Nature Communications

https://doi.org/10.1038/s41467-026-69678-5

Article in Press

Engineered Un1Cas12f1 for multiplex genome editing with enhanced activity and targeting scope

Received: 15 August 2025

Accepted: 4 February 2026

Cite this article as: Huo, Y., Mei, J., Zhang, D. *et al.* Engineered Un1Cas12f1 for multiplex genome editing with enhanced activity and targeting scope. *Nat Commun* (2026). https://doi.org/10.1038/s41467-026-69678-5

Yanan Huo, Jiale Mei, Dan Zhang, Bing Yan, Dexin Zhang, Chao Dong, Shuming Yin, Meizhen Liu, Xinyan Wang, Dan Chen, Yuting Guan, Gaojie Song, Bing Du, Yongming Wang, Zongli Zheng, Hong Liu, Dali Li, Lei Yang & Liren Wang

Engineered Un1Cas12f1 for multiplex genome editing with enhanced activity and targeting scope

Yanan Huo[1,†], Jiale Mei[1,†], Dan Zhang[1,†], Bing Yan[1], Dexin Zhang[2], Chao Dong[1], Shuming Yin[1], Meizhen Liu[1], Xinyan Wang[1], Dan Chen[1], Yuting Guan[1], Gaojie Song[1], Bing Du[1], Yongming Wang[3], Zongli Zheng[4], Hong Liu[5], Dali Li[1,6*], Lei Yang[7*], Liren Wang[1*]

[1] Shanghai Frontiers Science Center of Genome Editing and Cell Therapy, Shanghai Key Laboratory of Regulatory Biology, School of Life Sciences, East China Normal University, Shanghai, 200241, China

[2] Department of Urology, Children's Hospital of Fudan University, Shanghai, 201102, China

[3] Shanghai Engineering Research Center of Industrial Microorganisms, Shanghai, 200438, China

[4] Department of Biomedical Sciences and Tung Biomedical Sciences Centre, City University of Hong Kong, Kowloon, Hong Kong SAR of the People's Republic of China

[5]Department of Dermatology, Xiangya Hospital & School of Life Sciences & Furong Laboratory, Central South University, Changsha, China

[6] Shanghai Academy of Natural Sciences (SANS), Shanghai, China

[7] Hangzhou Institute of Medicine (HIM), Chinese Academy of Sciences, Hangzhou, 310018, China

[†] These authors contributed equally

[*] Corresponding authors emails: Dali Li: dlli@bio.ecnu.edu.cn; Lei Yang: yanglei@him.cas.cn; Liren Wang: lrwang@bio.ecnu.edu.cn

Abstract

The compact CRISPR-Cas12f system is promising for AAV-delivered gene therapy, but its application has been constrained by restrictive PAM recognition (e.g., TTTR) and suboptimal editing efficiency. Through bacterial library screening and mammalian cell validation, we engineer evoCas12f, an optimized variant incorporating five key mutations, that dramatically expands PAM recognition to NTNR/NYTR. This advancement reduces median distance between two neighbouring PAM sites to 2 nucleotides in the human genome. It also demonstrates 1.4-fold enhanced activity at TTTR sites compared to wild-type Un1Cas12f1, achieving up to 91% editing efficiency. Remarkably, evoCas12f enables efficient generation of homozygous mutations in F0 generation mice, even at non-canonical PAM sites. We further adapt this system for robust transcriptional activation and precise base editing with a well-defined editing window. As a compact yet highly efficient platform, evoCas12f represents a significant advance in CRISPR technology, enabling multiplexed editing for high-resolution targeting applications and expanding possibilities for therapeutic genome engineering.

Introduction

Originally discovered in bacteria and archaea as an adaptive immune mechanism, the CRISPR-Cas system has since evolved and been adapted into revolutionary tools for genome editing[1–3]. Among the six types of the CRISPR systems belonging to two classes, types II, V, and VI are composed of a single effector nuclease that can be flexibly guided by engineered single guide RNAs (sgRNAs), and they attract more attention for multiplex applications[4–6]. Compared to the widely used type II Cas9 system, which primarily recognizes G-rich protospacer adjacent motifs (PAMs), the type V CRISPR-Cas12 system relies on T-rich or a variety of PAMs with high specificity[7,8]. Although SpCas9 has been engineered for unconstrained genome targeting, its large size still limits its application due to in vivo delivery challenges[9–11]. In recent years, the identification of Cas12 family flourished and at least 14 distinct subtypes have been discovered with a wide range of protein sizes from 400 to 1500 amino acids (aa)[12,13], including recently reported highly compact Cas12 systems (less than 700 aa), such as Cas12m (~600 aa), Cas12n (400-700 aa), and Cas12f (400–700 aa)[14–16]. These miniature Cas enzymes have attracted a lot of attention since their small size is appropriate for packaging into adenovirus-associated virus (AAV) vectors, which are the most successful delivery system for in vivo gene therapy[17–20], despite having a limited packaging capacity of approximately 4.7 kb[21]. However, most of these compact Cas12 proteins demonstrate only modest or even infrequent editing activity in their natural state in mammalian cells.

