



Published in final edited form as:

Nat Biotechnol. 2019 September ; 37(9): 1070–1079. doi:10.1038/s41587-019-0193-0.

Continuous evolution of base editors with expanded target compatibility and improved activity

Benjamin W. Thuronyi^{1,2}, Luke W. Koblan^{1,2,†}, Jonathan M. Levy^{1,2,†}, Wei-Hsi Yeh^{1,2,3}, Christine Zheng², Gregory A. Newby^{1,2}, Christopher Wilson^{1,2}, Mantu Bhaumik⁴, Olga Shubina-Oleinik⁵, Jeffrey R. Holt⁵, David R. Liu^{1,2,6,*}

¹Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA.

²Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts, USA.

³Program in Speech and Hearing Bioscience and Technology, Harvard Medical School, Boston, Massachusetts, USA.

⁴Department of Neurology, F.M. Kirby Neurobiology Center, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA.

⁵Departments of Otolaryngology and Neurology, F.M. Kirby Neurobiology Center, Boston Children's Hospital and Harvard Medical School, Boston, Massachusetts, USA.

⁶Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts, USA.

Base editors use DNA-modifying enzymes targeted with a catalytically impaired CRISPR protein to precisely install point mutations. Here we develop phage-assisted continuous evolution of base editors (BE-PACE) to improve their editing efficiency and target sequence

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

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

Present address: B.W.T.: Department of Chemistry, Williams College, Williamstown, Massachusetts, USA.

†These authors contributed equally

Author Contributions

B.W.T. designed the research, designed and constructed plasmids, and performed PACE and bacterial experiments. J.M.L. and L.W.K. designed and performed HEK cell experiments and analyzed data. J.M.L. designed and performed *APOE* editing experiments. W.H.Y. designed and performed barigo editing experiments. C.Z. constructed plasmids and performed bacterial experiments for selection development. G.A.N. designed and constructed plasmids for the HEK cell experiment for WFS1 editing. C.W. designed and performed ancestral sequence reconstruction. M.B., O.S.-O., and J.R.H. contributed *barigo* mouse cells. D.R.L. designed and supervised the research. B.W.T., L.W.K., and D.R.L. wrote the manuscript. All authors contributed to editing the manuscript.

Competing Financial Interests Statement

D.R.L. is a consultant and co-founder of Beam Therapeutics, Editas Medicine, and Pairwise Plants, companies that use genome editing. D.R.L., B.W.T. and C.W. have filed patent applications on aspects of this work.

Statistics. No statistical tests were used.

Life Sciences Reporting Summary.

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

Data Availability Statement.

Key plasmids from this work will be available from Addgene (depositor: David R. Liu) and other plasmids are available upon request. All unmodified reads for sequencing-based data in the manuscript will be available from the NCBI Sequence Read Archive, accession number PRJNA511456. Figs. 4b, 5, 6, Supplementary Table 3, Supplementary Figs. 8–14, 16, and 18–22 are based on processing of sequencing data. Protein sequences used for Supplementary Fig. 17 are supplied as Supplementary Data 1.

compatibility. We used BE-PACE to evolve new cytosine base editors (CBEs) that overcome target sequence context constraints of canonical CBEs. One evolved CBE, evoAPOBEC1-BE4max, is up to 26-fold more efficient at editing GC, a disfavored context for wild-type APOBEC1 deaminase, while maintaining efficient editing in all other sequence contexts tested. Another evolved deaminase, evoFERNY, is 29% smaller than APOBEC1 and edits efficiently in all tested sequence contexts. We also evolved a CBE based on CDA1 deaminase with much higher editing efficiency at difficult target sites. Finally, we used data from evolved CBEs to illuminate the relationship between deaminase activity, base editing efficiency, editing window width, and byproduct formation. These findings establish a system for rapid evolution of base editors and inform their use and improvement.

Genome editing has revolutionized the life sciences and entered clinical trials to treat genetic diseases.¹ The use of programmable nucleases to generate double-stranded DNA breaks (DSBs) followed by homology-directed repair can introduce a wide variety of modifications but is inefficient in non-dividing cells, and is typically accompanied by an excess of unwanted insertions and deletions (indels), translocations, or other chromosomal rearrangements.² Base editing directly modifies target DNA bases in living cells and has become widely used to correct or install point mutations in organisms ranging from bacteria to human embryos.³ Base editors use a catalytically impaired Cas9 to open a single-stranded DNA loop at a specified genomic site (Fig. 1a). Bases within the editing window (typically ~5 nt wide) in this region are modified by a tethered base-modification enzyme that only accepts single-stranded DNA. Two classes of base editors have been developed to date: cytosine base editors (CBEs) convert C•G to T•A, and adenine base editors (ABEs) convert A•T to G•C.³

CBEs such as BE3,⁴ BE4,⁵ and BE4max (a state-of-the-art CBE)⁶ use the cytidine deaminase APOBEC1 to modify target Cs.^{4,7} In cells that express them robustly, CBEs can edit some sites with high (~50%) efficiency,^{6,8} but perform poorly at others. One factor that limits the most commonly used CBEs is the native sequence context preference of APOBEC1, which deaminates GC motifs poorly.^{4,9} While a GC target positioned in the center of the editing window may be edited efficiently by APOBEC1-based CBEs,⁹ other GCs are not.^{4,10} CBEs incorporating different cytidine deaminases can edit GC targets more efficiently; for example, a CBE based on the CDA1 deaminase⁷ edited GC₃ in the HEK3 site (numbering shown in Fig. 1a) more efficiently than the corresponding APOBEC1-based CBE (20% vs. 2%).⁵ Non-APOBEC1 CBE alternatives, however, showed lower average performance than APOBEC1-CBE across a variety of targets in human cells.⁵

Cytosine base editing would therefore benefit from new CBE variants that edit with high efficiency regardless of target sequence context. Such CBEs would be especially useful for applications that involve challenging target sites or cell types, multiplexed base editing, or large-scale screening.¹¹ Moreover, a general platform to tailor base editor properties would enable the development of editors ideally suited for specific applications, which differ widely in their requirements for efficiency, sequence compatibility, and tolerance for unwanted editing.

Given the complexity of CBEs, we envisioned harnessing phage-assisted continuous evolution (PACE)^{12,13} to generate editors with improved target sequence context compatibility and higher activity. PACE performs dozens of generations of mutation, selection, and replication per day and has facilitated dramatic alterations of protein function.^{13–24} Here we describe a PACE system for evolving base editors (BE-PACE), and its application to generate CBEs with high editing activity on both GC and non-GC targets.

Results

Development of a genetic circuit that responds to base editing

During PACE, an activity of interest is coupled to the propagation of M13 bacteriophage that encode a biomolecule with that activity (Fig. 1b). To achieve this coupling, the desired activity—here, cytosine base editing—is linked to the expression of gene III, which is required to produce infectious progeny phage, using a genetic circuit encoded in *E. coli* host cells. In BE-PACE, this circuit must be activated by a single-base change, respond to small numbers of editing events, and turn on rapidly enough to support phage propagation under continuous dilution. To meet these requirements, we designed a gene circuit in which cytosine deamination occurs on the transcription template strand to revert an inactivating mutation in a protein (Fig. 2a). This design allows a transcription-level response independent of DNA replication and repair, since *E. coli* RNA polymerase accepts uracil-containing templates.²⁵ To amplify editing signal, we chose T7 RNA polymerase (T7 RNAP) as the target gene for editing and placed expression of gene III and a luciferase reporter under control of a T7 promoter. We inactivated T7 RNAP by fusing a proteolytic degradation tag (degron) to its C-terminus,²⁶ which targets the enzyme for degradation and disrupts its folding and catalytic mechanism (Supplementary Fig. 1).²⁷ The degron is linked to T7 RNAP through a TGG tryptophan codon (template strand CCA) such that deamination of either or both template strand cytidines results in a W884STOP codon substitution in mRNA that removes the degron. This design allows the bases upstream of the target to be freely varied, changing the sequence context for deamination (Fig. 2a).

We implemented this circuit using two plasmids, one encoding T7 RNAP and the other encoding the gIII-luciferase operon and a guide RNA targeting the T7 RNAP C-terminus. We began with a GTCCA editing target to match the native TC context preference of APOBEC1 and to minimize selection stringency. We tested the circuit using plasmid-encoded BE2 in a luciferase reporter assay (Fig. 2b). Activation was dependent on all components of the system, including fusion of APOBEC1 to dCas9. The circuit's signal amplification, and thus selection stringency during PACE, can be tuned by adjusting the expression of T7 RNAP. We assayed a series of ribosome binding site and promoter strengths and chose a combination that provided strong circuit activation (>500-fold) (Supplementary Fig. 2). Despite the strong activation observed with plasmid-encoded BE2, phage encoding BE2 (Supplementary Discussion 1) did not measurably activate the circuit within 3 h (Supplementary Fig. 3). A discrete overnight phage propagation assay (Fig. 2c) showed that although the circuit was competent to propagate phage and was dependent on base editing activity, phage encoding BE2 did not enrich strongly enough to support PACE, which typically requires 10-fold enrichment in this assay.

Split base editor phage support an optimized PACE selection

To achieve robust, base editing-dependent phage propagation, we increased the copy number of the gIII-containing plasmid. We also cloned BE2 into an evolved phage backbone (“generation 2”) containing 41 mutations that previously emerged from extensive PACE¹⁴ and observed that reporter gene expression increased (Supplementary Fig. 4) and BE2 phage propagation improved by as much as 100-fold (Supplementary Table 1). However, these phage still enriched <10-fold.

We speculated that the large size of BE2 phage might be impeding their propagation. While the M13 capsid can be extended to accommodate an arbitrarily large quantity of DNA,²⁸ phage infectivity or replication might be influenced by genome size.²⁹ To address this possibility, we designed a modified scheme in which only part of the base editor is encoded by phage, and the remainder is supplied by a host-cell plasmid, with each half fused to a fast-splicing split intein³⁰ such that full-length base editor protein forms after infection (Fig. 3a).

We tested two base editor intein splits. In split A, the phage encode the deaminase and its downstream linker, while dCas9 and UGI are supplied by the host cell. In split B, the deaminase and part of dCas9 is encoded in the phage, and the rest of dCas9 and UGI are provided by the host cell. Both versions produced active base editor in an intein- and guide RNA-dependent manner, but split A led to faster circuit activation (Fig. 3b and Supplementary Table 1). Combining split A with the other optimizations (referred to hereafter as the BE-PACE circuit) resulted in ~10-fold overnight phage propagation in a host-cell culture and >1,000-fold selectivity for base editor phage (Supplementary Table 1).

Importantly, the BE-PACE circuit can be used to assess deaminase kinetics in cells in a CBE context based on expression of the luciferase reporter. Because the circuit responds to deaminated cytosine directly *via* transcription, the activation rate of the circuit as measured by the rate of change of luminescence over time should reflect the rate of cytosine deamination.

We tested BE-PACE using the low-stringency selection circuit (TCC1, Supplementary Table 2) in a competitive continuous propagation experiment (Fig. 3c). A small amount of APOBEC1–intein phage was seeded along with a large excess of phage encoding red fluorescent protein (RFP). The phage titer initially plunged as the non-replicating RFP phage were diluted out of the lagoon, but by 45 h, the titer of APOBEC1–intein phage recovered and no trace of RFP phage was detected by PCR. These results validate CBE activity-dependent phage propagation during BE-PACE.

BE-PACE generates improved deaminases

To address the sequence context limitations of APOBEC1—which strongly favors TC and disfavors GC targets^{4,9,31}—we constructed a low-stringency circuit (GCC1, Supplementary Table 2) with an AGC₄C₅A target that requires editing of a GC to support maximal phage propagation. Initial activity of APOBEC1–intein phage as measured by luciferase assay on this GC target was negligible. Therefore, we sought to first increase overall APOBEC1 activity through PACE on a series of TC₄C₅ circuits with increasing stringency (Fig. 4a,

Supplementary Discussion 2, and Supplementary Fig. 5). After successive PACE on the TCC1 and TCC2 circuits, we isolated mutant phage that showed weak but measurable activity on the GCC1 circuit. Several mutations, including predominant A165S and F205S substitutions, were similar to mutations that arose during the PACE of APOBEC1 to maximize its soluble expression.²³ Further PACE on either GCC1 or TCC3 led to top-performing phage clones that showed up to 28-fold improvements in apparent activity when tested on the GCC1 circuit by luciferase assay (Supplementary Fig. 6).

