ARTICLES
https://doi.org/10.1038/s41587-020-0414-6
nature biotechnology

Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors

Jordan L. Doman[1,2,3,4], Aditya Raguram[1,2,3,4], Gregory A. Newby[1,2,3] and David R. Liu[1,2,3*]

Cytosine base editors (CBEs) enable targeted C•G-to-T•A conversions in genomic DNA. Recent studies report that BE3, the original CBE, induces a low frequency of genome-wide Cas9-independent off-target C•G-to-T•A mutation in mouse embryos and in rice. Here we develop multiple rapid, cost-effective methods to screen the propensity of different CBEs to induce Cas9-independent deamination in *Escherichia coli* and in human cells. We use these assays to identify CBEs with reduced Cas9-independent deamination and validate via whole-genome sequencing that YE1, a narrowed-window CBE variant, displays background levels of Cas9-independent off-target editing. We engineered YE1 variants that retain the substrate-targeting scope of high-activity CBEs while maintaining minimal Cas9-independent off-target editing. The suite of CBEs characterized and engineered in this study collectively offer ~10–100-fold lower average Cas9-independent off-target DNA editing while maintaining robust on-target editing at most positions targetable by canonical CBEs, and thus are especially promising for applications in which off-target editing must be minimized.

CBEs are genome-editing agents consisting of a cytidine deaminase fused to a catalytically impaired Cas9 protein and one or more copies of a uracil glycosylase inhibitor (UGI)[1,2]. Deamination of cytosine within a base-editing activity window (canonically, protospacer positions ~4–8, counting the protospacer-adjacent motif (PAM) as positions 21–23) in the single-stranded DNA loop displaced by the Cas9 guide RNA generates uracil, which is partially protected from base excision by the UGI. Selective nicking of the opposite DNA strand biases cellular DNA repair to replace the nonedited strand, resulting in the conversion of a target C•G base pair (bp) to a T•A bp (refs. [1–3]). CBEs have achieved high levels of single-nucleotide polymorphism (SNP) conversion with low levels of indels in numerous cell types and organisms, including animal models of human genetic diseases[3–7].

Similar to other Cas9-directed genome-editing tools, base editors can bind to off-target genomic loci that have high sequence homology to the target protospacer. A subset of these Cas9-dependent off-target binding events can lead to base editing[1,8–11], which can be minimized by using Cas9 variants with higher DNA specificity, or by delivering base editors as transient protein–RNA complexes rather than expressing them from longer-lived DNA constructs[11].

In addition to Cas9-dependent off-target base editing, deamination from Cas9-independent binding of a base editor's deaminase domain to DNA represents a distinct type of off-target base editing. Yang, Gao and their respective coworkers recently reported that when overexpressed in mouse embryos and rice, BE3, the original CBE, induces random genome-wide mutations at average frequencies of 5×10^{-8} per bp and 5.3×10^{-7} per bp, respectively[12,13]. These off-target edits likely arise from the intrinsic DNA affinity of BE3's deaminase domain, independent of the guide RNA-programmed DNA binding of Cas9 (refs. [12,13]). Ye and coworkers subsequently demonstrated that CBEs can also induce Cas9-independent off-target mutations in human induced pluripotent stem cells[14]. Unlike Cas9-dependent off-target editing, Cas9-independent deamination occurs at different loci between cells, making it difficult to characterize by targeted high-throughput sequencing. Extensive whole-genome sequencing (WGS) experiments such as those performed by Yang, Gao, Ye and their respective coworkers are low-throughput, expensive and time-intensive, limiting their use for evaluating and engineering CBE variants with decreased Cas9-independent deamination activity. Here we describe the development of methods to efficiently evaluate the propensity of a base editor to cause Cas9-independent deamination, and the application of these methods to identify and engineer CBE variants that minimize Cas9-independent DNA editing.

Results

Bacterial rifampin resistance assay. First, we assayed Cas9-independent deamination by CBEs in bacteria using a rifampin resistance assay. Measuring resistance to the antibiotic rifampin has previously been used to characterize the activity and mutagenicity of proteins expressed in *E. coli*[15–19]. Deaminase-catalyzed C•G-to-T•A mutations in the *rpoB* gene render *E. coli* resistant to rifampin. We hypothesized that cells transformed with a plasmid encoding a CBE with Cas9-independent deamination activity would become resistant to rifampin at a frequency that reflects the magnitude of this activity. To simultaneously assess the on-target activity of the base editor, we also transformed a second plasmid encoding a chloramphenicol acetyltransferase with an inactivating T•A-to-C•G point mutation, together with a guide RNA that directs the CBE to revert this point mutation. Base editors with higher on-target activity more effectively rescue chloramphenicol resistance[8]. Survival rates on chloramphenicol thus reflect on-target editing efficiency, while survival rates on rifampin reflect Cas9-independent deamination activity (Fig. 1a).

To validate this assay, we measured the chloramphenicol and rifampin resistance of bacteria transformed with wild-type APOBEC1, the deaminase used in BE3, and the catalytically inactive E63A mutant of APOBEC1 in three different architectures: as free deaminases, as deaminase–dCas9–UGI fusions or as deaminase–dCas9

[1]Merkin Institute of Transformative Technologies in Healthcare, Broad Institute of Harvard and MIT, Cambridge, MA, USA. [2]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA, USA. [3]Howard Hughes Medical Institute, Harvard University, Cambridge, MA, USA. [4]These authors contributed equally: Jordan L. Doman, Aditya Raguram. *e-mail: drliu@fas.harvard.edu

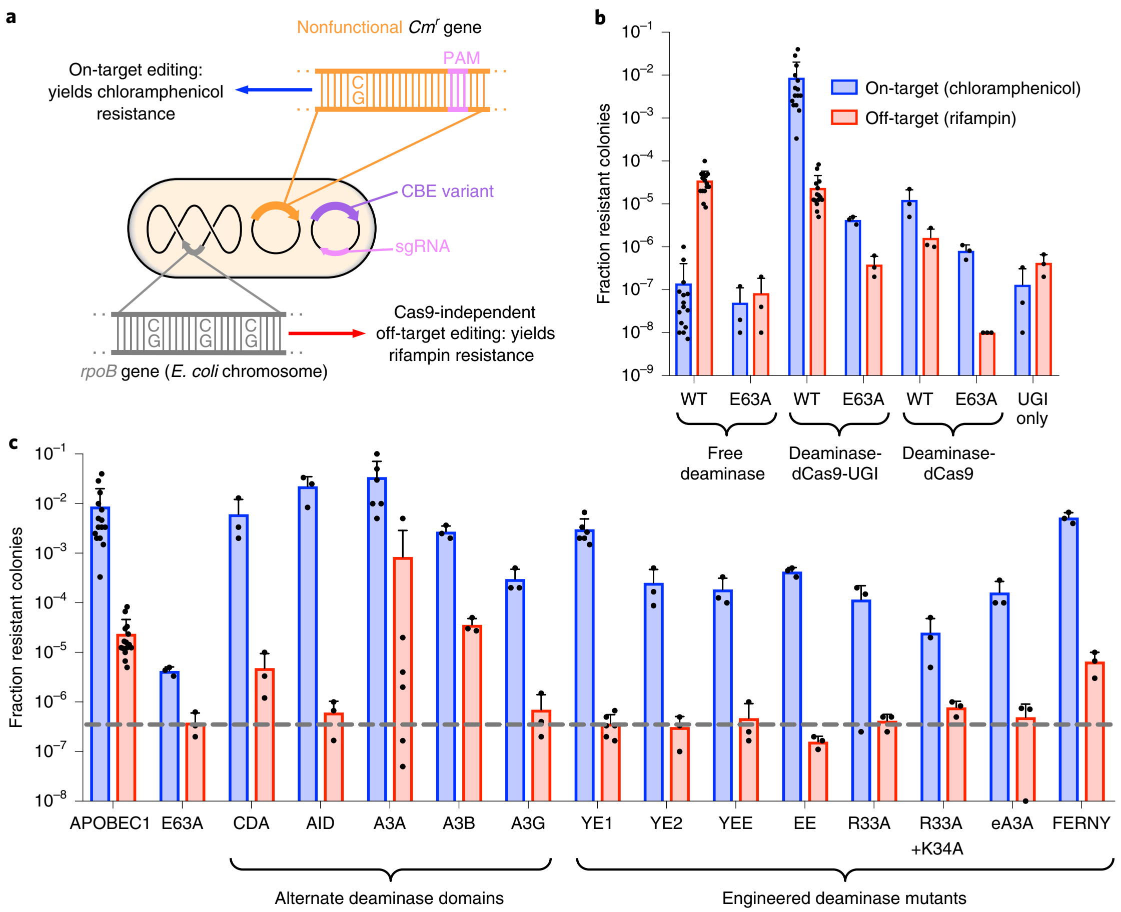

Fig. 1 | On-target and Cas9-independent off-target DNA editing in *E. coli*. a, Experimental design. **b**, Assay validation. Fraction of resistant colonies was calculated relative to number of *E. coli* plated on maintenance antibiotics. Data are shown as individual data points and mean ± s.e.m. for $n = 3$ or $n = 15$ bacterial colonies. E63A refers to a catalytically inactivated APOBEC1 E63A mutant. **c**, Performance of CBEs that use alternative deaminases. All constructs are in the deaminase-dCas9–UGI (BE2) architecture. The gray dotted line indicates the background level of rifampin resistance of the inactive APOBEC1 E63A deaminase-BE2 control. Data are shown as individual data points and mean ± s.e.m. for $n = 3$ or $n = 15$ bacterial colonies. WT, wild type.

fusions lacking the UGI domain (Fig. 1b). We used dCas9 instead of Cas9 nickase for bacterial assays because *E. coli* lack the nick-directed mismatch repair pathway that enables improved editing by Cas9 nickase CBEs in mammalian cells[20]. Compared with the background resistance levels of the untethered inactive APOBEC1 E63A construct, untethered active APOBEC1 induced a 1,000-fold increase in rifampin resistance and a tenfold increase in chloramphenicol resistance. The APOBEC1–dCas9–UGI base editor yielded the same level of rifampin resistance as that of untethered APOBEC1, but a 250-fold higher level of chloramphenicol resistance compared with its inactive counterpart. These data are consistent with high on-target activity of CBEs and with Cas9-independent off-target mutagenesis.

To confirm that rifampin resistance was accompanied by CBE-induced mutations, we sequenced the *rpoB* gene of rifampin-resistant colonies and observed primarily C•G-to-T•A mutations (Supplementary Fig. 1). The inactive APOBEC1 E63A–dCas9 fusion resulted in rifampin resistance levels equivalent to the background of the assay, suggesting that dCas9 alone does not contribute to these off-target mutations. Compared with the APOBEC1–dCas9–UGI fusion, both the APOBEC1–dCas9 fusion and the APOBEC1 E63A–dCas9-UGI exhibited a substantial decrease in rifampin resistance. These results suggest that both the deaminase domain of the base editor and the UGI domain can contribute to Cas9-independent off-target mutagenesis.

To minimize Cas9-independent deamination, we focused on the deaminase domain for two reasons. First, the rifampin resistance frequency from expression of APOBEC1 alone was 100-fold higher than the average rifampin resistance from expression of UGI alone (Fig. 1b). Second, when we analyzed the off-target DNA sequencing data from Yang and coworkers[12], we found a strong 5′ T preference among edited cytosines (Supplementary Fig. 2). This preference suggests that APOBEC1, which has a preference for deaminating 5′ T$\underline{\textbf{C}}$ substrates[21], is primarily responsible, as opposed to UGI, which is not known to cause a sequence context bias among C•G-to-T•A mutations.