Un1Cas12f1, the first characterized member of the Cas12f family, was originally identified as a programmable nuclease targeting single-stranded DNA (ssDNA) and subsequently demonstrated to cleave double-stranded DNA[16,22]. Un1Cas12f1 shows great promise for genome editing due to its compact size and high specificity[23]. Later, its cleavage activity has been improved through sgRNA scaffold optimization and protein engineering[24,25]. Following the success of Un1Cas12f1 optimization, several other members of the Cas12f family, which recognize various PAM sequences, have been characterized. These include AsCas12f1 (PAM: NTTR), SpaCas12f1 (PAM: NTTY), RhCas12f1 (PAM: NCCD), CnCas12f (PAM: NCCD), and OsCas12f1 (PAM:

YTTH)[22,26–29]. Leveraging their small size for efficient delivery, some of these Cas12f members have been adapted to develop base editors and transcriptional regulatory tools potentially for in vivo gene therapy[24,30–33]. Although both the Cas12f enzymes and guide RNAs have undergone engineering to improve their editing efficiency, their activity remains suboptimal. Additionally, the targeting scope of engineered Cas12f variants is limited by their complex PAM sequence requirements. Therefore, there is an urgent need for highly efficient miniature Cas proteins with broad PAM compatibility to unlock their full potential across diverse genome editing applications.

To expand the targeting scope, we establish bacterial positive library selection and endogenous target screening to develop evoCas12f, a variant containing five amino acid substitutions that exhibits a broader targeting scope and higher editing efficiency. evoCas12f expands the PAM recognition range from the original TTTR to a more permissive NTNR or NYTR spectrum. Compared to wild-type Un1Cas12f1, evoCas12f demonstrates an average 12.5-fold increase in cleavage efficiency across all permissible PAMs, achieving editing rates of up to 91% in mammalian cells and generating homozygous mouse model in F0 generation at non-TTTR PAM targets. Notably, evoCas12f shows superior editing activity compared to previously engineered Cas12f variants, such as enAsCas12a and other Cas12f family members, under similar PAM contexts. Furthermore, we employ the strategy of using engineered sgRNA to recruit effector proteins to develop adenine base editors (ABEs), cytosine base editors (CBEs), and transcriptional activators based on evoCas12f. Collectively, these findings demonstrate that evoCas12f is a compact, highly efficient, and versatile genome-editing platform with a broadened targeting scope.

Results

Relaxing the PAM recognition profile using a point mutation library

Mutations within the PAM-interacting region are likely to alter the PAM recognition spectrum of a Cas effector[9–11,34–40]. To explore this, we constructed a plasmid library containing single-point saturation mutations across the PAM-interacting regions (K129-S168, K196-I207, Q242-K249) of Un1Cas12f1 whose expression was induced by IPTG. A previously reported bacterial positive selection system[41] was leveraged for the screening (Fig. 1a). In this system, a screening plasmid containing a target with a specific PAM sequence followed by an arabinose-inducible *ccdB* suicide gene was first introduced into *E. coli.* During the screening process, the mutant library was electroporated into the competent *E. coli* cells, induced by IPTG, and plated on LB agar containing arabinose. The survival of *E. coli* depended on the ability of the Cas12f variant to recognize the PAM and cleave the target sequence, thus preventing the expression of the suicide gene. To validate this system, we electroporated WT-Un1Cas12f1 into cells containing the TTTG-PAM selection plasmid or negative control plasmid (no target sequence is included). As expected, the growth of *E. coli* containing the negative control plasmid was completely inhibited after arabinose induction. Cleavage of the selection plasmid by WT-Un1Cas12f1 restored the growth of competent cells (Supplementary Fig. 1 a and b). When the TTTG-PAM was replaced with non-canonical PAMs, *E. coli* survival decreased very dramatically, confirming that WT-Un1Cas12f1 exhibits very minimal activity on non-TTTG PAMs in bacteria (Supplementary Fig. 1c).

Next, we electroporated the Un1Cas12f1 library into four batches of competent cells, each harboring one of the NTTR-PAMs. Colonies that grew on LB agar containing arabinose were harvested the following day and analyzed by Next-Generation Sequencing (NGS). We cloned 38 Un1Cas12f1 variants that showed over 100-fold enrichment into eukaryotic-expression plasmids (Fig. 1b). Their editing efficiencies in HEK293T cells were evaluated at three endogenous sites of each NTTR PAM. We discovered three point mutations (D143S, L152S and Q244R) that exhibited the highest editing efficiencies at (T/C/A/G)TTR PAM targets respectively (Fig. 1c and Supplementary Data 1). To develop mutants with broader PAM compatibility, we integrated

D143S, L152S and Q244R to develop Cas12f-SSR. Then we tested 14 endogenous targets with NTTR PAMs to evaluate its performance. The results demonstrated that Cas12f-SSR increased its average efficiency from 5.0% to 24.2% on VTTR PAMs, and an average 1.6-fold improvement at TTTR sites compared to WT-Un1Cas12f1 (Fig. 1d).