The GCC1 and TCC3 populations independently evolved H122L and D124N. To evaluate the contributions of these and other mutations to GCC editing activity we cloned evolved deaminases into a standardized phage backbone (“generation 3”) isolated after 210 h of PACE and tested them on circuits containing all four ANCCA targets in the luciferase assay (Fig. 4b). The results for wild-type APOBEC1 recapitulated its known sequence context preferences.³¹ Results for the PACE-evolved deaminases show that the D124N and especially H122L mutations dramatically improve activity on non-TC targets, reducing the TCC/GCC activity ratio from 27-fold in wild-type APOBEC1 to between 1.9 and 0.9 in evolved clones.

Ancestral sequence reconstruction of deaminases⁶ could provide promising starting points for BE-PACE. We chose five ancestral sequences from nodes within our APOBEC phylogeny⁶ (Supplementary Fig. 7) and constructed corresponding deaminase–intein phage. We excluded from these sequences the N- and C-termini of rat APOBEC1, which have low sequence similarity to other APOBEC deaminases and are implicated in functions irrelevant to base editing.^{32,33} After subjecting a mixture of all five genotypes to BE-PACE on the TCC1 circuit, (Fig. 4a and Supplementary Fig. 5) we obtained improved phage clones from the ancestral node 656 sequence, which we named “FERNY” for its five N-terminal amino acids (Supplementary Fig. 7). We further evolved FERNY–intein phage on the TCC3 circuit (Fig. 4a), resulting in H102P and D104N mutations at positions corresponding to H122 and D124 in APOBEC1. These mutations substantially improved apparent activity on the GCC target (Fig. 4b), further implicating these positions as sequence compatibility determinants.³⁴

Wild-type CDA1–intein phage showed much higher starting activity on both TCC and GCC targets in the luciferase assay than APOBEC1–intein phage (Fig. 4b), in agreement with yeast mutational assays.³⁵ We sought to improve CDA1 activity further through three stages of BE-PACE with increasing stringency (Fig. 4a and Supplementary Fig. 5), resulting in conserved mutations including A123V. All tested variants exhibited increased apparent activity by luciferase assay (Fig. 4b).

Evolved deaminases improve base editing in mammalian cells

To determine whether the apparent activity improvements in the luciferase assay translated into improved mammalian cell base editing, we subcloned a panel of evolved deaminase variants into the BE4max architecture⁶ and transfected them into HEK293T cells along with guide RNAs targeting five genomic sites previously shown to undergo efficient editing. Under optimal plasmid dosing and conditions (Fig. 4b, Supplementary Figs. 8–10, and

Supplementary Table 3), we observed that editing efficiency at the center of the activity window reaches a maximum of ~60–80% (“plateau levels”), likely limited by CBE-independent factors such as transfection efficiency or cellular DNA repair processes (Discussion). Notably, editing at positions away from the center of the activity window was improved for all evolved BEs, and editing values at these positions correlated with luciferase assay activity (Supplementary Fig. 11).

Among APOBEC1 CBEs, evolved mutations including H122L and D124N resulted in a striking improvement in base editing of GC targets. For example, editing of GC₃ at the HEK3 site rose from very low ($2.3 \pm 0.42\%$) to plateau levels ($58 \pm 3.7\%$), a 25-fold improvement; editing of GC₃ at HEK4 increased from $5.0 \pm 0.41\%$ to $64 \pm 5.2\%$; and editing of GC₈ at HEK4 increased from $12 \pm 2.1\%$ to $58 \pm 5.6\%$. The unevolved FERNY-CBE exhibited higher activity on GC targets than APOBEC1-CBE (e.g. 20% at HEK3 GC₃, Fig. 4b), and the H102P D104N double mutant CBE achieved plateau levels at GC₃ ($70 \pm 4.8\%$, Fig. 4b).

In the BE4max context, the wild-type CDA1-CBE showed a much wider editing window than was reported for Target-AID⁷ (which places CDA1 C-terminal to Cas9), resulting in plateau levels of editing ($63 \pm 9.2\%$) across protospacer positions 3–8 (Fig. 4b). Evolved CDA1 CBEs showed further broadening of the editing window to include positions such as HEK3 AC₉, EMX1 GC₁₀, HEK2 GC₁₁, and RNF2 TC₁₂.

Based on these results, we selected one high-performing evolved variant of each deaminase to characterize in depth: evoAPOBEC1 (clone 330–1), evoFERNY (164–1), and evoCDA1 (184–1). We created fully codon-optimized⁶ BE4max variants and tested their editing activity using optimal dosing levels (Fig. 5a) and in a dose titration experiment (Supplementary Fig. 12). We also characterized a panel of 24 CBE variants to dissect the roles of evolved mutations, confirming the role of H122 and D124 mutations in GC activity of the evolved APOBEC-family CBEs, and finding additive improvements from the three mutations in evoCDA1-BE4max (Supplementary Discussion 3 and Supplementary Figs. 13–16). Interestingly, the critical H122L D124N mutations evolved during APOBEC1 PACE are present in <1% of 1,189 naturally occurring APOBEC sequences and often co-occur (Supplementary Fig. 17).

EvoAPOBEC1-BE4max dramatically outperforms the state-of-the-art CBEs BE4max and AncBE4max at GC targets, while showing similar or higher activity at non-GC targets (Fig. 5a). The base editing window of evoAPOBEC1-BE4max is very similar to that of BE4max (Fig. 5b). Despite its 29% smaller size compared to APOBEC1, FERNY performs comparably to APOBEC1 in CBEs at non-GC targets and more effectively on GC targets (Fig. 5a). EvoFERNY-BE4max shows further improvement on GC targets, giving similar or higher editing levels compared to evoAPOBEC1-BE4max (Fig. 5a).

The editing windows of CDA1-BE4max and evoCDA1-BE4max are very different from that of BE4max (Fig. 5b). Half-maximal editing efficiency, which spans protospacer positions 3–8 for both BE4max and evoAPOBEC1-BE4max, covers positions 1–9 for CDA1-BE4max and expands to positions 1–13 for evoCDA1-BE4max (Fig. 5b). EvoCDA1-BE4max showed

editing at HEK3 GC₁₄ and RNF2 TC₁₂, target positions at which all other CBEs resulted in almost no activity (Supplementary Fig. 12). However, all editors, including evoCDA1-BE4max, mediated nearly identical editing levels at protospacer positions 4–6 at each dose.

Recently, APOBEC3A (A3A) has been reported to support high-activity cytosine base editing in a BE3 architecture^{9,36,37}. CBEs containing either wild-type A3A or the W98Y mutant were shown to edit a GC₄ target with higher efficiency than BE3.⁹ We constructed BE4max variants with the A3A and A3A-W98Y deaminases (these showed mutagenicity in *E. coli*; see Supplementary Discussion 4) and compared them with the evolved CBEs (Supplementary Fig. 18). A3A-BE4max generally showed properties similar to evoCDA1-BE4max, with plateau editing across a wide window and efficient editing of GC targets. We did not observe any effect from the W98Y mutation (Supplementary Fig. 18). These features of A3A-BE4max establish that it is complementary to, but distinct from, the GC-compatible evolved CBEs with standard editing window widths.

Evolved base editors greatly outperform previous base editors on disease-relevant targets

Although efficiently editable target sites in cooperative cell lines are useful for CBE characterization, evolved CBEs will be most useful for applications targeting sites or cell types that are difficult to edit. To assess the potential impact of PACE-evolved CBEs for these more challenging applications, we used evolved CBEs to edit three disease-relevant sites in primary cells and cell lines (Supplementary Table 4).

First, we tested correction of a pathogenic transmembrane channel-like 1 (TMC1) point mutation that causes recessive hearing loss.³⁸ In the *baringo* mouse, a Y182C mutation caused by an A to G mutation in exon 8 of *TMC1* causes profound deafness.³⁸ We nucleofected primary embryonic fibroblasts from *baringo* mice with plasmids encoding CBEs and a guide RNA targeting *TMC1* GC₈ (Fig. 6a). BE4max showed low, variable editing of GC₈, generating 0.05 to 9.6% alleles with the C182Y conversion and no other amino acid changes (“desired alleles”; Supplementary Tables 4 and 6), while CDA1-BE4max generated 9.1±0.8% desired alleles. Surprisingly, despite their improved activity on edge-of-window GC targets, evoAPOBEC1-BE4max and evoFERNY-BE4max also resulted in low conversion to desired alleles (3.3±1.0% and 7.2±0.7%) at this in-window target, while FERNY-BE4max performed better (17.0±1.7%). By contrast, evoCDA1-BE4max and A3A-BE4max gave the highest conversion to desired alleles (24±5.7% and 36±3.6%) which included silent bystander mutations across wide windows (Fig. 6a, Supplementary Tables 3 and 4). EvoCDA1-BE4max showed a 2.6-fold improvement over CDA1-BE4max. Notably, A3A-BE4max, but no other CBEs, showed editing far outside the protospacer, with up to 7% C-to-T conversion at GC₋₁₂ and CC₋₁₁ (Supplementary Table 3), suggesting a trade-off between activity and a well-defined editing window.

Second, we examined whether our evolved variants could convert *APOE4* (R112 R158) alleles, associated with high Alzheimer’s disease risk, to risk-neutral *APOE3* (C112 R158) or protective *APOE2* (C112 C158) alleles.³⁹ Such conversion requires editing to install R112C, which has not been achieved efficiently due to the GC target context and the lack of an appropriately positioned canonical protospacer-adjacent motif (PAM). We nucleofected immortalized *APOE4* mouse astrocytes⁴ with plasmids encoding CBEs and guide RNAs

targeting either R112 or R158 (Fig. 6b). The R158C edit (GC₅) was made efficiently (39–77%) by all tested CBEs. However, CBEs with wide activity windows also edited the non-silent GC₁₂, so that the most efficient conversion to desired alleles was from FERNY- and evoFERNY-BE4max (59±7.4% and 61±9.4%). To target R112 efficiently, we used CBEs incorporating an engineered SpCas9-NG variant (BE4max-NG) and a guide RNA with an AGT PAM.⁴⁰ EvoFERNY-BE4max-NG gave high conversion (41±8.5%) to desired alleles, as did several CBEs, while wide-window editors (evoCDA1-BE4max-NG and A3A-BE4max-NG) gave high editing of the target GC₅, but also editing of GC₁₁ (non-silent) or even GC₁₄ (silent) (Fig. 6b). Using a TGG PAM, evoCDA1-BE4max and A3A-BE4max gave the highest conversion to desired alleles, with evoCDA1-BE4max outperforming CDA1-BE4max by 14-fold (Fig. 6b). Editing levels were low for the other CBEs. EvoAPOBEC1-BE4max could also edit R112 using a GC₆ guide RNA with the non-canonical CAG PAM, outperforming BE4max by 4.8-fold (Supplementary Fig. 19).

Finally, we generated a HEK293T cell line model for Wolfram syndrome 1, an autosomal recessive disease caused by mutations in *WFS1*.⁴¹ Clean installation of the Wolfram syndrome-associated *WFS1* Q668STOP mutation⁴² requires editing of GC₄ without concomitant non-silent editing of GC₁. EvoAPOBEC1-BE4max and evoFERNY-BE4max resulted in 25±9.8% and 24±13% conversion to the clean Q668STOP allele, respectively, compared to 7.8±2.8% for BE4max (Fig. 6c). The wider editing windows of CDA1-BE4max and evoCDA1-BE4max led to nearly identical editing of GC₄ and the bystander GC₁ (Fig. 6c), resulting in only 1.0±0.45% and 0.61±0.18% conversion to the clean Q668STOP allele respectively. A3A-BE4max showed an even wider editing window that included GC₁₄, giving 10±2.5% conversion to the clean Q668STOP allele.

