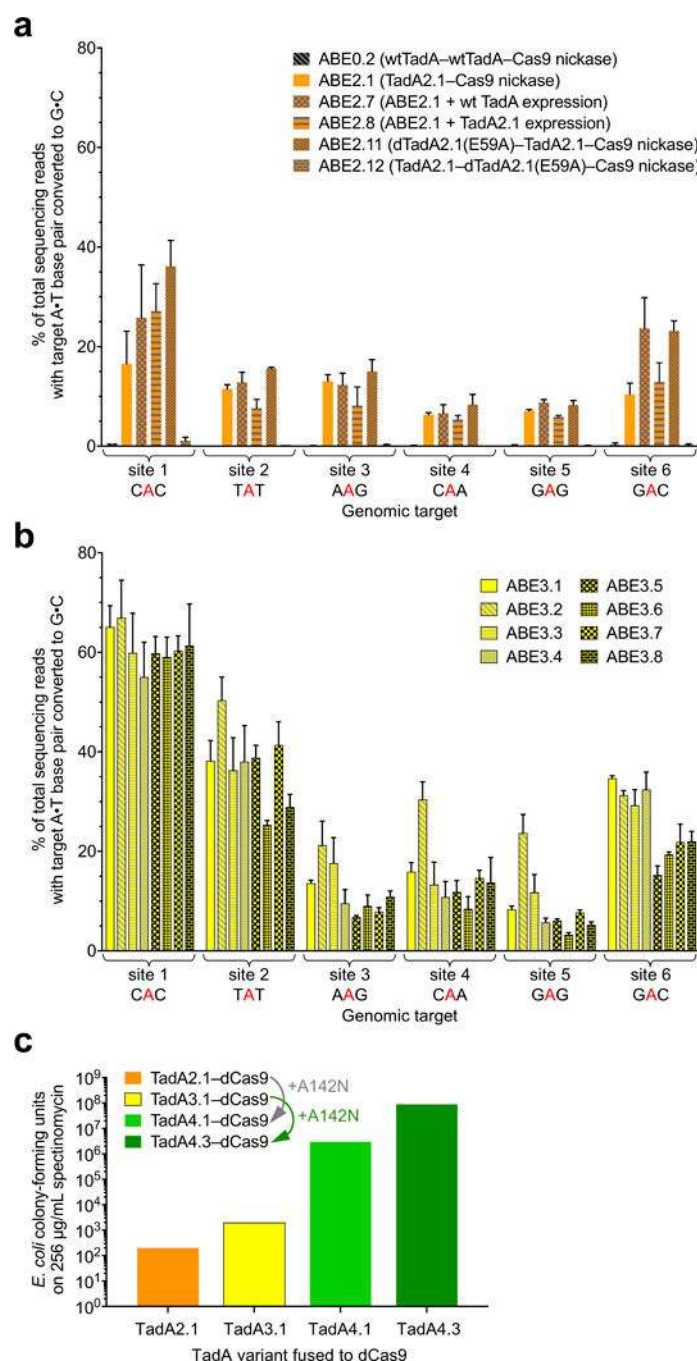
Extended Data Figure E4. Base editing efficiencies of additional ABE2 and ABE3 variants, and the effect of adding A142N to TadA*–dCas9 on antibiotic selection survival in *E. coli*
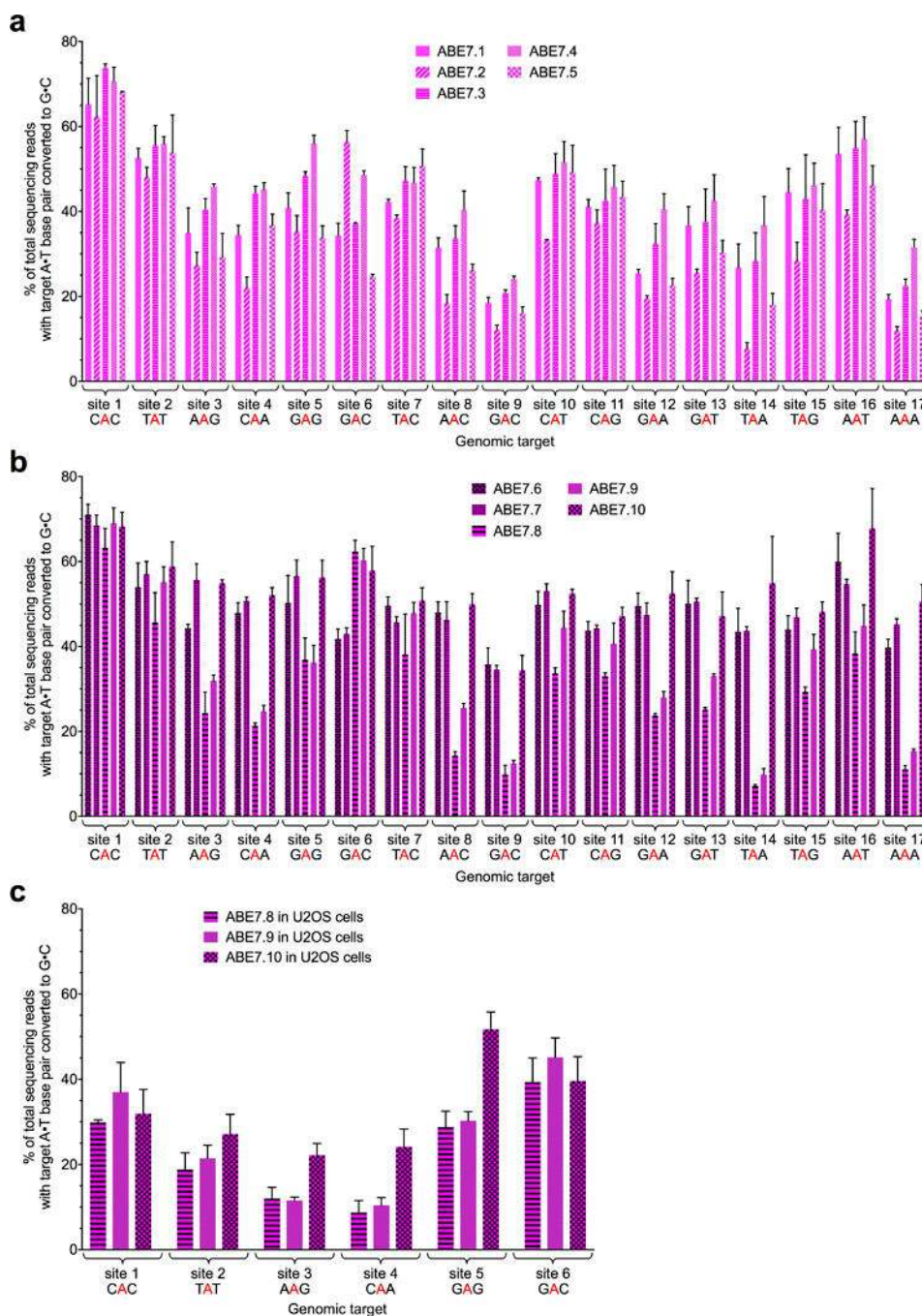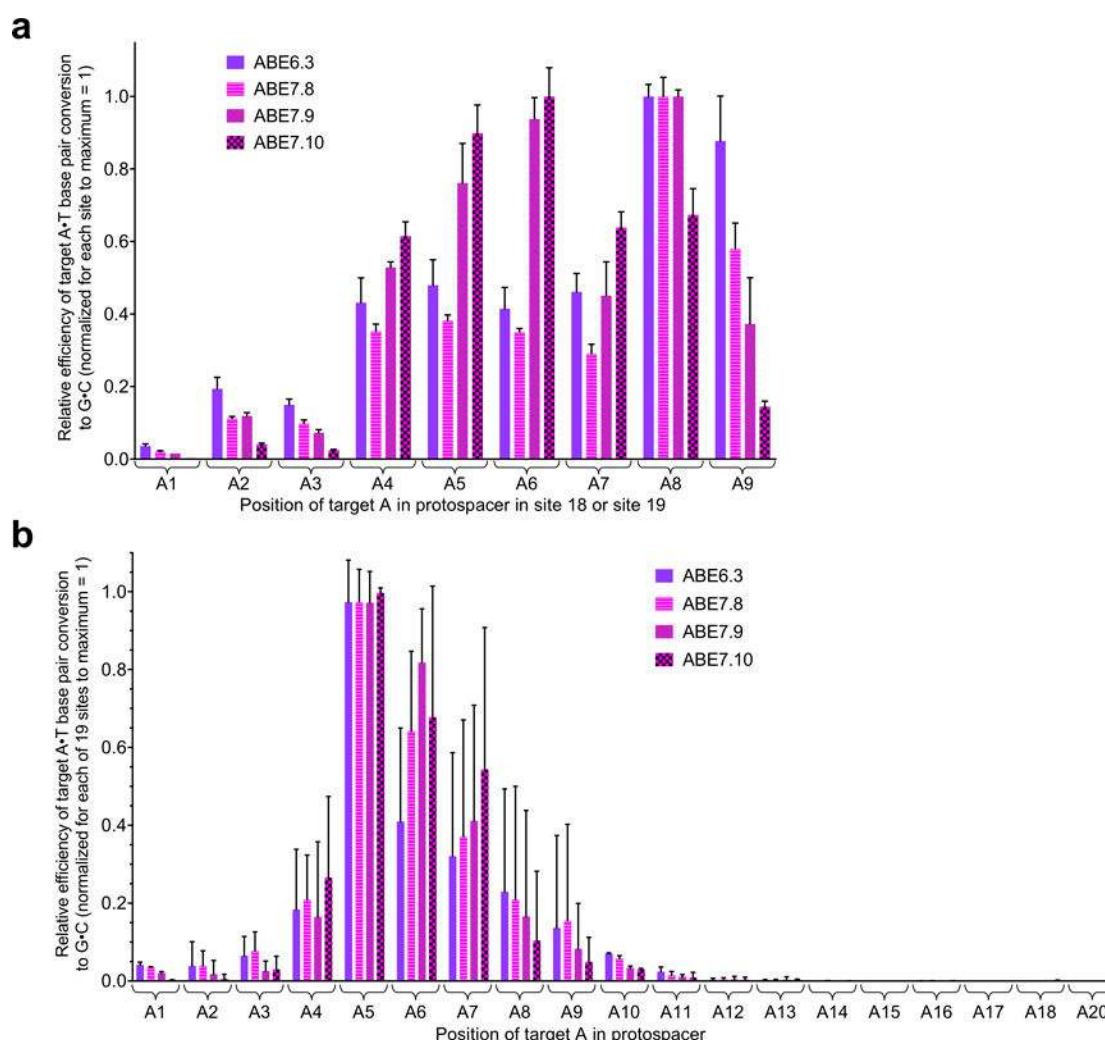
a, A•T to G•C base editing efficiencies in HEK293T cells at six human genomic target DNA sites of ABE2 variants with different engineered dimeric states. A control ABE variant containing two wild-type TadA domains and