To further relax the PAM requirement at the second position of the TTTR PAM, we conducted a second round of screening against TVTR PAMs (V = A/C/G) through the same methods used above. We discovered the N133S variant exhibited increased activity at two of three endogenous targets with TATR PAMs, while E151V outperformed other variants on TGTR PAMs (Fig. 1 e and f). Then we incorporated N133S and E151V into Cas12f-SSR and resulted in the Cas12f-SSRSV (hereafter referred as evoCas12f). Next, using 24 endogenous sites with NTTR or TNTR PAM in HEK293T cells, we tried to evaluate whether evoCas12f had expanded PAM recognition. Notably, evoCas12f showed cleavage efficiencies ranged from 14.79% to 71.41% across all these targets, with a broader PAM compatibility compared to WT-Un1Cas12f1 (Fig. 1g).

Arginine (Arg) substitutions are often used to strengthen DNA–protein interactions. Motivated by the benefit of Q244R, we introduced individual Arg substitutions at N133, D143, L152, and E151 and also evaluated combinatorial variants. However, the results indicated that D143S, L152S, N133S, and E151V each had generally higher editing efficiency than their to-arginine counterparts at the corresponding PAM endogenous sites. (Supplementary Fig. 2 a-d) Moreover, the five-Arg construct (N133R/D143R/E151R/L152R/Q244R) exhibited consistently lower activity on NTTR and TNTR PAM sites (Supplementary Fig. 2e). To investigate potential mechanisms, we analyzed the structure of Un1Cas12f1-sgRNA-dsDNA (PDB: 7C7L) and performed corresponding in silico point mutations. Our analysis revealed that Q244R, extends deeper into the DNA minor groove, forming hydrophilic interactions and salt bridges with phosphate groups to enhance editing efficacy and broaden PAM compatibility via non-specific affinity (Supplementary Fig. 3). Among the non-arginine mutations, N133S may promote DNA strand separation near the cleavage site; D143S and E151V may function by neutralizing acidic residues to alleviate electrostatic repulsion near the PAM-interacting DNA backbone; while L152S may improve steric compatibility with

surrounding polar residues to harmonize the local environment (Supplementary Fig. 3; see Discussion).

Characterization of the activity and PAMs of evoCas12f

To further characterize the activity of evoCas12f on more relaxed PAM targets, we tested 40 endogenous sites containing NNTR PAMs in HEK293T cells (Supplementary Data 1). The editing efficiencies of WT-Un1Cas12f1 were below 5% at most NVTR sites and less than 10% at the VTTR PAM (V=A/G/C) sites (Fig. 2a and Supplementary Fig. 4). In contrast, evoCas12f exhibited higher editing efficiency across all NNTR sites, with an average increase of 15.4-fold (Fig. 2 a and d). Notably, WT-Un1Cas12f1 displayed extremely low efficiency at GCTG-2 site (0.2%), while evoCas12f achieved 19.93% indel efficiency. Statistically, evoCas12f exhibits a 20-fold increase in activity on NVTR PAM targets (Fig. 2a).

To validate the PAM preference of evoCas12f statistically, we employed an HEK293T cell line harboring a GFP reporter system[42]. This system includes a GFP gene preceded by a target sequence with a N_5 PAM after the ATG start codon, resulting in a frameshift mutation. Cleavage by the Cas12f protein introduces indels in the protospacer, potentially restoring the GFP reading frame (Supplementary Fig. 5). As expected, WT-Un1Cas12f1 exhibited a TTTR PAM, while evoCas12f displayed a more relaxed NTNR PAM, maintaining a slight bias toward 'T' at position 2 (Fig. 2b). It is worth noting that the second thymidine in the PAM was not strictly required, as demonstrated in Fig 2a Supplementary Fig. 4.

To address evoCas12f's mild preference for T at PAM position 2, structural analysis identified Y146 and Q197 as contacting the second T of the PAM (Supplementary Fig. 6a). We conducted single amino acid saturation mutagenesis at each residue and tested editing at 6 VVTR sites. The results showed that any substitution at Y146 abolished its activity, whereas Q197K increased its activity at 4/6 VTTR PAM sites (Supplementary Fig. 6b). We then evaluated Q197K at 40 NNTR endogenous loci and found that Q197K boosts activity on NCTR PAMs, has minimal effect on NRTR, and reduces activity on NTTR to near WT Un1Cas12f1 levels (Supplementary Fig. 6c).

Collectively, Q197K does not expand evoCas12f's overall NNTR targeting range but shifts the preference of the 2nd position from T to C. Thus evoCas12f (Q197K) is best considered a specialized variant for NCTR targets.

