Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions

Citation

Published Version

Permanent link

Terms of Use

Share Your Story

Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions

Y. Bill Kim[1,2], **Alexis C. Komor**[1,2], **Jonathan M. Levy**[1,2], **Michael S. Packer**[1,2], **Kevin T. Zhao**[1,2], and **David R. Liu**[1,2,3,*]

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

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

[3]Broad Institute of Harvard and MIT, Cambridge, MA, 02142

Abstract

Base editing is a recently developed approach to genome editing that uses a fusion protein containing a catalytically defective *Streptococcus pyogenes* Cas9, a cytidine deaminase, and an inhibitor of base excision repair to induce programmable, single-nucleotide changes in the DNA of living cells without generating double-strand DNA breaks, without requiring a donor DNA template, and without inducing an excess of stochastic insertions and deletions[1]. Here we report the development of five new C→T (or G→A) base editors that use natural and engineered Cas9 variants with different protospacer-adjacent motif (PAM) specificities to expand the number of sites that can be targeted by base editing by 2.5-fold. Additionally, we engineered new base editors containing mutated cytidine deaminase domains that narrow the width of the apparent editing window from approximately 5 nucleotides to as little as 1 to 2 nucleotides, enabling the discrimination of neighboring C nucleotides that would previously be edited with comparable efficiency, thereby doubling the number of disease-associated target Cs that can be corrected preferentially over nearby non-target Cs. Collectively, these developments substantially increase the targeting scope of base editing and establish the modular nature of base editors.

Keywords

Base editing; genome editing; CRISPR; Cas9; protein engineering; genetic disease; single-nucleotide polymorphism

CRISPR-Cas9 nucleases have been widely used to mediate targeted genome editing[2]. In most genome editing applications, Cas9 forms a complex with a single guide RNA (sgRNA) and induces a double-stranded DNA break (DSB) at the target site specified by the sgRNA sequence. Cells primarily respond to this DSB through the non-homologuous end-joining (NHEJ) repair pathway, which results in stochastic insertions or deletions (indels) that can cause frameshift mutations that disrupt the gene. In the presence of a donor DNA template with a high degree of homology to the sequences flanking the DSB, gene correction can be achieved through an alternative pathway known as homology directed repair (HDR)[3,4]. Unfortunately, under most non-perturbative conditions HDR is inefficient, dependent on cell state and cell type, and dominated by the formation of indels[3,4]. As most of the known genetic variations associated with human disease are point mutations[5], methods that can more efficiently and cleanly make precise point mutations are needed.

We recently described base editing, which enables replacement of a target base pair with a different base pair in a programmable manner without inducing DSBs[1]. The first examples of base editing use a fusion protein between a catalytically inactivated (dCas9) or nickase form of *Streptococcus pyogenes* Cas9 (SpCas9), a cytidine deaminase such as APOBEC1, and an inhibitor of base excision repair such as uracil glycosylase inhibitor (UGI) to convert cytidines into uridines within a five-nucleotide window specified by the sgRNA[1]. Our third-generation base editor, BE3, converts C:G base pairs to T:A base pairs, including disease-relevant point mutations, in a variety of cell lines with higher efficiency and lower indel frequency than what can be achieved using other genome editing methods[1]. Subsequent studies have validated the base editing approach to genome editing in a variety of settings[6,7,8,9,10].

Efficient editing by BE3 requires the presence of an NGG PAM that places the target C within a five-nucleotide window near the PAM-distal end of the protospacer (positions 4–8, counting the PAM as positions 21–23)[1]. This PAM requirement substantially limits the number of sites in the human genome that can be efficiently targeted by BE3, as many sites of interest lack an NGG 13- to 17- nucleotides downstream of the target C. Moreover, the high activity of BE3 results in conversion of all Cs within the editing window to Ts, which can potentially introduce undesired changes to the target locus[1]. Here we report new C:G to T:A base editors that address both of these limitations and thereby substantially expand the targets suitable for base editing.

We hypothesized that any Cas9 homolog that binds DNA and forms an "R-loop" complex[11] containing a single-stranded DNA bubble could in principle be converted into a base editor. These new base editors would expand the number of targetable loci by allowing non-NGG PAM sites to be edited. The Cas9 homolog from *Staphylococcus aureus* (SaCas9) is considerably smaller than SpCas9 (1,053 vs. 1,368 residues), can mediate efficient genome editing in mammalian cells, and requires an NNGRRT PAM[12]. We replaced the nickase form of SpCas9 with that of SaCas9 in BE3 to generate APOBEC1–SaCas9n–UGI (SaBE3), and transfected HEK293T cells with plasmids encoding SaBE3 and sgRNAs targeting six human genomic loci (Fig. 1a). After 3 d, we used high-throughput DNA sequencing (HTS) to quantify base editing efficiency. SaBE3 enabled C to T base editing of target Cs at a variety of genomic sites in human cells, with very high conversion efficiencies

(approximately 50–75% of total DNA sequences converted from C to T, without enrichment for transfected cells) (Fig. 1b). The efficiency of SaBE3 on NNGRRT-containing target sites in general exceeded that of BE3 on NGG-containing target sites[1]. Perhaps due greater solvent exposure of the strand not paired with the guide RNA[13], SaBE3 can also result in detectable base editing at target Cs at positions outside of the canonical BE3 activity window (Fig. 1b). In comparison, BE3 showed greatly reduced editing for the same non-NGG target sites (0–11% editing), consistent with the known PAM requirement of SpCas9 (Supplementary Fig. 1a)[14]. These data show that SaBE3 can mediate base editing at sites not accessible to BE3.