no evolved TadA* domains (ABE0.2) did not result in A•T to G•C editing at the six genomic 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 Figure E1 for ABE genotypes and architectures. **c**, Colony-forming units on 2xYT agar with 256 μg/mL 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) reflect the mean and s.d. of three independent biological replicates performed on different days.
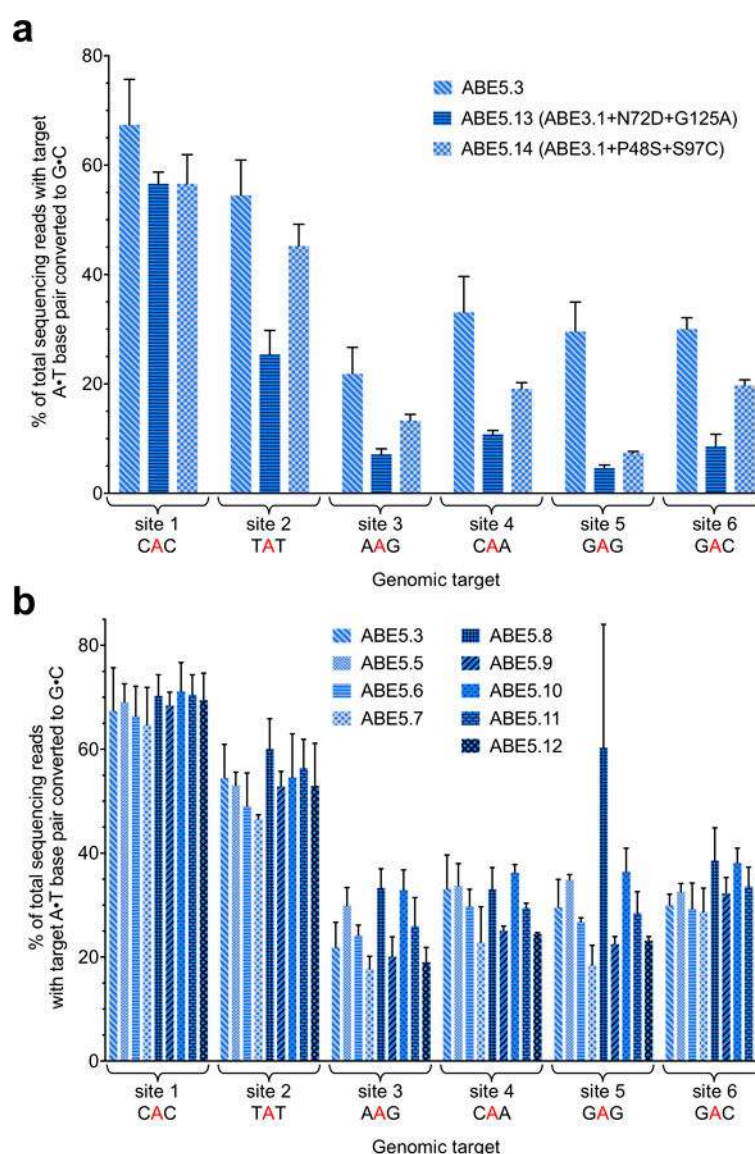
Extended Data Figure E5. Base editing efficiencies of additional ABE5 variants