To further evaluate the activity of evoCas12f on NTNR PAM targets, we tested additional 40 endogenous sites. For the ATNR, CTNR, and GTNR PAM targets, evoCas12f exhibited comparable average editing efficiencies (22.42%, 20.69%, and 21.24% respectively), representing an ~11-fold increase compared to the WT control (Fig. 2c). Among all endogenous targets with NNTR and NTNR PAM tested, evoCas12f demonstrated a 12.5-fold average increase in editing efficiency compared to WT-Un1Cas12f1 (Fig. 2d). Based on results of GFP reporter assay and endogenous cleavage experiments, we defined evoCas12f's PAM specificity as 5'-NTNR-3' and 5'-NYTR-3' (Y=C/T), although some of the NATR and NGTR sites also showed efficient (>20%) editing (Fig. 2b, Supplementary Fig. 7). We further analyzed the distribution of PAM sequences suitable for evoCas12f or WT-Un1Cas12f1 in the human genome. evoCas12f increased the potential cleavage sites by 13.3-fold, reducing the median distance between adjacent cleavage sites to 2 bp, whereas WT-Un1Cas12f1 primarily targeted sites spaced >30 bp apart (Fig. 2e).

Comparing editing activity of evoCas12f with other engineered Cas12 proteins

Next, we compared evoCas12f's cleavage efficiency with other engineered Cas12 proteins. Compared to enAsCas12a (1308aa), which recognizes expanded PAMs from TTTV to TTYN, VTTN, and TRTV sequence, evoCas12f exhibited comparable cleavage efficiency (69.0% VS 70.7%), indicating its activity was on a level with the large size Cas12 variant (Fig. 2 f and g). We further evaluated evoCas12f against other miniature Cas12 members including CasMINI[24], CWCas12f[30] variants, AsCas12f variants[43], enOsCas12f1[27] and enRhCas12f1[27]. As shown in Supplementary Fig. 8, across 47 endogenous targets spanning five compact Cas12f families, evoCas12f displayed generally superior or comparable activity. Notably, evoCas12f achieved higher cleavage activity at all matched targets than other engineered Cas12f variants recognizing T-rich PAMs. It also showed superior activity at 4 of 6 target sites compared to enRhCas12f1, which binds the 5′-CCD (D = A/G/T), a non-T-rich PAM (Supplementary Fig. 8e).

Compared with other engineered Cas12 variants, evoCas12f combines a compact structure, broadened targeting range, and enhanced cleavage activity, highlighting its potential for applications.

Generating Un1Cas12f1-ABE via RNA-protein recruitment modules

The compact size of Un1Cas12f1 makes it an ideal basal module for CRISPR-derived tools, such as base editors and epigenome regulators. Although several Cas12f related compact base editors and transcriptional regulatory tools have been reported[24,30–33,43], their activity, editing window and targeting flexibility remain suboptimal. In our preliminary experiments, fusing TadA8e directly to the N- or C-terminus of wild type dUn1Cas12f1 yielded only limited editing efficiency (Supplementary Fig. 9). We reasoned that Cas12f likely functions through dimerization, potentially obstructing TadA8e from accessing the substrate as it also functions as a dimer.

To address this limitation, we adopted an alternative strategy by incorporating an MS2 aptamer into the sgRNA stem loop to recruit the MCP-TadA8e fusion protein (Fig. 3a). Four constructs engineered with TadA8e at varying positions relative to MCP and dUn1Cas12f1 were tested. All constructs induced A-to-G conversions within the target site (Fig. 3 b and c). Among them, construct C1 showed the highest editing efficiency at the A_3 position of the two sites tested (Fig. 3c). Meanwhile, we screened optimal insertion sites for the MS2 aptamer and found that sgRNA with the MS2 in stem loop 4 yielded the highest editing efficiency (Fig. 3 d and e). However, adding a second MS2 aptamer in stem loop 4 did not further enhance editing efficiency (Fig. 3f). To optimize the linker connecting MCP and TadA8e, we generated the MCP–TadA8e constructs comprising a central XTEN linker flanked by various repeats of the flexible “GSSG” motif (Supplementary Table 3). However, in 3 endogenous sites with TTTR PAMs, we observed no obvious difference in editing efficiencies or in editing windows among these constructs (Supplementary Fig. 10). Thus we picked our initial constructs bearing a 32 aa linker in the following experiment. Previous studies have demonstrated that introducing mutations into TadA8e can substantially reduce off-target and bystander editing[44–46]. We therefore introduced mutations into TadA8e of the C1 construct, generating seven variants, as indicated in Fig. 3g. These variants

were evaluated at four target sites and the variant dCas12f-P2A-MCP-TadA8e (F148A) showed both high editing efficiency and a narrow editing window mainly restricted to the A_3 position (Fig. 3g). This construct was selected for subsequent studies and designated Un1Cas12f1-ABE.

Characterization of evoCas12f-derived base editors and epigenome editor

To generate highly efficient and PAM-expanded miniature base editors, we replaced WT-dUn1Cas12f1 with devoCas12f, creating evoCas12f-ABE. At 11 endogenous sites with non-TTTR PAMs, evoCas12f-ABE demonstrated on average, a 20-fold increase in A-to-G conversion rates (5.96%–41.75%) versus Un1Cas12f1-ABE (0.14%–5.48%) (Fig. 4 a and b). Importantly, evoCas12f-ABE maintained a narrow editing window while achieving broader PAM compatibility (Fig. 4c and Supplementary Fig. 11).