We sought to further expand the targeting range of base editors by applying recently engineered Cas9 variants that expand or alter PAM specificities. Joung and coworkers recently reported three SpCas9 mutants that accept NGA (VQR-Cas9), NGAG (EQR-Cas9), or NGCG (VRER-Cas9) PAM sequences[15] as well as an engineered SaCas9 variant containing three mutations (SaKKH-Cas9) that relax its PAM requirement to NNNRRT[16]. We replaced the SpCas9 portion of BE3 with these four Cas9 variants to produce VQR-BE3, EQR-BE3, VRER-BE3, and SaKKH-BE3, which should target NGAN, NGAG, NGCG, and NNNRRT PAMs respectively. We transfected HEK293T cells with plasmids encoding these constructs and sgRNAs targeting several genomic loci for each new base editor, and measured C to T base conversions using HTS.

SaKKH-BE3 edited sites with NNNRRT PAMs with efficiencies up to 62% of treated, non-enriched cells (Fig. 1b,c). As expected, SaBE3 was unable to efficiently edit targets containing PAMs that were NNHRRT (where H = A, C, or T) (Fig. 1c). VQR-BE3, EQR-BE3, and VRER-BE3 exhibited more modest, but still substantial base editing efficiencies of up to 50% of treated, non-enriched cells at genomic loci with the expected PAM requirements with an editing window similar to that of BE3 (Fig. 1d–f). Base editing efficiencies of VQR-BE3, EQR-BE3, and VRER-BE3 in general closely paralleled the reported PAM requirements of the corresponding Cas9 nucleases; for example, EQR-BE3 was unable to efficiently edit targets containing NGAH PAM sequences (Fig. 1e). Consistent with the known PAM requirements of SpCas9[14], BE3 was unable to efficiently edit sites with NGA or NGCG PAMs (0–3% efficiency) (Supplementary Fig. 1b). To confirm that the five new base editors functioned in multiple mammalian cell types, we assessed their performance in U2OS cells and observed robust editing, albeit with slightly lower editing and/or transfection efficiency. (Supplementary Fig. 2a). Collectively, the properties of SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, and VRER-BE3 establish that base editors behave in a modular fashion that facilitates our ability to repurpose Cas9 homologs and engineered variants for base editing.

We examined the off-target activity of the altered-PAM base editors in human cells. We selected two on-target loci for each new editor that have been previously analyzed for Cas9 off-target cleavage and sequenced 7 off-targets for the SaBE3 constructs and 10 off-targets for the SpBE3 constructs[15,16] (Supplementary Fig. 3a). Consistent with our previous study[1], we detected off-target base editing by SaBE3 and SaKKH-BE3 at a subset of known Cas9 off-target cleavage loci containing an appropriately placed target C and that conform to their PAM requirements (Supplementary Fig. 3b). In contrast, we observed substantially less off-

target base editing from VQR-BE3 or EQR-BE3 at ten known off-target loci of VQR SpCas9[15], suggesting that these base editors may offer enhanced specificity (Supplementary Fig. 3c).

Next, we sought to develop base editors with altered activity window widths. All Cs within the 5-nucleotide activity window of BE3 are typically converted to Ts with comparable efficiency[1]. The ability to modulate the width of this window is useful when it is important to edit only a subset of Cs present in the activity window, such as cases in which the target C is adjacent to other Cs that if changed would result in undesired mutations to the gene of interest.

We previously noticed that the length of the linker between APOBEC1 and dCas9 modulates the number of bases that are accessible by APOBEC1 *in vitro*[1]. In HEK293T cells, however, varying the linker length did not significantly modulate the width of the editing window, suggesting that in the complex cellular milieu, the relative orientation and flexibility of dCas9 and the cytidine deaminase are not strongly determined by linker length (Supplementary Fig. 4). We hypothesized that truncating the 5′ end of the sgRNA might narrow the base editing window by reducing the length of single-stranded DNA accessible to the deaminase upon formation of the RNA-DNA heteroduplex. We co-transfected HEK293T cells with plasmids encoding BE3 and sgRNAs of different spacer lengths targeting several loci with multiple Cs in the editing window. Although for some target loci, truncated guide RNAs with 16- or 17-base spacers showed narrowed editing windows, we observed no consistent changes in the editing window width when using truncated sgRNAs with 15- to 19-base spacers (Supplementary Fig. 5).

