

Base editing with a Cpf1–cytidine deaminase fusion

Xiaosa Li^{1–3,6}, Ying Wang^{4,6}, Yajing Liu^{1–3,6}, Bei Yang^{5,6} , Xiao Wang^{1–3}, Jia Wei⁴, Zongyang Lu^{1–3}, Yuxi Zhang¹, Jing Wu¹, Xingxu Huang¹, Li Yang⁴  & Jia Chen¹ 

The targeting range of CRISPR–Cas9 base editors (BEs) is limited by their G/C-rich protospacer-adjacent motif (PAM) sequences. To overcome this limitation, we developed a CRISPR–Cpf1-based BE by fusing the rat cytosine deaminase APOBEC1 to a catalytically inactive version of *Lachnospiraceae* bacterium Cpf1. The base editor recognizes a T-rich PAM sequence and catalyzes C-to-T conversion in human cells, while inducing low levels of indels, non-C-to-T substitutions and off-target editing.

The BEs developed by combining apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like (APOBEC) or activation-induced deaminase (AID) cytidine deaminase family members with the CRISPR–Cas system have been used for targeted C-to-T base editing in various species^{1–5}. However, the use of the Cas9 nickase (nCas9) as the deaminase fusion partner in the most active currently used BEs increases the frequency of unwanted insertions and deletions (indels) and non-C-to-T base substitutions^{1–3}, and limits editing of regions with G/C-rich PAM sequences^{6,7}.

Cpf1 (Cas12a) is another Cas protein that differs from Cas9 in several ways: Cpf1 requires a T-rich PAM sequence (TTTV) for target-DNA recognition^{8,9}; the guide RNA for Cpf1 (CRISPR RNA (crRNA)) is shorter than that for Cas9 (single guide RNA (sgRNA)); and the Cpf1-cleavage site is located distal and downstream relative to the PAM sequence in spacer DNA, rather than proximal and upstream as for Cas9 (refs. 10,11). Cpf1 also induces less off-target (OT) cleavage genome wide than does Cas9 (refs. 12–14).

To make use of the beneficial properties for base editing, we fused rat APOBEC1 to either catalytically inactive *Acidaminococcus* sp. Cpf1 (dAsCpf1) or catalytically inactive *L. bacterium* Cpf1 (dLb-Cpf1) together with uracil DNA glycosylase inhibitor (UGI), thus creating two dCpf1-based BEs: dAsCpf1-BE0 and dLbCpf1-BE0 (Supplementary Fig. 1a). We first tested their editing potential in an *Escherichia coli* plasmid-derived episomal shuttle-vector reporter system (Supplementary Fig. 1a) that has been shown to be a sensitive tool for probing base substitutions in human cells¹⁵.

dLbCpf1-BE0 induced a high level of C-to-T base editing in the target regions (the editing frequency of single cytosines ranged from 44% to 74%), whereas dAsCpf1-BE0 did not show detectable base editing under similar conditions (Supplementary Fig. 1b,c). Hence, we used dLb-Cpf1-BEs in the rest of this study, and we refer to them as dCpf1-BEs for simplicity. We also found that crRNAs with spacers ranging from 19 nt to 27 nt showed similar editing frequencies (Supplementary Fig. 2).

Next, we analyzed the performance of dCpf1-BE0 at endogenous genomic sites in mammalian cells. As expected, dCpf1-BE0 also induced base editing at targeted genomic sites and exhibited a 6–37% C-to-T-editing frequency (mean 20%, counting the highest editing frequency of single cytosines in each target, Supplementary Fig. 3a,b). Deletion of the SV40 internal nuclear localization sequence (iNLS) between dCpf1 and UGI dramatically decreased the base-editing efficiency (Supplementary Fig. 3b–d, comparison of dCpf1-BE0ΔiNLS to dCpf1-BE0), a result consistent with a recent finding that the number of NLSs affects gene-editing efficiency¹⁶. Because a longer linker can increase the base-editing efficiency, as reported for Cas9-based BEs¹⁷, this iNLS might also function as a longer linker in dCpf1-BE0. Further experiments are required to test this possibility. As expected, dCpf1-BE with an additional copy of N-terminal SV40NLS increased the base-editing efficiencies at all tested genomic loci (Supplementary Fig. 4a–d; $P = 3 \times 10^{-11}$).

To evaluate its general efficacy, we tested dCpf1-BE at 12 target sites with the TTTV PAM sequence and three target sites with the TTTT PAM sequence (Fig. 1a). Among the 12 target sites with the TTTV PAM sequence, dCpf1-BE induced base editing (highest single C-to-T-conversion frequency ranging from 11% to 46%; mean 22%) at ten sites and induced inefficient base editing (frequency <5%; mean 3%) at two sites. At sites with the TTTT PAM sequence, Cpf1-BE induced relatively low efficiencies (~10% editing at two sites and no detectable editing for the other site). These results were consistent with a previous result showing that Cpf1 prefers a TTTV PAM sequence¹⁸. The main editing window of dCpf1-BE ranged from positions 8 to 13, counting the base next to the PAM as position 1 (Supplementary Fig. 5a), and dCpf1-BE scarcely induced C-to-T base editing at the cytosines following a guanosine (Fig. 1a), in agreement with findings from previous studies showing that APOBEC1 cannot efficiently deaminate cytosines with a 5'-guanosine^{1,17}. Notably, because dCpf1 is used in dCpf1-BEs, unwanted indels were not generally induced, and high fractions of C-to-T conversions were achieved at tested sites (Supplementary Figs. 3e, 4e and 5b,c). Similarly, dCpf1-BE induced base editing in another human cell line, U2OS, at all examined sites (highest single C-to-T-conversion frequency ranging from 10% to 33%; mean 20%), while inducing low levels of unwanted indels and non-C-to-T substitutions (Supplementary Fig. 6).