The evolved CBEs thus showed greatly improved editing outcomes relative to previously reported state-of-the-art CBEs at all three disease-relevant targets, including in primary fibroblasts and immortalized astrocytes. These results highlight the improved GC editing and narrow windows of evoAPOBEC1-BE4max and evoFERNY-BE4max, as well as the value of a panel of CBEs, which maximizes the likelihood of finding a CBE that suits the requirements of a given application.

Off-target base editing and indel formation in mammalian cells by evolved CBEs

Changes in deaminase properties may affect off-target as well as on-target editing. We examined editing in HEK293T cells by evolved CBEs at several known Cas9 off-target binding sites^{43–45}, each having 2–4 mismatches with their respective guide RNAs (Fig. 5c and Supplementary Fig. 20). Like BE3 and BE4,^{4,44} BE4max and AncBE4max edited these off-target sites at a range of efficiencies, from ~40–55% (HEK4-OT1 and HEK4-OT4) to <0.2% (HEK2-OT2). The off-target profiles of evoAPOBEC1-BE4max and evoFERNY-BE4max resembled that of BE4max, other than showing increased editing in GC contexts.

CDA1-BE4max exhibited higher off-target editing than BE4max, giving measurable editing at all sites tested (Fig. 5c). EvoCDA1-BE4max showed still higher editing, up to 12-fold more than CDA1-BE4max among Cs in protospacer positions 1–9, and ranging from 0.3% to >40% editing at off-target sites for which BE4max editing was undetectable (Fig. 5c and Supplementary Fig. 20). In agreement with a previous report,⁹ we found that A3A-BE4max

also exhibits high off-target editing activity (Supplementary Fig. 20). These observations support the expectation that higher activity deaminases lead to higher off-target editing.

We examined indel formation by CBEs at all tested sites. We observed divergent behavior between the well-edited sites in HEK293T cells and the poorly edited sites in other cell types. At many, but not all, well-edited sites, evolved CBEs generated higher levels of indels compared to starting CBEs (Supplementary Fig. 21). BE4max generated $2.6 \pm 0.52\%$ indels across all five sites, while evoAPOBEC1-BE4max and evoFERNY-BE4max generated between 2.5 and 15% indels. While CDA1-BE4max generated 0.8 to 8.7% indels, evoCDA1-BE4max generated 6.8 to 20% indels, and A3A-BE4max showed similar levels to evoCDA1-BE4max (Supplementary Fig. 21).

By contrast, at the TMC1, APOE, and WFS1 disease-relevant sites, indels were much lower ($<1.1\%$ for evoAPOBEC1-BE4max, $<3.9\%$ for evoFERNY-BE4max, and 3.8 to 9.1% for evoCDA1-BE4max). The ratio of in-window editing:indels was generally similar for evolved and starting CBEs at these sites (Supplementary Fig. 21) and at poorly-edited off-target sites (Supplementary Table 3). Together, these observations suggest that deaminase activity correlates with both indel rate and editing efficiency, but indels are not subject to the same limits as editing.

Discussion

We used BE-PACE to evolve three new families of base editors with enhanced sequence context compatibility, higher activity, and broadened editing windows. The suitability of a CBE for a particular application depends on that application's requirements for efficiency, specificity, and indel formation. Generating and testing a panel of state-of-the-art CBEs therefore offers the best chance of success. For sites in which a standard editing window (approximately five bases wide) is desirable, we recommend evoAPOBEC1-BE4max and evoFERNY-BE4max over existing CBEs. These editors show no apparent bias against GC targets, and edit at least as efficiently as BE4max or AncBE4max at all sites tested in this study. EvoFERNY-BE4max edits GC motifs with slightly higher efficiency than evoAPOBEC1-BE4max (Fig. 6a, Fig. 4b) and contains a deaminase that is 29% smaller, while evoAPOBEC1-BE4max shows slightly higher activity at non-GC targets (Fig. 5a). EvoFERNY-BE4max may prove especially useful for viral delivery applications constrained by payload size. EvoAPOBEC1- and evoFERNY-BE4max did not exhibit significant changes in editing window width or off-target editing compared to BE4max. We note that for some sites in which bystander bases are in a GC sequence context, BE4max (or, for TC targets, eA3A-BE3⁹) may be preferable to CBEs with unbiased sequence context preferences.

We recommend evoCDA1-BE4max for carefully chosen applications suited to its characteristics. At well-edited sites, this CBE generates higher indel levels without increasing editing beyond plateau levels, instead showing an expanded window. By contrast, at poorly edited sites, it shows increased in-window editing levels without increased indels. In all cases, evoCDA1-BE4max shows increased off-target editing. These considerations suggest that evoCDA1-BE4max (and other high-activity CBEs such as A3A-BE4max,

Supplementary Discussion 5) should be applied when off-target and bystander editing are not concerns and high efficiency is paramount.

One set of such applications is high-throughput screening with base editing,³ which typically relies on a single CBE to efficiently edit many target sites. Bystander and off-target edits are unlikely to invalidate loss-of-function screening hits. High-activity CBEs are also well-suited for plant genome editing³, a field that emphasizes phenotypic screening over genotyping and has historically relied on untargeted mutagenesis.^{46,47} Finally, high-activity CBEs such as evoCDA1-BE4max provide an option to achieve otherwise inaccessible on-target editing levels for a difficult-to-edit site or cell type.

The findings in this study inform a conceptual model of cytosine base editing in mammalian cells (Fig. 6d, Supplementary Discussion 6, and Supplementary Fig. 22). Previous studies demonstrated that increases in base editor protein expression lead to more efficient editing across all window positions.^{6,8} We initially hypothesized that improvements to deaminase activity would have similar effects. Instead, evolved CBEs showed unexpected behavior that diverged between two classes of target sites. For the first class—sites already known to be edited with high efficiency by unevolved CBEs (including HEK2, HEK3, HEK4, EMX1, RNF2, HEK4-OT4, APOE R158, APOE 112/AGT, and WFS1)—evolved CBEs showed no improvement in plateau editing levels (~60–80%) but exhibited broader editing windows (across all sequence contexts for evoCDA1, but only on GC targets for evoAPOBEC1 and evoFERNY) (Fig. 4b, 5a, Supplementary Fig. 12). For the second class—sites that showed poor editing by unevolved CBEs (including TMC1, APOE R112/TGG, APOE R112/CAG and off-target sites other than HEK4-OT4)—evolved CBEs instead showed improved editing at peak window positions and narrower windows (Fig. 6, Supplementary Fig. 20). We refer to these two classes as “well-edited sites” and “poorly edited sites.”

Our observations suggest that several cell-type, cell-state, site-specific, and deaminase factors contribute to editing efficiency. The rate of deamination at a given target site is determined by factors including the ability of Cas9 to generate and maintain an R-loop, the accessibility and effective concentration of each position in the R-loop with respect to the deaminase (summarized as “exposure of target C to deaminase” in Fig. 6d), and the sequence context-dependent K_m and k_{cat} of the deaminase for each position in the R-loop (summarized as “deaminase activity on the target sequence” in Fig. 6d). While target base deamination once the R-loop is formed is controlled by the properties of the fused deaminase enzyme, editor-extrinsic factors control the fate of the deaminated bases and consequently limit the efficiency of base editing (horizontal green lines in Fig. 6d).

These factors explain the different behavior of CBEs at well-edited versus poorly edited sites (Fig. 6d). At well-edited sites, CBEs with higher-activity deaminases (such as evoCDA1) are able to deaminate cytidines at a broader range of protospacer positions—even those with lower intrinsic accessibility—during each Cas9 binding event. This ability leads to a wider window of plateau editing levels across more positions. Our group previously observed the converse effect, in which mutations that reduce the catalytic activity of APOBEC1 lead to activity window narrowing without reducing editing at peak window positions.¹⁰ The fact that even lower-activity deaminase CBEs achieve plateau editing levels for some protospacer

positions suggests that at well-edited sites, most Cas9 domain binding events already lead to successful conversion of highly accessible bases. Further increases in deaminase efficiency do not improve editing at such positions (since factors other than deaminase activity limit product formation) but can increase flux through other pathways such as off-target base editing, bystander editing, or indel formation (Supplementary Fig. 21).

Cytidines at positions with low accessibility do not show plateau editing levels because CBEs do not deaminate them during every Cas9 binding event. Since deamination frequency, rather than cellular repair, limits editing at such positions, the sequence context preferences of the deaminase contribute to observed editing outcomes and can control which positions are efficiently edited.^{4,9} Thus, the strong context preferences of APOBEC1 give BE4max different editing windows for different sequence contexts: positions 3–8 for non-GC targets, but positions 5–7 for GC targets (Figs. 5a, 6b, and 6c). By contrast, evoAPOBEC1-BE4max has an activity window of positions ~3–8 across sequence contexts (Fig. 5b and 6).

Poorly edited sites show entirely different properties from well-edited sites. Our model suggests that these differences arise because cytosines in poorly edited sites are less available for deamination (Fig. 6d, top blue curve). Because deamination events do not occur frequently even at the peak accessibility position, editing never reaches plateau levels, and sequence context-dependent deaminase activity influences editing outcomes across all protospacer positions (Fig. 6d, bottom blue curves; see Fig. 6a).

Our model suggests further lines of investigation to illuminate and improve base editing outcomes, including identification of cellular repair processes that lead to plateau levels of editing, and characterization of the factors that distinguish well-edited and poorly edited sites. We anticipate that the base editor PACE platform described here will continue to diversify the repertoire of base editors that expand the application scope of base editing and deepen our understanding of the determinants of base editing outcomes.

Online Methods

General methods and molecular cloning.

Antibiotics (Gold Biotechnology) were used at the following working concentrations: carbenicillin, 50 µg/mL; spectinomycin, 50 µg/mL; chloramphenicol, 40 µg/mL; kanamycin, 30 µg/mL. HyClone water (GE Healthcare Life Sciences) was used for PCR reactions and cloning. For all other experiments, water was purified using a MilliQ purification system (Millipore). Q5 Hot Start High-Fidelity 2× Master Mix (New England BioLabs) was used for diagnostic PCRs and Phusion U Hot Start DNA polymerase (Thermo Fisher Scientific) was used for all other PCRs.

Plasmids were cloned by USER assembly⁴⁹ or Golden Gate assembly⁵⁰. For USER cloning, 10–30 °C melt temperature junctions were used and constructs assembled by digesting at 37 °C for 15 min followed by transformation into chemically competent cells. For Golden Gate assembly, LguI (SapI isoschizomer, Life Technologies) or BsaIHFv2 (New England BioLabs) were used as the type IIS restriction enzymes along with T4 DNA ligase (New

England BioLabs). Overhangs for BsaI cloning were selected from a validated set.^{51,52} Typical assemblies contained final concentrations of ~0.5–2 ng kb⁻¹ μL⁻¹ plasmids, with a ~2:1 ratio of donor to acceptor plasmids. Assemblies were either thermally cycled (~3–5 min 37°C, 3–5 min 16 °C for 10–50 cycles) or isothermal at 37 °C⁵² for between 30 min and 18 h followed by transformation into chemically competent cells.

Synthetic genes were obtained as gBlock gene fragments from Integrated DNA Technologies. Codon-optimized sequences for human cell expression were obtained from Genscript. Plasmids were cloned and amplified using TOP10 cells (Thermo Fisher Scientific). Plasmid DNA was isolated using the Qiagen Spin Miniprep Kit according to manufacturer instructions. All constructs assembled using PCR were fully sequence-verified using Sanger sequencing (Quintara Biosciences), while constructs assembled using Golden Gate cloning were sequence-verified across all assembly junctions. A full list of plasmids used in this work is given in Supplementary Table 5. Protospacer sequences for guide RNA plasmids are described in Supplementary Table 4.