As an alternative approach, we envisioned that mutations to the deaminase domain might narrow the width of the editing window through at least two possible mechanisms. First, because the high activity of APOBEC1 likely contributes to the deamination of multiple Cs per DNA binding event[1,17,18], mutations that reduce the catalytic efficiency of the deaminase domain of a base editor might prevent it from catalyzing successive rounds of deamination before dissociating or being displaced from DNA. Once any C:G to T:A editing event has taken place, the sgRNA no longer perfectly matches the target DNA sequence and re-binding of the base editor to the target locus should be less favorable. Second, some mutations may alter substrate binding, the conformation of bound DNA, or substrate accessibility to the active site in ways that reduce tolerance for non-optimal presentation of a C to the deaminase active site. We sought to test both strategies to discover new base editors that distinguish among multiple cytidines within the original editing window.

Given the absence of an available APOBEC1 structure, we identified several mutations previously reported to modulate the catalytic activity of APOBEC3G, a cytidine deaminase from the same family that shares 42% sequence similarity of its active site-containing domain to that of APOBEC1[19]. We incorporated the corresponding APOBEC1 mutations into BE3 and evaluated their effect on base editing efficiency and editing window width in HEK293T cells at two C-rich genomic sites containing Cs at positions 3, 4, 5, 6, 8, 9, 10, 12, 13, and 14 (site A); or containing Cs at positions 5, 6, 7, 8, 9, 10, 11, and 13 (site B). For analysis purposes, we define the "editing window width" as the number of nucleotide

positions at a given site for which editing efficiency exceeds the half-maximal value for that target site. The editing window width of BE3 for the two C-rich genomic sites tested was approximately 4 (site A) and 6 (site B) nucleotides (Fig. 2a).

Mutating residues that are essential for deaminase activity, such as R118A, led to dramatic loss of base editing efficiency (Supplementary Fig. 6). However, we identified other mutations that narrowed the editing window while maintaining substantial editing efficiency (Fig. 2a, Supplementary Fig. 6). W90 in APOBEC1, corresponding to W285 in APOBEC3G, is predicted to be crucial for the formation of a hydrophobic active site in APOBEC1[17,18]. APOBEC3G mutant W285A has undetectable deaminase activity[17,18], and BE3 W90A similarly shows almost no base editing activity (Supplementary Fig. 6). We hypothesized that W90Y or W90F might decrease the hydrophobicity of the active site while maintaining catalytic activity. Indeed, W90Y and W90F only modestly decreased base editing activity while narrowing the editing window width at sites A and B to 3 or 2 nucleotides, respectively (Fig. 2a, Supplementary Fig. 6). These results demonstrate that mutations of the cytidine deaminase domain can narrow the width of the base editing window.

Next we sought to test if mutations to residues involved in APOBEC1 substrate binding could also alter the editing window width. R126 in APOBEC1 is predicted to interact with the phosphate backbone of ssDNA[18]. Previous studies have shown that mutation of the corresponding residue to Ala in APOBEC3G (R320A) decreases apparent activity by at least 5-fold in a rifampin resistance assay[18]. Interestingly, when introduced into APOBEC1 in BE3, R126A and R126E maintained activity comparable to that of BE3 at the most strongly edited central positions (C5 and C6 for site A, C6 and C7 for site B), while decreasing editing activity at other positions (Fig. 2a, Supplementary Fig. 6). Both of these two mutations narrowed the width of the editing window at site A and site B to 3 nucleotides (Fig. 2a, Supplementary Fig. 6). R132 is also near R126 and could influence single-stranded DNA accessibility. R132E led to a decrease in editing efficiency but also narrowed the width of the editing window to 3 nucleotides for both sites A and B. (Supplementary Fig. 6).

We combined W90Y/F, R126E, and R132E, the three mutations that narrowed the editing window without drastically reducing base editing activity, into doubly and triply mutated base editors. Combining the R126E and W90F mutations modestly narrowed the editing window compared to the W90F mutant, while combining W90F with R132E resulted in a dramatic loss of activity (Supplementary Fig. 6). The double mutant W90Y+R126E, however, resulted in a base editor (YE1-BE3) with BE3-like maximal editing efficiencies, but substantially narrowed editing window width of approximately 2 nucleotides for both site A and site B. (Fig. 2a). The W90Y+R132E base editor (YE2-BE3) exhibited modestly lower editing efficiencies, but also narrowed the editing window width to approximately 2 nucleotides for both site A and site B (Supplementary Fig. 6). The R126E+R132E double mutant (EE-BE3) showed similar maximal editing efficiencies and editing window width as YE2-BE3 (Supplementary Fig. 6). The triple mutant W90Y+R126E+R132E (YEE-BE3) exhibited 2.9-fold lower average maximal editing yields but very little editing beyond the C6 position and an editing window width of approximately 2 and 1 nucleotides for site A and site B, respectively (Supplementary Fig. 6). These data taken together indicate that

mutations in the cytidine deaminase domain can strongly affect editing window widths, in some cases with minimal or only modest effects on editing efficiency.