Many CBE variants with alternate deaminase domains have now been reported[1,22–33]. We measured the chloramphenicol and rifampin resistance of *E. coli* transformed with virtually all previously reported CBEs, starting with naturally occurring APOBEC1, AID, CDA, APOBEC3A, APOBEC3B and APOBEC3G deaminases (Fig. 1c and Supplementary Fig. 3). *E. coli* transformed with CBEs that use CDA, APOBEC3A and APOBEC3B exhibited rifampin resistance levels that were comparable to or higher than the rifampin resistance arising from the original APOBEC1 base editor, consistent with our recent characterization of high editing activity from CDA- and APOBEC3A-derived CBEs[30]. In contrast, APOBEC3G and AID base editors produced substantially lower levels of rifampin resistance, suggesting that they generate less Cas9-independent deamination in bacteria.

Next, we expanded our panel of deaminases to include engineered deaminase variants that we and others previously developed for base-editing applications. We created APOBEC1 variants W90Y + R126E (YE1), W90Y + R132E (YE2), R126E + R132E (EE) and W90Y + R126E + R132E (YEE) to narrow the on-target base-editing window[31]. In addition, Joung and coworkers engineered APOBEC1 R33A and APOBEC1 R33A + K34A to have lower off-target RNA editing[32]. Joung and coworkers also designed an engineered APOBEC3A (eA3A) to have a strict 5′ T sequence context requirement[33]. Finally, we recently reported FERNY, a truncated, ancestrally reconstructed deaminase, which lacks an RNA-binding motif that could mediate nonspecific interactions with nucleic acids[30]. Promisingly, most of these engineered CBEs yielded substantially lower rifampin resistance levels. In particular, eA3A, YE1, YE2, EE, YEE, R33A and R33A + K34A APOBEC1 variants all resulted in rifampin resistance frequencies equivalent to that of the inactive APOBEC1 E63A–dCas9–UGI control (Fig. 1c and Supplementary Fig. 3). These results indicate that several base editor variants have much lower Cas9-independent deamination in *E. coli*, consistent with their original design goals of lower deamination activity[31,32] or increased requirements for deamination[33].

Bacterial thymidine kinase toxicity assay. Numerous studies have shown that cytidine deaminases exhibit strong sequence context preferences; for example, while APOBEC1 prefers 5′-T<u>C</u> substrates, A3G prefers 5′-C<u>C</u> substrates and AID prefers 5′-G<u>C</u> substrates[23,24,29,30]. To ensure that the results of the rifampin assay were not skewed by the sequence contexts of particular cytosines whose mutagenesis led to rifampin resistance, we sought to recapitulate the rifampin assay using a different selectable target gene with a different set of cytosines (and thus a different set of 5′ base contexts) that yield resistance when deaminated. To do this, we inserted a single copy of the herpes simplex virus thymidine kinase (HSV-TK) gene into the *E. coli* chromosome. HSV-TK leads to toxicity in the presence of the nucleoside analog 6-(β-D-2-deoxyribofuranosyl)-3,4-dihydro-8H-pyrimido-[4,5-c][1,2]oxazin-7-one (dP) (ref. [34]). We reasoned that off-target C•G-to-T•A mutations in the HSV-TK gene that inactivate the enzyme would lead to survival on dP. Indeed, while the dynamic range of this assay was narrower than that of the rifampin assay, we observed the same trends: APOBEC1, A3A, A3B and CDA induced more mutagenesis, whereas most other CBEs induced levels of dP resistance comparable to background (Supplementary Fig. 4). Sequencing the HSV-TK gene confirmed various resistant alleles caused by C•G-to-T•A mutations (Supplementary Fig. 4). The consistency between the rifampin and HSV-TK resistance assays suggests that sequence context bias plays a minimal role in the above results.

Human cell orthogonal R-loop assay. Next, we developed assays for Cas9-independent deamination by CBEs in human cells that are not dependent on WGS. Since the results of our bacterial assays, as well as the findings of Yang, Gao and their respective coworkers[12,13], suggest that the frequency of stochastic Cas9-independent deamination by BE3 is well below the ~0.1% detection limit of practical high-throughput DNA sequencing experiments, we developed an assay that magnifies Cas9-independent off-target deamination at specific loci that can be monitored by targeted high-throughput sequencing. All of the deaminases used in CBEs to date deaminate single-stranded DNA or RNA efficiently, but not double-stranded nucleic acids, and recent reports detailing Cas9-independent deamination noted that the observed mutations were enriched in transcribed regions of the genome[12,13]. We therefore reasoned that generating long-lived single-stranded DNA at specific positions would create artificially high Cas9-independent deamination levels that could be detected by targeted amplicon sequencing.

To evaluate the ability of different base editors to deaminate cytosines in single-stranded DNA regions unrelated to their on-target loci, we cotransfected HEK293T cells with plasmids encoding an SpCas9-based CBE, an SpCas9 on-target guide RNA (sgRNA), a catalytically inactive *Staphylococcus aureus* Cas9 (dSaCas9) and an SaCas9 sgRNA targeting a genomic locus unrelated to the on-target site (Fig. 2a,b). We generated all editors for mammalian cell experiments using the current 'BE4max' architecture, with optimized codon usage, optimized nuclear localization signals (NLSs) and the optimized structure of NLS–deaminase–Cas9 nickase–UGI–UGI–NLS (Supplementary Sequences)[4,35]. Deamination of cytosines in the R-loop formed by dSaCas9 should occur in a CBE-dependent, but SpCas9 guide RNA-independent, manner. Indeed, high-throughput sequencing of six dSaCas9 loci 3 d after plasmid cotransfection resulted in off-target deamination by APOBEC1-based BE4 (ref. [4,35]) that was easily detected by targeted DNA sequencing (0.4–25%), and was independent of the on-target SpCas9 guide RNA (Fig. 2c). Encouragingly, A3A-BE4 (ref. [30]), a CBE that uses APOBEC3A, demonstrated substantially higher off-target deamination of dSaCas9-generated R-loops relative to BE4 (Fig. 2c and Supplementary Fig. 5a), consistent with its higher frequency of generating resistant colonies in the bacterial rifampin assay (Fig. 1c), and with the previously reported high degree of mutagenicity of APOBEC3A in human cells[36]. These results collectively suggest that in trans deamination within R-loops generated by an orthogonal Cas9 homolog can be used to assess the propensities of SpCas9-derived CBEs to mediate Cas9-independent deamination.

To identify base editor variants that exhibit reduced Cas9-independent deamination relative to BE4 in human cells, we evaluated the same panel of 14 deaminase domains (APOBEC1, CDA, AID, APOBEC3A, eA3A, APOBEC3B, APOBEC3G and FERNY; and APOBEC1 mutants YE1, YE2, YEE, EE, R33A and R33A + K34A)[1,22–33] in the BE4max architecture for their ability to deaminate dSaCas9-induced R-loops in trans. Base editors with narrowed on-target DNA-editing windows such as YE1-BE4, YE2-BE4 and EE-BE4, or with reduced RNA editing propensities such as R33A-BE4, again exhibited substantially reduced Cas9-independent DNA deamination compared with BE4 (Fig. 2d and Supplementary Fig. 5b). Indeed, YEE-BE4 and R33A + K34A-BE4 displayed nearly undetectable levels of Cas9-independent deamination in this assay. Nearly all of the other CBE variants assayed displayed comparable or higher levels of Cas9-independent deamination relative to BE4 for at least a subset of off-target cytosines within SaCas9-induced R-loops. Compared with BE4, CBEs derived from CDA, AID and FERNY exhibited higher levels of Cas9-independent deamination at 5′-G<u>C</u> substrates, as expected given their higher activity on 5′-G<u>C</u> sequences than APOBEC1 (refs. [23,24,30]). Likewise, eA3A-BE4 and A3G-BE4 displayed moderate to high levels of Cas9-independent deamination at 5′-T<u>C</u>R and 5′-C<u>C</u> substrates, respectively, also consistent with their known sequence context preferences[29,33]. All transfected constructs had similar effects on cell viability (Supplementary Fig. 6), which indicates that cell viability is not a confounding factor in this assay. These trends agree with the results of the rifampin resistance and thymidine kinase assays, with the exception of AID-BE4: our results show higher amounts of off-target editing by AID-BE4 in mammalian cells compared with in *E. coli*. This observation is consistent with previous studies that show higher AID activity in human cells compared with bacteria, potentially due to protein–protein interaction partners or post-translational modifications to the enzyme[17,37]. These data suggest that R33A-BE4, YE1-BE4, YE2-BE4, EE-BE4, YEE-BE4 and R33A + K34A-BE4 are especially promising CBE variants for applications in which Cas9-independent off-target editing must be minimized.

In vitro kinetics assay. We hypothesized that a primary determinant of Cas9-independent deamination propensity is the catalytic efficiency of the enzyme. Ideal CBE deaminases should inefficiently catalyze deamination of substrates that are present at low

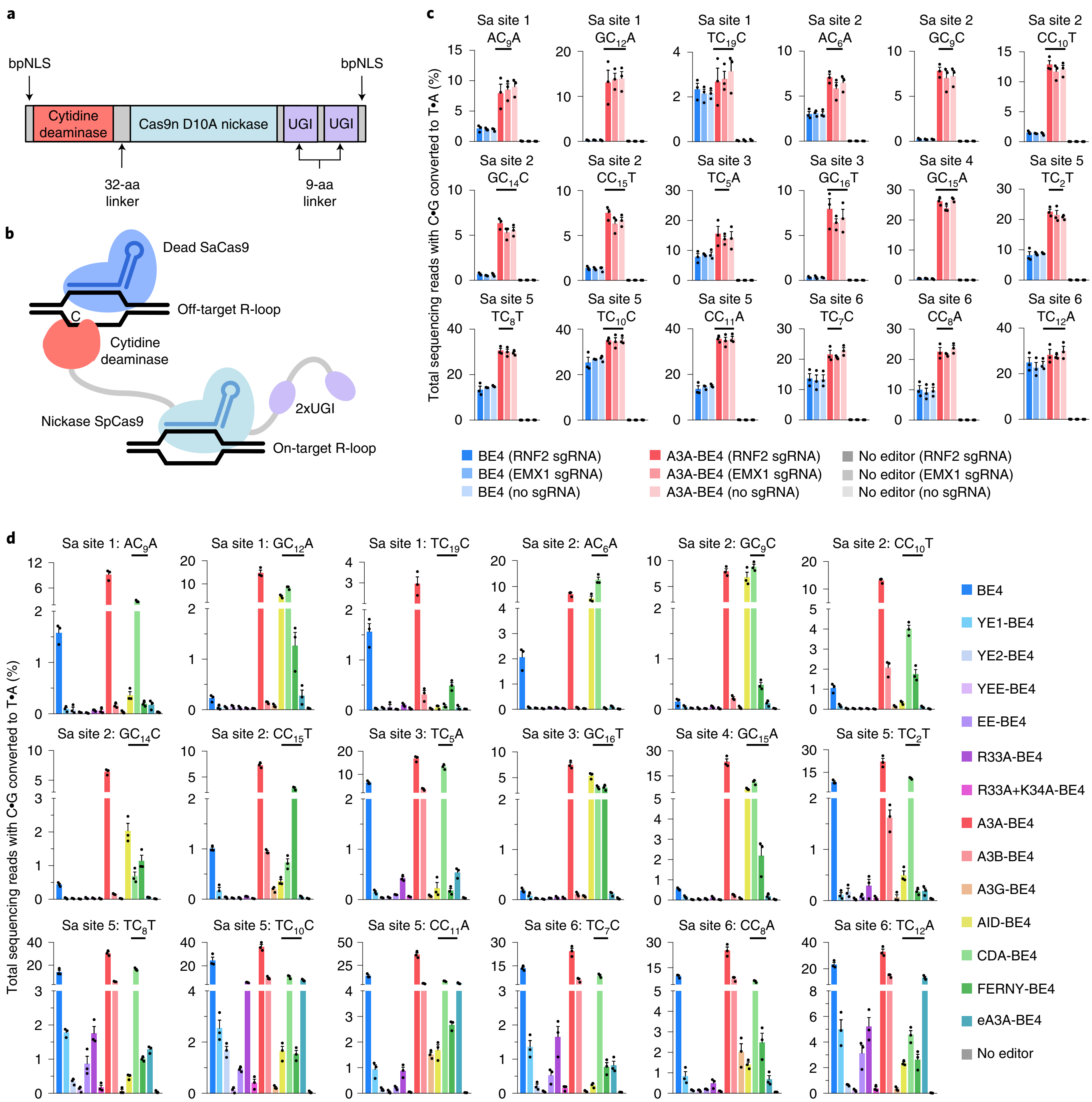

Fig. 2 | Cas9-independent deamination by CBEs in HEK293T cells. a, BE4max architecture for all CBE constructs used in mammalian cell experiments. **b**, Schematic of Cas9-independent deamination of cytosines within dSaCas9-induced R-loops by SpCas9 CBEs. **c**,**d**, Cas9-independent off-target C•G-to-T•A editing frequencies detected by targeted high-throughput sequencing of six dSaCas9 loci following cotransfection with SpCas9-targeted CBEs. Each subplot shows the observed C•G-to-T•A conversion of a single underlined cytosine and its immediate sequence context. Transfections in **c** were performed with one of two SpCas9 sgRNAs (targeting the *RNF2* or *EMX1* genomic loci) or no SpCas9 sgRNA, with on-target editing controls shown in Supplementary Fig. 5a. Transfections in **d** were performed with one SpCas9 sgRNA targeting the *RNF2* genomic locus, with on-target editing controls shown in Supplementary Fig. 5b. For **c** and **d**, data are shown as individual data points and means ± s.e.m. for $n=3$ independent biological replicates performed on different days. aa, amino acid; bpNLS, bipartite NLS; Sa, *S. aureus*.