¹School of Life Science and Technology, ShanghaiTech University, Shanghai, China. ²Shanghai Institute of Biochemistry and Cell Biology, Chinese Academy of Sciences, Shanghai, China. ³University of Chinese Academy of Sciences, Beijing, China. ⁴Key Laboratory of Computational Biology, CAS-MPG Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, University of Chinese Academy of Sciences, Chinese Academy of Sciences, Shanghai, China. ⁵Shanghai Institute for Advanced Immunochemical Studies, ShanghaiTech University, Shanghai, China. ⁶These authors contributed equally to this work. Correspondence should be addressed to X.H. (huangxx@shanghaitech.edu.cn), L.Y. (liyang@picb.ac.cn) or J.C. (chenjia@shanghaitech.edu.cn).

Received 13 September 2017; accepted 15 February 2018; published online 19 March 2018; doi:10.1038/nbt.4102

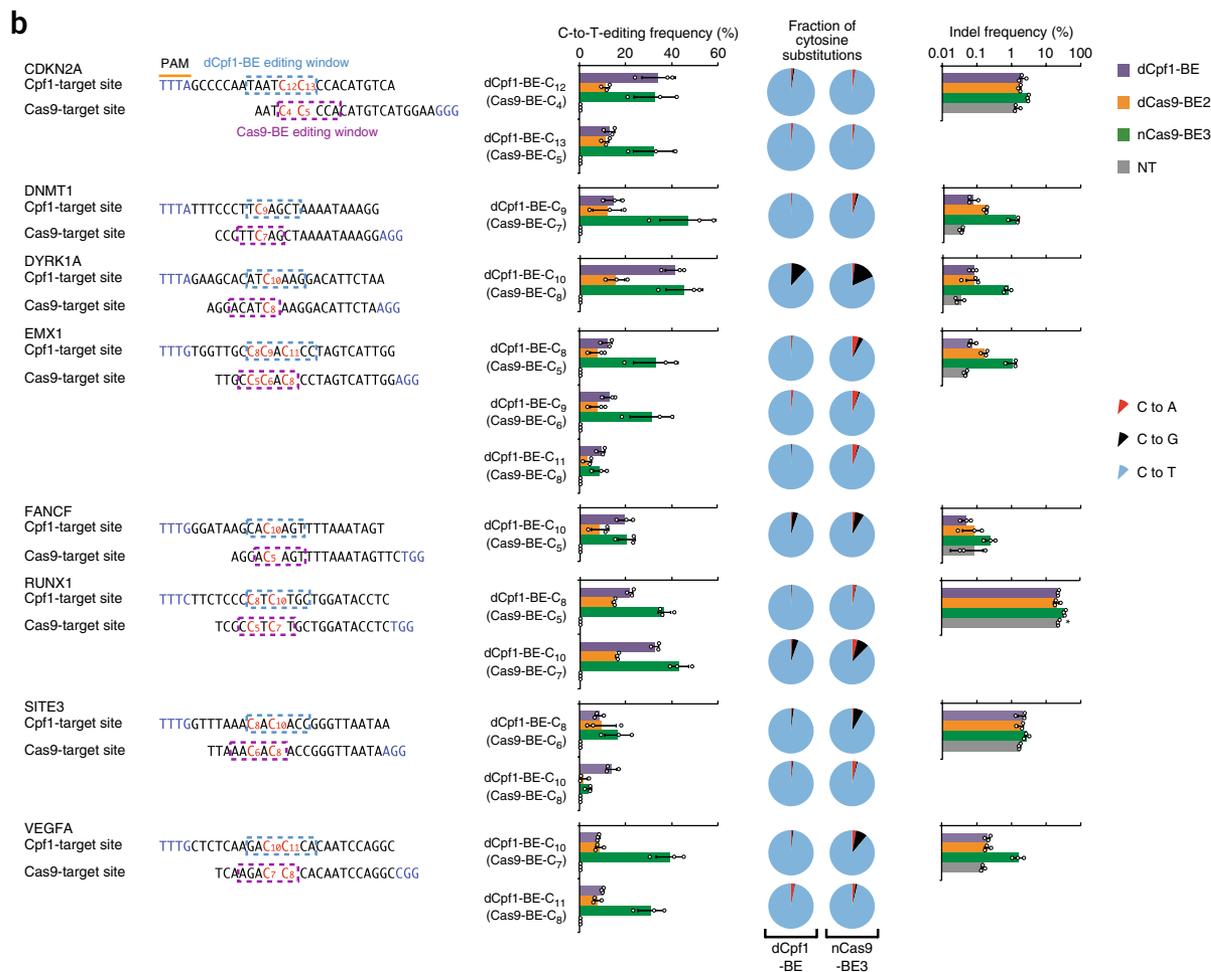
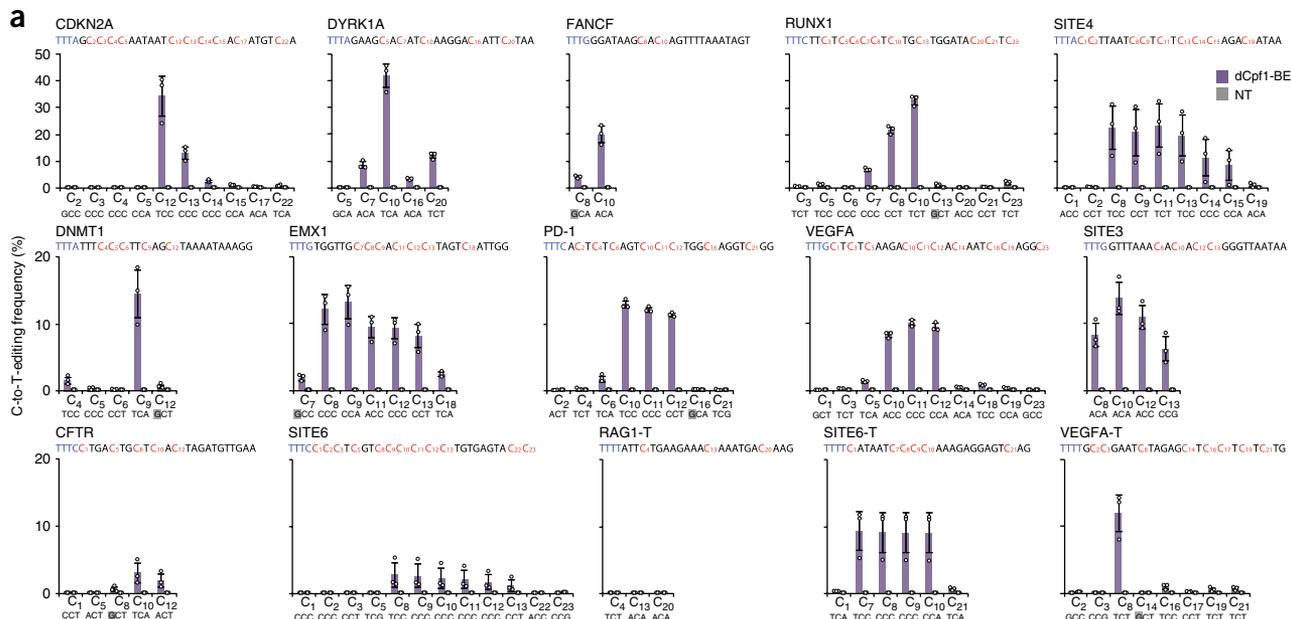