Next we compared the base editing outcomes of BE3, YE1-BE3, YE2-BE3, EE-BE3, and YEE-BE3 in HEK293T cells targeting four well-studied human genomic sites that contain multiple Cs within the BE3 activity window[1]. These target loci contained target Cs at positions 4 and 5 (HEK293 site 3), positions 4 and 6 (HEK293 site 2), positions 5 and 6 (EMX1), or positions 6, 7, 8, and 11 (FANCF). BE3 exhibited little preference for editing any of the Cs within the activity window of position 4–8 (< 1.2 fold average base preference for HEK293 sites 2, 3, and EMX1). In contrast, YE1-BE3 exhibited a 1.3-fold preference for editing C5 over C4 (HEK site 3), 2.7-fold preference for C6 over C4 (HEK site 2), 2.0-fold preference for C5 over C6 (EMX1), and 1.5-fold preference for C6 over C7 (FANCF) (Fig. 2b). YE2-BE3 and EE-BE3 exhibited greater positional specificity, averaging 2.4-fold preference for editing C5 over C4 (HEK site 3), 9.8-fold preference for C6 over C4 (HEK site 2), 2.9-fold preference for C5 over C6 (EMX1), and 2.5-fold preference for C7 over C6 (FANCF) (Fig. 2b). YEE-BE3 showed the greatest positional selectivity, with a 2.9-fold preference for editing C5 over C4 (HEK site 3), 31-fold preference for C6 over C4 (HEK site 2), 7.6-fold preference for C5 over C6 (EMX1), and 8.0-fold preference for C6 over C7 (FANCF) (Fig. 2b). We then assessed editing of YE1-BE3, EE-BE3, YE2-BE3, and YEE-BE3 in U2OS cells, and found similar alteration of base editing window widths (Supplementary Fig. 2b) as we observed in HEK293T cells. These findings establish that mutant base editors with narrowed editing windows can discriminate between adjacent Cs, even when both nucleotides are within the BE3 editing window.

We analyzed by HTS the product distributions of the above four mutants and BE3 to assess their apparent processivity. BE3 generated predominantly T4-T5 (HEK site 3), T4-T6 (HEK site 2), and T5-T6 (EMX1) products in treated HEK293T cells, resulting in, on average, 7.4-fold more products containing two Ts, than products containing a single T (Fig. 2c). In contrast, YE1-BE3, YE2-BE3, EE-BE3, and YEE-BE3 showed substantially higher preferences for singly edited C4-T5, C4-T6, and T5-C6 products at these sites (Fig. 2c). YE1-BE3 yielded products with an average single-T to double-T product ratio of 1.4. YE2-BE3 and EE-BE3 yielded products with an average single-T to double-T product ratio of 4.3 and 5.1, respectively (Fig. 2c). Consistent with the above results, the YEE-BE3 triple mutant strongly favored single-T products by an average of 14-fold across the three genomic loci. (Fig. 2c). Interestingly, for the target site in which only one C is within the target window (HEK293 site 4, at position C5), all four mutants exhibited comparable editing efficiencies as BE3 (Supplementary Fig. 7). These findings together indicate that these BE3 mutants have decreased apparent processivity and can favor the conversion of only a single C at target sites containing multiple Cs within the BE3 editing window. These data also suggest a positional preference of C5 > C6 > C7 ≈ C4 for these mutant base editors, although we note that this preference could differ depending on the target sequence.

To test if a narrow-window base editor maintains its editing window profile over time following a single treatment, we treated HEK293T cells with YE1-BE3, EE-BE3, YE2-BE3, or YEE-BE3 and analyzed by HTS the distribution of base editing within the protospacer sequence 3 d, 6 d, or 9 d after treatment. We observed no substantial change in the positional

distribution of base editing over 9 d at any of the four genomic loci tested (Supplementary Fig. 8)

Next, we examined the off-target activity of the base editors with narrowed activity windows. Our previous work established that BE3 exhibits off-target base editing activity at a subset of Cas9 off-target cleavage loci[1]. We assayed by HTS 21 known off-target loci corresponding to three on-target loci. We observed on average 3.6-fold less efficient off-target base editing per on-target editing event of the BE3 mutants compared to BE3 for the 21 off-target sites (Supplementary Fig. 9). This improvement in DNA specificity may arise from a reduction in the residence time of base editors at these off-target loci that further impedes deamination with catalytically impaired APOBEC1 domains.

Finally, we combined the two innovations described in this work. The window-modulating mutations in APOBEC1 were applied to VQR-BE3, allowing editing with a narrowed activity window and greater positional selectivity of target sites containing an NGA PAM (Supplementary Fig. 10a). Not all window-narrowing mutations could be productively combined with altered PAM base editors, however. When window-narrowing mutations were installed in SaKKH-BE3, for example, we observed a decrease in base editing efficiency without any obvious change in the width of the activity window, perhaps arising from differences in the substrate accessibility of this base editor compared with that of BE3 and its variants (Supplementary Fig. 10b).