concentrations (such as Cas9-independent off-target sites) but efficiently deaminate on-target substrates when presented at high effective local concentration due to DNA binding of the tethered Cas9 domain. To test this hypothesis, we purified three different CBE proteins and measured their catalytic efficiencies (k_{cat}/K_m ratios) in vitro for a 5′-Cy3-labeled single-stranded DNA (ssDNA) oligonucleotide that contained a single cytosine and was unrelated to the sgRNA present in the reaction. To measure reaction velocities, we quantified uracil-containing product formation by gel densitometry following USER enzyme treatment[38]. YE1–dCas9–UGI and APOBEC3A–dCas9–UGI have k_{cat}/K_m values for ssDNA that are 69-fold lower and 1.3-fold higher, respectively, than that of

APOBEC1–dCas9–UGI (Supplementary Fig. 7). These findings are consistent with the results of the orthogonal R-loop assays in Fig. 2, as well as the rifampin resistance assays in Fig. 1, and support a model in which CBEs with higher k_{cat}/K_m values for ssDNA have a greater propensity for Cas9-independent deamination in cells.

Human cell ssDNA deamination assay. As an additional independent assay of Cas9-independent deamination by CBEs in mammalian cells, we also measured intracellular deamination frequencies from BE4, A3A-BE4, YE1-BE4, YEE-BE4 and R33A + K34A-BE4 of a cotransfected 164-mer ssDNA oligonucleotide containing 35 cytosines in HEK293T cells, in light of previous reports that endogenous deaminases can induce mutagenesis in transfected ssDNA oligonucleotides (Supplementary Fig. 8)[39]. We observed that A3A-BE4 showed 4.4-fold higher Cas9-independent off-target editing compared with BE4, while YE1-BE4, YEE-BE4 and R33A + K34A-BE4 showed 1.7-, 3.2- and 1.4-fold lower average Cas9-independent off-target editing relative to BE4 at the twelve 5′-T<u>C</u> cytosines present in the oligonucleotide that were deaminated above background (Supplementary Fig. 8), again concordant with findings from the other assays. Taken together, the results from the rifampin and HSV-TK resistance assays in bacteria, orthogonal R-loop assay in human cells, kinetic assay in vitro and ssDNA deamination assay in human cells are consistent with a model in which CBEs with deaminases that have a low intrinsic catalytic efficiency (k_{cat}/K_m) for cytosine-containing ssDNA substrates exhibit lower Cas9-independent off-target deamination.

Comparison of adenine and cytosine base editors. Previous studies that detected Cas9-independent off-target DNA editing by CBEs did not detect off-target editing induced by the canonical adenine base editor, ABE[12,13], so ABE should produce minimal off-target editing in our assays. Indeed, in the rifampin and HSV-TK resistance assays, ABE induced background levels of resistance, and in the orthogonal R-loop and intracellular ssDNA deamination assays, ABE induced only very low levels of off-target A•T-to-G•C editing (Supplementary Fig. 9). Therefore, low off-target activity as assessed by the methods developed in this study is consistent with low off-target activity as assessed by previous WGS studies[12,13].

Each deaminase domain tested has a distinct on-target editing and off-target editing profile, which is shown in Fig. 3a,b. Of the CBEs that we identified as being especially promising for minimizing Cas9-independent editing, YE1-BE4 and R33A-BE4 offer the best balance between decreased off-target editing and robust on-target activity (Fig. 3b and Supplementary Fig. 10). Meanwhile, YE2-BE4, EE-BE4, R33A + K34A-BE4 and YEE-BE4 produce even lower off-target editing but with a significant decrease in average on-target activity tested across six sites (Fig. 3b and Supplementary Fig. 10).

WGS of treated human cells. To further validate that our methods are representative of genome-wide Cas9-independent off-targets, we performed WGS of HEK293T cells treated with BE4, YE1-BE4 or a Cas9 D10A nickase control. Four days following transfection with an sgRNA plasmid and a plasmid encoding BE4, YE1-BE4 or nCas9(D10A) cotranslationally fused to GFP, we isolated the top ~25% of GFP-positive cells by flow cytometry, diluted single cells into individual wells and grew them into clonal populations for 16 d before genomic DNA extraction. This approach ensured that CBE-derived off-target mutations would be present at high allele frequencies within the clonal samples derived from a single CBE-treated cell (Supplementary Fig. 11). We performed WGS at an average depth of 77× on all samples and determined all single-nucleotide variants (SNVs) present in each sample using the intersection of variants called by three algorithms (Supplementary Fig. 11 and Supplementary Tables 3 and 4). To restrict our analysis to SNVs that were generated following CBE treatment, we filtered out SNVs that were present in the original clonal population of cells before CBE treatment.

WGS results revealed that BE4, but not YE1, produced significantly more C•G-to-T•A SNVs than the Cas9 nickase-only negative control (Fig. 3c). These observations confirmed the findings of Yang, Gao and their respective coworkers that CBEs containing wild-type APOBEC1 produce off-target C•G-to-T•A SNVs in a Cas9-independent manner. We also found that BE4-treated samples contained more non-C•G-to-T•A SNVs than YE1 or nickase samples (Fig. 3d and Supplementary Fig. 12), consistent with previous reports that deaminase overexpression in HEK293 cells leads to overall increased SNVs of all types[40]. The frequency of BE4-mediated off-target edits that we observed (8.0×10^{-6} per bp) was also much higher than either of the previously reported values (5×10^{-8} per bp and 5.3×10^{-7} per bp reported by Yang and Gao, respectively). This difference likely arises from different delivery methods, our sorting of cells to isolate those that express CBEs most highly, our clonal expansion approach to maximizing SNV detection sensitivity and the different cell types used. Importantly, the above WGS results confirmed the findings of other assays: YE1 exhibits significantly reduced Cas9-independent off-target editing compared with BE4; indeed, YE1 treatment did not lead to statistically significant differences relative to the Cas9 nickase-only control (Fig. 3d and Supplementary Fig. 12).

Engineering CBE variants with minimal off-target activity and expanded targeting scope. Because YE1 and the CBE variants that we assessed to have minimal Cas9-independent off-target activity all exhibit narrowed on-target DNA-editing windows[31] (YE1-BE4, YE2-BE4, YEE-BE4 and EE-BE4), or a specific DNA sequence context requirement[32] (R33A + K34A-BE4), we sought to expand the targeting scope of these CBEs to increase their overall utility. We tested whether these deaminases are compatible with SpCas9-NG, one of two recently reported Cas9 variants that recognize a broadened NG PAM[41,42], and found that YE1, and to a lesser extent YE2, YEE, EE and R33A + K34A, maintained compatibility with SpCas9-NG nickase (Fig. 4a). YE1-NG expands the targeting scope of CBEs while maintaining substantially decreased Cas9-independent off-target activity (Supplementary Fig. 13).

Next, we replaced the SpCas9 nickase domain of YE1-BE4, YE2-BE4, YEE-BE4, EE-BE4 and R33A + K34A-BE4 with CP1028, a circularly permuted SpCas9 variant[43]. We recently reported that some circularly permutated Cas9 variants can widen or shift the on-target editing window of CBEs and ABEs[4,43]. Indeed, in HEK293T cells at a variety of endogenous loci, we observed that YE1-BE4-CP1028, YE2-BE4-CP1028 and EE-BE4-CP1028 exhibit base-editing activity windows shifted towards the PAM compared with that of nonpermuted YE1-BE4 (Fig. 4b and Supplementary Fig. 14). Collectively, YE1-BE4 and YE1-BE4-CP1028 enable targeting of nearly all cytosines present in the original base-editing activity window of BE4, with the exception of sites that contain long multi-C repeats (Supplementary Fig. 15). In addition, YEE-BE4-CP1028 and R33A + K34A-BE4-CP1028 were also active at a subset of sites tested and showed shifted editing windows at those sites (Fig. 4b).

Variants such as YEE-BE4 and R33A + K34A-BE4 are intriguing in that they offer extremely low, if any, off-target deamination in our orthogonal R-loop assay, but they are only active at a subset of on-target sites. To further increase the target sequence compatibility of R33A + K34A-BE4, which exhibits a 5′-T**<u>C</u>** requirement for base editing, we incorporated H122L and D124N, two mutations that we recently found during the continuous evolution of APOBEC1 to enable efficient deamination of 5′-G**<u>C</u>** substrates[30]. The resulting R33A + K34A + H122L + D124N-BE4 variant (referred to as AALN-BE4) indeed changed the profile of targetable C's relative to the original R33A + K34A variant, enabling editing of some positions that were not accessed by R33A + K34A-BE4