Figure 1 Base editing mediated by dCpf1-BE. (a) Determination of dCpf1-BE-induced base-editing frequency at every single cytosine in the indicated spacer region. The dCpf1-BE showed inefficient C-to-T base editing at the cytosines following a guanosine (shaded gray). The cytosines were counted with the base proximal to the PAM setting as position 1. (b) Comparison of base editing mediated by dCpf1-based and Cas9-based BEs. The C-to-T-editing frequencies of the indicated cytosines, the fractions of cytosine substitutions and the indel frequencies were individually determined at the indicated genomic target sites under different conditions. The target-site sequences and editing windows of dCpf1-BE and Cas9-BEs are shown. NT, nontransfected. Asterisk denotes an unusually high basal indel frequency (or amplification, sequencing or alignment artifact) at the examined RUNX1 site in the nontransfected 293FT cells. Data are shown as means \pm s.d. from three independent experiments.

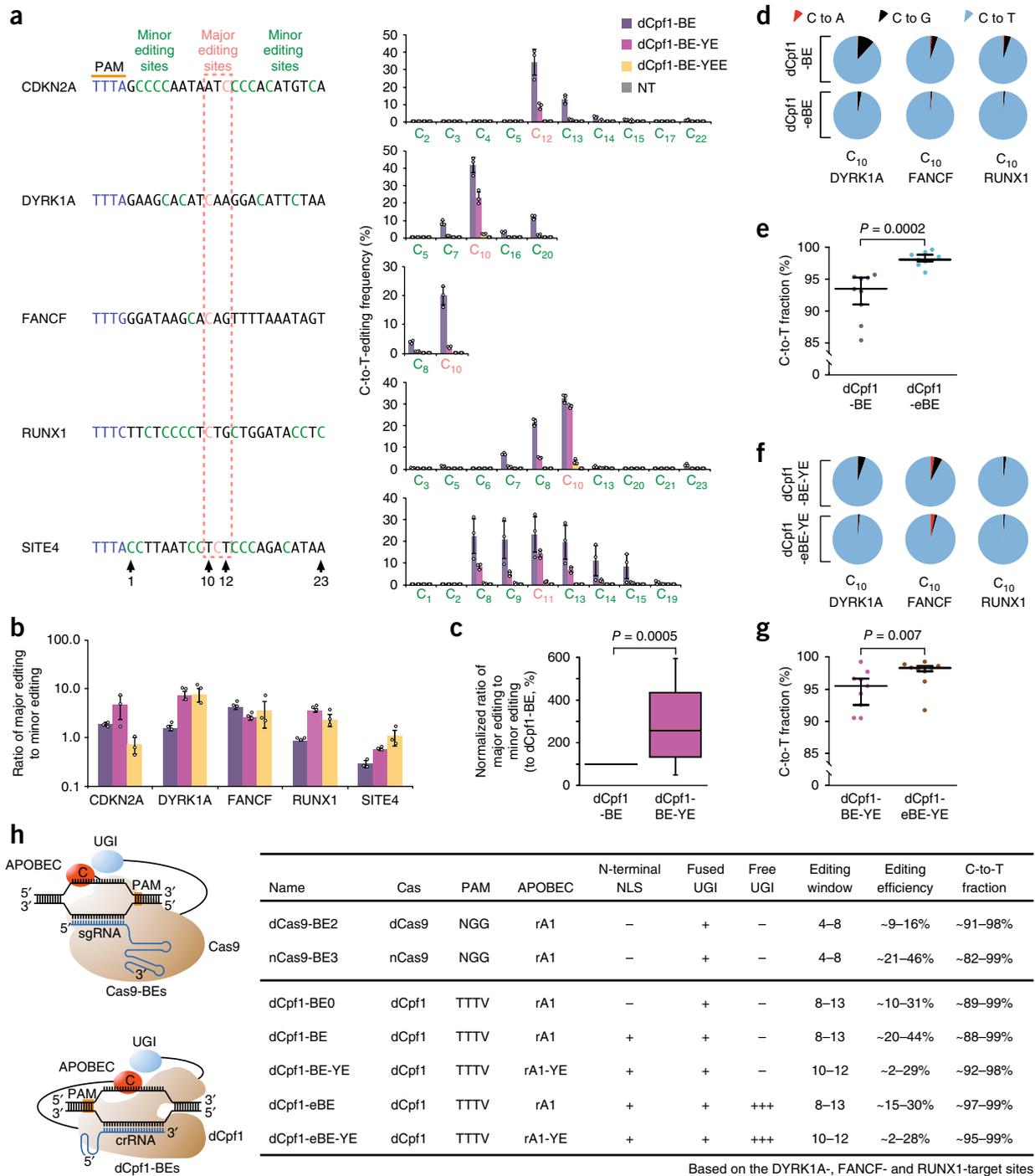


Figure 2 Improvements in dCpf1-BE. **(a)** Mutating APOBEC1 in dCpf1-BE to narrow down the editing window. The C-to-T-editing frequency at every single cytosine was individually determined in the indicated genomic target sites under different conditions. The target-site sequences and the narrowed editing windows of dCpf1-BE are shown. The major editing sites (C₁₀-C₁₂) are in salmon, and the minor editing sites (C₁-C₉ and C₁₃-C₂₃) are in green. Means \pm s.d. from three independent experiments are shown in **a** and **b**. **(b)** The ratios of major editing to minor editing, determined at the indicated genomic target sites. **(c)** Statistical analysis of the normalized ratios of major editing to minor editing, with those induced by dCpf1-BE set to 100%. dCpf1-BE-YE induced a significantly higher ratio of major editing to minor editing. The