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 Figure E1 for ABE genotypes and architectures. Values and error bars reflect the mean and s.d. of three independent biological replicates performed on different days.

Extended Data Figure E6. Base editing efficiencies of ABE7 variants at 17 genomic sites

A•T to G•C base editing efficiencies in HEK293T cells at 17 human genomic target DNA sites of ABE7.1-7.5 (**a**), and ABE7.6-7.10 (**b**). See Extended Data Figure E1 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 ABE7.8-7.10. The lower editing efficiencies observed in U2OS cells compared with HEK293T cells are consistent with transfection efficiency differences between the two cell lines; we observed transfection efficiencies of ~40–55% in U2OS cells under the conditions used in this study, compared to ~65–80% in HEK293T cells. Values and error bars reflect the mean and s.d. of three independent biological replicates performed on different days.

Extended Data Figure E7. 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 A at each of these positions. Values are normalized to the maximum observed efficiency at each of the two sites for each ABE = 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 to the maximum observed efficiency at each of the

19 sites for each ABE = 1. For (a) and (b), values and error bars reflect the mean and s.d. of three independent biological replicates performed on different days.
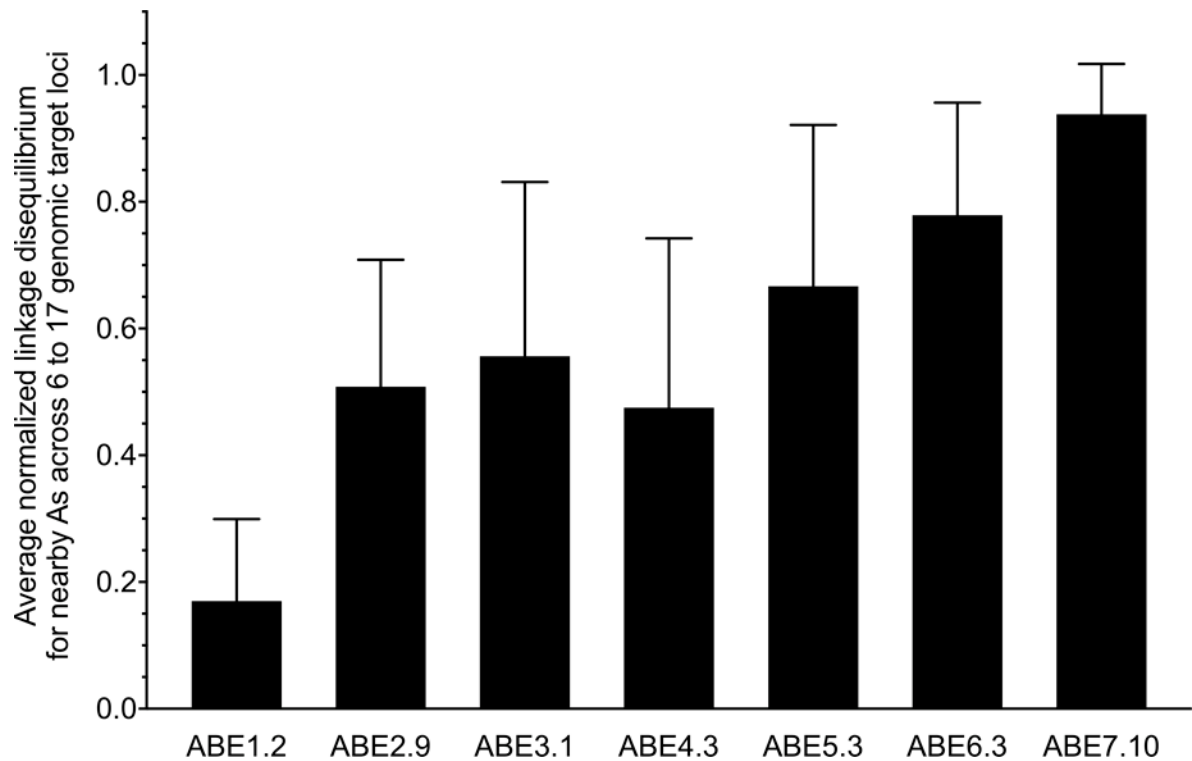
Extended Data Figure E8. Rounds of evolution and engineering increased ABE processivity
The calculated mean normalized linkage disequilibrium (LD) between nearby target As at 6 to 17 human genomic target DNA sites for the most active ABEs emerging from each round of evolution and engineering. Higher LD values indicate that an ABE is more likely to edit an A if a nearby A in the same DNA strand (the same sequencing read) is also edited. LD values are normalized from 0 to 1 in order to be independent of editing efficiency. Values and error bars reflect the mean and s.d. of normalized LD values from three independent biological replicates performed on different days.