Constructs were designed using ribosome binding site (RBS) series⁵³ and insulated promoter series⁵⁴ and gene-specific RBSs were designed using the Ribosome Binding Site calculator (www.denovodna.com).⁵⁵ Guide RNA protospacer sequences for *E. coli* experiments were checked to avoid *E. coli* toxicity.⁵⁶

Preparation and transformation of chemically competent cells and strain storage.

Strain S2060²⁰ was used in all experiments, including luciferase assays, phage propagation tests and PACE. Chemically competent cells were prepared essentially as described.⁵⁷ An overnight culture was diluted 50-fold into LB media (United States Biologicals) and grown at 37 °C with shaking at 230 r.p.m. to OD₆₀₀ ~0.3–0.6. Cells were cooled on ice and pelleted by centrifugation at 4,000 *g* for 10 min at 4 °C. The cell pellet was then resuspended by gentle stirring in ice-cold LB media (~1 mL per 10 mL culture volume) and an equal volume of 2x TSS (LB media supplemented with 10% v/v DMSO, 20% w/v PEG 3350, and 40 mM MgCl₂) was added. The cell suspension was mixed thoroughly by inversion, aliquoted and frozen in liquid nitrogen, then stored at –80 °C until use. To transform cells, 50 μL of competent cells thawed on ice was added to a prechilled mixture of plasmid(s) (~25–50 ng for single and 100–500 ng for double transformations) and 50 μL KCM solution (100 mM KCl, 30 mM CaCl₂, and 50 mM MgCl₂ in water). The mixture was heat shocked at 42 °C for 90 s and LB media (400 μL) was added. Cells were allowed to recover at 37 °C with shaking at 230 r.p.m. for 1 h then spread on LB media with 1.5% agar (United States Biologicals) plates containing the appropriate antibiotic(s), and incubated at 37 °C for 16–18 h. To prepare –80 °C stocks, a single colony was picked into LB media containing the appropriate antibiotic(s) and grown for 16–18 h. 1 mL of this overnight culture was mixed with 333 μL sterile 60% w/w glycerol in a 2 mL screw-top tube and frozen at –80 °C.

Bacterial luciferase assays.

S2060 cells containing plasmids of interest were prepared as described above. Plasmids encoding CBEs were always freshly transformed and grown for no more than 18 h before inoculation (Fig. 2b and Supplementary Figs. 1, 2); for each biological replicate, a single

colony was picked and inoculated in Davis Rich Medium (DRM)¹³ (prepared from US Biological CS050H-001/CS050H-003) in a 96-well deep well plate (Eppendorf) fitted with porous sealing film. Strains not containing CBE-encoding plasmids (Figs. 2c, 3b and 4b, and Supplementary Figs. 3 and 6) were either freshly transformed and inoculated in the same way, or inoculated from -80°C stocks. For each biological replicate prepared from a single -80°C stock, a separate overnight culture was independently inoculated. Plates were incubated at 37°C with shaking (230 r.p.m.) overnight. Cultures were diluted 50-fold into 1 mL fresh DRM with antibiotics, then grown for ~ 1.5 h (for phage-expressed base editor) or 2 h (for plasmid-expressed base editor) in deep-well plates. For phage assays, 135 μL of each host cell culture was mixed with 15 μL of high-titer phage stock ($>10^{10}$ pfu/mL) in a clear-bottom black-walled 96-well assay plate (Costar). A biological replicate constituted a host cell culture grown independently from a colony or -80°C stock mixed with a single clonal phage stock for each phage genotype tested. For plasmid-expressed base editor assays, cultures were induced with arabinose (10 mM unless otherwise indicated) and grown for a further 3h, then transferred to assay plates (150–200 μL per well). Absorbance at 600 nm (OD_{600}) and luminescence (integration time 500 ms, no attenuation) were monitored using an Infinite M1000 Pro microplate reader (Tecan) with temperature set to 37°C . For kinetic assays, readings were made every 3.5 minutes during the monitoring period and the plate was shaken for 20 s between reads (double orbital, 168 r.p.m.). OD_{600} -normalized luminescence values were obtained by dividing raw luminescence by background-subtracted 600 nm absorbance. The background value was set to the 600 nm absorbance of wells containing DRM only. For phage-based CBE assays, the slope of OD_{600} -normalized luminescence vs. time (s) was calculated by least squares linear regression over a span of at least eight time points from between 2.25 and 3 h post-infection, with the range of time points chosen such that regression gave a Pearson correlation coefficient R^2 of >0.9 for the majority of conditions having slopes of at least 0.5 luminescence units $\text{OD}_{600}^{-1} \text{ s}^{-1}$. A single time range was used to calculate slopes across all conditions within a single assay. For Fig. 3b, data for each of the four sets of host cell biological replicates were collected on the same day. For Fig. 4b (top panel), data for each of the three sets of host cell biological replicates were collected on different days.

Plaquing.

Phage were plaqued on S2060²⁰ host cells containing plasmid pJC175e (activity-independent propagation),¹³ plasmids pJC175e + pDB016¹⁵ (for negative selection against T7 RNAP activity; this plaquing strain was used routinely for plaquing samples from PACE experiments), plasmid pT7-AP¹³ (to check for the presence of T7 RNAP recombinants), or no plasmid (to check for the presence of gene III recombinant phage). To prepare a cell stock for plaquing, overnight culture of host cells (fresh or stored at 4°C for up to ~ 1 week) was diluted 50-fold in DRM containing appropriate antibiotic(s) and grown for 3–5 h at 37°C , then stored at 4°C for up to ~ 2 weeks. Serial dilutions of phage (10-fold) were made in LB medium or water. To prepare plates, molten 2xYT medium agar (1.5% agar, 55°C) was mixed with Bluo-gal (10% w/v in DMSO) to a final concentration of 0.04% Bluo-gal. The molten agar mixture was pipetted into quadrants of quartered Petri dishes (1.5 mL per quadrant) or wells of a 12-well plate (~ 1.2 mL per well) and allowed to set. To prepare top agar, DRM (30 mL) was warmed by microwave heating for 15 s and 15 mL of molten

(55 °C) 2xYT medium agar (1.5%) was added to give 45 mL top agar (0.5% agar final). Top agar was maintained tightly capped at 55 °C for up to 1 week. To plaque, cell stock (75–100 μ L) and phage (10 μ L) were mixed in 2 mL library tubes (VWR), and 55° C top agar added (400 or 900 μ L for 12-well plate or Petri dish, respectively), then the mixture was immediately pipetted (without mixing) onto the solid agar medium in one well of a 12-well plate or one quadrant of a quartered Petri dish. Top agar was allowed to set undisturbed (10 min at room temperature), then plates or dishes were incubated (without inverting) at 37 °C overnight inside an unsealed plastic bag (to prevent desiccation).

Phage propagation assays

(Fig. 2c and Supplementary Table 1). Host cells in DRM were prepared as described above for luciferase assays and grown for ~1.5 h following dilution. Cells were diluted 2-fold with previously titrated phage stocks to a final concentration of 10^6 plaque forming units per mL and a volume of 1 mL and grown overnight in a 96-well deep well plate (Eppendorf) fitted with porous sealing film. The cultures were centrifuged (3,600 *g*, 10 minutes) to remove cells and the supernatants titrated as previously described.¹⁴ Fold enrichment was calculated by dividing the titer of phage propagated on host cells by the titer of phage at the same input concentration shaken overnight in DRM without host cells.

Phage-assisted continuous evolution

(Fig. 4a and Supplementary Fig. 5). Unless otherwise noted, PACE apparatus, including lagoons, chemostats, pumps and media, were prepared and used as previously described¹⁴.

S2060 host cells containing the appropriate plasmids were freshly transformed with MP6 or DP6⁴⁸ and plated on 2xYT medium with 1.5% agar supplemented with 0.5% (w/v) glucose and appropriate antibiotics. To verify the function of the mutagenesis plasmid, single colonies were resuspended in 50 μ L DRM and serially diluted 10-fold; 1 μ L of each dilution was plated on 2xYT medium with 1.5% agar containing either 0.5% (w/v) glucose or 10 mM arabinose and incubated overnight at 37 °C. The same colony stock dilutions (10^5 - 10^8 fold) were added to DRM containing antibiotics (5 mL) in 13 mL tubes and grown overnight at 37 °C with shaking. All colony stocks routinely showed robust growth on glucose-supplemented plates and zero growth (regardless of dilution level) on arabinose-supplemented plates, indicating uniform induction of mutagenesis. The entire overnight culture with the lowest visible cell density was used to inoculate a chemostat (80 mL), which was grown to OD₆₀₀ ~0.6–1 then maintained under continuous dilution with fresh DRM at 1–1.5 volumes/h to keep cell density roughly constant. Lagoons were initially filled with DRM, then continuously diluted with chemostat culture for at least 2 h prior to seeding with phage.

In the APOBEC1 stage 1 and 2 PACE experiments, stock solution of arabinose (1 M) was pumped directly into lagoons (10 mM final) as previously described,^{14,23} with or without the appropriate concentration of aTc present in the stock solution to give the indicated final concentrations. Syringes containing aTc solution were covered in aluminum foil and work was conducted so as to minimize light exposure of tubing and lagoons. In all other PACE experiments, chemostat culture being pumped to lagoons was mixed using a Y junction with

1 M arabinose at a flow ratio of 100 volumes/h culture:1 volume/h arabinose, giving a ~10 mM final concentration of arabinose. Mixing with arabinose occurred before tubing was split to feed into each connected lagoon. In these experiments, aTc (800 ng/mL final) was added directly to the appropriate lagoons, either concurrently with phage seeding or at a later time-point, and was allowed to dilute over time.

Lagoons were seeded at a starting titer of $\sim 10^7$ pfu/mL. Dilution rate was adjusted by modulating lagoon volume (5–20 mL) and/or culture inflow rate (10–20 mL/hr). Lagoons were sampled at indicated times (usually every 24 h) by removal of culture (500 μ L) by syringe through the waste needle. Samples were centrifuged at 13,500 *g* for 2 minutes and the supernatant removed and stored at 4 °C. Titers were evaluated by plaquing on S2060 + pJC175e + pDB016 and recombinant phage titers assessed by plaquing on S2060 + AP-T7 A¹³. Phage genotypes were assessed from pool samples or single plaques by diagnostic PCR (0.5 μ L phage + 14.5 μ L PCR mixture) using primers BT-52F (5' - GTCGCGCAACTATCGGTATCAAGCTG) and BT-52R2 (5' - AGTAAGCAGATAGCCGAACAAAGTTACCAGAAGGAAAC) and a two-step program (98° for 3 min, followed by 25 or 30 cycles of 98° for 10 s and 72 °C for 90 s, followed by 72° for 2 min). The resulting PCR products were subjected to 1% agarose gel electrophoresis (shown in Fig. 3c) and/or Sanger sequencing.

Phage titers were determined by plaquing on pJC175e pDB16 S2060 host cells, which allow activity-independent propagation but negatively select against T7 RNAP recombinant phage. The presence of T7 RNAP or gene III recombinant phage was monitored by plaquing on S2060 cells containing pT7-AP and no plasmid.

Details of PACE seed phage, aTc dosing, and workup are in Supplementary Fig. 5.

HEK cell culture, transfections and genomic DNA extraction.