The five base editors with altered PAM specificities described in this study together increase the number of disease-associated mutations in the ClinVar database that can in principle be corrected by base editing by 2.5-fold (Fig. 3a). Similarly, the development of base editors with narrowed editing windows approximately doubles the fraction of ClinVar entries with a properly positioned NGG PAM that can be corrected by base editing without comparable modification of a non-target C (Fig. 3b).

In summary, we have expanded substantially the targeting scope of base editing by developing base editors that use Cas9 variants with different PAM specificities, and by developing a collection of deaminase mutants with varying editing window widths. The modularity of base editing established in this study suggests that the advances described here should also be applicable to other fusions of Cas9 variants linked to nucleotide conversion enzymes.

ONLINE METHODS

Cloning

PCR was performed using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs). Plasmids for BE and sgRNA were constructed using USER cloning (New England Biolabs) from previously reported plasmids[1]. DNA vector amplification was carried out using NEB 10beta competent cells (New England Biolabs). Site-directed mutagenesis of APOBEC1 variants was done using blunt-end ligation. Briefly, a primer with an overhang containing the desired point mutation was used to amplify the appropriate vector plasmid by

PCR. KLD enzyme mix (New England Biolabs) was used to phosphorylate and circularize the PCR product prior to transformation.

Cell culture

HEK293T (ATCC CRL-3216) and U2OS (ATCC HTB-96) cells were cultured in Dulbecco's Modified Eagle's Medium plus GlutaMax (ThermoFisher) supplemented with 10 % (v/v) fetal bovine serum (FBS), at 37 °C with 5 % CO_2.

Transfections

HEK293T cells seeded on 48-well collagen-coated BioCoat plates (Corning) were transfected at approximately 70% confluency. 750 ng of BE and 250 ng of sgRNA expression plasmids were transfected using 1.5 μl of Lipofectamine 2000 (ThermoFisher Scientific) per well according to the manufacturer's protocol. 500 ng of BE and 250 ng of sgRNA expression plasmids were transfected into U2OS cells using a Lonza 4D-Nucleofector with the DN-100 program according to the manufacturer's protocols.

High-throughput DNA sequencing of genomic DNA samples

Transfected cells were harvested after 3 d. The genomic DNA was isolated using the Agencourt DNAdvance Genomic DNA Isolation Kit (Beckman Coulter) according to the manufacturer's instructions. Genomic regions of interest were amplified by PCR with flanking HTS primer pairs listed in the Supplementary Information. PCR amplification was carried out with Phusion hot-start II DNA polymerase (ThermoFisher) according to the manufacturer's instructions. PCR products were purified using RapidTips (Diffinity Genomics). Secondary PCR was performed to attach sequencing adaptors. The products were gel-purified and quantified using the KAPA Library Quantification Kit-Illumina (KAPA Biosystems). Samples were sequenced on an Illumina MiSeq as previously described[1].

Data analysis

Nucleotide frequencies were assessed using a previously described MATLAB script[1]. Briefly, the reads were aligned to the reference sequence via the Smith-Waterman algorithm. Base calls with Q-scores below 30 were replaced with a placeholder nucleotide (N). This quality threshold results in nucleotide frequencies with an expected theoretical error rate of 1 in 1,000.

Analyses of base editing processivity were performed using a custom python script (Supplementary Note 1). This program trims sequencing reads to the 20 nucleotide protospacer sequence as determined by a perfect match for the 7 nucleotide sequences that should flank the target site. These targets were then consolidated and sorted by abundance to assess the frequency of base editing products.

Bioinformatic analysis of the ClinVar database of human disease-associated mutations was performed in a manner similar to that previously described but with small adjustments (Supplementary Note 2)[1]. These adjustments enable the identification of targets with PAMs of customizable length and sequence. In addition, this improved script includes a priority

ranking of target C positions (C5 > C6 > C7 >C8 ≈ C4), thus enabling the identification of target sites in which the on-target C is either the only cytosine within the window, or is placed at a position with higher predicted editing efficiency than any non-target C within the editing window.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

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

2. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol. 2014; 32:347–355. [PubMed: 24584096]

3. Cong L, et al. Multiplex genome engineering using CRISPR/Cas systems. Science. 2013; 339:819–823. [PubMed: 23287718]

4. Ran FA, et al. Genome engineering using the CRISPR-Cas9 system. Nat Protocols. 2013; 8:2281–2308. [PubMed: 24157548]

5. Landrum MJ, et al. ClinVar: public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2015; 44:D862–D868. [PubMed: 26582918]

6. Nishida K, et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science. 2016; 353 aaf8729-1-8.

7. Ma Y, et al. Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat Methods. 2016; 13:1029–1035. [PubMed: 27723754]

8. Hess GT, et al. Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat Methods. 2016; 13:1036–1042. [PubMed: 27798611]

9. Yang L, et al. Engineering and optimising deaminase fusions for genome editing. Nat Comm. 2016; 7:13330.

10. Li J, Sun Y, Du J, Zhao Y, Xia L. Generation of targeted point mutations in rice by a modified CRISPR/Cas9 system. Mol Plant. 2017