RNA	A position	% of RNA sequencing reads with a mutation at the specified A			
		ABE-treated 1	ABE-treated 2	Untreated 1	Untreated 2
beta-actin	112	0.11	0.11	0.14	0.21
beta-actin	118	0.41	0.40	0.44	0.33
beta-actin	137	0.25	0.29	0.42	0.34
beta-actin	239	0.22	0.25	0.27	0.22
beta-catenin	21	0.77	0.84	1.18	2.49
GAPDH	33	0.28	0.43	0.40	0.42
GAPDH	66	0.30	0.48	0.46	0.50
GAPDH	74	0.25	0.36	0.37	0.35
GAPDH	75	0.30	0.49	0.46	0.51
GAPDH	78	0.45	0.56	0.51	0.50
GAPDH	86	0.34	0.54	0.50	0.54
GAPDH	105	0.51	0.77	0.67	0.72
GAPDH	112	0.22	0.35	0.25	0.29
GAPDH	123	0.26	0.42	0.41	0.47
GAPDH	127	0.20	0.32	0.35	0.33
GAPDH	137	0.15	0.24	0.26	0.28
GAPDH	172	0.28	0.41	0.46	0.52
GAPDH	210	0.77	1.27	2.17	2.34
GAPDH	211	0.81	1.35	2.25	2.42
GAPDH	213	0.76	1.19	2.02	2.15
RB1	37	0.26	0.22	0.28	0.19
RB1	50	0.16	0.38	1.06	1.95
RB1	86	0.05	0.16	0.28	0.36
RB1	88	0.11	0.16	0.26	0.40
RB1	96	1.22	1.27	2.11	2.21
RB1	114	0.50	0.55	0.57	0.52
RB1	219	0.90	0.84	0.94	0.78
MN1	16	0.65	0.93	0.48	0.65
MN1	211	0.56	0.59	0.24	0.82
MN1	212	3.88	4.89	3.39	5.12
MN1	215	1.53	1.77	1.11	1.28
RSLD1	196	0.16	0.29	0.19	0.18

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...
MN1 amplicon (position 1-30): CTGGTTGACCAATCTGTGGTGAATAGTGGA...
RSLD1 amplicon (position 1-30): TTGGCTTTCCAAATCAGTGGGTCTGACTTG...

Extended Data Figure E9. Analysis of cellular RNAs in ABE7.10-treated cells compared with untreated cells

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), and two cellular RNAs with sequence homology to the tRNA anticodon loop that is the native substrate of *E. coli* TadA (encoding MN1 and RSLD1), were amplified and analyzed by HTS. Within each amplicon, the mutation frequency at each adenine position for which the mutation rate was ≥0.2% 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.

Extended Data Figure E10. High-throughput DNA sequencing analysis of HEK293T cells treated with five late-stage ABE variants and an sgRNA targeting -198T in the promoter of *HBG1 and HBG2*

One representative replicate is shown of DNA sequences at the *HBG1* (a) and *HBG2* (b) promoter targets. ABE-mediated base editing installs a -198T→C mutation on the strand

complementary to the one shown in the sequencing data tables. Data from untreated HEK293T cells are shown for comparison.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Krokan HE, Drabløs F, Slupphaug G. Uracil in DNA–occurrence, consequences and repair. Oncogene. 2002; 21:8935–8948. DOI: 10.1038/sj.onc.1205996 [PubMed: 12483510]

2. Lewis CA, 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. 2016; 113:8194–8199. DOI: 10.1073/pnas.1607580113 [PubMed: 27382162]

3. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016; 533:420–424. DOI: 10.1038/nature17946 [PubMed: 27096365]

4. Nishida K, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science. 2016; 353

5. Komor AC, 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. 2017; 3:eaao4774. [PubMed: 28875174]

6. Komor AC, Badran AH, Liu DR. Editing the genome without double-stranded DNA breaks. ACS Chem Biol. 2017

7. Kim YB, et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat Biotechnol. 2017; 35:371–376. DOI: 10.1038/nbt.3803 [PubMed: 28191901]

8. Rees HA, et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. Nat Commun. 2017; 8:15790. [PubMed: 28585549]

9. Satomura A, et al. Precise genome-wide base editing by the CRISPR nickase system in yeast. Sci Rep. 2017; 7:2095. [PubMed: 28522803]

10. Lu Y, Zhu J-K. Precise Editing of a Target Base in the Rice Genome Using a Modified CRISPR/Cas9 System. Mol Plant. 2017; 10:523–525. DOI: 10.1016/j.molp.2016.11.013 [PubMed: 27932049]

11. Zong Y, et al. Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. Nat Biotechnol. 2017; 35:438–440. DOI: 10.1038/nbt.3811 [PubMed: 28244994]

12. Zhang Y, et al. Programmable base editing of zebrafish genome using a modified CRISPR-Cas9 system. Nat Commun. 2017; 8:118. [PubMed: 28740134]

13. Billon P, et al. CRISPR-mediated base editing enables efficient disruption of eukaryotic genes through induction of STOP codons. Mol Cell. 2017; 67:1068–1079.e1064. DOI: 10.1016/j.molcel.2017.08.008 [PubMed: 28890334]

14. Kuscu C, et al. CRISPR-STOP: gene silencing through base-editing-induced nonsense mutations. Nat Methods. 2017; 14:710–712. DOI: 10.1038/nmeth.4327 [PubMed: 28581493]

15. Kim K, et al. Highly efficient RNA-guided base editing in mouse embryos. Nat Biotechnol. 2017; 35:435–437. DOI: 10.1038/nbt.3816 [PubMed: 28244995]

16. Chadwick AC, 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. 2017; 37:1741–1747. [PubMed: 28751571]

17. Liang P, et al. Correction of β-thalassemia mutant by base editor in human embryos. Protein & Cell. 2017; 12:61–12. DOI: 10.1007/s13238-017-0475-6

18. Li G, et al. Highly efficient and precise base editing in discarded human tripronuclear embryos. Protein & Cell. 2017; 532:289–284. DOI: 10.1007/s13238-017-0458-7

19. Tang W, Hu JH, Liu DR. Aptazyme-embedded guide RNAs enable ligand-responsive genome editing and transcriptional activation. Nat Commun. 2017; 8:15939. [PubMed: 28656978]

20. 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. 2008; 377:1015–1023. DOI: 10.1016/j.jmb.2008.01.033 [PubMed: 18304575]

21. Zheng Y, Lorenzo C, Beal PA. DNA editing in DNA/RNA hybrids by adenosine deaminases that act on RNA. Nucleic Acids Res. 2017; 45:3369–3377. DOI: 10.1093/nar/gkx050 [PubMed: 28132026]