median, interquartile range (IQR) and 1.5 \times IQR are shown. $n = 15$ independent samples from 3 independent experiments. P values from one-tailed Student's t test are shown in **c**, **e** and **g**. **(d-g)** The addition of free UGI enhances the purity of editing outcomes induced by dCpf1-BEs. The fractions **(d,f)** and statistical analyses **(e,g)** of cytosine substitutions at the indicated editing positions under the indicated conditions are shown. dCpf1-eBE and dCpf1-eBE-YE induced significantly purer C-to-T-editing outcomes than did dCpf1-BE and dCpf1-BE-YE **(e,g)**. Data are shown as median and IQR. $n = 9$ independent samples from 3 independent experiments. **(h)** Summary of Cas9-based and dCpf1-based BEs. Left, schematic diagrams illustrating the complexes of Cas9-BE-sgRNA-target DNA and dCpf1-BE-crRNA-target DNA. Right, list of relevant features in Cas9-based and dCpf1-based BE systems. Comparisons are based on base editing at DYRK1A-, FANCF- and RUNX1-target sites.

Furthermore, we examined possible OT base editing induced by dCpf1-BE at 40 predicted OT sites¹⁹ for eight crRNAs (five OT sites per crRNA) and found OT base editing for one crRNA at three sites (Supplementary Fig. 7). Although previous findings have shown that Cpf1 nuclease has a high specificity of gene editing^{12–14}, and we also found herein that dCpf1-BE induced a low level of base editing at *in silico*-predicted OT sites, other experimentally characterized Cpf1 OTs can be further examined in the dCpf1-BE system to test specificity. It will also be interesting to compare the specificity of dCpf1-BE with high-fidelity Cas9-based BE3 (HF-BE3)²⁰ and to use an engineered Cpf1 nuclease with improved specificity²¹ for additional Cpf1 BEs. Finally, assaying a 44-nt region outside the spacer sequence, we rarely detected C-to-T base conversions (Supplementary Fig. 8).