11. Jiang F, et al. Structures of a CRISPR–Cas9 R-loop complex primed for DNA cleavage. Science. 2016; 351:867–71. [PubMed: 26841432]

12. Ran FA, et al. *In vivo* genome editing using *Staphylococcus aureus* Cas9. Nature. 2015; 520:186–191. [PubMed: 25830891]

13. Nishimasu H, et al. Crystal Structure of *Staphylococcus aureus* Cas9. Cell. 2015; 162:1113–1126. [PubMed: 26317473]

14. Zhang Y, et al. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. Sci Rep. 2014; 4

15. Kleinstiver BP, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015; 523:481–485. [PubMed: 26098369]

16. Kleinstiver BP, et al. Broadening the targeting range of *Staphylococcus aureus* CRISPR-Cas9 by modifying PAM recognition. Nat Biotechnol. 2015; 33:1293–1298. [PubMed: 26524662]

17. Holden LG, et al. Crystal structure of the anti-viral APOBEC3G catalytic domain and functional implications. Nature. 2008; 452:121–124.

18. Chen KM, et al. Structure of the DNA deaminase domain of the HIV-1 restriction factor. APOBEC3G. Nature. 2008; 452:116–119. [PubMed: 18288108]

19. Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA Editing Enzyme APOBEC1 and Some of Its Homologs Can Act as DNA Mutators. Molecular Cell. 2002; 10:1247–1253. [PubMed: 12453430]
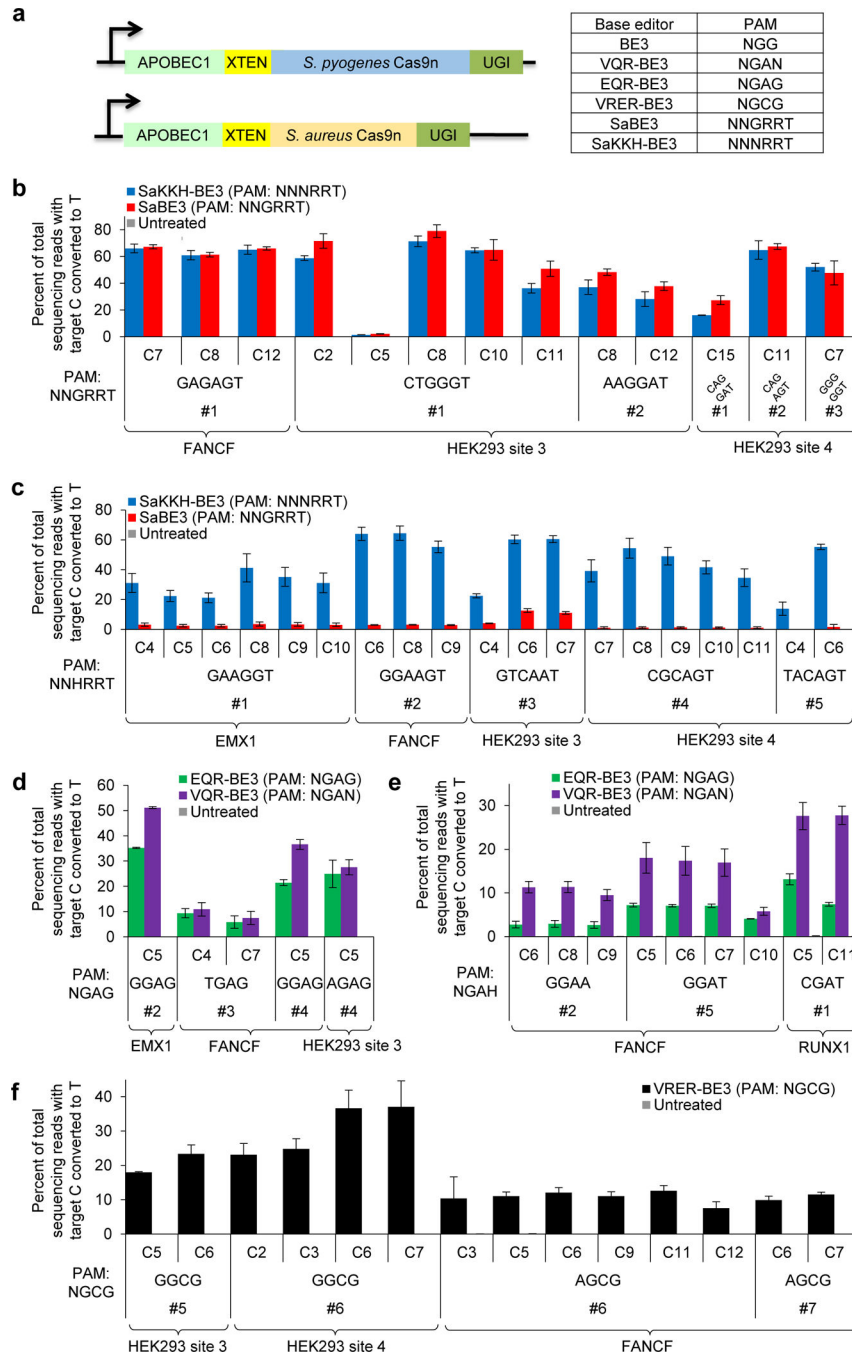
Figure 1. SaBE3, SaKKH-BE3, VQR-BE3, EQR-BE3, and VRER-BE3 mediate efficient base editing at target sites containing non-NGG PAMs in human cells