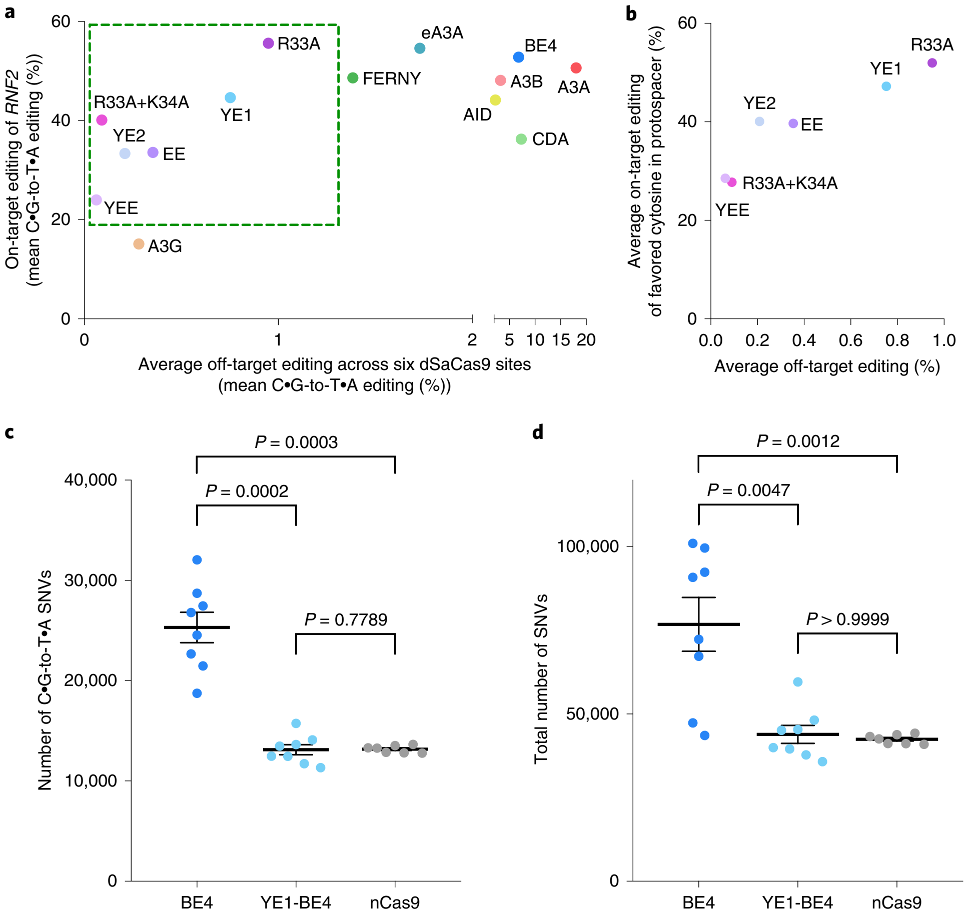

Fig. 3 | YE1 balances efficient on-target editing with greatly decreased Cas9-independent editing as confirmed by WGS. a, On-target editing versus average off-target editing for all CBEs in this study. The *y* axis reflects the mean on-target control editing at the on-target *RNF2* locus used in the orthogonal R-loop assay, and the *x* axis reflects the mean off-target editing for six orthogonal R-loops. The green box indicates CBE variants that have substantially decreased Cas9-independent off-target editing but retain appreciable on-target activity. See Supplementary Fig. 5 and Fig. 2d for mean values and s.e.m. at individual sites. **b**, Average maximum on-target and average off-target editing for constructs with decreased Cas9-independent editing events. The *y* axis reflects average editing across six on-target protospacers of the most highly edited cytosine within that protospacer. The *x* axis reflects the average off-target editing in the orthogonal R-loop assay. See Supplementary Fig. 10 and Fig. 2d for mean values and s.e.m. at individual sites. **c**, Number of C•G-to-T•A SNVs relative to the initial parent sample detected by WGS. **d**, Total number of SNVs relative to the initial parent sample detected by WGS. For both **c** and **d**, each dot represents the number of SNVs called in a clonal population of cells relative to the parent sample. Each clonal population was derived from a single GFP-positive cell that was isolated after flow sorting HEK293T cells transfected with a CBE–P2A–GFP construct for the GFP-positive cells. Horizontal lines and error bars indicate mean number of SNVs ±s.e.m. for $n=8$ (BE4 and YE1) or $n=7$ (nCas9). *P* values were calculated using the two-sided Mann–Whitney *U*-test.

(Supplementary Fig. 16). Importantly, the AALN variant maintains the minimized levels of Cas9-independent deamination shown by R33A + K34A-BE4, and circularly permuted variants likewise displayed Cas9-independent deamination levels equivalent to or lower than their unpermuted counterparts (Supplementary Figs. 17 and 18). This result indicates that deaminases with the lowest number of off-target edits can be engineered to enhance their targeting scope without disrupting their minimal off-target editing profile.

Next, we assessed whether the CBEs that exhibit minimal Cas9-independent deamination have altered propensities to generate other unwanted editing outcomes, such as indels and Cas9-dependent off-target DNA base editing. We observed that all of these variants (YE1-BE4, YE2-BE4, YEE-BE4, EE-BE4, R33A + K34A-BE4, R33A-BE4, AALN-BE4, the CP1028 variants of the first five of these variants and the Cas9-NG variants of the same five CBEs), induce lower or comparable levels of indels relative to BE4 across all on-target genomic sites tested in this study (Supplementary Fig. 19). Moreover, all seven CBE variants (YE1-BE4, YE1-BE4-CP1028, YE2-BE4, EE-BE4, YEE-BE4, R33A + K34A-BE4 and AALN-BE4) showed much lower levels of Cas9-dependent off-target DNA editing than BE4 when tested at 20 genomic sites previously identified by genome-wide unbiased identification of DSBs enabled by sequencing (GUIDE-seq)[44] to be the most highly edited off-target substrates of SpCas9 nuclease for three target loci (Supplementary Fig. 20). In addition, we note that YE1-BE3, R33A-BE3 and R33A + K34A-BE3 were recently found to exhibit substantially reduced levels of transcriptome-wide Cas9-independent RNA off-target editing compared with BE3 (refs. [32,45]). We confirmed that these variants exhibit decreased Cas9-independent off-target editing of three abundant RNA transcripts and found that YEE also shows decreased RNA off-target editing (Supplementary Fig. 21). These results collectively indicate that the CBEs that minimize Cas9-indepdendent off-target editing do not suffer from higher levels of other forms of unwanted editing; in general, they give rise to fewer indels, less Cas9-dependent DNA off-target editing and less RNA off-target editing.

Collectively, the expanded targeting capabilities of engineered YE1 variants (YE1-BE4, YE1-BE4-CP1028 and YE1-BE4-NG)

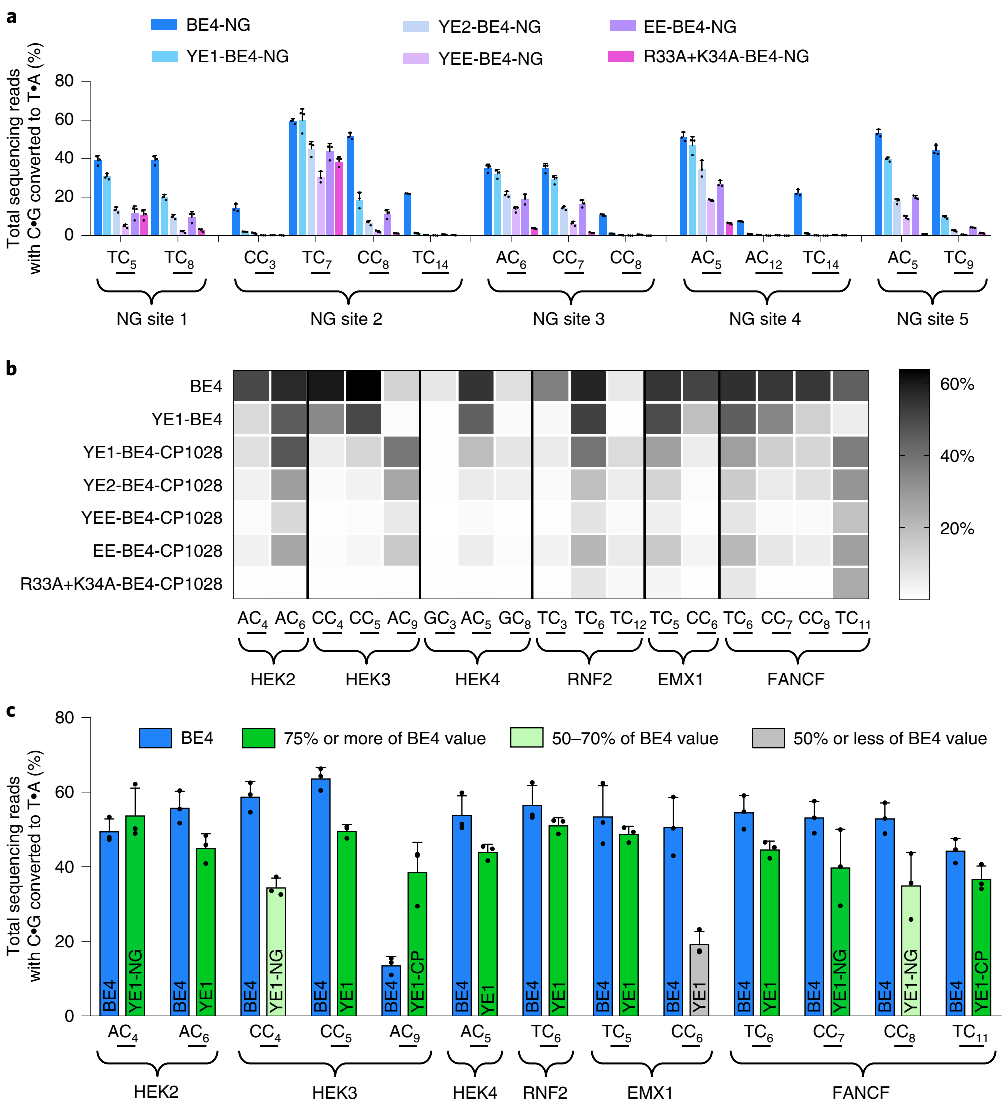

Fig. 4 | Expanding the utility of CBEs with decreased Cas9-independent off-targets through protein engineering. a, On-target base editing by different Cas9-NG-targeted CBEs across five different sites with NG PAMs (non-NGG) in HEK293T cells. Individual data points and mean values ± s.e.m. are shown for $n = 3$ independent biological replicates. **b**, On-target base editing by circularly permuted CBEs across six sites in HEK293T cells. Intensities reflect the mean editing value from three biological replicates, which are shown in Supplementary Fig. 14. **c**, Summary of the on-target editing of BE4 and engineered YE1 variants across six test loci. Blue bars indicate BE4 values for particular cytosines, and green bars represent YE1 variants that can be used to target those cytosines. The name of each YE1 variant used is indicated on the bar. Shading reflects the efficiency that each cytosine can be edited by a YE1 variant when compared with BE4. Individual data points and mean values ± s.e.m. for $n = 3$ independent biological replicates are shown.

enable targeting, in principle, of 65% of pathogenic SNPs in ClinVar that can be corrected by a C•G-to-T•A edit, compared with the only 19% that can be targeted by SpCas9-YE1max alone (Supplementary Fig. 22). The known pathogenic SNPs that can be targeted by these engineered CBEs include the vast majority (~80%) of pathogenic SNPs that can be targeted with the most broadly targetable current-generation BE4max variants, and far outnumber the SNPs targetable by SpCas9-BE4max alone, the most widely used CBE (Supplementary Fig. 22).

Analysis of expression levels and off-target editing by ribonucleoproteins. Finally, we explored how base editor expression and exposure contribute to Cas9-independent off-target editing. Western blots of BE4, YE1-BE4, YE1-BE4-NG and A3A-BE4 in HEK293T cells revealed that the expression levels of YE1-BE4 and YE1-BE4-NG were comparable to that of BE4. However, A3A-BE4 had drastically reduced expression compared with the other three editors, in stark contrast with its higher levels of off-target editing (Supplementary Fig. 23). We then transfected a variant of BE4 with only one, as opposed to two, nuclear localization signals. This construct should have lower levels of CBE trafficked to the nucleus, and therefore lower effective dosing. Indeed, we saw decreased off-target editing in the R-loop assay when we included only one NLS (Supplementary Fig. 24). This collection of experiments conveys that while expression of a base editor influences Cas9-independent off-target editing, it cannot fully explain the propensity of an editor to perform Cas9-independent deamination.

These results also suggested that limiting the time of exposure to the base editor through protein delivery might decrease Cas9-independent off-target editing. Therefore, we delivered a 1× NLS-BE4 construct into HEK293T cells as a protein–RNA complex and measured levels of orthogonal R-loop deamination: average Cas9-independent deamination decreased 21-fold relative to plasmid delivery, while retaining similar on-target editing efficiencies

(Supplementary Fig. 24). Therefore, even if a specific target can only be edited to an acceptable level by a BE4-like CBE that uses a deaminase with a high k_{cat}/K_m, protein delivery may still provide a path forward to minimize Cas9-independent off-target editing.

Discussion

The assays developed and applied in this study enable rapid and cost-effective profiling of base editors for Cas9-independent deamination of DNA in bacteria and mammalian cells, and complement in vivo methods such as those performed by Yang, Gao and their respective coworkers[12,13]. The WGS data collected in this study validate that these assays are representative of genome-wide off-target DNA mutagenesis rates, and suggest that those CBEs that show low off-target editing in these assays should exhibit low levels of genome-wide off-targets. We anticipate that the assays used here will provide a valuable means of evaluating many CBE variants much more efficiently and with much lower costs than experiments that require extensive WGS[12,13].