We then compared evoCas12f-ABE with other Cas12f-based ABEs, including dCasMINI-ABE[24], eCas12f-ABE[33], TaRGET-ABE-C3.1[30] and miniABE[31]. Comparison was performed at six endogenous sites with TTTR PAM (Fig. 4d and Supplementary Data 1), which is the optimal PAM for all adenine base editors above. evoCas12f-ABE, eCas12f-ABE and dCasMINI-ABE showed higher A-to-G frequencies than other compact ABEs (Fig. 4d). Although the efficiencies of dCasMINI-ABE and eCas12f-ABE were comparable to evoCas12f-ABE at these targets, evoCas12f-ABE displayed the narrowest editing window. TaRGET-ABE-C3.1, an ABE derived from CWCas12f, showed lower editing efficiency but a broader editing window compare to evoCas12f-ABE, while another Un1Cas12f1 derived miniABE demonstrated minimal activity (Fig. 4d). To further characterize evoCas12f-ABE, we analyzed its A-to-G conversion efficiencies across positions in the above 11 non-TTTR PAM and 6 TTTR PAM targets. The editing window of evoCas12f-ABE is predominantly localized to positions A_3 and A_4 (Fig. 4e). Overall, evoCas12f-ABE demonstrated superior performance in both efficiency and editing specificity compared to other compact Cas12f-based ABEs. These results suggest that the RNA-protein recruitment strategy confers an additional advantage by reducing bystander editing relative to direct fusion architectures (Fig. 4f). Next we compared Cas-independent DNA off-target editing of evoCas12f-ABE and other compact Cas12f-based ABEs using the R-loop assay[47]. All ABEs,

except for eCas12f-ABE, induced low (<3%) Cas-independent off-target editing at all six orthogonal SaCas9 sites, with evoCas12f-ABE displaying the highest on-target editing (Fig. 4g).

To generate a miniature CBE, we replaced TadA8e with our previously reported hA3A-D131P, a highly efficient and accurate variant of human APOBEC3A[48]. We compared C-to-T conversion efficiencies at different positions across two TTTR PAM targets. miniCBE[31] produced the highest C-to-T conversion rate (~10%) at position C_3, with minimal editing observed at other positions. In contrast, evoCas12f-CBE produced much higher editing efficiency, reaching up to 40% (Fig. 4h). Furthermore, evoCas12f-CBE exhibited superior C-to-T conversion efficiencies at non-canonical PAM targets compared to miniCBE (Fig. 4i).

The CRISPR-based transcriptional activation system has been shown to be a promising strategy for gene therapy[49,50]. Encouraged by the success of generating evoCas12f-ABE/CBE, we developed a transcriptional activator via replacing the TadA8e with transcriptional activation domain of P65 and HSF1 (Fig. 4j). We first optimized the relative positions of P65, HSF1, MCP and dUn1Cas12f1 domains, as well as the copy numbers of MS2, assessing activation efficiency using a TRE3G-GFP HEK293T reporter cell line (Supplementary Fig. 12 a-d). The optimal configuration—sgRNA with 5′-MS2+S2-MS2 (Fig. 3d and Supplementary Fig. 12 a and b) and construct with TA-N2 architecture—achieved the highest GFP-positive rate (81.53 ± 0.64%) (Supplementary Fig. 12 c and d). Subsequently, we replaced the dUn1Cas12f1 component with the devoCas12f, generating an evoCas12f based transcriptional activator—evoCas12fa. We compared the activation activity of four Cas12f transcriptional activation systems (evoCas12fa, SminiCRa[32], dCasMINI-VPR[24], and AsCas12f-HKRA-MPH[43]) in the TRE3G-GFP HEK293T reporter cell line. dCasMINI-VPR achieved the highest GFP-positive rate, reaching 88%, followed by evoCas12fa, at 72.8%, slightly higher than that of AsCas12f-HKRA-MPH (71.6%), while the activation efficiency of SminiCRa was the lowest, at only 56.4% (Fig. 4k). To validate evoCas12fa at endogenous loci, we selected 6 target sites across 4 genes: *ASCL1*, *CD2*, *HBB2*, and *IFNG*. 4 of 6 sites were with TTTR PAMs which were previously used to activate *ASCL1* and *IFNG*[24]. The other 2 sites were intentionally designed to be preceded by non-canonial PAMs (ATCA, and

TTCA). Compared with TA-N2 architecture, evoCas12fa demonstrated much higher transcriptional activation activity at the sites with TTTR PAMs, with the highest activation of 4734 fold. Notably, although evoCas12fa showed reduced efficiency at ASCL1-TTCA and HBB-ATCA sites compared to other TTTR PAM targets, the wild type activator (TA-N2) hardly induced gene activation at these loci (Supplementary Fig. 12e). Thus, evoCas12fa outperforms the wild type activation activator (TA-N2) in both efficiency and targeting range.