22. Kim J, et al. Structural and kinetic characterization of *Escherichia coli* TadA, the wobble-specific tRNA deaminase. Biochemistry. 2006; 45:6407–6416. DOI: 10.1021/bi0522394 [PubMed: 16700551]

23. Wolf J, Gerber AP, Keller W. tadA, an essential tRNA-specific adenosine deaminase from *Escherichia coli*. EMBO J. 2002; 21:3841–3851. [PubMed: 12110595]

24. Matthews MM, et al. Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. Nat Struct Mol Biol. 2016; 23:426–433. DOI: 10.1038/nsmb.3203 [PubMed: 27065196]

25. Grunebaum E, Cohen A, Roifman CM. Recent advances in understanding and managing adenosine deaminase and purine nucleoside phosphorylase deficiencies. Curr Opin Allergy Clin Immunol. 2013; 13:630–638. DOI: 10.1097/ACI.0000000000000006 [PubMed: 24113229]

26. Gerber AP, Keller W. An adenosine deaminase that generates inosine at the wobble position of tRNAs. Science. 1999; 286:1146–1149. [PubMed: 10550050]

27. Fukui K. DNA mismatch repair in eukaryotes and bacteria. J Nucleic Acids. 2010; 2010

28. Shi K, et al. Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B. Nat Struct Mol Biol. 2017; 24:131–139. DOI: 10.1038/nsmb.3344 [PubMed: 27991903]

29. Macbeth MR, et al. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. Science. 2005; 309:1534–1539. [PubMed: 16141067]

30. Losey HC, Ruthenburg AJ, Verdine GL. Crystal structure of *Staphylococcus aureus* tRNA adenosine deaminase TadA in complex with RNA. Nat Struct Mol Biol. 2006; 13:153–159. DOI: 10.1038/nsmb1047 [PubMed: 16415880]

31. Lau AY, Wyatt MD, Glassner BJ, Samson LD, Ellenberger T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. Proc Natl Acad Sci USA. 2000; 97:13573–13578. DOI: 10.1073/pnas.97.25.13573 [PubMed: 11106395]

32. Sebastian Vik E, et al. Endonuclease V cleaves at inosines in RNA. Nat Commun. 2013; 4:2271. [PubMed: 23912683]

33. Paquet D, et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. Nature. 2016; 533:125–129. DOI: 10.1038/nature17664 [PubMed: 27120160]

34. Kim D, et al. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. Nat Biotechnol. 2017; 35:475–480. DOI: 10.1038/nbt.3852 [PubMed: 28398345]

35. Tsai SQ, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat Biotechnol. 2015; 33:187–197. DOI: 10.1038/nbt.3117 [PubMed: 25513782]

36. Traxler EA, et al. A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. Nat Med. 2016; 22:987–990. DOI: 10.1038/nm.4170 [PubMed: 27525524]

37. Wienert B, et al. KLF1 drives the expression of fetal hemoglobin in British HPFH. Blood. 2017; 130:803–807. DOI: 10.1182/blood-2017-02-767400 [PubMed: 28659276]

38. Alexander J, Kowdley KV. HFE-associated hereditary hemochromatosis. Genet Med. 2009; 11:307–313. [PubMed: 19444013]

39. Badran AH, et al. Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. Nature. 2016; 533:58–63. DOI: 10.1038/nature17938 [PubMed: 27120167]

40. Müller KM, et al. Nucleotide exchange and excision technology (NExT) DNA shuffling: a robust method for DNA fragmentation and directed evolution. Nucleic Acids Res. 2005; 33:e117–e117. DOI: 10.1093/nar/gni116 [PubMed: 16061932]

41. Esvelt KM, Carlson JC, Liu DR. A system for the continuous directed evolution of biomolecules. Nature. 2011; 472:499–503. DOI: 10.1038/nature09929 [PubMed: 21478873]