The many deaminases and CBEs characterized and generated in this study collectively form a landscape of base-editing options with different on-target and off-target editing characteristics, plotted in Fig. 3a. Given this landscape, and the fact that the 5×10^{-8} per bp mutation rate attributed to Cas9-independent deamination by BE3 in mouse embryos[12] is lower than the observed rate of spontaneous mutation in many mammalian somatic cell types in vivo[46–50], the optimal choice of base editor depends strongly on a given application's on-target sequence context, on-target PAM availability, target tissue type and the extent to which minimizing low levels of Cas9-independent deamination is critical. For applications in which off-target editing must be strictly minimized, we recommend YE1-BE4, YE2-BE4, YEE-BE4, EE-BE4, R33A + K34A-BE4, YE1-BE4-CP1028, YE1-BE4-NG and AALN-BE4 variants, each of which offer ~10–100-fold lower levels of Cas9-independent off-target DNA editing (Figs. 1–3), ~5–50-fold lower levels of Cas9-dependent off-target DNA editing (Supplementary Fig. 20) and lower or similar levels of indel formation (Supplementary Fig. 19), while maintaining ~50–90% of average on-target DNA-editing levels (Figs. 3a,b and 4c) relative to BE4max. Additionally, base editor exposure may also be limited to achieve lower off-target editing. Collectively, the diverse targeting capabilities of this suite of CBEs, especially those that utilize the YE1 deaminase domain, enable high-fidelity base editing at the vast majority of previously accessible target sites with efficiencies approaching those of BE4 (Supplementary Fig. 22 and Fig. 4c).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-020-0414-6.

Received: 24 November 2019; Accepted: 7 January 2020;
Published online: 10 February 2020

References

1. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420–424 (2016).
2. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell* **168**, 20–36 (2017).
3. Rees, H. A. & Liu, D. R. Base editing: precision chemistry on the genome and transcriptome of living cells. *Nat. Rev. Genet.* **19**, 770–788 (2018).
4. Paz Zafra, M. S. et al. Optimized base editors enable efficient editing in cells, organoids and mice. *Nat. Biotechnol.* **36**, 888–893 (2018).
5. Kim, K. et al. Highly efficient RNA-guided base editing in mouse embryos. *Nat. Biotechnol.* **35**, 435–437 (2017).
6. Zhang, Y. et al. Programmable base editing of zebrafish genome using a modified CRISPR–Cas9 system. *Nat. Commun.* **8**, 118 (2017).
7. Zong, Y. et al. Precise base editing in rice, wheat and maize with a Cas9–cytidine deaminase fusion. *Nat. Biotechnol.* **35**, 438–440 (2017).
8. Gaudelli, N. M. et al. Programmable base editing of A*T to G*C in genomic DNA without DNA cleavage. *Nature* **551**, 464–471 (2017).
9. Kim, D. et al. Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. *Nat. Biotechnol.* **35**, 475–480 (2017).
10. Liang, P. et al. Genome-wide profiling of adenine base editor specificity by EndoV-seq. *Nat. Commun.* **10**, 67 (2019).
11. Rees, H. A. et al. Improving the DNA specificity and applicability of base editing through protein engineering and protein delivery. *Nat. Commun.* **8**, 15790 (2017).
12. Zuo, E. S. et al. Cytosine base editor generates substantial off-target single-nucleotide variants in mouse embryos. *Science* **364**, 289–292 (2019).
13. Jin, S. Z. et al. Cytosine, but not adenine, base editors induce genome-wide off-target mutations in rice. *Science* **364**, 292–295 (2019).
14. McGrath, E. et al. Targeting specificity of APOBEC-based cytosine base editor in human iPSCs determined by whole genome sequencing. *Nat. Commun.* **10**, 5353 (2019).
15. Badran, A. H. & Liu, D. R. Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nat. Commun.* **6**, 8425 (2015).
16. Garibyan, L. Use of the rpoB gene to determine the specificity of base substitution mutations on the *Escherichia coli* chromosome. *DNA Repair* **2**, 593–608 (2003).
17. Harris, R. S., Petersen-Mahrt, S. K. & Neuberger, M. S. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol. Cell* **10**, 1247–1253 (2002).
18. Kohli, R. M. et al. A portable hot spot recognition loop transfers sequence preferences from APOBEC family members to activation-induced cytidine deaminase. *J. Biol. Chem.* **284**, 22898–22904 (2009).
19. Lee, H., Popodi, E., Tang, H. & Foster, P. L. Rate and molecular spectrum of spontaneous mutations in the bacterium *Escherichia coli* as determined by whole-genome sequencing. *Proc. Natl Acad. Sci. USA* **109**, E2774–E2783 (2012).
20. Fukui, K. DNA mismatch repair in eukaryotes and bacteria. *J. Nucleic Acids* **2010**, 260512 (2010).
21. Saraconi, G. S., Sala, C., Mattiuz, G. & Conticello, S. G. The RNA editing enzyme APOBEC1 induces somatic mutations and a compatible mutational signature is present in esophageal adenocarcinomas. *Genome Biol.* **15**, 417 (2014).
22. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. *Science* **353**, aaf8729–aaf8729 (2016).
23. Ma, Y. et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nat. Methods* **13**, 1029–1035 (2016).
24. Hess, G. T. et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nat. Methods* **13**, 1036–1042 (2016).
25. Wang, X. et al. Efficient base editing in methylated regions with a human APOBEC3A–Cas9 fusion. *Nat. Biotechnol.* **36**, 946–949 (2018).
26. Coelho, M. A. et al. BE-FLARE: a fluorescent reporter of base editing activity reveals editing characteristics of APOBEC3A and APOBEC3B. *BMC Biol.* **16**, 150 (2018).
27. 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 Res.* **46**, e84 (2018).
28. Martin, A. S. et al. A panel of eGFP reporters for single base editing by APOBEC-Cas9 editosome complexes. *Sci. Rep.* **9**, 497 (2019).
29. Liu, Z. et al. Highly precise base editing with CC context-specificity using engineered human APOBEC3G-nCas9 fusions. *bioRxiv* https://www.biorxiv.org/content/10.1101/658351v1 (2019).
30. Thuronyi, B. W. K. et al. Continuous evolution of base editors with expanded target compatibility and improved activity. *Nat. Biotechnol.* **37**, 1070–1079 (2019).
31. Kim, Y. B. et al. Increasing the genome-targeting scope and precision of base editing with engineered Cas9–cytidine deaminase fusions. *Nat. Biotechnol.* **35**, 371–376 (2017).
32. Grunewald, J. et al. Transcriptome-wide off-target RNA editing induced by CRISPR-guided DNA base editors. *Nature* **569**, 433–437 (2019).
33. Gehrke, J. M. et al. An APOBEC3A–Cas9 base editor with minimized bystander and off-target activities. *Nat. Biotechnol.* **36**, 977–982 (2018).
34. Tashiro, Y., Fukutomi, H., Terakubo, K., Saito, K. & Umeno, D. A nucleoside kinase as a dual selector for genetic switches and circuits. *Nucleic Acids Res.* **39**, e12 (2011).
35. Koblan, L. W. et al. Improving cytidine and adenine base editors by expression optimization and ancestral reconstruction. *Nat. Biotechnol.* **36**, 843–846 (2018).
36. Chan, K. et al. An APOBEC3A hypermutation signature is distinguishable from the signature of background mutagenesis by APOBEC3B in human cancers. *Nat. Genet.* **47**, 1067–1072 (2015).

37. Eto, T., Kinoshita, K., Yoshiwaka, K., Muramatsu, M. & Honjo, T. RNA-editing cytidine deaminase Apobec-1 is unable to induce somatic hypermutation in mammalian cells. *Proc. Natl Acad. Sci. USA* **100**, 12895–12898 (2003).
38. Carpenter, M. A. et al. Methylcytosine and normal cytosine deamination by the foreign DNA restriction enzyme APOBEC3A. *J. Biol. Chem.* **287**, 34801–34808 (2012).
39. Lei, L. et al. APOBEC3 induces mutations during repair of CRISPR–Cas9-generated DNA breaks. *Nat. Struct. Mol. Biol.* **25**, 45–52 (2018).
40. Akre, M. K. et al. Mutation processes in 293-based clones overexpressing the DNA cytosine deaminase APOBEC3B. *PLoS ONE* **11**, e0155391 (2016).
41. Nishimasu, H. S. et al. Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).
42. Hu, J. H. et al. Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).
43. Huang, T. P. et al. Circularly permuted and PAM-modified Cas9 variants broaden the targeting scope of base editors. *Nat. Biotechnol.* **37**, 626–631 (2019).
44. Tsai, S. Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33**, 187–197 (2015).
45. Zhou, C. et al. Off-target RNA mutation induced by DNA base editing and its elimination by mutagenesis. *Nature* **571**, 275–278 (2019).
46. Hazen, J. L. et al. The complete genome sequences, unique mutational spectra, and developmental potency of adult neurons revealed by cloning. *Neuron* **89**, 1223–1236 (2016).
47. Milholland, B. et al. Differences between germline and somatic mutation rates in humans and mice. *Nat. Commun.* **8**, 15183 (2017).
48. Dong, X. et al. Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat. Methods* **14**, 491–493 (2017).
49. Lynch, M. Evolution of the mutation rate. *Trends Genet.* **26**, 345–352 (2010).
50. Rahbari, R. et al. Timing, rates and spectra of human germline mutation. *Nat. Genet.* **48**, 126–133 (2016).

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

Methods

Cloning. All plasmids for this study were created using either Uracil-Specific Excision Reagent (USER) cloning or kinase, ligase, DpnI (KLD) cloning as described previously[1]. DNA was amplified using PhusionU Green Multiplex PCR Master Mix (Thermo Fisher Scientific). Mach1 (Invitrogen) or Turbo (New England BioLabs) chemically competent *E. coli* were used for plasmid construction.

Preparation and transformation of chemically competent *E. coli*. Commercially available chemically competent BL21 *E. coli* (New England BioLabs) were transformed with a plasmid harboring an inactivated chloramphenicol resistance gene. Transformed cells were plated on LB medium + 1.5% agar supplemented with maintenance antibiotic (kanamycin, 30 µg ml^{-1}). The following day, a single colony was picked and grown overnight in 2× YT medium supplemented with maintenance antibiotic. The overnight culture was diluted 100-fold into 50 ml of 2× YT medium supplemented with maintenance antibiotic and grown at 37 °C with shaking at 230 r.p.m. to optical density $(OD)_{600} \approx 0.4$–0.6. Cells were collected by centrifugation at 3,400*g* for 10 min at 4 °C. The cell pellet was resuspended by gentle stirring in 2.5 ml of cold LB medium followed by 2.5 ml of 2× TSS (LB medium supplemented with 5% v/v dimethylsulfoxide, 10% w/v PEG 3350 and 20 mM $MgCl_2$). After thorough resuspension, cells were aliquoted, frozen on dry ice and stored at −80 °C until use.

To transform cells, 100 ml of competent cells thawed on ice were added to a prechilled mixture of plasmid (2 µl) in 98 µl of KCM solution (100 mM KCl, 30 mM $CaCl_2$ and 50 mM $MgCl_2$ in H_2O). The mixture was incubated on ice for 20 min and heat shocked at 42 °C for 75 s followed by addition of 500 µl of SOC medium (New England BioLabs). Cells were recovered at 37 °C with shaking at 230 r.p.m. for 1 h, streaked on 2× YT medium + 1.5% agar plates containing the appropriate antibiotics and incubated at 37 °C for 14–16 h.