See Figs. 4b, 5 and 6c, Supplementary Table 3 and Supplementary Figs. 8–14, 16 and 20. HEK293T culturing conditions, transfections for both single dose and titration experiments, and genomic DNA extraction were all conducted as previously described.⁶ Briefly, HEK293T cells were seeded into 48-well Poly-D-Lysine-coated plates (Corning 354509) at 30,000 cells/well. 1 day after plating, cells were transfected by Lipofectamine 2000 (Thermo Fisher) with 750 ng of base editor plasmid, 250 ng of guide RNA plasmid, and 10 ng of fluorescent protein expression plasmid as a transfection control following the manufacturer's directions. For titration experiments, base editor plasmid was replaced by equal mass of pUC to maintain the ratio of DNA to Lipofectamine 2000. Cells were cultured for 3 days before genomic DNA was extracted by replacement of culture media with 100 μ L lysis buffer (10 mM Tris-HCl, pH 7.5, 0.05% SDS, 25 μ g/mL proteinase K (NEB) and 37 °C incubation for 1 hour. Proteinase K was inactivated by 30-minute incubation at 80 °C. Replicates constituted independent transfections of separate splits of cells, performed on the same day or on different days.

Nucleofection of baringo MEFs and ApoE4 astrocytes

(Fig. 6a, 6b, Supplementary Table 3 and Supplementary Fig. 21). MEF cells were cultivated until confluent, then pooled. Replicates were performed on the same day using three

separate nucleofections followed by cultivation in separate wells. Each nucleofection contained 400 ng base editor plasmid and 100 ng guide RNA plasmid. Transfection programs were optimized following manufacturer's instructions (CZ-167, P4 Primary Cell 4D-Nucleofector® X Kit, Lonza). Cells were harvested for genomic DNA extraction after ~96 h. ApoE4-expressing astrocytes were diluted to 200,000 astrocytes per 20-µL reaction and nucleofected with 750 ng base editor plasmid and 250 ng guide RNA plasmid (program EN-150, SF Cell Line 4D Nucleofector® X Kit, Lonza). Reactions were diluted to 100 µL with pre-warmed media, and half of the resulting solution was plated in 12-well dishes. After 72h, genomic DNA was extracted with 300 µL lysis buffer.

Baringo PMEF generation.

Baringo females at 3–4 weeks of age were treated with single intra-peritoneal injection of 5 U each of pregnant mare's serum gonadotropin (Prospec) followed by human chorionic gonadotropin (Sigma) after 44–45 h and paired with baringo males. The following morning, females were examined for copulatory plugs to confirm matings and marked as 0.5 dpc. At day 13.5 females were sacrificed by CO₂ inhalation followed by cervical dislocation. Embryos were harvested in PBS under aseptic conditions. To harvest primary embryonic fibroblasts, each embryo was eviscerated and head was removed. The remaining parts of each embryo were minced to prepare single-cell suspensions and treated with 0.25% Trypsin-EDTA (Gibco) at 37 °C for 10 minutes, followed by centrifugation for 10 minutes. Pellets were resuspended in growth media containing DMEM, 10% FBS, penicillin-streptomycin (100 U/mL) and plated on 15-cm tissue culture plates, then incubated at 37 °C until confluent. The baringo colony is maintained *ad libitum* and all animal procedures are approved by the Children's Hospital IACUC in compliance with relevant ethical regulations.

Library preparation and high-throughput sequencing.

DNA sequencing libraries of edited sites were prepared as previously described.^{4,6,10} PCR1 amplifications for new sites (TMC1, WFS1, APOE, off-target sites) were carried out by qPCR and amplification was stopped within the exponential phase to prevent overamplification and PCR bias, but were otherwise prepared for sequencing identically to previously tested sites. Primers used are shown in Supplementary Table 6.

High-throughput sequencing data analysis.

Editing was quantified using CRISPResso2,⁵⁸ available at <http://crispresso.pinellolab.partners.org>, using the guide and amplicon sequences given in Supplementary Table 6 and with the following options: `base_editor_output`, `quantification_window_size` 28, `quantification_window_center` -13, `plot_window_size` 24. Reads were not quality filtered. Reads containing both indels as well as base edits were counted toward base editing and towards indels. Read mapping for cytosines in and near the editing window were compiled from the amplicon nucleotide percentage summary output. Insertions and deletions were quantified using CRISPResso2 quantification of editing frequency output. The number of reads with substitutions only were subtracted from the number of modified reads and the result divided by the number of total mapped reads and multiplied by 100. For quantification of conversion to desired alleles (Figure 6a–c), the percentage counts of aligned reads around the target site that matched the sequences given in

Supplementary Table 6, all of which contain the targeted coding mutation with no other non-silent mutations or indels, were summed for each replicate from the CRISPResso allele table.

Ancestral sequence reconstruction

(Supplementary Fig. 7). We performed ancestral sequence reconstruction using our previously reported alignment and phylogenetic tree of the APOBEC protein family.⁶ Briefly, 468 APOBEC protein sequences were collected from the UniProt database aligned with MAFFT and a phylogenetic tree was built from the alignment using IQ-Tree⁵⁹ and the best fitting model (JTT + F + R5). Sequences at internal nodes were inferred using the FastML package⁶⁰ given the tree, alignment and substitution model.

Structure-guided alignment and homology model

(Supplementary Figs. 15 and 17). A BLAST⁶¹ search was performed with the protein sequences P41238, H2P4E7, E1BTD6, and H2P4E9 (UNIPROT) against the Protein Data Bank archive (PDB). A match is considered a suitable template if it displays a minimum of 50% sequence identity with the source sequence and a minimum coverage of 70%. These criteria were chosen to limit the selected templates to close homologues whose alignment with the source is non-ambiguous. Ten structures resulted for structure-guided alignment. These templates were used to refine the alignment of 1,189 APOBEC homologs previously described⁶ using the Expresso algorithm which is part of the T-Coffee software package.⁶² The resulting alignment was manually searched for the presence of the LxN motif in the recognition loop.

Homology models of APOBEC1, FERNY and CDA were generated by using the I-TASSER web server.^{63,64} The quality of the models predicted by I-TASSER was assigned a TM-score of 0.62 \pm 0.14, 0.85 \pm 0.08 and 0.74 \pm 0.11 respectively. Typically, a TM-score value above 0.5 implies a correct topology for the model.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank B. Fu and C. Canavan for assistance with plasmid construction and assays; H. Rees, T. Wang, J. Bessen, A. Badran, and P. Lichter for helpful discussion; and K. Clement for CRISPResso2 support. This work was supported by U.S. NIH U01 AI142756, RM1 HG009490, R01 EB022376, and R35 GM118062, St. Jude Collaborative Research Consortium, DARPA HR0011-17-2-0049, the Ono Pharma Foundation, and HHMI. L.W.K. is an NSF Graduate Research Fellow and was supported by NIH Training Grant T32 GM095450. O.S.O. and J.R.H. were supported by NIH DC013521. C.Z. was supported by the Harvard College Research Program. C.W. is the Marion Abbe Fellow of the Damon Runyon Cancer Research Foundation (DRG-2343-18).

References

1. Cornu TI, Mussolino C & Cathomen T Refining strategies to translate genome editing to the clinic. *Nature Medicine* 23, 415–423, doi:10.1038/nm.4313 (2017).
2. Webber BR et al. Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *bioRxiv*, 1–23, doi:10.1101/482497 (2018).

3. Rees HA & Liu DR Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature Reviews Genetics* 19, 770–788, doi:10.1038/s41576-018-0059-1 (2018).
4. 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* 533, 420–424, doi:10.1038/nature17946 (2016). [PubMed: 27096365]
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. *Science Advances* 3, eaao4774, doi:10.1126/sciadv.aao4774 (2017). [PubMed: 28875174]
6. Koblan LW et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nature Biotechnology* 36, 843–846, doi:10.1038/nbt.4172 (2018).
7. Nishida K et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* 353, aaf8729–aaf8729, doi:10.1126/science.aaf8729 (2016). [PubMed: 27492474]
8. Zafra MP et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nature Biotechnology* 36, 888, doi:10.1038/nbt.4194 (2018).
9. Gehrke JM et al. An APOBEC3A-Cas9 base editor with minimized bystander and off-target activities. *Nature Biotechnology* 36, 977–982, doi:10.1038/nbt.4199 (2018).
10. Kim YB et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. *Nature Biotechnology* 35, 371–376, doi:10.1038/nbt.3803 (2017).
11. Najm FJ et al. Orthologous CRISPR–Cas9 enzymes for combinatorial genetic screens. *Nature Biotechnology* 36, 179–189, doi:10.1038/nbt.4048 (2018).
12. Badran AH & Liu DR In vivo continuous directed evolution. *Current Opinion in Chemical Biology* 24, 1–10, doi:10.1016/j.cbpa.2014.09.040 (2015). [PubMed: 25461718]
13. Esvelt KM, Carlson JC & Liu DR A system for the continuous directed evolution of biomolecules. *Nature* 472, 499–503, doi:10.1038/nature09929 (2011). [PubMed: 21478873]
14. Badran AH et al. Continuous evolution of *Bacillus thuringiensis* toxins overcomes insect resistance. *Nature* 533, 58–63, doi:10.1038/nature17938 (2016). [PubMed: 27120167]
15. Bryson DI et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nature Chemical Biology* 13, 1253–1260, doi:10.1038/nchembio.2474 (2017). [PubMed: 29035361]
16. Carlson JC, Badran AH, Guggiana-Nilo DA & Liu DR Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology* 10, 216–222, doi:10.1038/nchembio.1453 (2014). [PubMed: 24487694]
17. Dickinson BC, Leconte AM, Allen B, Esvelt KM & Liu DR Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proceedings of the National Academy of Sciences of the United States of America* 110, 9007–9012, doi:10.1073/pnas.1220670110/-DCSupplemental (2013). [PubMed: 23674678]
18. Dickinson BC, Packer MS, Badran AH & Liu DR A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nature Communications* 5, 5352, doi:10.1038/ncomms6352 (2014).
19. Hu JH et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556, 57–63, doi:10.1038/nature26155 (2018). [PubMed: 29512652]
20. Hubbard BP et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nature Chemical Biology* 12, 939–942, doi:10.1038/nmeth.3515 (2015).
21. Leconte AM et al. A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes. *Biochemistry* 52, 1490–1499, doi:10.1021/bi3016185 (2013). [PubMed: 23360105]
22. Packer MS, Rees HA & Liu DR Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nature Communications* 8, 956, doi:10.1038/s41467-017-01055-9 (2017).
23. Wang T, Badran AH, Huang TP & Liu DR Continuous directed evolution of proteins with improved soluble expression. *Nature Chemical Biology* 14, 972–980, doi:10.1038/s41589-018-0121-5 (2018). [PubMed: 30127387]