42. Kwart D, Paquet D, Teo S, Tessier-Lavigne M. Precise and efficient scarless genome editing in stem cells using CORRECT. Nat Protocols. 2017; 12:329–354. DOI: 10.1038/nprot.2016.171 [PubMed: 28102837]
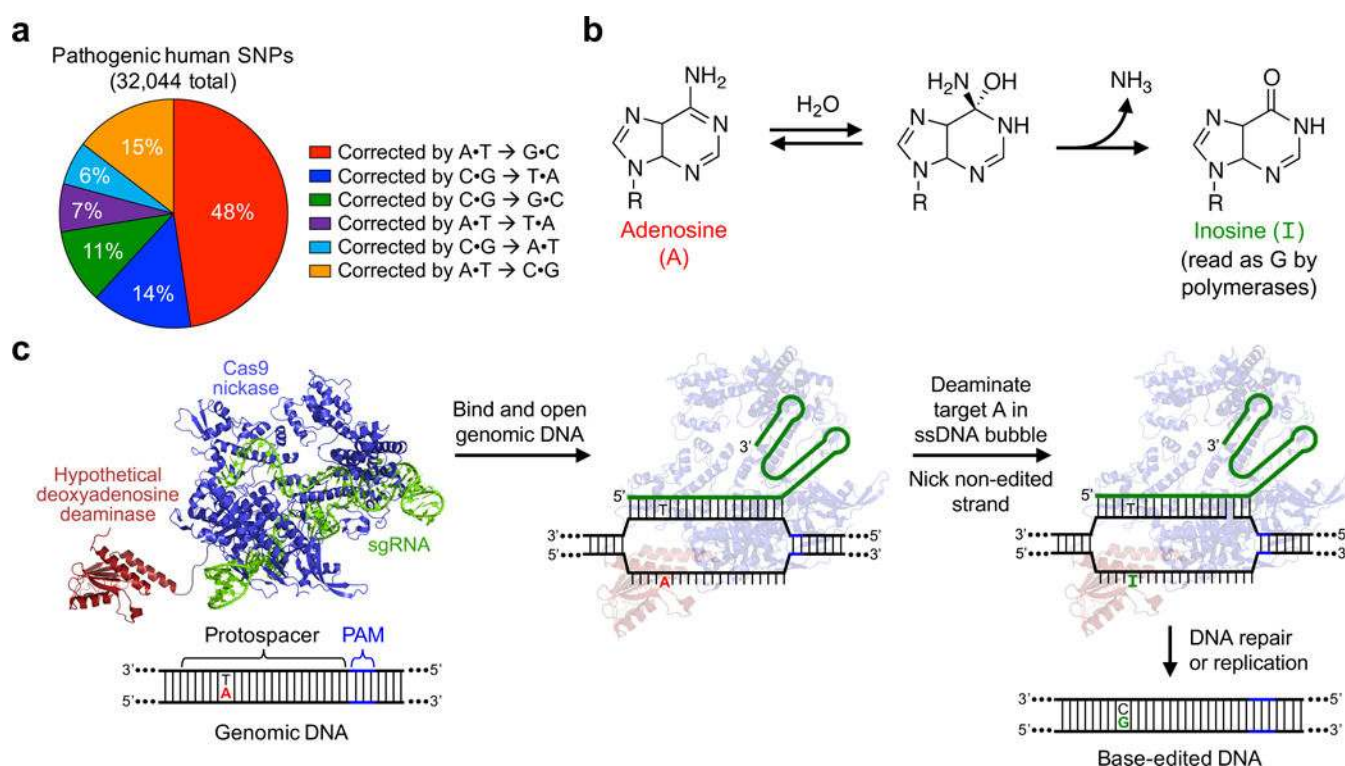
Figure 1. Scope and overview of base editing by an A•T to G•C base editor (ABE)
a, Base pair changes required to correct pathogenic human SNPs in the ClinVar database. **b**, The deamination of adenosine (A) forms inosine (I), which is read as guanosine (G) by polymerase enzymes. **c**, ABE-mediated 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 catalyzes A to I formation 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.
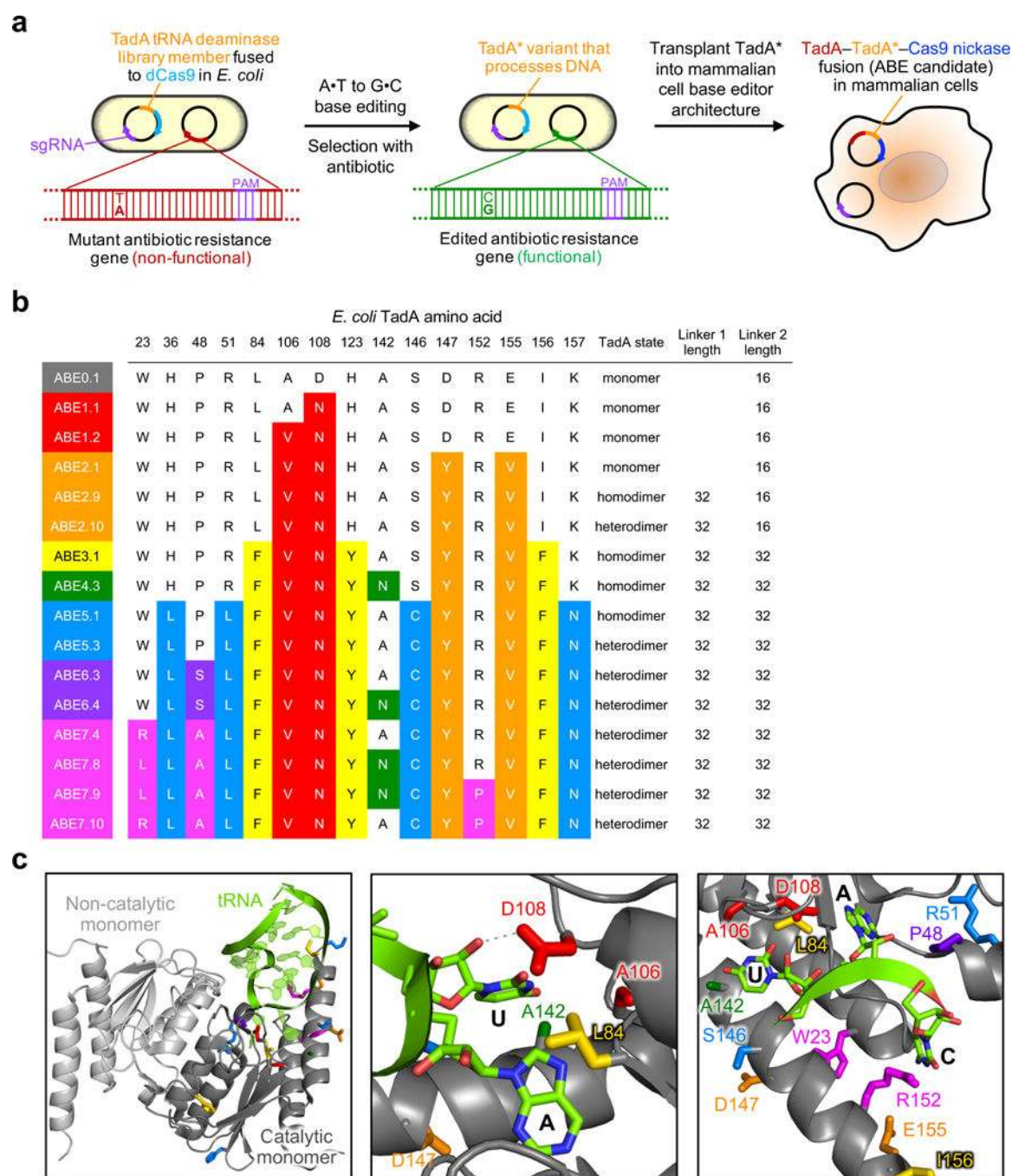
Figure 2. Protein evolution and engineering of ABEs

a, Strategy to evolve a DNA deoxyadenosine deaminase starting from TadA. A library of *E. coli* harbors 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 Figure E1. The dimerization state (monomer, TadA*–TadA* homodimer, or wild-type TadA–evolved TadA* heterodimer) and linker