Precise correction of pathogenic mutations by evoCas12f-ABE and evoCas12f-CBE

To demonstrate the potential of evoCas12f-ABE/CBE, we performed an in-silico survey of ClinVar pathogenic SNVs under editor-specific activity-window and PAM constraints. PAM expansion with evoCas12f increased the ABE targetable fraction from 1.84% to 26.47% (≈14.4-fold) and the CBE targetable fraction from 3.21% to 54.01% (≈16.8-fold) (Fig. 5a). Thus, evoCas12f-ABE/CBE markedly enlarges the pool of variants theoretically correctable relative to WT-Un1Cas12f1 based editors.

To investigate the therapeutic potential of evoCas12f-ABE and evoCas12f-CBE, We designed four sgRNAs with non-TTTR PAMs to correct pathogenic SNVs in *ACVRL1* (associated with hereditary hemorrhagic)[49], *LAMA2* (associated with muscular dystrophy)[50], *GAA* (associated with Pompe disease)[51] and *NOTCH3* (associated with cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy)[52]. Delivery of evoCas12f-ABE or evoCas12f-CBE with their respective sgRNAs into stable pathogenic SNV cell lines achieved precise correction with efficiencies of 23.1% (*ACVRL1*), 30.5% (*LAMA2*), 26.3% (*GAA*), and 34.7% (*NOTCH3*) (Fig. 5b). The observed editing efficiencies and specificity, coupled with their flexible PAM requirements, underscore the promise of these tools for gene therapy.

Efficient generation of genetic disease model via evoCas12f

To explore the potential of evoCas12f *in vivo*, we targeted the tyrosinase (*Tyr*) gene to generate a mouse model for albinism. We designed 8 sgRNAs in exon 1 of mouse *Tyr* gene and tested them in NIH/3T3 cells (Fig. 5c and Supplementary Data 1). evoCas12f exhibited dramatically higher

cleavage activity than WT-Un1Cas12f1 across all targets with diverse PAMs, particularly at Tyr-sg1, sg3, sg4, and sg6, which were minimally edited by WT-Un1Cas12f1. Among these, Tyr-sg7 demonstrated the highest cleavage efficiency (Fig. 5d) and was selected for embryo microinjection. A mixture of in vitro-transcribed sgRNA and evoCas12f mRNA was microinjected into one-cell-stage embryos of C57BL/6J mice. Of the 26 F0 mice generated, 17 carried modified *Tyr* alleles, with editing efficiencies ranging from 6.3% to 99.9% (Fig. 5e). Three albino mice displayed near-homozygous phenotypes (Fig. 5f). Genotyping confirmed that 5 of them contained over 90% mutation rates, and 10 of them showed about 50% editing efficiency (Supplementary Table 4). Notably, all 17 mice with modified *Tyr* alleles exhibited ≤4 types of indels, suggesting that evoCas12f achieves highly efficient cleavage and completes editing by the 2-cell stage. These results highlight the utility of evoCas12f for rapid and precise generation of animal models.

Specificity analysis and optimization of evoCas12f

To evaluate the specificity of evoCas12f, we firstly performed sgRNA mismatch assays using sgRNAs with single or dual-base mismatches across the protospacer region. The mismatch tolerance pattern of evoCas12f closely resembled those of WT-Un1Cas12f1 and CasMINI. evoCas12f activity was notably compromised when single mismatches occurred at the positions 1-5 or 18-20 within the protospacer, or when dual mismatches were present at positions 19-20. However, the activities of all Cas12f variants were fully abolished when a mismatch occurred at positions 6-17 within the protospacer (Fig. 6a).

Next, we conducted GUIDE-seq[51] to further evaluate the off-target events of evoCas12f at two endogenous sites (Fig. 6b). We quantified off-target editing sites where reads exceeded 1% of the total. evoCas12f demonstrated higher total read counts than WT-Un1Cas12f1, indicating enhanced enzymatic activity. At the TTTG-12 site, both WT-Un1Cas12f1 and evoCas12f exhibited only background-level off-target editing, confirming evoCas12f's preserved fidelity. However, at the ATTA-4 site, evoCas12f showed elevated off-target activity, evidenced by a higher percentage of off-target reads. Further sequence analysis revealed that mismatches predominantly occurred at positions 1-5 and 18-20 downstream of the PAM, consistent with our mismatch assay results.

Additionally, PAM analysis at off-target sites revealed a T > C preference at the second base, aligning with our endogenous targets. These results underscore the importance of selecting targets with unique sequences or ≥3 mismatches within the seed region to minimize off-target effects in precision applications.

While evoCas12f broadens the target range and improves editing efficiency, its elevated off-target activity remains a concern. Prior studies reported that disrupting non-specific contacts between target DNA and positively charged or polar residues in Cas proteins enhances specificity[52,53]. Inspired by this and based on the Un1Cas12f1 crystal structure, we generated 26 evoCas12f variants carrying single mutation and tested them at known off-target sites. 6 of the variants showed improved specificity and comparable activity over evoCas12f at the CLIC4 (TTTA-21) target site (Supplementary Fig. 13). To further improve specificity, we introduced two or three of these six mutations into evoCas12f and evaluated the on-target/off-target efficiency. Most of the combined mutants retained on-target activity while substantially reduced off-target effects (Fig. 6c). Compared to evoCas12f, introducing K179A/Y232A mutation in evoCas12f (evoCas12f-2A) reduced off-target editing and increased on-target activity, while additional R258A substitution (evoCas12f-3A) showed similar effects (Fig. 6c), and it was further confirmed through testing of three endogenous target sites (Fig. 6d). These data suggest that introduction of additional mutations could substantially reduce potential off-target effects for more precise editing.