24. Roth T, Woolston B, Stephanopoulos G & Liu DR Phage-assisted evolution of *Bacillus methanolicus* methanol dehydrogenase 2. *ACS Synthetic Biology* 8, 796–806, doi:10.1021/acssynbio.8b00481 (2019). [PubMed: 30856338]
25. Raindlóvá V et al. Influence of major-groove chemical modifications of DNA on transcription by bacterial RNA polymerases. *Nucleic Acids Research*, gkw171–113, doi:10.1093/nar/gkw171 (2016).
26. Karzai AW, Roche ED & Sauer RT The SsrA-SmpB system for protein tagging, directed degradation and ribosome rescue. *Nature Structural Biology* 7, 449–455, doi:10.1038/75843 (2000). [PubMed: 10881189]
27. Lykke-Andersen J & Christiansen J The C-terminal carboxy group of T7 RNA polymerase ensures efficient magnesium ion-dependent catalysis. *Nucleic Acids Research* 26, 5630–5635, doi:10.1093/nar/26.24.5630 (1998). [PubMed: 9837993]
28. Rakonjac J, Bennett NJ, Spagnuolo J, Gagic D & Russel M Filamentous bacteriophage: biology, phage display and nanotechnology applications. *Current Issues in Molecular Biology* 13, 51–76 (2011). [PubMed: 21502666]
29. Zinder ND & Boeke JD The filamentous phage (Ff) as vectors for recombinant DNA--a review. *Gene* 19, 1–10 (1982). [PubMed: 6292041]
30. Iwai H, Züger S, Jin J & Tam P-H Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Letters* 580, 1853–1858, doi:10.1016/j.febslet.2006.02.045 (2006). [PubMed: 16516207]
31. Beale RCL et al. Comparison of the Differential Context-dependence of DNA Deamination by APOBEC Enzymes: Correlation with Mutation Spectra in Vivo. *Journal of Molecular Biology* 337, 585–596, doi:10.1016/j.jmb.2004.01.046 (2004). [PubMed: 15019779]
32. Navaratnam N et al. *Escherichia coli* cytidine deaminase provides a molecular model for ApoB RNA editing and a mechanism for RNA substrate recognition. *Journal of Molecular Biology* 275, 695–714, doi:10.1006/jmbi.1997.1506 (1998). [PubMed: 9466941]
33. Salter JD, Bennett RP & Smith HC The APOBEC Protein Family: United by Structure, Divergent in Function. *Trends in Biochemical Sciences* 41, 578–594, doi:10.1016/j.tibs.2016.05.001 (2016). [PubMed: 27283515]
34. Kohli RM et al. A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *Journal of Biological Chemistry* 284, 22898–22904, doi:10.1074/jbc.M109.025536 (2009). [PubMed: 19561087]
35. Lada AG et al. Mutator effects and mutation signatures of editing deaminases produced in bacteria and yeast. *Biochemistry (Moscow)* 76, 131–146 (2011). [PubMed: 21568845]
36. St Martin A et al. A fluorescent reporter for quantification and enrichment of DNA editing by APOBEC–Cas9 or cleavage by Cas9 in living cells. *Nucleic Acids Research* 9, 229–210, doi:10.1093/nar/gky332 (2018).
37. Wang X et al. Efficient base editing in methylated regions with a human APOBEC3A-Cas9 fusion. *Nature Biotechnology* 36, 946–949, doi:10.1038/nbt.4198 (2018).
38. Manji SSM, Miller KA, Williams LH & Dahl H-HM Identification of three novel hearing loss mouse strains with mutations in the *Tmc1* gene. *The American Journal of Pathology* 180, 1560–1569, doi:10.1016/j.ajpath.2011.12.034 (2012). [PubMed: 22330676]
39. Liu C-C, Liu C-C, Kanekiyo T, Xu H & Bu G Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nature reviews. Neurology* 9, 106–118, doi:10.1038/nrneurol.2012.263 (2013). [PubMed: 23296339]
40. Nishimasu H et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science*, eaas9129–9128, doi:10.1126/science.aas9129 (2018).
41. Rigoli L, Bramanti P, Di Bella C & De Luca F Genetic and clinical aspects of Wolfram syndrome 1, a severe neurodegenerative disease. *Pediatric Research* 83, 921–929, doi:10.1038/pr.2018.17 (2018). [PubMed: 29774890]
42. Hardy C et al. Clinical and molecular genetic analysis of 19 Wolfram syndrome kindreds demonstrating a wide spectrum of mutations in *WFS1*. *American Journal of Human Genetics* 65, 1279–1290, doi:10.1086/302609 (1999). [PubMed: 10521293]

43. Gaudelli NM et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* 551, 464–471, doi:10.1038/nature24644 (2017). [PubMed: 29160308]
44. Rees HA et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nature Communications* 8, 15790, doi:10.1038/ncomms15790 (2017).
45. Tsai SQ et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nature Biotechnology* 33, 187–197, doi:10.1038/nbt.3117 (2014).
46. Scheben A & Edwards D Towards a more predictable plant breeding pipeline with CRISPR/Cas-induced allelic series to optimize quantitative and qualitative traits. *Current Opinion in Plant Biology* 45, 218–225, doi:10.1016/j.pbi.2018.04.013 (2018). [PubMed: 29752075]
47. Urnov FD, Ronald PC, biotechnology D. C. N. & 2018 A call for science-based review of the European court's decision on gene-edited crops. *Nature Biotechnology* 36, 800–802, doi:10.1038/nbt.4252 (2018).
48. Badran AH & Liu DR Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nature Communications* 6, 8425, doi:10.1038/ncomms9425 (2015).
49. Cavaleiro AM, Kim SH, Seppälä S, Nielsen MT & Nørholm MH H. Accurate DNA Assembly and Genome Engineering with Optimized Uracil Excision Cloning. *ACS Synthetic Biology* 4, 1042–1046, doi:10.1021/acssynbio.5b00113 (2015). [PubMed: 26263045]
50. Engler C, Kandzia R & Marillonnet S A One Pot, One Step, Precision Cloning Method with High Throughput Capability. *Plos One* 3, e3647–3647, doi:10.1371/journal.pone.0003647 (2008). [PubMed: 18985154]
51. Lee ME, DeLoache WC, Cervantes B & Dueber JE A Highly Characterized Yeast Toolkit for Modular, Multipart Assembly. *ACS Synthetic Biology*, 150501080052002, doi:10.1021/sb500366v (2015).
52. Potapov V et al. Comprehensive Profiling of Four Base Overhang Ligation Fidelity by T4 DNA Ligase and Application to DNA Assembly. *ACS Synthetic Biology* 7, 2665–2674, doi:10.1021/acssynbio.8b00333 (2018). [PubMed: 30335370]
53. Ringquist S et al. Translation initiation in *Escherichia coli*: sequences within the ribosome-binding site. *Molecular Microbiology* 6, 1219–1229 (1992). [PubMed: 1375310]
54. Davis JH, Rubin AJ & Sauer RT Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Research* 39, 1131–1141, doi:10.1093/nar/gkq810 (2011). [PubMed: 20843779]
55. Salis HM The ribosome binding site calculator. *Methods in Enzymology* 498, 19–42, doi:10.1016/B978-0-12-385120-8.00002-4 (2011). [PubMed: 21601672]
56. Cui L et al. A CRISPRi screen in *E. coli* reveals sequence-specific toxicity of dCas9. *Nature Communications*, 1–10, doi:10.1038/s41467-018-04209-5 (2018).
57. Chung CT & Miller RH Preparation and storage of competent *Escherichia coli* cells. *Methods in Enzymology* 218, 621–627 (1993). [PubMed: 8510550]
58. Clement K et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. *Nature Biotechnology* 37, 224–226, doi:10.1038/s41587-019-0032-3 (2019).
59. Nguyen LT, Schmidt HA, von Haeseler A & Minh BQ IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32, 268–274, doi:10.1093/molbev/msu300 (2015). [PubMed: 25371430]
60. Ashkenazy H et al. FastML: a web server for probabilistic reconstruction of ancestral sequences. *Nucleic Acids Res* 40, W580–584, doi:10.1093/nar/gks498 (2012). [PubMed: 22661579]
61. Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ Basic local alignment search tool. *J Mol Biol* 215, 403–410, doi:10.1016/S0022-2836 (1990). [PubMed: 2231712]
62. Notredame C, Higgins DG & Heringa J T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302, 205–217, doi:10.1006/jmbi.2000.4042 (2000). [PubMed: 10964570]
63. Roy A, Kucukural A & Zhang Y I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 5, 725–738, doi:10.1038/nprot.2010.5 (2010). [PubMed: 20360767]

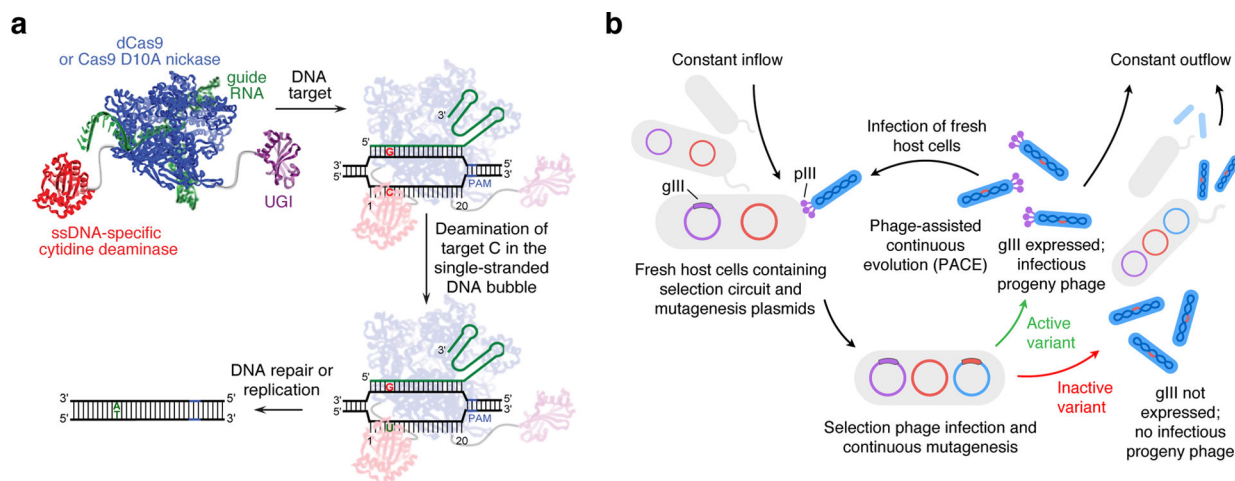
64. Yang J & Zhang Y Protein Structure and Function Prediction Using I-TASSER. *Curr Protoc Bioinformatics* 52, 5 8 1–15, doi:10.1002/0471250953.bi0508s52 (2015). [PubMed: 26678385]