a, Base editor architectures using *S. pyogenes* and *S. aureus* Cas9, and recently characterized Cas9 variants with alternate or relaxed PAM requirements[12,13]. **b–f**, HEK293T cells were treated with the base editor variants shown as described in the Methods. The percentage of total DNA sequencing reads (with no enrichment for transfected cells) with C converted to T at the target positions indicated are shown. The PAM sequence of each target tested is shown below the X-axis. The charts show the results for SaBE3 and SaKKH-BE3 at
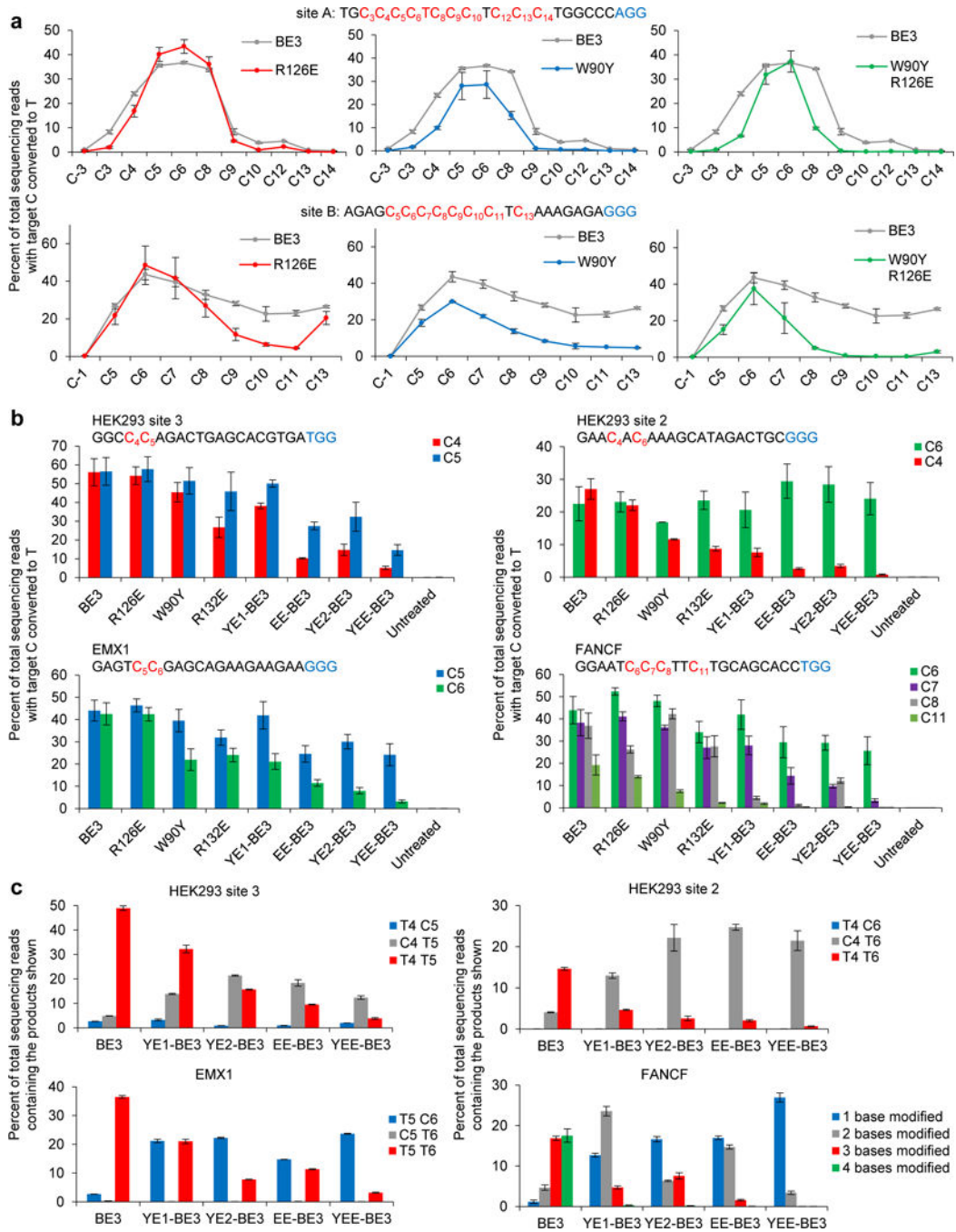
genomic loci with NNGRRT PAMs (**b**), SaBE3 and SaKKH-BE3 at genomic loci with NNHRRT PAMs (**c**), VQR-BE3 and EQR-BE3 at genomic loci with NGAG PAMs (**d**) and with NGAH PAMs (**e**), and VRER-BE3 at genomic loci with NGCG PAMs (**f**). Values and error bars reflect the mean and standard deviation of at least two biological replicates.