Discussion

In this study, we engineered the evoCas12f variant by introducing five amino acid substitutions (N133S/D143S/E151V/L152S/Q244R), which conferred distinct PAM specificities enabling recognition of expanded PAM motifs (5'-NTNR-3' and 5'-NYTR-3'). Compared to the original TTTR PAM recognition, this modification increased the targeting scope by over 13-fold (Fig. 2e), reducing the median distance between two neighbouring PAM sites to 2 bp in the human genome. Notably, evoCas12f not only broadened the targeting range but also enhanced editing efficiency, demonstrating 1.4-fold higher activity than wild-type at TTTR PAM sites. Furthermore, evoCas12f

can be readily adapted for efficient base editing and programmable transcriptional activation. Importantly, this engineered nuclease maintained robust editing activity in mouse embryos even at non-canonical PAM target sites, highlighting its versatility for diverse genome engineering applications.

Since the original editing efficiency of miniature Cas12f is low, previous studies focused on increasing the activity mainly through arginine substitutions to strengthen DNA-protein interactions[24,27,54]. While in this study, we conducted two rounds of targeted random mutation library-based screening in *E.coli*, and identified 71 Un1Cas12f1 mutations capable of cleaving non-TTTR PAM targets. Although we did not expect activity enhancement, evoCas12f achieved 1.4-fold higher efficiency at native TTTR PAMs and up to 91% editing efficiency in mammalian cells. Structural analysis revealed novel non-arginine mutations such as D143S, E151V, and N133S, which enhance activity through distinct mechanisms: D143S replaces a repulsive acidic residue (D143) with serine downstream of the PAM motif, likely alleviating electrostatic repulsion between the negatively charged residue and the phosphate backbone of dsDNA. E151V may function similarly as D143, which resolve the repulsion force between the acidic E151 and the phosphate group of the second nucleotide in PAM. N133S resides in a space close to the region where the dsDNA is broken which may facilitate strand separation or pairing with the sgRNA; L152S, although not directly interacting with the dsDNA, may increase the thermal stability of the Cas12f protein by harmonize better with the surrounding polar residues such as Q108. This dual enhancement underscores the superiority of unbiased mutagenesis over targeted substitution, enabling synergistic improvements in flexibility and activity. It could be the reason why these combined mutations not only expanded the targeting scope but also enhanced the activity.

Miniature CRISPR-derived editors, including base editors and epigenetic regulators, are important tools for in vivo therapies[27,30,31,43]. dCasMINI-ABE, TaRGET-ABE-C3.1, and eCas12f-ABE, which directly fuse 2*TadA to the N- or C-terminus of Cas12f, enable stable A-to-G conversion but exhibit a broad editing window, which may lead to bystander editing in practical applications (Fig. 4d). miniABE inserts TadA into Cas12f, partially narrows the editing window

but suffers from substantially reduced efficiency, likely due to disrupted Cas12f dimer binding to the target site. In our preliminary experiments, fusing a single TadA8e monomer to the N- or C-terminus of Cas12f failed to achieve efficient A-to-G conversion (Supplementary Fig. 9). We hypothesize that such fusion architectures may perturb the Cas12f dimer:gRNA:DNA ternary complex and hinder target base accessibility to the catalytic pocket of the TadA8e domain. To address this, we used the MS2-MCP tethering strategy to develop evoCas12f-derived gene editing tools with preserved complex integrity. The resulting editors combine broad targeting scope and superior activity. Notably, the evoCas12f-ABE achieves relative high editing efficiency with an narrow window (A_3 – A_4), outperforming all reported Cas12f-based ABEs and positioning it as a valuable tool for precision therapy. Our evoCas12f-CBE exhibits activity across two primary editing windows (C_3–C_4 and C_{16}–C_{18}), suggesting distinct deaminase accessibility patterns at target sites. Further studies to reveal underlying mechanisms might help to engineer and refine the editing window toward precision optimization. Critically, the MS2-MCP recruitment strategy was also successful in developing potent transcriptional activators (PHM-evoCas12f), achieving up to 4734-fold mRNA upregulation. We believe that the evoCas12f platform, with its high targeting density (1 site per 2 bp), will help to raise the likelihood of identifying potent sites and enabling multiplexed gRNA–mediated activation[55–57].