Next, we compared the editing efficiency of dCpf1-BE with those of different Cas9-BEs at eight target sites where the editing windows of dCpf1-BE (position 8–13) and Cas9-BEs (position 4–8) overlap. dCpf1-BE generally induced higher editing frequencies than did dCas9-BE2 at the 14 commonly editable cytosines, and it reached or exceeded the editing level induced by nCas9-BE3 at 5 of the 14 editable cytosines (Fig. 1b, C-to-T-editing frequency). However, we cannot rule out the possibility that the greater number of NLSs and the longer linker between dCpf1 and UGI might have contributed to the performance of dCpf1-BE compared with nCas9-BE3. At the other nine commonly editable cytosines, dCpf1-BE induced lower base-editing levels than did nCas9-based BE3 (Fig. 1b). Notably, in all cases, dCpf1-BE induced fewer indels and non-C-to-T substitutions than did nCas9-BE3 (Fig. 1b and Supplementary Fig. 5d; $P = 5 \times 10^{-10}$). Notably, recent efforts have been applied to enhance nCas9-based BEs (eBE and BE4) by using additional UGIs to decrease indels and non-C-to-T conversions^{17,22}. Although no conclusive reports to date have described a Cpf1 nickase, such enzymes would be developed in the future and might be useful to increase base-editing efficiency.

To further narrow the 6-nt editing window of dCpf1-BE (positions 8–13; Supplementary Fig. 5a) and to decrease multiple C-to-T base conversions (Supplementary Fig. 9), we introduced mutations (W90Y and R126E) into the APOBEC domain that have previously been shown to narrow the editing window in nCas9-BE3 (ref. 6) (Supplementary Fig. 10a). In four of the five tested genomic loci, dCpf1-BE-YE retained ~30–90% of the original editing efficiency at its highly preferred editing positions (major editing sites, positions 10–12; Fig. 2a) but showed markedly lower editing efficiencies elsewhere in the spacer region (minor editing sites, positions 1–9 and 13–23; Fig. 2a), thus leading to increased ratios of major editing to minor editing (Fig. 2b). The ratios of major editing to minor editing induced by dCpf1-BE-YE, after normalization to dCpf1-BE (Supplementary Fig. 10b), increased approximately two- to threefold (Fig. 2c; $P = 0.0005$). As expected, dCpf1-BE-YE also yielded a higher fraction of single C-to-T substitutions than did dCpf1-BE when two or more cytosines were in the editing window (Supplementary Fig. 10c,d). dCpf1-BE-YEE with a third mutation (R132E) decreased the base-editing frequency at all editing positions to near-background levels (Fig. 2a). Similarly to dCpf1-BE, both dCpf1-BE-YE and dCpf1-BE-YEE scarcely induced unwanted indels (Supplementary Fig. 10e). Thus, base editing was able to be specifically narrowed to a 3-nt window (positions 10–12) by dCpf1-BE-YE, although the editing efficiency was also strongly decreased.

Fewer non-C-to-T substitutions were induced by dCpf1-BE than by nCas9-BE3 (Fig. 1b), but these substitutions were still noticeable at some editing sites (DYRK1A-C₁₀, FANCF-C₁₀ and RUNX1-C₁₀; Fig. 1b, fractions of cytosine substitutions). Coexpressing extra UGI proteins has been shown to substantially decrease these unintended

non-C-to-T substitutions^{17,22}. We therefore added three copies of 2A-UGI sequences to the 3' end of the dCpf1-BE coding region to construct dCpf1-eBE (Supplementary Fig. 11a). As expected, the formation of non-C-to-T substitutions was suppressed in dCpf1-eBE-mediated editing (Fig. 2d). As a result, the fraction of C-to-T substitutions was further enhanced (Fig. 2e; $P = 0.0002$) while the editing efficiencies remained largely unchanged (Supplementary Fig. 11b, comparison of dCpf1-eBE and dCpf1-BE). Similarly, the C-to-T fraction was also increased in dCpf1-eBE-YE-mediated base editing (Fig. 2f,g, $P = 0.007$, and Supplementary Fig. 11d) with little influence on the editing efficiency (Supplementary Fig. 11e). Moreover, both dCpf1-eBE and dCpf1-eBE-YE induced almost undetected indels at all examined genomic loci (Supplementary Fig. 11c,f).

In summary, we developed a series of CRISPR-Cpf1-based BEs that can perform targeted base editing with very low levels of indel formation and non-C-to-T substitutions (Fig. 2h), and can facilitate base editing in A/T-rich regions. In the future, we expect that other Cpf1 enzymes (for example, FnCpf1 which recognizes a TTN PAM⁸) or engineered Cpf1 mRNA and crRNA²³ may be used to further enhance dCpf1-BEs.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the [online version of the paper](#).

ACKNOWLEDGMENTS

This work was supported by grants 2014CB910600 (L.Y.) from MOST; grants 31600619 (B.Y.), 31600654 (J.C.), 31471241 (L.Y.), 31730111 (L.Y.) and 91540115 (L.Y.) from NSFC; and grants 16PJ1407000 (J.C.) and 16PJ1407500 (B.Y.) from the Shanghai Municipal Science and Technology Commission.

AUTHOR CONTRIBUTIONS

J.C., X.H. and L.Y. conceived, designed and supervised the project. J.C. managed the project. X.L. and Y.L. performed most experiments on plasmid construction and cell culture with the help of X.W., Z.L., Y.Z. and J. Wu. J. Wei. prepared libraries for deep sequencing, and Y.W. performed bioinformatics analyses, supervised by L.Y. J.C., B.Y. and L.Y. wrote the paper with input from all authors.