Rifampin assay. Chemically competent *E. coli* harboring a plasmid encoding an inactivated chloramphenicol resistance gene were transformed with a plasmid encoding a base editor + guide RNA. Transformed cells were plated on maintenance antibiotics (30 µg ml^{-1} kanamycin, 50 µg ml^{-1} spectinomycin, with no chloramphenicol). The following day, colonies were picked and grown overnight in Davis rich medium (DRM) and maintenance antibiotics. Overnight cultures were diluted 1:100 into DRM and maintenance antibiotics and grown at 37 °C with shaking at 230 r.p.m. When cells reached $OD_{600} = 0.5$, 5 mM rhamnose was added to induce base editor expression. After 18 h, 700 µl of each culture was centrifuged at 3,400*g* for 10 min and the cell pellet was resuspended in 150 µl of total DRM. Serial dilutions in H_2O of each resuspended culture were plated on three different conditions in parallel: (1) 2× YT agar + 30 µg ml^{-1} kanamycin + 50 µg ml^{-1} spectinomycin + 20 mM glucose, (2) 2× YT agar + 30 µg ml^{-1} kanamycin + 50 µg ml^{-1} spectinomycin + 20 mM glucose + 100 µg ml^{-1} rifampin or (3) 2× YT agar + 30 µg ml^{-1} kanamycin + 50 µg ml^{-1} spectinomycin + 20 mM glucose + 10 µg ml^{-1} chloramphenicol. Surviving colonies were counted following an incubation at 37 °C for 24 h after plating. To obtain survival rates, the number of colonies in the chloramphenicol or rifampin conditions was divided by the number of colonies counted on the maintenance antibiotic plate.

Sanger sequencing of *rpoB* mutations from rifampin-resistant colonies. Rifampin-resistant colonies were picked into 10 µl of H_2O and heated at 95 °C for 10 min, followed by PCR using primers AB1678 (5′-AATGTCAAATCCGTGGCGTGAC) and AB1682 (5′-TTCACCCGGATACATCTCGTCTTC). Each fragment was sequenced twice using primers AB1680 (5′-CGGAAGGCACCGTAAAAGACAT) and AB1683 (5′-CGTGTAGAGCGTGCGGTGAAA).

HSV-TK assay. Lambda red recombineering was performed as described previously[51] to chromosomally integrate a single copy of the HSV-TK gene under a constitutive promoter and β-lactamase into the *tonB* locus of BL21 *E. coli*. The resulting strain was transformed with a plasmid encoding a base editor + guide RNA. Transformed cells were plated on plasmid maintenance antibiotics (50 µg ml^{-1} carbenicillin, 50 µg ml^{-1} spectinomycin). The following day, colonies were picked and grown overnight in DRM and maintenance antibiotics. Overnight cultures were diluted 1:100 into DRM and maintenance antibiotics and grown at 37 °C with shaking at 230 r.p.m. When cells reached $OD_{600} = 0.5$, 5 mM rhamnose was added to induce base editor expression. After 18 h, 700 µl of each culture was centrifuged at 3,400*g* for 10 min and the cell pellet was resuspended in 150 µl of total DRM. Serial dilutions in H_2O of each resuspended culture were plated on two different conditions in parallel: (1) 2× YT agar + 50 µg ml^{-1} carbenicillin + 50 µg ml^{-1} spectinomycin + 20 mM glucose or (2) 2× YT agar + 50 µg ml^{-1} carbenicillin + 50 µg ml^{-1} spectinomycin + 20 mM glucose + 10 µM dP. Surviving colonies were incubated at 37 °C for 24 h after plating, then counted. To obtain survival rates, the number of colonies in the dP condition was divided by the number of colonies counted on the maintenance antibiotic plate.

Sanger sequencing of HSV-TK mutations from dP-resistant colonies. dP-resistant colonies were picked into 10 µl of H_2O and heated at 95 °C for 10 min, followed by PCR using primers AR393 (5′-AGGCAGTGGGATTGTGGTG) and AR394 (5′-CGGTCAGCATTAATATTGAAGTGTGG). Each fragment was sequenced three times using primers AB301 (5′-ATAAAGTTGCAGGACCACTTCT), AR341 (5′- GCAAGCAGCCCGTAAAC) and AR392 (5′-CGTACGTCGGTTGCTATG).

Analysis of BE3-induced point mutations in mouse embryos reported by Yang and coworkers. Using the genomic locations of all C•G-to-T•A SNVs reported by Yang and coworkers in Supplementary Tables 6 and 7 of their recent work[12], the flanking sequences (20 bp on either side) were extracted from the mouse mm10 reference genome (NCBI accession code GCA_000001635.2). These flanking sequences were aligned, fixing the mutant cytosine in each case at position 21, and the resulting alignment was used to produce a sequence logo using WebLogo v.3.6.0 (ref. [52]). The custom Python script used for this analysis is included in Supplementary Note 1.

Cell culture. HEK293T cells were maintained in DMEM + GlutaMAX (Life Technologies) supplemented with 10% (v/v) fetal bovine serum. Cells were cultured at 37 °C with 5% carbon dioxide and were confirmed to be negative for mycoplasma by testing with MycoAlert (Lonza Biologics).

Mammalian cell transfections. HEK293T cells were seeded in a 48-well, poly-D-lysine-coated plate (Corning) and transfected at 70% confluence. Plasmids were prepared for transfection using either a ZymoPURE II midi prep kit (Zymo Research Corporation) or a Qiagen midi prep kit (Qiagen). For on-target editing experiments, 750 ng of base editor plasmid and 250 ng of guide RNA plasmid were cotransfected into HEK293T cells using 1.5 µl of Lipofectamine 2000 (Thermo Fisher Scientific) per well as directed by the manufacturer. As a transfection control, 20 ng of pmaxGFP transfection control plasmid (Lonza Biologics) was used. For orthogonal R-loop assays to measure off-target editing, 200 ng of SpCas9 guide RNA plasmid, 200 ng of SaCas9 guide RNA plasmid, 300 ng of base editor plasmid and 300 ng of dSaCas9 plasmid were cotransfected into HEK293T cells using 1.5 µl of Lipofectamine 2000. For controls involving no base editor or no sgRNA, pUC19 DNA was used to maintain the total quantity of transfected DNA at 1,000 ng. For the intracellular oligonucleotide deamination experiment, 750 ng of base editor plasmid, 250 ng of guide RNA plasmid and 1 pmol of ssDNA oligonucleotide (Integrated DNA Technologies) were cotransfected into HEK293T cells using 1.5 µl of Lipofectamine 2000.

High-throughput sequencing of genomic DNA. Genomic DNA was sequenced using methods previously described[1]. Briefly, genomic DNA was isolated from HEK293T cells 3 d after transfection. Cells were washed with PBS and then lysed with 150 µl of lysis buffer consisting of 10 mM Tris-HCl (pH 7), 0.05% SDS and 25 µg ml^{-1} Proteinase K (Thermo Fisher Scientific) at 37 °C for 1 h and then heat inactivated at 80 °C for 30 min. Following lysis, 1 µl of the genomic DNA lysate was used as input for the first of two PCR reactions. Genomic loci were amplified using a PhusionU PCR kit (Life Technologies). PCR1 primers for genomic loci are listed in Supplementary Table 1 under the HTS_fwd and HTS_rev columns. Thirty cycles of PCR were performed for all loci with an annealing temperature of 61 °C and an extension time of 30 s. For sequencing of the cotransfected ssDNA oligonucleotide, 22 cycles of PCR1 were performed. PCR1 products were confirmed on a 2% agarose gel. Then, 1 µl of PCR1 was used as an input for PCR2 to install Illumina barcodes. PCR2 was conducted using a Phusion HS II kit (Life Technologies). Following PCR2, samples were pooled and gel extracted in a 2% agarose gel using a Qiaquick Gel Extraction Kit (Qiagen). Library concentration was quantified using the Qubit High-Sensitivity Assay Kit (Thermo Fisher Scientific). Samples were sequenced on an Illumina MiSeq instrument (paired-end read, read 1: 200–280 cycles, read 2: 0 cycles) using an Illumina 300 v2 Kit (Illumina).

High-throughput sequencing data analysis. Sequencing reads were demultiplexed using the MiSeq Reporter (Illumina) and fastq files were analyzed using Crispresso2 (ref. [53]). Representative analysis input and usage are described in Supplementary Note 2. Prism 8 (GraphPad) was used to generate dot plots and bar plots of these data. Base-editing values are representative of $n = 3$ independent biological replicates performed at different times, generally by different researchers, with the mean ± s.e.m. shown.

Protein expression and purification for in vitro assays. Base editor purification was performed as described previously[11], with a few modifications. BL21DE3* (Thermo Fisher Scientific) chemically competent *E. coli* were transformed with a plasmid encoding N-terminally 6× His-tagged base editor under control of an isopropyl-β-D-thiogalactoside (IPTG)-induced T7 promoter. Individual colonies were picked and grown in 1 l of 2× YT medium until $OD_{600} \approx 0.7$–0.8. Cells were cold shocked on ice for 1–2 h, then induced with 1 mM IPTG (Gold Biotechnology) and grown for a further 12–16 h at 16 °C with shaking at 220 r.p.m. Cells were collected by centrifugation at 6,000*g* for 20 min and the resulting cell pellet was resuspended in 25 ml of high-salt buffer (100 mM Tris–Cl pH 8.0, 1 M NaCl, 5 mM tris(2-carboxyethyl)phosphine (TCEP; Sigma-Aldrich), 20% glycerol) supplemented with 0.4 mM phenylmethane sulfonyl fluoride (PMSF; Sigma-Aldrich) and EDTA-free protease inhibitor pellet (Roche; 1 pellet per 50 ml of lysis buffer used). Cells were lysed by sonication (6 min total, 3 s on, 3 s off) and the lysate was cleared by centrifugation at 22,000g for 20 min. The cleared lysate was incubated with 1.5 ml of TALON Cobalt resin (Clontech) with rotation at 4 °C for 1–2 h. The resin was washed two times with 15 ml of cold high-salt buffer and bound protein was eluted in medium-salt buffer (100 mM Tris-HCl pH 8.0, 0.5 M NaCl, 20% glycerol, 5 mM TCEP) supplemented with 200 mM imidazole. The isolated protein was then buffer-exchanged with low-salt buffer and concentrated using an Amicon Ultra-15 centrifugal filter unit (100,000 molecular weight cut-off). The isolated protein was further purified on a 5-ml Hi-Trap HP SP (GE Healthcare) cation exchange column using an Akta Pure chromatography system. Protein-containing fractions were pooled and concentrated using an Amicon Ultra-15 centrifugal filter unit (100,000 molecular weight cut-off). Proteins were quantified using Quick Start Bradford reagent (Bio-Rad) using BSA standards (Bio-Rad) and stored short-term at 4 °C.

Protein purity was characterized by SDS–PAGE analysis. Briefly, proteins were denatured at 95 °C for 10 min in Laemmli sample loading buffer (Bio-Rad) supplemented with 2 mM dithiothreitol (Sigma-Aldrich) and separated by electrophoresis at 200 V for 40 min on a Bolt 4–12% Bis-Tris Plus (Thermo Fisher Scientific) precast gel in Bolt MES SDS running buffer (Thermo Fisher Scientific). Gels were stained with InstantBlue reagent (Expedeon) for 1 h and washed several times with H_2O before imaging with a G:Box Chemi XRQ (Syngene).