length (in amino acids) are also listed. **c**, Three views of the *E. coli* TadA deaminase (PDB 1Z3A) aligned with *S. aureus* TadA (not shown) complexed with tRNA$^{\text{Arg2}}$ (PDB 2B3J). The U$\underline{\text{A}}$C anticodon loop of the tRNA is the native substrate of wild-type TadA.
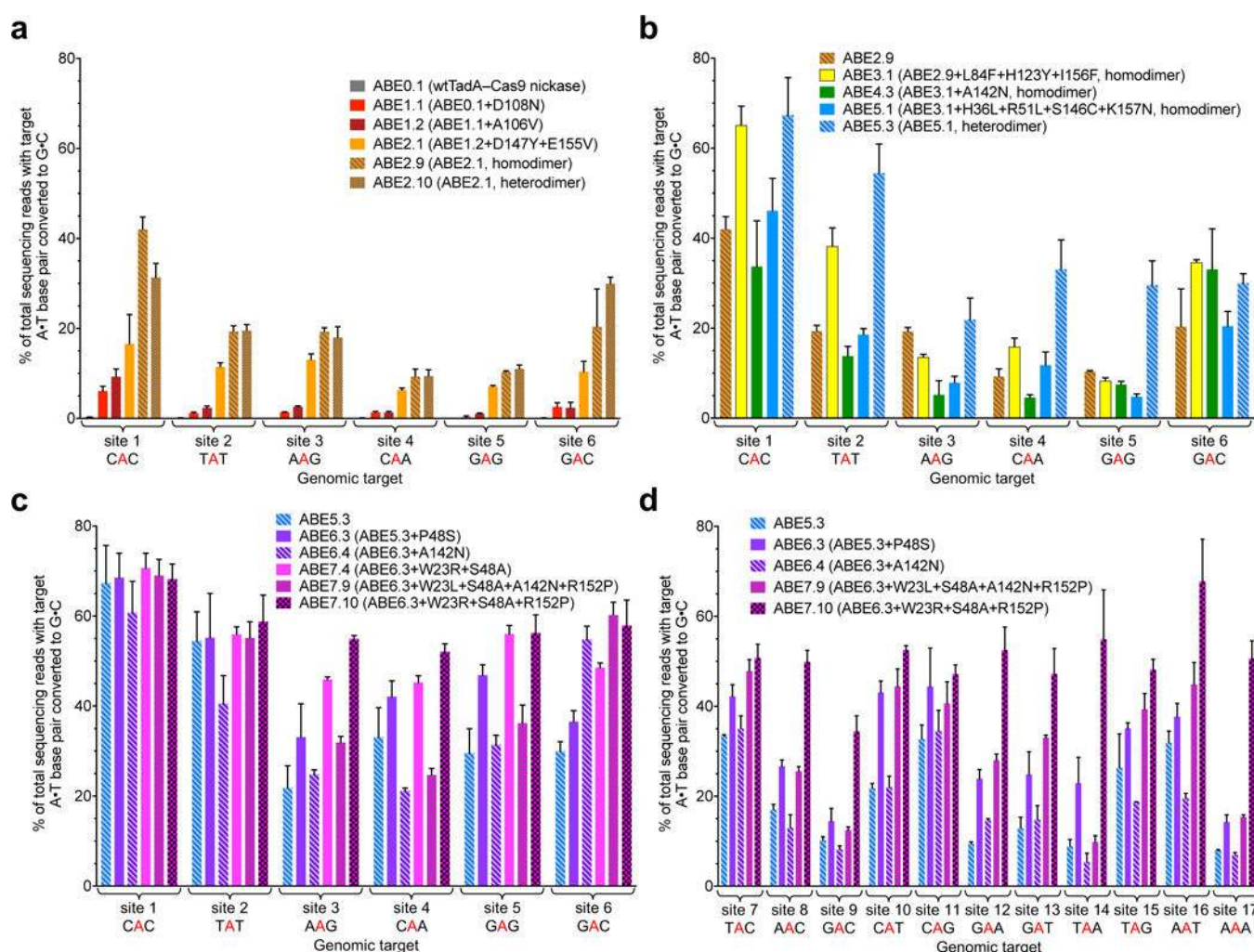
Figure 3. Evolved ABEs mediate A•T to G•C base editing at human genomic DNA sites
A•T to G•C base editing efficiencies in HEK293T cells of round 1 and round 2 ABEs (**a**),
round 3, round 4, and round 5 ABEs (**b**), and round 6 and round 7 ABEs (**c**) at six human
genomic DNA sites. **d**, Editing efficiencies in HEK293T cells of round 6 and round 7 ABEs
at an expanded set of human genomic sites. Values 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. ABEs in (b) are homodimers except ABE5.3; ABEs in (c)
and (d) are all heterodimers.

Untreated — Site 2 (Indel% 0.04)

	G₁	A₂	G₃	T₄	A₅	T₆	G₇	A₈	G₉	G₁₀	C₁₁	A₁₂	T₁₃	A₁₄	G₁₅	A₁₆	C₁₇	T₁₈	G₁₉	C₂₀
A	0.0	100.0	0.0	0.0	99.8	0.0	0.1	99.9	0.1	0.0	0.0	99.9	0.0	100.0	0.0	99.9	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	100.0	0.1	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9
G	99.9	0.0	100.0	0.0	0.2	0.0	99.8	0.0	99.9	99.9	0.0	0.0	0.0	0.0	100.0	0.1	0.0	0.0	100.0	0.1
T	0.0	0.0	0.0	100.0	0.0	100.0	0.0	0.1	0.1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0

ABE7.10 — Site 2 (Indel% 0.1)

	G₁	A₂	G₃	T₄	A₅	T₆	G₇	A₈	G₉	G₁₀	C₁₁	A₁₂	T₁₃	A₁₄	G₁₅	A₁₆	C₁₇	T₁₈	G₁₉	C₂₀
A	0.1	99.9	0.0	0.0	38.5	0.0	0.2	97.5	0.0	0.0	0.0	99.8	0.0	100.0	0.1	100.0	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	100.0	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	0.0	61.5	0.0	99.7	2.5	99.9	99.9	0.0	0.1	0.0	0.0	99.9	0.0	0.0	0.0	99.9	0.0
T	0.1	0.0	0.0	100.0	0.0	99.9	0.1	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0	100.0	0.1	0.1

ABE7.9 — Site 2 (Indel% 0.04)

	G₁	A₂	G₃	T₄	A₅	T₆	G₇	A₈	G₉	G₁₀	C₁₁	A₁₂	T₁₃	A₁₄	G₁₅	A₁₆	C₁₇	T₁₈	G₁₉	C₂₀
A	0.1	99.8	0.0	0.0	40.8	0.0	0.2	87.2	0.0	0.0	0.0	99.9	0.0	100.0	0.0	100.0	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	100.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	99.9
G	99.9	0.2	100.0	0.0	59.2	0.0	99.8	12.7	100.0	100.0	0.0	0.1	0.0	0.0	100.0	0.0	0.0	0.0	100.0	0.0
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

Untreated — Site 6 (Indel% 0.02)