Figure 2. Base editors with mutant cytidine deaminase domains exhibit narrowed editing windows

a–c, HEK293T cells were transfected with plasmids expressing mutant base editors and an appropriate sgRNA. Three days after transfection, genomic DNA was extracted and analyzed by high-throughput DNA sequencing at the indicated loci. The percentage of total DNA sequencing reads (without enrichment for transfected cells) with C changed to T at the target positions indicated are shown for the EMX1 site, HEK293 site 3, FANCF site, HEK293 site 2, site A, and site B loci. **a**, Cytidine deaminase mutations narrow the width of

the editing window. See Supplementary Figure 6 for the characterization of additional mutations. **b**, Effect of cytidine deaminase mutations that narrow editing window width on genomic loci containing multiple Cs within the canonical BE3 editing window (positions 4–8). Combining mutations has an additive effect on narrowing the editing window. YE1 = W90Y + R126E; EE = R126E + R132E; YE2 = W90Y + R132E; YEE = W90Y + R126E + R132E. **c**, YE1-BE3, YE2-BE3, EE-BE3, and YEE-BE3 alter the product distribution of base editing, producing predominantly singly-modified products, in contrast with BE3, even when multiple Cs are present in positions 4–8. Values and error bars reflect the mean and standard deviation of at least two biological replicates.
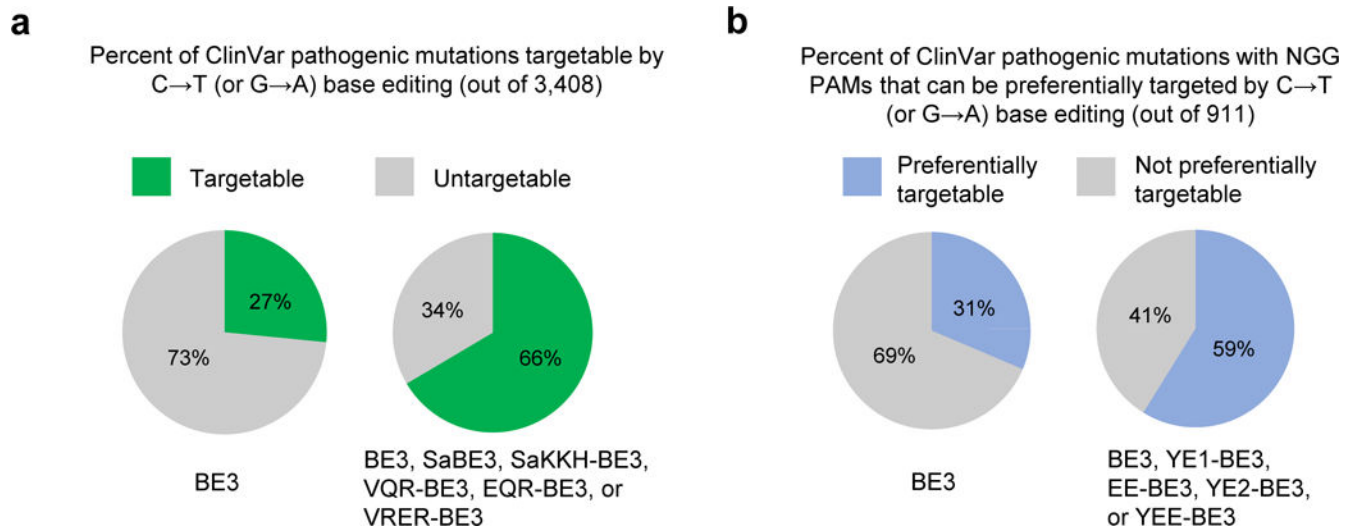
Figure 3. Genetic variants in ClinVar that can be corrected in principle by the base editors developed in this work

The NCBI ClinVar database of human genetic variations and their corresponding phenotypes (see Ref. 5) was searched for genetic diseases that in theory can be corrected by base editing. **a**, The base editors with altered PAM specificities developed in this study substantially increases targetable loci among all pathogenic T→C or A→G mutations in the ClinVar database. **b**, Improvement in base editing targeting scope among all pathogenic T→C or A→G mutations in the ClinVar database through the use of base editors with narrowed activity windows. BE3 is assumed to edit Cs in positions 4–8 with comparable efficiency as shown in Fig. 2. YE1-BE3, EE-BE3, YE2-BE3, and YEE-BE3 are assumed to edit with a preference of C5 > C6 > C7 ≈ C4. The blue fractions denote the proportion of pathogenic T→C or A→G mutations that can be edited by BE3 without comparable editing of other Cs (left), or that can be edited by BE3 or by one of BE3 mutants without comparable editing of other Cs (right).