In vitro deamination assays. A 5′-Cy3-labeled ssDNA oligonucleotide (5′-Cy3-ATTATTATTATTTCTATTTATTTATTTATTT) was purchased as an HPLC-purified oligonucleotide from Integrated DNA Technologies. All reactions were performed in reaction buffer[11] (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM dithiothreitol, 0.1 mM EDTA, 10 mM $MgCl_2$) with concentrations of 5′-Cy3-labeled oligonucleotide varying from 0.2 to 100 µM and concentrations of each purified base editor protein that were >20-fold lower than the substrate concentration assayed in each case. Base editor proteins were incubated at room temperature for 5 min with a nontargeting sgRNA added in a 1:1 molar ratio. Subsequently, the 5′-Cy3-labeled oligonucleotide was added to the appropriate concentration and the reactions were incubated at 37 °C for 30 min. Reactions were stopped by the addition of buffer PB (100 µl; Qiagen) and isopropanol (25 µl) and purified on a MinElute spin column (Qiagen), eluting in 15 µl of CutSmart buffer (New England BioLabs). USER enzyme (1.5 U; New England BioLabs) was added to the purified ssDNA and incubated at 37 °C for 1 h. Then, 10 µl of the resulting solution was combined with 10 µl of loading buffer (0.09 M tris(hydroxymethyl)aminomethane, 0.09 M sodium tetraborate, 10 mM EDTA pH 8.0, 10 M urea, 20% sucrose, 0.1% SDS) and loaded on a 10% Tris-borate-EDTA-urea gel (Bio-Rad) that was pre-run in 0.5× Tris-borate-EDTA buffer for 15 min at 180 V. The cleaved uracil-containing products were resolved from the uncleaved cytosine-containing starting material by electrophoresis for 30 min at 180 V, and the gel was imaged on a GE Typhoon FLA 7000 imager. The ratio of product to substrate bands was quantified by densitometry using ImageJ and used to calculate initial reaction velocities. Nonlinear regression was performed using Prism 8 (GraphPad) to fit these data to the Michaelis–Menten equation to determine k_{cat} and K_m values. Calculation of the propagated error in the k_{cat}/K_m ratio from the individual errors in the k_{cat} and K_m parameters estimated by the regression is described in Supplementary Note 3.

Transfection and fluorescence-activated cell sorting for WGS samples. HEK293T cells were transfected with 750 ng of CBE–P2A–GFP constructs and 250 ng of *RNF2*-targeting guide RNA as described in the mammalian cell transfections section of Methods. Four wells were transfected for each tested CBE or control (Cas9 nickase instead of a CBE). Four days after transfection, cells were trypsinized with 50 µl of trypsin per well, and resuspended in 200 µl of DMEM (50% FBS (v/v), 100 U ml^{-1} penicillin/streptomycin). Wells of the same editor were pooled, and cells were filtered through a cell-strainer cap (VWR International). Flow sorting was performed on a FACS Aria II (BD Biosciences) sorter using BDFACS Diva software. Cells were gated on forward/side scatter and then gated for GFP signal compared with an untransfected negative control. Cells were then gated on fluorescence intensity. Intensity gates were set to contain the top 28% of GFP-positive YE1–P2A–GFP cells, which corresponded to the top 30% of GFP-positive BE4–P2A–GFP cells and the top 45% of GFP-positive Cas9 nickase–P2A–GFP-positive cells (see Supplementary Note 5). Approximately 70,000 cells were collected for each sample in bulk. Of these, about 20,000 cells were sequenced for bulk on-target editing efficiency at the *RNF2* locus. The remaining cells were diluted to a concentration of 6 cells per ml (equivalent to 0.9 cells per well) in DMEM (10% FBS (v/v), 100 U ml^{-1} penicillin/streptomycin). Then, 150 µl of this diluted mixture was pipetted into each well of a 96-well plate. Wells were monitored daily to ensure that each population of cells came from only a single cell. Cells were split into a 48-well, poly-ᴅ-lysine-coated plate (Corning) and grown for 16 d before collecting.

WGS sample preparation. Cells were lysed using a DNA Agencourt Advance (Beckman Coulter) according to the manufacturer instructions. Briefly, 100 µl of lysis buffer (95 µl of Beckman lysis buffer, 2.5 µl of proteinase K (Thermo Fisher) and 2.5 µl of 1 M dithiothreitol) was added to each well and incubated for 5 min at 37 °C. Lysate was then transferred to PCR strips and incubated at 55 °C for 1 h. Then, 50 µl of Beckman Binding Buffer 1 (Beckman Coulter) was added, and samples were incubated for 2 min before the addition of magnetic beads contained in Beckman Binding Buffer 2 (Beckman Coulter). Samples were incubated for 5 min and then placed on a magnetic plate for 10 min. Supernatant was removed, and beads were washed twice with 70% ethanol. DNA was then resuspended in 50 µl of elution buffer. Samples were placed on a magnetic plate, and the supernatant containing the purified DNA was removed and transferred to fresh tubes. DNA yields were quantified with a Nanodrop. Libraries were created using a KAPA HyperPrep Plus kit according to the manufacturer instructions. Next, 800 ng of purified DNA per sample was diluted to a total volume of 35 µl in 10 mM Tris-HCl (pH 8). Then, 5 µl of KAPA frag buffer and 10 µl of KAPA frag enzyme were added to each reaction. Samples were placed in a precooled PCR block and then heated to 37 °C for 12 min. Immediately after 12 min, samples were placed on ice, and 7 µl of End Repair and A-tailing buffer and 3 µl of End Repair and A-tailing enzyme mix were added immediately to each sample. Samples were mixed and then heated at 20 °C for 30 min and then 65 °C for 30 min in a thermocycler with the lid temperature set to 85 °C. Following this incubation, 10 µl of DNA ligase, 30 µl of DNA ligation buffer and 10 µl of 15 µM KAPA Adapter primers were added. This mixture was then incubated at 20 °C for 15 min. A postligation cleanup was performed by adding 88 µl of KAPA Pure beads to the adapter mix. After a 10-min incubation, beads were collected on a magnetic plate, and the supernatant was discarded. Samples were washed twice with 200 µl of 80% ethanol. Beads were dried for 4 min, and 55 µl of elution buffer was added. After incubation, 50 µl of purified DNA was removed from beads.

WGS library size selection and quality control. A size selection was performed on the purified library. A 0.5× cut was performed to remove fragments greater than 1 kb: 25 µl of KAPA Pure beads were added to the eluted library, incubated and placed on a magnet. The supernatant was collected and saved. Then, 10 µl of fresh KAPA Pure beads were added to the supernatant to perform a 0.7× second cut. After incubation, libraries were placed on a magnet and the supernatant was removed and discarded. Beads were washed twice with 200 µl of 80% ethanol and then dried for 4 min. Next, 40 µl of 10 mM Tris-HCl (pH 8) was added to the beads to elute the final library. Each individual genome was quantified using the KAPA Quantification kit as described previously[1]. Library length was determined using an Agilent High Sensitivity DNA Kit and an Agilent 2100 Electrophoresis Bioanalyzer according to the manufacturer instructions. Mean fragment length for final libraries was approximately 700 bp.

WGS and data analysis. Sequencing was performed at the Broad Institute Genomics Platform on an Illumina NovaSeq 6000 using two S4 flow cells. Initial data processing and read alignment were performed by the Broad Institute Genomics Platform. Reads were demultiplexed and aligned to the hg19 (b37) reference genome using BWA-MEM (v.0.7.7) (ref. [54]). Aligned bams were sorted and optical duplicates were marked using Picard tools (v.2.21). Base quality recalibration was performed using GATK (v.3.4). All subsequent analyses were performed using the FAS RC Cannon high-performance computing cluster (Harvard University). Sequencing coverage was calculated using mosdepth (v.0.2.6) (ref. [55]). Variant calling was conducted on every sample independently using three algorithms, GATK HaplotypeCaller (v.4.1.3.0) (ref. [56]), freebayes (v.1.3.1) (ref. [57]) and VarScan (v.2.4.3) (ref. [58]), assuming a ploidy of four and a minimum alternate allele read frequency of 0.1 to call an SNV. SNVs were called on the GATK-recommended genomic intervals[56] that exclude highly repetitive regions such as centromeres and telomeres. bcftools (v.1.9) was used to find the intersection of the variants called by all three algorithms to generate high-confidence variant calls. For all treated samples, bcftools was used to filter out variants in the treated sample that were present in the parent to retain only de novo variants that arose post treatment with base editors. bcftools was also used to filter out variants present at allele frequencies greater than 0.5 as previously reported[40] to restrict analysis to variants that likely arose as a result of base editor treatment. Finally, bcftools was used to exclude variants that exhibited at least one of the following poor-quality metrics based on the GATK vcf annotations: QD < 2, FS > 60, SOR > 3, MQ < 40, MQRankSum < −5. These final, high-confidence variant calls for each treated sample were used for all downstream analyses.

RNA off-target editing analysis. HEK293T cells were transfected with 750 ng of plasmid encoding editors and 250 ng of guide RNA plasmid as described above. Cells were lysed 48 h after transfection using the RNeasy kit (Qiagen) following the manufacturer instructions. Briefly, medium was aspirated, and cells were washed with ice-cold PBS. To lyse, 350 µl of RLT buffer was added to each well. Cells were pipetted vigorously and then transferred to a DNA eliminator column. Columns were spun at 8,000g for 30 s, and 350 µl of 70% ethanol was added to the flow through, which was then applied to an RNeasy spin column. The mixture was centrifuged at 8,000g for 30 s. The column was then washed with 700 µl of RW1 buffer and then twice with 500 µl of RPE buffer. The membrane was dried by centrifuging at 8,000g for 1 min. Purified RNA was eluted with 40 µl of RNase-free water, and 2 µl of RNase-OUT (Fisher Scientific) was added. Complementary DNA was generated using SuperScript IV (Thermo Fisher Scientific). Next, 2 µl of purified RNA was combined with 1 µl of dNTPs, 1 µl of a poly T primer and 9 µl of RNase-free water. The mixture was heated to 65 °C for 5 min and then placed on ice for 1 min. Then, 4 µl of 5× superscript buffer, 1 µl of SuperScript IV reverse transcriptase, 1 µl of 0.1 M dithiothreitol and 1 µl of RNase-OUT were added. Two additional reactions were also performed, and reverse transcriptase was not added, as a control for genomic DNA contamination. Reverse transcription reactions were heated to 50 °C for 10 min, then to 80 °C for 10 min and then placed on ice. RNAse H (1 µl) was added, and the samples were heated to 37 °C for 20 min to degrade RNA. Then, 1 µl of this reaction was used as a template for the first PCR of amplicon sequencing: the remaining protocol is identical to that used for genomic DNA sequencing (see the high-throughput sequencing of genomic DNA section in Methods). Primers used for each cDNA amplicon and amplicon sequences are listed in Supplementary Table 5. The no-reverse-transcriptase controls were also subjected to Miseq prep, and it was ensured that there were negligible read counts for these samples.

Western blot analysis. HEK293T cells were transfected with 750 ng of plasmid encoding C-terminal 3× HA-tagged base editors and 250 ng of guide RNA plasmid as described above. Cells were lysed 48 h post transfection at 4 °C for 30 min in RIPA buffer (Thermo Fisher) supplemented with 1 mM PMSF (Sigma-Aldrich) and EDTA-free protease inhibitor pellet (Roche; 1 pellet per 50 ml lysis buffer used). Lysates were cleared by centrifugation at 12,000 r.p.m. for 20 min. Total protein concentration was quantified using Quick Start Bradford reagent (Bio-Rad) using BSA standards (Bio-Rad). Protein extracts were denatured at 95 °C for 10 min in Laemmli sample loading buffer (Bio-Rad) supplemented with 2 mM dithiothreitol (Sigma-Aldrich) and were separated by electrophoresis at 180 V for 40 min on a Bolt 4–12% Bis-Tris Plus (Thermo Fisher Scientific) precast gel in Bolt MES SDS running buffer (Thermo Fisher Scientific). In each well, 10 µg of total protein was loaded. Transfer to a PVDF membrane was performed using an iBlot 2 Gel Transfer Device (Thermo Fisher Scientific) according to the manufacturer's protocols. The membrane was cut in half at the 75-kDa marker and each half was blocked separately in Odyssey Blocking Buffer (LI-COR) in Tris-buffered saline (TBS) for 1 h at room temperature with rocking. The high-molecular-weight half was incubated with rabbit anti-HA (Cell Signaling Technologies 3724S; 1:1,000 dilution) in SuperBlock Blocking Buffer (Thermo Fisher Scientific) at 4 °C overnight with rocking. The low-molecular-weight half was incubated with rabbit anti-GAPDH (Cell Signaling Technologies 5174S; 1:1,000 dilution) in SuperBlock Blocking Buffer (Thermo Fisher Scientific) at 4 °C overnight with rocking. The membranes were washed twice with TBST (TBS + 0.5% Tween-20) for 10 min each time at room temperature, then incubated with goat anti-rabbit 680RD (LI-COR 926-68071) diluted 1:10,000 in SuperBlock for 1 h at room temperature. The membrane was washed as before and imaged using an Odyssey Imaging System (LI-COR).