COMPETING INTERESTS

The authors declare no competing interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. *Nature* **533**, 420–424 (2016).
- Nishida, K. *et al. Science* **353**, aaf8729 (2016).
- Kim, K. *et al. Nat. Biotechnol.* **35**, 435–437 (2017).
- Zong, Y. *et al. Nat. Biotechnol.* **35**, 438–440 (2017).
- Li, G. *et al. Protein Cell* **8**, 776–779 (2017).
- Kim, Y.B. *et al. Nat. Biotechnol.* **35**, 371–376 (2017).
- Kleistiver, B.P. *et al. Nature* **523**, 481–485 (2015).
- Zetsche, B. *et al. Cell* **163**, 759–771 (2015).
- Yamano, T. *et al. Cell* **165**, 949–962 (2016).
- Stella, S., Alcón, P. & Montoya, G. *Nature* **546**, 559–563 (2017).
- Swarts, D.C., van der Oost, J. & Jinek, M. *Mol. Cell* **66**, 221–233.e224 (2017).
- Kim, D. *et al. Nat. Biotechnol.* **34**, 863–868 (2016).
- Kleistiver, B.P. *et al. Nat. Biotechnol.* **34**, 869–874 (2016).
- Yan, W.X. *et al. Nat. Commun.* **8**, 15058 (2017).
- Chen, J., Miller, B.F. & Furano, A.V. *eLife* **3**, e02001 (2014).
- Staahl, B.T. *et al. Nat. Biotechnol.* **35**, 431–434 (2017).
- Komor, A.C. *et al. Sci. Adv.* **3**, eaao4774 (2017).
- Kim, H.K. *et al. Nat. Methods* **14**, 153–159 (2017).
- Bae, S., Park, J. & Kim, J.S. *Bioinformatics* **30**, 1473–1475 (2014).
- Rees, H.A. *et al. Nat. Commun.* **8**, 15790 (2017).
- Gao, L. *et al. Nat. Biotechnol.* **35**, 789–792 (2017).
- Wang, L. *et al. Cell Res.* **27**, 1289–1292 (2017).
- Li, B. *et al. Nat. Biomed. Eng.* **1**, 0066 (2017).

ONLINE METHODS

Plasmid construction. pST1374-Lb-Cpf1-NLS was commercially synthesized. Two primer sets (LB_D971A_F/LB_R4635 and LB_D971A_R/LB_F2096) were used to amplify the D832A-containing fragment LbCpf1-D832A. Then two primer sets (LB_E1006A_F/LB_E1006A_R and LB_D1225A_F/LB_D1225A_R) were used to introduce the mutations E925A and D1148A. The D832A-, E925A- and D1148A-containing dLbCpf1 was cloned into PstI- and ApaI-linearized pST1374-LbCpf1-NLS with a Clone Express (Vazyme, C112-02) plasmid-recombination kit to generate the dLbCpf1 expression plasmid pST1374-dLbCpf1-NLS. Two primer sets (LB_BE3_F1/LB_BE3_R1 and LB_BE3_F2/CPF_BE3_fu_R2) were used to amplify the dLbCpf1-SV40 NLS-UGI fragment, which was cloned into SmaI- and PmeI-linearized pCMV-BE3 to generate the dLbCpf1-BE0 (dCpf1-BE0) expression vector pCMV-Apobec1-XTEN-dLbCpf1(D832A/E1006A/D1125A)-SV40NLS-SGGS-UGI-SV40NLS.

pST1374-As-Cpf1-NLS was commercially synthesized. Two primer sets (AS_D917A_F/AS_R4871 and AS_D917A_R/AS_F2155) were used to amplify the D908A-containing fragment AsCpf1-D908A. Then two primer sets (AS_E1006A_F/AS_E1006A_R and AS_D1225A_F/AS_D1225A_R) were used to introduce the mutations E993A and D1235A. The D908A-, E993A- and D1235A-containing fragment dAsCpf1 was cloned into PstI- and ApaI-linearized pST1374-AsCpf1-NLS to generate the dAsCpf1 expression plasmid pST1374-dAsCpf1-NLS. Two primer sets (As_BE3_F1/As_BE3_R1 and As_BE3_F2/CPF_BE3_fu_R2) were used to amplify the dAsCpf1-SV40 NLS-UGI fragment, which was cloned into SmaI- and PmeI-linearized pCMV-BE3 to generate the dAsCpf1-BE0 expression vector pCMV-Apobec1-XTEN-dAsCpf1(D908A/E993A/D1235A)-SV40NLS-SGGS-UGI-SV40NLS.

Oligonucleotides (L079_LbCpf1scaffold_for/L080_LbCpf1scaffold_rev and L081_AsCpf1scaffold_for/L082_AsCpf1scaffold_rev) were annealed and ligated into BsaI- and EcoRI-linearized pGL3-U6-sgRNA-PGK-puromycin (Addgene, 51133) to generate the Lb-crRNA and As-crRNA expression vectors pLb-Cpf1-pGL3-U6-sgRNA and pAs-Cpf1-pGL3-U6-sgRNA.

Oligonucleotides supF_Cpf1_sg1_FOR/supF_Cpf1_sg1_REV, supF_Cpf1_sg2_FOR/supF_Cpf1_sg2_REV, supF_Cpf1_sg3_FOR/supF_Cpf1_sg3_REV or other pairs of oligonucleotides with different lengths were annealed and ligated into BsaI-linearized pLb-Cpf1-pGL3-U6-sgRNA or pAs-Cpf1-pGL3-U6-sgRNA to generate the expression vectors for the Lb-crRNAs or As-crRNAs targeting the *SupF* gene in the shuttle vector pSP189.