	G₁	G₂	A₃	T₄	T₅	G₆	A₇	C₈	C₉	C₁₀	A₁₁	G₁₂	G₁₃	C₁₄	C₁₅	A₁₆	G₁₇	G₁₈	G₁₉	C₂₀
A	0.0	0.0	100.0	0.0	0.0	0.1	99.8	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
C	0.0	0.0	0.0	0.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0	0.0	0.0	100.0	99.9	0.0	0.0	0.0	0.0	100.0
G	100.0	100.0	0.0	0.0	0.0	99.9	0.2	0.0	0.0	0.0	0.0	100.0	100.0	0.0	0.0	0.0	100.0	100.0	100.0	0.0
T	0.0	0.0	0.0	100.0	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

ABE7.10 — Site 6 (Indel% 0.05)

	G₁	G₂	A₃	T₄	T₅	G₆	A₇	C₈	C₉	C₁₀	A₁₁	G₁₂	G₁₃	C₁₄	C₁₅	A₁₆	G₁₇	G₁₈	G₁₉	C₂₀
A	0.0	0.0	92.8	0.0	0.0	0.0	37.1	0.0	0.0	0.0	97.5	0.0	0.0	0.0	0.0	100.0	0.1	0.0	0.0	0.0
C	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	100.0	100.0	0.0	0.1	0.0	0.0	100.0
G	100.0	100.0	7.2	0.0	0.0	100.0	62.9	0.0	0.0	0.0	2.5	100.0	100.0	0.0	0.0	0.0	99.9	100.0	100.0	0.0
T	0.0	0.0	0.0	100.0	100.0	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

ABE7.9 — Site 6 (Indel% 0.03)

	G₁	G₂	A₃	T₄	T₅	G₆	A₇	C₈	C₉	C₁₀	A₁₁	G₁₂	G₁₃	C₁₄	C₁₅	A₁₆	G₁₇	G₁₈	G₁₉	C₂₀
A	0.0	0.0	98.6	0.0	0.0	0.0	36.5	0.1	0.0	0.0	99.8	0.1	0.0	0.0	0.0	100.0	0.0	0.0	0.0	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	100.0	99.9	0.0	0.0	0.0	0.0	99.9
G	100.0	100.0	1.4	0.0	0.0	100.0	63.5	0.0	0.0	0.0	0.2	99.9	100.0	0.0	0.0	0.0	99.9	100.0	100.0	0.0
T	0.0	0.0	0.0	100.0	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	0.1

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. 22,746-111,215 sequencing reads were used at every position.
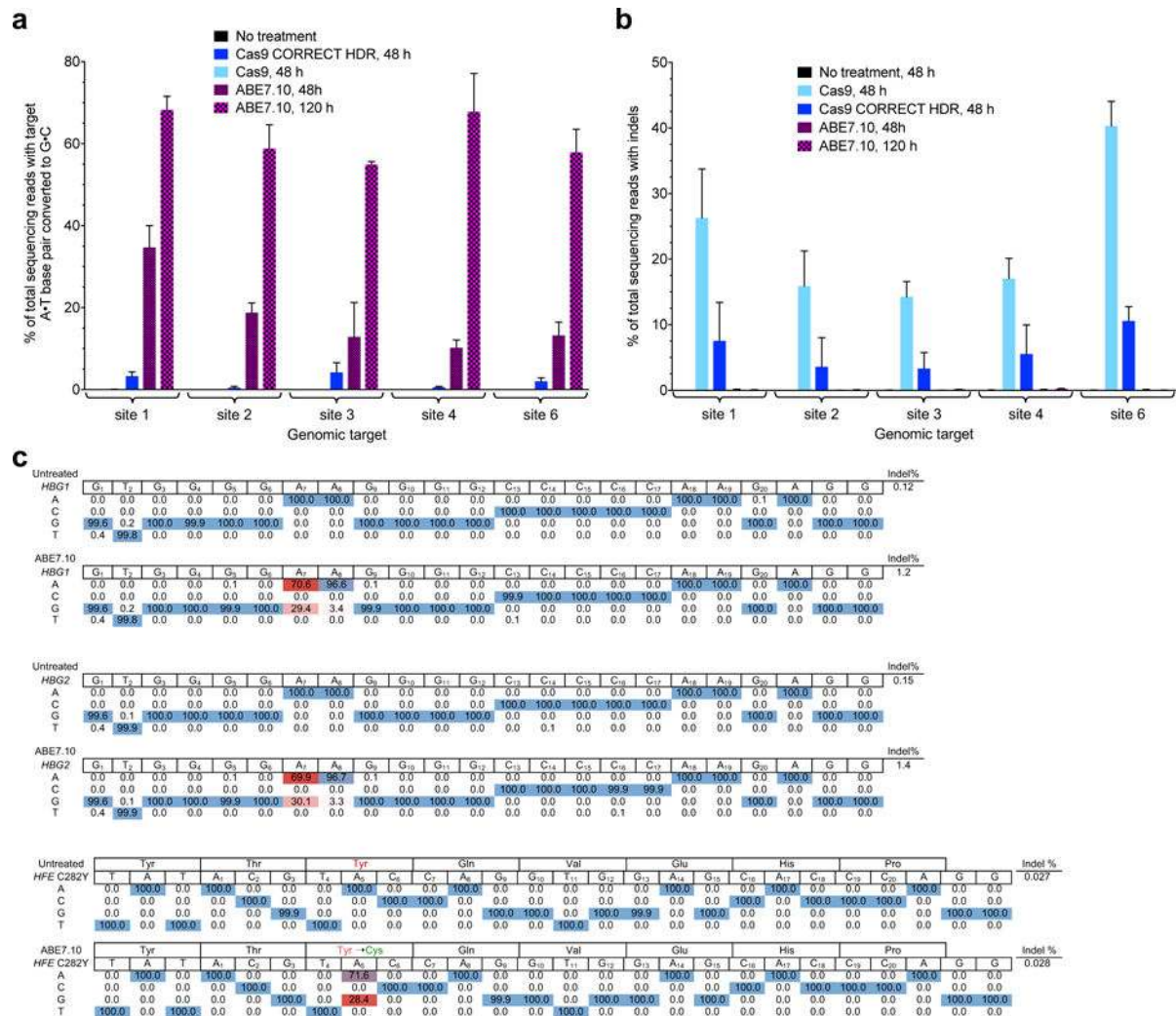
Figure 5. Comparison of ABE7.10-mediated base editing and Cas9-mediated HDR, and application of ABE7.10 to two disease-relevant SNPs

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[33]) 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 A is at protospacer positon 7. Bottom: ABE7.10-mediated reversion of the C282Y mutation in the *HFE* gene in LCL cells. The target A is at protospacer position 5.