Cell viability assay. HEK293T cells were seeded in a 96-well, clear-bottomed black plate (Corning) and transfected at 70% confluence with 200 ng of base editor plasmid, 40 ng of guide RNA plasmid and 0.5 µl of Lipofectamine 2000 (Thermo Fisher Scientific) per well. At 48 or 72 h post transfection, cell viability was measured using the CellTiter-Glo Reagent (Promega) according to the manufacturer's protocol. Luminescence was measured using an Infinite M1000 Pro microplate reader (Tecan).

Protein nucleofections. To compare on-target editing and off-target editing at orthogonal R-loops using DNA or ribonucleoprotein (RNP) delivery of base editors, cells were first lipofected as described above with the respective plasmids, supplemented to 1,000 ng total with pUC19 plasmid if base editor plasmid was not included. At 24 h after lipofection, to allow time for expression of SaCas9 and formation of the R-loop, cells that were treated only with dSaCas9 and orthogonal guide RNA plasmids were trypsinized in 50 µl of TrypLE express (Life Technologies) per well for 5 min at 37 °C. Cells were suspended and trypsin was quenched with an equal volume of fresh medium. Cells were counted in a Countess II cell counter (Thermo Fisher Scientific) and 200,000 cells per protein nucleofection sample were apportioned into a single tube. These cells were centrifuged for 8 min at 100*g*, the supernatant was discarded and cells were resuspended in 10 µl of nucleofection solution per 200,000 cells, supplemented as described by the manufacturer (Lonza, SF Cell Line 4D-Nucleofector X Kit S). RNP solutions were prepared by adding 100 pmol of chemically modified sgRNA (Synthego) to 10 µl of supplemented nucleofection solution per sample. Then, 94 pmol of BE4 protein (expressed and purified by Aldevron, and provided as a generous gift from Prof. Mark Osborn) was added to a final volume of 12 µl, and RNP complexes were formed by incubation at room temperature for 5 min. Next, 12 µl of RNP solution was mixed with 200,000 cells in 10 µl of nucleofection solution per sample, and added to a Nucleocuvette (Lonza, SF Cell Line 4D-Nucleofector X Kit S). Cells were nucleofected in a Lonza 4D Nucleofector using program CM-130 according to the manufacturer's instructions. Immediately following nucleofection, cells were recovered for 5 min by adding 80 µl of prewarmed medium. Then, 30 µl of recovered cells from each sample were diluted to a final volume of 250 µl in fresh medium and incubated at 37 °C for 2 more days before extraction of genomic DNA from all samples, including those treated only with DNA. Three different splits of cells were used in triplicate samples for each treatment.

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

Data availability

High-throughput sequencing and whole-genome sequencing data are deposited in the NCBI Sequence Read Archive (PRJNA553240). Plasmids used in this study are available from Addgene. Amino acid sequences of all base editors in this study are provided in the Supplementary Sequences.

Code availability

The script used to analyze the SNVs reported by Yang and coworkers[12] is provided in Supplementary Note 1. The script and parameters used for running CRISPResso2 analyses are provided in Supplementary Note 2. The script used for calculating the number of pathogenic SNPs targetable by CBEs is provided in Supplementary Note 4.

References

51. Thomason, L. C. et al. Recombineering: genetic engineering in bacteria using homologous recombination. *Curr. Protoc. Mol. Biol.* **106**, 1 16 11–39 (2014).
52. Crooks, G. E. et al. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).
53. Clement, K. et al. CRISPResso2: accurate and rapid analysis of genome editing data from nucleases and base editors. *Nat. Biotechnol.* **37**, 224–226 (2019).
54. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
55. Pedersen, B. S. & Quinlan, A. R. Mosdepth: quick coverage calculation for genomes and exomes. *Bioinformatics* **34**, 867–868 (2018).
56. Van der Auwera, G. A. et al. From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11–33 (2013).
57. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. Preprint at *arXiv* https://arxiv.org/abs/1207.3907 (2012).
58. Koboldt, D. C. et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.* **22**, 568–576 (2012).

Acknowledgements

We thank B. Thuronyi for discussions; K. Zhao, T. Huang and S. Miller for providing guide RNA plasmids; and B. Mok for providing an sgRNA for in vitro experiments. M. Osborn, Aldevron and the Kidz1stFund provided the BE4 protein used for protein delivery experiments, and K. Tian and S. Pandey provided assistance with experiments. We thank D. Court for providing pSIM5, which was used for recombineering the HSV-TK gene onto the *E. coli* chromosome. We thank T. Mason and E. LaRoche for assistance with WGS library preparation and FAS Research Computing at Harvard University for computational resources. This work was supported by US NIH grant nos. U01 AI142756, RM1 HG009490, R35 GM118062 and HHMI; the Bill and Melinda Gates Foundation; and the St. Jude Collaborative Research Consortium. A.R. is an NSF Graduate Research Fellow and was supported by NIH Training Grant no. T32 GM095450. J.L.D. is supported by the Hertz Foundation and the NSF GRFP fellowship. G.A.N. is a Howard Hughes Medical Institute Fellow of the Helen Hay Whitney Foundation. Flow cytometry was supported by Cancer Center Support (core) Grant no. P30-CA14051 from the NCI.

Author contributions

J.L.D., A.R. and D.R.L. designed the research. A.R. and J.L.D. performed experiments. G.A.N. performed the BE4 protein delivery off-target experiment. All authors contributed to writing the manuscript.

Competing interests

D.R.L. is a consultant and cofounder of Editas Medicine, Pairwise Plants, Beam Therapeutics and Prime Medicine, companies that use genome editing. J.L.D., A.R. and D.R.L. through the Broad Institute have filed patent applications on aspects of this work.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0414-6.

Correspondence and requests for materials should be addressed to D.R.L.

Reprints and permissions information is available at www.nature.com/reprints.

nature research

Corresponding author(s):	David R. Liu
Last updated by author(s):	12/20/2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	High-throughput sequencing data was collected on Illumina Miseq instruments. Whole-genome sequencing was collected on Illumina Novaseq 6000 instruments using two S4 flow cells. Flow sorting was performed on a FACS Aria II (BD Biosciences) sorter using BDFACS Diva software.

Data analysis	High-throughput sequencing data was analyzed using CRISPResso2 and custom scripts. Sequencing reads were demultiplexed using the MiSeq Reporter (Illumina) and fastq files were analyzed using Crispresso2. Prism 8 (GraphPad) was used to generate dot plots and bar plots of these data. For WGS analysis, initial data processing and read alignment was performed by the Broad Institute Genomics Platform. Reads were demultiplexed and aligned to the hg19 (b37) reference genome using BWA-MEM (v0.7.7). Aligned bams were sorted and optical duplicates were marked using Picard tools (v2.21). Base quality recalibration was performed using GATK (v3.4). All subsequent analyses were performed using the FAS RC Cannon high-performance computing cluster (Harvard University). Sequencing coverage was calculated using mosdepth (v0.2.6). We conducted variant calling on every sample independently using three algorithms, GATK HaplotypeCaller (v4.1.3.0), freebayes (v1.3.1), and VarScan (v2.4.3), assuming a ploidy of four and a minimum alternate allele read frequency of 0.1 to call an SNV. We called SNVs on the GATK-recommended genomic intervals57 that exclude highly repetitive regions such as centromeres and telomeres. We used bcftools (v1.9) to find the intersection of the variants called by all three algorithms in order to generate high-confidence variant calls. For all treated samples, we used bcftools to filter out variants in the treated sample that were present in the parent in order to retain only de novo variants that arose post treatment with base editors. We also used bcftools to filter out variants present at allele frequencies greater than 0.5 as previously reported in order to restrict analysis to variants that likely arose as a result of base editor treatment. Finally, we used bcftools to exclude variants that exhibited at least one of the following poor quality metrics based on the GATK vcf annotations: QD < 2, FS > 60, SOR > 3, MQ < 40, MQRankSum < -5. These final, high-quality variant calls for each treated sample were used for all downstream analyses. Further details and references and provided in the Methods. For the analysis of Yang and coworkers' data, genomic locations of all C•G-to-T•A SNVs reported by Yang and coworkers in tables S6 and S7 of their recent work13 the flanking sequences (20 base pairs on either side) were extracted from the mouse mm10 reference genome [GCA_000001635.2]. These flanking sequences were aligned, fixing the mutant cytosine in each case at position 21, and the resulting alignment was used to produce a sequence logo using WebLogo 3.6.0. The custom Python script used for this analysis is included in Supplementary Note 1.

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

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

All editor plasmids have been submitted to Addgene for public release. High-throughput sequencing reads and whole-genome sequencing reads are deposited in the NCBI Sequence Read Archive (PRJNA553240).

Field-specific reporting

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

☒ Life sciences □ Behavioural & social sciences □ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	All cell samples were evaluated in at least biological triplicates (n >= 3). In vitro biochemical experiments were performed 3 independent times.
Data exclusions	No data was excluded.
Replication	Biological triplicate experiments were done with distinct aliquots of cells at intervals ranging from weeks to months between experiments and performed by up to three different researchers. All findings have been replicated.
Randomization	Not relevant to these experiments.
Blinding	Not relevant to these experiments.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
□	☒ Antibodies
□	☒ Eukaryotic cell lines
☒	□ Palaeontology
☒	□ Animals and other organisms
☒	□ Human research participants
☒	□ Clinical data

Methods

n/a	Involved in the study
☒	□ ChIP-seq
□	☒ Flow cytometry
☒	□ MRI-based neuroimaging

Antibodies

Antibodies used	Rabbit anti-HA (Cell Signaling Technologies 3724SCell Signaling Technologies 3724S; 1:1000 dilution in SuperBlock Blocking Buffer (ThermoFisher Scientific)), Rabbit anti-GAPDH (Cell Signaling Technologies 5174S, 1:1000 dilution in SuperBlock Blocking Buffer (ThermoFisher Scientific)), Goat anti-rabbit 680RD (LI-COR 926-68071, 1:10,000 dilution in SuperBlock Blocking Buffer (ThermoFisher Scientific))
Validation	Rabbit anti-HA: validated by manufacturer by western blotting against HA-FoxO4 or HA-Akt3 from HeLa cell extract. Rabbit anti-GAPDH: validated by manufacturer by western blotting against whole cell lysates from various human cell lines. Goat anti-rabbit: validated by manufacturer for western blot detection by dot blot against serum proteins from different species.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	ATCC HEK293T (CRL-3216)
Authentication	Cells were authenticated by the supplier by STR analysis.
Mycoplasma contamination	HEK293T cells tested negative for mycoplasma as detailed in the Methods.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were grown in tissue culture and then transfected as described in the methods. 4 days following transfection, cells were trypsinized, resuspended in antibiotic-containing media, and then filtered to remove debris.
Instrument	FACS Aria II
Software	BD FACS DIVA software
Cell population abundance	Post-sort fractions were slightly different between constructs. The abundances of GFP-positive cells were 35.4%, 40.5%, and 47.9% for YE1-P2A-GFP, BE4-P2A-GFP, and the nickase-P2A-GFP constructs, respectively. The intensity gate then yielded the following relative to the GFP-positive parent sample: 27.7%, 30.1%, 44.9% for YE1-P2A-GFP, BE4-P2A-GFP, and nickase-P2A-GFP, respectively.
Gating strategy	Negative controls (cells not transfected with GFP plasmid) were used to establish GFP +/- gates. An initial gate was drawn to collect GFP positive cells based on this negative control. Of the GFP-positive population, a gate was set to collect the top ~27% of GFP-positive cells. This gate was used for all subsequent samples, such that all samples were gated based on the same fluorescence intensity. Please see Methods and Supp. Note. 5 for more information.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.