Two primer sets (LB_BE3_F1/LB_R and UGI_F/CPF_BE3_fu_R2) were used to amplify the dLbCpf1-SGGS-UGI fragment, which was cloned into the SmaI- and PmeI-linearized dLbCpf1-BE0 (dCpf1-BE0) expression vector to generate the dLbCpf1-BE0ΔNLS expression vector pCMV-Apobec1-XTEN-dLbCpf1(D832A/E1006A/D1125A)-SGGS-UGI-SV40NLS.

The primer set 1xNLS_pcrF/1xNLS_pcrR was used to amplify the fragment NLS-Apobec1 from pCMV-BE3, and the gel-purified NLS-Apobec1 fragment was ligated into the SmaI- and NotI-linearized dCpf1-BE0 expression vector to generate the dCpf1-BE expression vector pCMV-SV40NLS-Apobec1-XTEN-dLbCpf1(D832A/E1006A/D1125A)-SV40NLS-SGGS-UGI-SV40NLS.

Two primer sets (APOBEC_W90Y_F1/1xNLS_pcrR and 1xNLS_pcrF/APOBEC_W90Y_R1) were used to amplify the W90Y-containing fragment APOBEC-Y with the primer set. Two primer sets (APOBEC_R126E_F/APOBEC_R126E_R and APOBEC_R132E_F/APOBEC_R132E_R) were used to introduce the mutations R126E and R132E. The APOBEC-YE and APOBEC-YEE fragments were ligated into the NotI- and SmaI-linearized dCpf1-BE expression vector to generate the respective dCpf1-BE-YE and dCpf1-BE-YEE expression vectors pCMV-SV40NLS-Apobec1(W90Y/R126E)-XTEN-dLbCpf1(D832A/E1006A/D1125A)-SV40NLS-SGGS-UGI-SV40NLS and pCMV-SV40NLS-Apobec1(W90Y/R126E/R132E)-XTEN-dLbCpf1(D832A/E1006A/D1125A)-SV40NLS-SGGS-UGI-SV40NLS.

The primer set LB_F2096/BE8.1_PmeI_ApaI_R was used to introduce the ApaI site into the dCpf1-BE expression vector to generate pCMV-dCpf1-BE-ApaI. The primer set ApaI_1T2AUGI_F/PmeI_3T2AUGI_R was used to amplify the 3 × 2A-UGI fragment from the commercially synthesized DNA fragment 3 × 2A-UGI, and the 3 × 2A-UGI fragment was ligated into PmeI- and ApaI-linearized pCMV-dCpf1-BE-ApaI to generate the dCpf1-eBE expression vector pCMV-SV40NLS-Apobec1-XTEN-dLbCpf1(D832A/E1006A/D1125A)-SV40NLS-SGGS-UGI-SV40NLS-T2A-UGI-SV40NLS-P2A-UGI-SV40NLS-T2A-UGI-SV40NLS.

Apobec1-YE fragments were ligated into the NotI- and SmaI-linearized dCpf1-eBE expression vector to generate the dCpf1-eBE-YE expression vector pCMV-SV40NLS-Apobec1(W90Y/R126E)-XTEN-dLbCpf1(D832A/E1006A/D1125A)-SV40NLS-SGGS-UGI-SV40NLS-T2A-UGI-SV40NLS-P2A-UGI-SV40NLS-T2A-UGI-SV40NLS.

Oligonucleotides hCDKN2A_cpf1_sg1_FOR/hCDKN2A_cpf1_sg1_REV were annealed and ligated into BsaI-linearized pLb-Cpf1-pGL3-U6-sgRNA to generate the crCDKN2A expression vector pcrCDKN2A. Oligonucleotides hCDKN2A_cpfsf_sg1_FOR/hCDKN2A_cpfsf_sg1_REV were annealed and ligated into BsaI-linearized pGL3-U6-sgRNA-PGK-puromycin to generate the sgCDKN2A expression vector psgCDKN2A. Other crRNA and sgRNA expression vectors were constructed in the same manner.

The sequences of the oligonucleotides used for plasmid construction are listed in **Supplementary Table 1**, and the sequences of plasmids are listed in **Supplementary Note 1**.

Cell culture and transfection. 293FT and U2OS cells from ATCC were maintained in DMEM (10566, Gibco/Thermo Fisher Scientific) with 10% FBS (16000-044, Gibco/Thermo Fisher Scientific) and were tested to exclude mycoplasma contamination.

For base editing in episomal shuttle vectors, 293FT cells were seeded in a six-well plate at a density of 5×10^5 cells/well and were transfected with 500 μl serum-free Opti-MEM containing 4 μl Lipofectamine LTX (Invitrogen/Life Technologies), 2 μl Lipofectamine Plus (Invitrogen/Life Technologies), 1 μg dLbCpf1-BE0 expression vector (or dAsCpf1-BE0 expression vector), 0.5 μg crRNA-expressing plasmid and 0.5 μg shuttle vector pSP189. After 48 h, the plasmids were extracted from the cells with a TIANprep Mini Plasmid Kit (DP103-A, Tiangen).