Author Manuscript

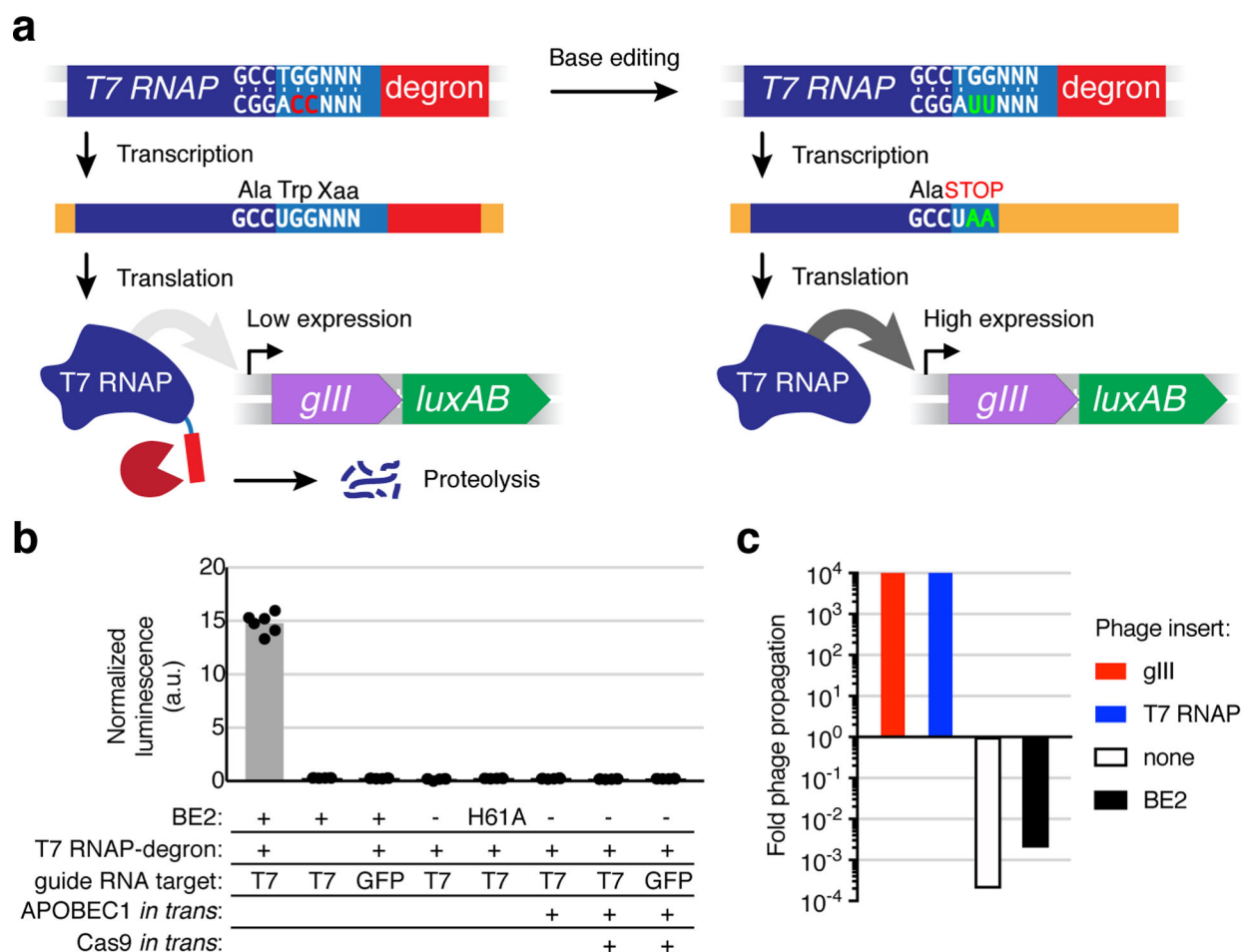
Author Manuscript

Author Manuscript

Author Manuscript

**Figure 1.**

Overview of base editing and PACE. **(a)** Cytosine base editing converts target C•G base pairs to T•A at guide RNA-specified DNA sequences. R-loop formation by the Cas9 domain exposes a small bubble of single-stranded DNA (bases numbered 1–20, with the PAM at positions 21–23) to the fused cytidine deaminase. Cytosines in this bubble are deaminated to uracil, which is protected from excision by uracil glycosylase inhibitor (UGI). To increase editing efficiency in some CBEs, the non-edited strand is nicked, stimulating cellular repair and replication to replace the original C•G base pair with T•A. **(b)** General PACE schematic. *E. coli* host cells contain a plasmid-based genetic circuit that links expression of gene III (gIII, encoding pIII) to the activity of the biomolecule of interest encoded in a modified M13 bacteriophage (blue). The production of infectious progeny phage requires expression of gene III, which only occurs in host cells infected with phage variants encoding the desired activity of interest. The phage genome is mutagenized continuously by a mutagenesis plasmid in the host cells.⁴⁸ Since the phage exist in a fixed-volume vessel (the lagoon) continuously diluted with host-cell culture, only those phage that propagate faster than the rate of dilution can persist and evolve in the lagoon.

**Figure 2.**

Design and validation of BE-PACE. **(a)** Schematic for the BE-PACE circuit. T7 RNA polymerase, required to express gene III and luciferase (translationally coupled *via* overlapping stop and start codons) from a T7 promoter, is fused through a Trp-containing linker to a C-terminal degron that causes the destruction of the protein. The Trp codon (TGG) provides the CBE target CCA on the transcriptional template strand. Deamination of either cytosine by a CBE converts the Trp codon to a STOP codon (UAG, UGA or UAA, arising from transcription of CUA, UCA or UUA on the template strand), preventing translation of the degron, restoring T7 RNA polymerase activity, and activating gene III expression. The 5' context for the deamination target can be varied by changing the linker sequence. **(b)** A luciferase assay shows that all components of the BE-PACE system are required for circuit activation. Leaky expression of BE2 during cell growth resulted in significant activation of the circuit even in the absence of induction. Guide RNA targets the CBE to either the T7 RNAP–degron linker or to green fluorescent protein. The H61A mutation in the base editor inactivates its APOBEC1 deaminase. Expression of APOBEC1 and dCas9 as separate polypeptides instead of BE2 does not activate the circuit. Dots represent biological replicates and bars represent mean values. **(c)** Discrete overnight phage propagation assays to test the prototype BE-PACE circuit. Phage containing the genes shown were mixed with an excess of host cells and allowed to propagate until the host cells grow to

saturation. The output phage titer was divided by the input titer to calculate fold phage propagation. Phage with activities that circumvent or short-circuit the selection (gIII, T7 RNAP) enrich strongly (10^4 -fold), while BE2 phage propagate weakly (output titer 500-fold lower than input titer) but more than empty phage (5,000-fold lower output titer).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

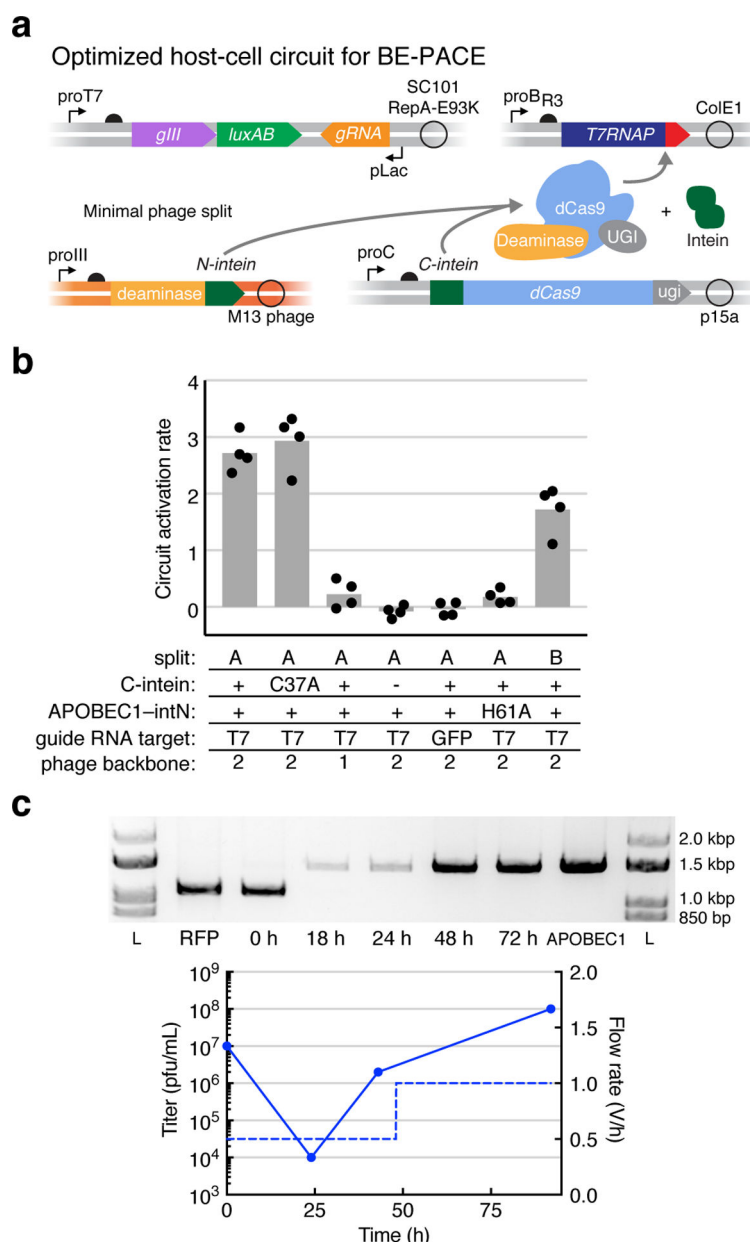
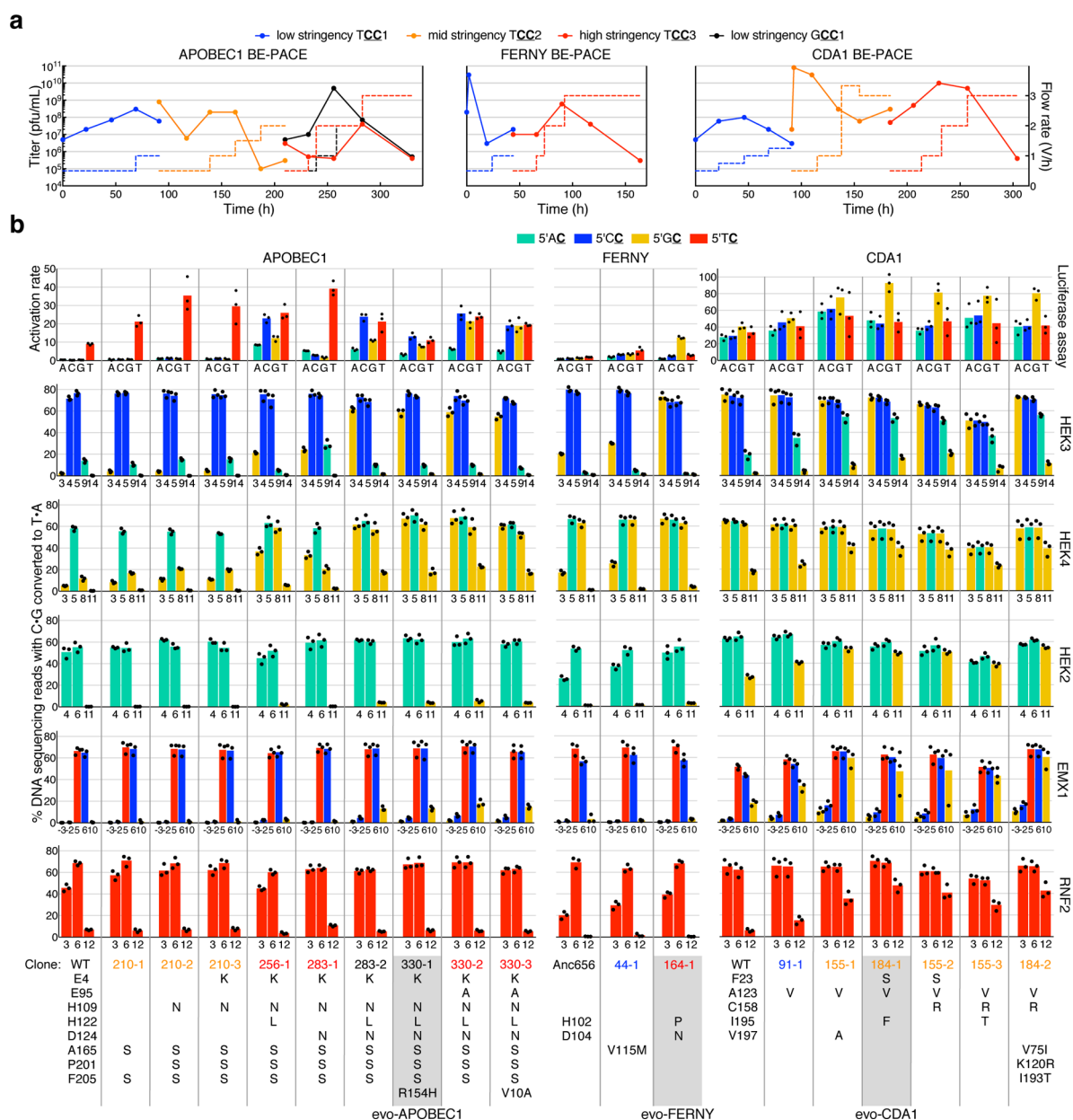