For base editing in genomic DNA, 293FT and U2OS cells were seeded in a 24-well plate at a density of 2×10^5 cells/well and transfected with 500 μl serum-free Opti-MEM containing 5.04 μl Lipofectamine LTX (Invitrogen/Life Technologies), 1.68 μl Lipofectamine Plus (Invitrogen/Life Technologies), 1 μg dCpf1-BE0 expression vector (dCpf1-BE0ΔNLS, dCpf1-BE, dCpf1-BE-YE, dCpf1-BE-YEE, dCpf1-eBE or dCpf1-eBE-YE expression vector, or pCMV-BE2 or pCMV-BE3) and 0.68 μg crRNA or sgRNA-expressing plasmid. After 72 h, the genomic DNA was extracted from the cells with QuickExtract DNA Extraction Solution (QE09050, Epicentre).

Blue-white colony screening. The plasmids extracted from transfected cells were digested with DpnI (which removes unreplicated input plasmid) and transformed into *E. coli* MBM7070 (*lacZ*^{uag-amber}) cells, which were grown on LB plates containing 50 μg/ml kanamycin, 1 mM IPTG and 0.03% blue-gal (Invitrogen/Life Technologies) at 37 °C overnight and then at room temperature for another day (for maximal color development). To determine the mutation spectrum, white colonies were randomly picked for Sanger sequencing. The sequences of mutant shuttle vectors are listed in **Supplementary Note 2**.

DNA-library preparation and sequencing. Target genomic sites were PCR-amplified with the high-fidelity DNA polymerase PrimeSTAR HS (Clontech) with primers flanking each examined sgRNA-target site. The PCR primers used to amplify target genomic sequences are listed in **Supplementary Table 2**. Indexed DNA libraries were prepared with a TruSeq ChIP Sample Preparation Kit (Illumina) with minor modifications. Briefly, the PCR products amplified from genomic-DNA regions were fragmented with a Covaris S220 ultrasonicator. The fragmented DNAs were then PCR amplified with a TruSeq ChIP Sample Preparation Kit (Illumina). After being quantified with a Qubit High-Sensitivity DNA Kit (Invitrogen), PCR products with different tags were pooled for deep sequencing by using Illumina HiSeq 2500 (1 × 100) or HiSeq X-10 (2 × 150) platforms at the CAS-MPG Partner Institute for Computational Biology Omics Core, Shanghai, China. Raw read qualities were evaluated with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>, v0.11.4). For paired-end sequencing, only R1 reads were used. Adaptor sequences and read sequences on both ends with Phred quality scores below 28 were trimmed. Trimmed reads were then mapped with the BWA-MEM algorithm (BWA v0.7.9a) to target sequences. After being piped up with samtools (v0.1.18), indels and base substitutions were further calculated.

Indel-frequency calculation. For Cpf1, indels were estimated in the aligned regions spanning from 3 nt upstream to 48 nt downstream of PAM sites (55 bp). For Cas9, indels were estimated in the aligned regions spanning from 8 nt upstream of the target site to 19 nt downstream of PAM sites (50 bp). Indel frequencies were subsequently calculated by dividing reads containing at least one inserted and/or deleted nucleotide by all the mapped reads at the same region. Counts of indel-containing reads and total mapped reads are listed in **Supplementary Table 3**.

Base-substitution calculation. Base substitutions were selected at each position of the examined sgRNA- (or crRNA-) target sites that mapped with at least 1,000 independent reads, and obvious base substitutions were observed at only the targeted base-editing sites. Counts of reads for each base and total

reads are listed in **Supplementary Table 4**. Base-substitution frequencies were calculated by dividing base-substitution reads by total reads.

Statistical analysis. *P* values were calculated with one-tailed Student's *t* tests.

Life Sciences Reporting Summary. Further information on experimental design is available in the **Life Sciences Reporting Summary**.

Data availability. The deep-sequencing data from this study have been deposited in the NCBI Gene Expression Omnibus (accession no. [GSE110136](#)) and the National Omics Data Encyclopedia (accession no. [NODEP00371765](#)). The data sets used in this study are provided in **Supplementary Tables 3** and **4**.

Life Sciences Reporting Summary

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

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

► Experimental design

1. Sample size

Describe how sample size was determined.

No statistical methods were used to predetermine sample size. More than twelve target sites were tested in this study, which was sufficient in the field.

2. Data exclusions

Describe any data exclusions.

No data were excluded

3. Replication

Describe whether the experimental findings were reliably reproduced.

The experimental findings in all figures were reproduced successfully

4. Randomization

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

Samples were not randomized

5. Blinding

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

The investigators were not blinded to group allocation

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

6. Statistical parameters

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

n/a Confirmed

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

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

► Software

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

7. Software

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

FastQC (v0.11.4), BWA (v0.7.9a), samtools (v0.1.18), KaleidaGraph (v4.5.3) and Microsoft Excel (v14.7.7)

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

► Materials and reagents

Policy information about [availability of materials](#)

8. Materials availability

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

Materials in this study are available for distribution following MTA

9. Antibodies

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

No antibodies were used

10. Eukaryotic cell lines

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

All cell lines were from ATCC

b. Describe the method of cell line authentication used.

No cell lines have were authenticated

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

All cell lines have been tested for mycoplasma contamination free by PCR methods

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

No commonly misidentified cell lines were used

► Animals and human research participants

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

11. Description of research animals

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

No animals were used

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

12. Description of human research participants

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

This study did not involve human research participants