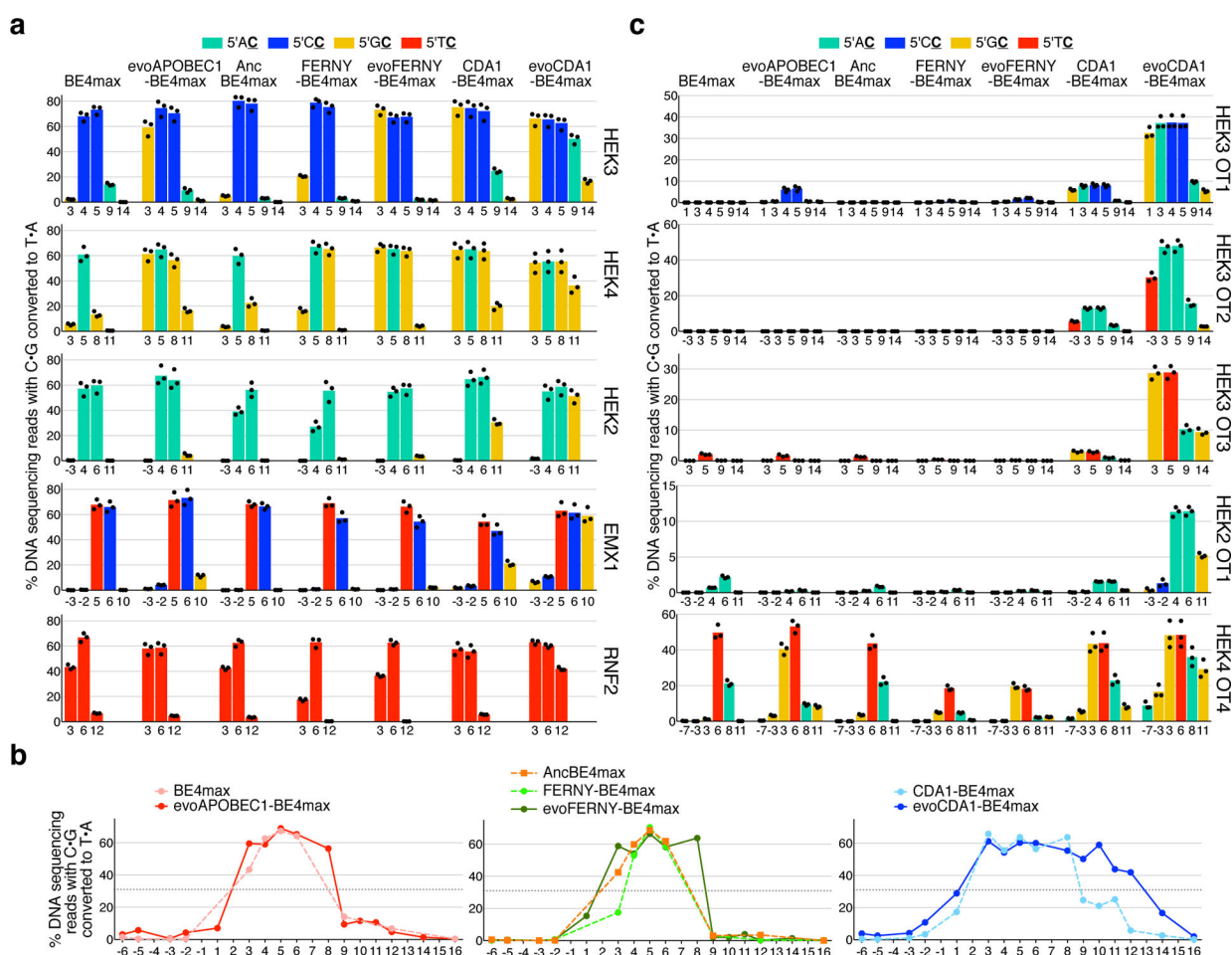
Figure 3. Design and validation of the split intein BE-PACE selection. **(a)** Plasmids (grey backbones) and phage (orange backbone) in the optimized BE-PACE host-cell selection circuit. In the minimal phage split, only the deaminase is encoded by phage, while dCas9 is encoded on a host plasmid. In the balanced phage split, part of dCas9 is encoded on the phage and the remainder on a host plasmid. In either split, full-length BE2 is reconstituted upon phage infection as a single polypeptide by *trans*-splicing inteins. **(b)** Luciferase assays show that split BE2 activates the circuit in an intein-, guide RNA-, and APOBEC1 activity-dependent manner. The C37A mutant of the C-intein disrupts splicing, but not association between the split intein components. Phage backbone numbers refer to which generation phage backbone was used (see text). Bar heights represent mean values for fitted slopes of luminescence per

OD₆₀₀ versus time (see Methods) and dots represent slopes for individual biological replicates. (c) BE-PACE competition experiment. Host cells contained the low-stringency selection circuit TCC1 and mutagenesis plasmid. A lagoon, continuously diluted with host cells, was seeded with 99.9% red fluorescent protein (RFP) phage (lower band) and 0.1% APOBEC1-intein phage (upper band). The phage population composition was monitored by PCR using primers flanking the phage insert. L denotes DNA size standard ladder. The total phage titer over time and the lagoon flow rate is shown on the graph at the bottom. This experiment was not repeated.

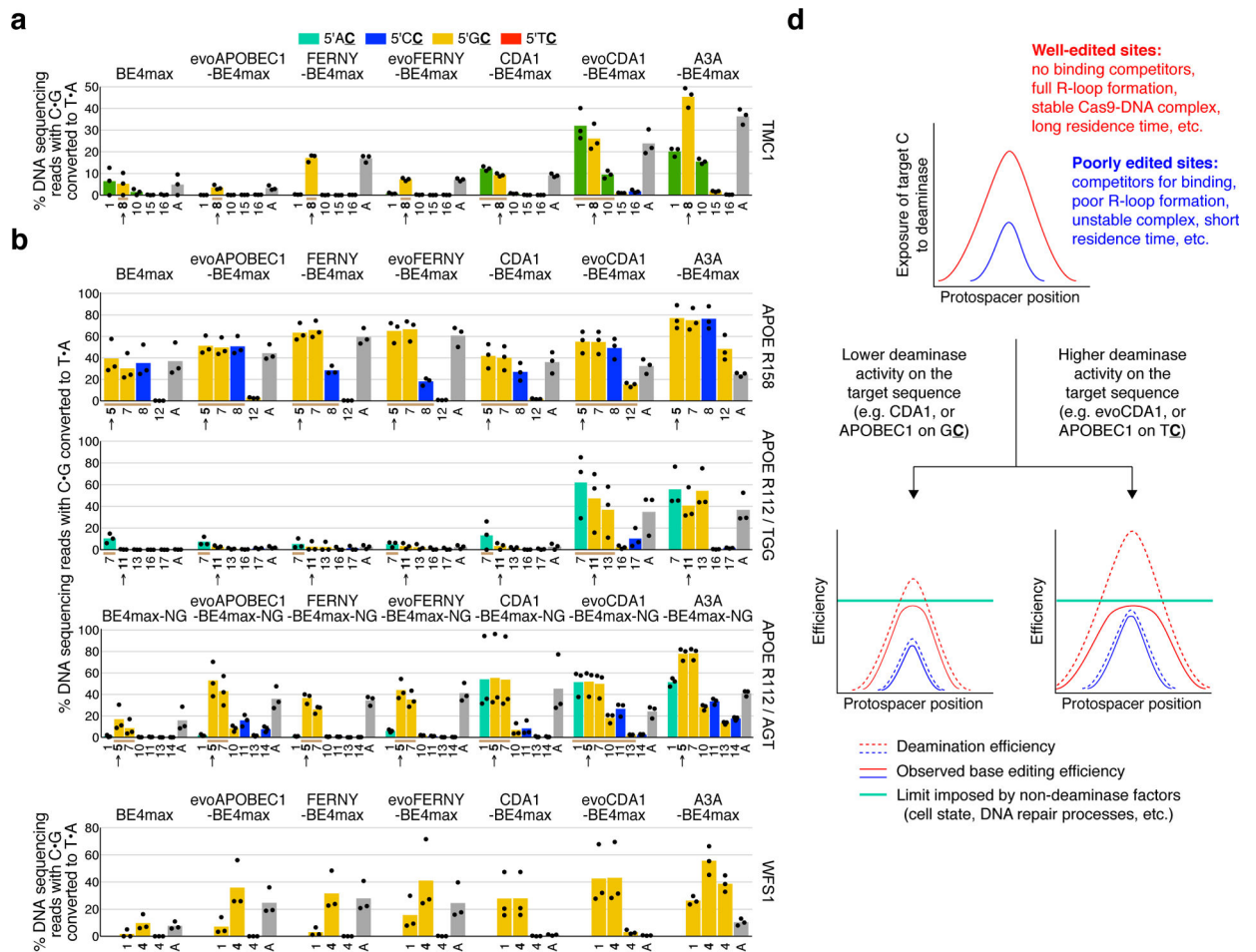
**Figure 4.**

BE-PACE of APOBEC1, FERNY, and CDA1, and characterization of evolved deaminase CBEs in mammalian cells. **(a)** PACE of APOBEC1, FERNY, and CDA1 deaminase phase as intein fusions. BE-PACE selection circuits are described in Supplementary Table 2. Solid lines and dots show phage titers (left axis) and dotted lines show flow rate (right axis) during BE-PACE. These experiments were not repeated. **(b)** Performance of evolved deaminases in the luciferase assay in bacteria (top panel) and in CBEs editing HEK293T cells (bottom panels) at five endogenous genomic sites. Target cytosines are color coded according to the base immediately 5' of the edited C. In the top panel, deaminases were tested on all four possible ANC_4C_5A target sites with N = A, C, G or T in low-stringency BE-PACE circuits (see Supplementary Table 2). In the lower panels, C•G-to-T•A base editing is shown for

cells transfected with each CBE (vertical columns, with evolved deaminase genotypes shown at the bottom) and each of five guide RNAs. Deaminases were not codon-optimized for human cell expression, but the remainder of the BE4max architecture was codon-optimized. Editing byproducts are shown in Supplementary Table 3. Base editing levels are shown for each edited C within the protospacer (positive X-axis numbers, with the PAM at positions 21–23) or upstream of the protospacer (negative X-axis numbers, with –1 being one base upstream). Genotypes for the wild-type deaminase (or the reconstructed ancestral sequence for node 656, “FERNY”, labeled Anc656) and for each mutant are given below each clone name. Evolved clones are named for the time point (in hours) at which they were isolated from PACE (Fig. 4a). The genotypes in grey are evoAPOBEC1-BE4max (left), evoFERNY-BE4max (middle), and evoCDA1-BE4max (right). Dots represent individual biological replicates and bars represent mean values.

**Figure 5.**

Base editing performance of evolved deaminase CBEs, all codon-optimized in the BE4max architecture, in mammalian cells. **(a)** Editing by wild-type and evolved deaminase CBEs for five endogenous genomic test sites in HEK293T cells. Target cytosines are color-coded by the base immediately 5' of the edited C as in Fig. 3b. Protospacer positions are specified by X-axis numbers. Dots represent individual biological replicates and bars represent mean values. **(b)** Base editing activity window plots showing mean C•G-to-T•A editing at all tested protospacer positions across six HEK293T sites. Target cytosines preceded by a 5' G are excluded for BE4max, AncBE4max, and FERNY-BE4max to avoid misrepresenting their editing windows due to sequence context preference. The dotted horizontal line represents half-maximal peak editing to approximate editing window width. **(c)** Off-target editing by wild-type and evolved deaminases for a selection of known off-target dCas9 binding sites for the HEK2, HEK3 and HEK3 guide RNAs. The data shown are for off-target sites amplified from the treated cells shown in corresponding panels of Fig. 4a. Protospacer positions are specified by X-axis numbers. Dots represent individual biological replicates and bars represent mean values. Data from all off-target sites examined are in Supplementary Fig. 20 and Supplementary Table 3.

**Figure 6.**

(a-c) Cytosine base editing at disease-relevant sites in primary mammalian cells and cell lines. Protospacer positions are specified by X-axis numbers, with the target C indicated by an arrow, and Cs within the expected editing window of each CBE (based on Fig. 5b) marked with horizontal brown lines. Dots represent individual biological replicates and bars represent mean values. (a) Editing the *TMC1* site to revert the Y182C mutation in primary embryonic fibroblasts from the *baringo* mouse model of recessive hearing loss. (b) Editing the Alzheimer's disease-associated *APOE4* allele into *APOE3'* and *APOE3* by installing R158C and R112C in immortalized mouse astrocytes. In (a) and (b), the percent of sequencing reads that contain the targeted coding mutation with no other non-silent mutations or indels is shown in grey and labeled "A". (c) Editing the Wolfram syndrome 1-associated *WFS1* gene to install Q1884STOP in HEK293T cells. The grey "A" bar shows the percent of reads with Q1884STOP and no other sequence changes. (d) Model for the relationship between site characteristics, deaminase activity, and editing outcomes in mammalian cells (see Discussion and Supplementary Discussion 6).