Our study addresses key limitations of compact CRISPR systems through protein engineering to simultaneously expand both targeting scope and editing efficiency. We further adapted the evoCas12f variant for base editing with a narrow editing window - a critical feature for AAV-based in vivo gene therapy applications. While LNP-mRNA delivered genome editing offers advantages including low immunogenicity, capacity for repetitive dosing, and reduced potential off-target effects, the AAV-based miniature genome editing system provides unique benefits: the ability to target diverse tissues and achieve long-term effects (particularly valuable for transcriptional activation applications). Our evoCas12f-based system serves as a versatile platform for genome editing, indicating particular promise for genome-wide screening and in vivo gene therapy applications.

Methods

Ethics statement

Our research complies with all relevant ethical regulations. All animal experiments were approved by the Animal Experiment Ethics Committee of East China Normal University (approval number: m20240305).

Plasmid Construction

Human codon-optimized Un1Cas12f1, ge4.1, and P65-HSF1 were synthesized by GENEWIZ. Point mutations in Un1Cas12f1 were introduced via PCR and cloned into the pCMV vector using ClonExpress MultiS One Step Cloning Kit (Vazyme). sgRNA oligos were subjected to a 5-minute denaturation at 95°C, followed by gradual cooling to room temperature to form double strands, which were then ligated into the U6-ge4.1-Ef1α-EGFP vector treated with BbSI restriction enzyme using T4 ligase. All sequences are listed in Supplementary Data 2. The Un1Cas12f1 mutation library oligo was synthesized by GenScript, and the library fragments were amplified using the NEBNext® Q5 Hot Start HiFi PCR kit and purified with the QIAquick gel extraction kit. The purified fragments were assembled with Gibson Assembly Master Mix (NEB) into the PET-28a plasmid backbone treated with restriction enzymes using. PAM-target sequence oligos were annealed using the same method for sgRNA and ligated to the p11-LacY-wtx1 vector treated with XbaI and HindIII restriction enzymes using T4 ligase. Lentiviral vector plasmids for generating stable HEK293T cell lines were constructed using the lentiCRISPR v2 (Addgene #52961), by inserting an ~80 bp gene fragment containing pathogenic SNVs from the ClinVar database and an Ef1a promoter-driven DsRed expression cassette.

sgRNA design

All sgRNAs targeting endogenous sites were listed in Supplementary Data 1 along with their PAM sequences. All endogenous sites with TTTR PAMs and a portion of endogenous sites with non-canonical PAM were previously validated[24,25,30,35], while the remaining targets including targets

for *Tyr* in animal studies were randomly designed. For enodogenous gene activation, sgRNAs are designed within the 250 base pairs relative to transcription start site.

Bacterial Positive Selection Assay

E.coli strain BW25141 (λDE3) was purchased from Bioneer. The screening plasmid containing an arabinose inducible *ccdb* gene was first electroporated into the *E.coli* and maintained under Ampicillin selection. The Un1Cas12f1 mutant library (including sgRNA) were electroporated into *E. coli* for screening. Cells were then incubated in pre-warmed 2×YT medium at 37°C for 1 hour to allow for full recovery. The entire cell suspension was added to a conical flask containing 100 mL of liquid 2×YT medium (supplemented with 50 μg/mL kanamycin and 0.5 mM IPTG) and incubated at 37°C with shaking at 250 rpm for 6 hours. Four milliliters of the culture were spread evenly onto four 25x25 cm square petri dishes (containing 2×YT solid medium, 50 μg/mL kanamycin, and 10 mM arabinose). Plates were incubated overnight at 37°C. The following day, colonies were collected, and plasmids were extracted using a HiPure BAC EF Maxi Kit (Megan). The extracted plasmids served as templates to amplify the library fragments, which were subsequently sequenced via next-generation sequencing (NGS) to identify mutant variants.

Cell Culture, Cell Transfection, and Genomic DNA Preparation

HEK293T (ATCC CRL-3216) and NIH/3T3 **(**ATCC CRL-1658) cell lines were purchased from ATCC. The GFP-expressing cell line was a gift from Professor Yongming Wang (Fudan University). All cells were cultured at 37°C in a 5% CO_2 atmosphere. Complete growth medium consisted of DMEM supplemented with 1% penicillin-streptomycin and 10% fetal bovine serum. For transfection, HEK293T cells were seeded at a density of 200,000 cells per well in 24-well plates with 500 μL of complete medium and incubated overnight. When cells reached ~80% confluency, 750 ng of pCMV-Un1Cas12f1 and 250 ng of U6-ge4.1-sgRNA-Ef1a-GFP were co-transfected using PEI the manufacturer's instructions (Polysciences). Three days post-transfection, GFP-positive cells were sorted using a FACSAria III (BD Biosciences) and collected (Supplementary Fig. 14). Genomic DNA was extract by adding 20 μL of Quick Extract™ DNA

Extraction Solution, followed by vortexing for 20 seconds, incubation at 65°C for 6 minutes, a second vortexing for 20 seconds, and final incubation at 98°C for 2 minutes.

RNA Preparation, Reverse Transcription, and qPCR

GFP-positive cells were sorted using a FACSAria III (BD Biosciences) and collected (Supplementary Fig. 14). Total RNA was extracted following the instructions of the RaPure Total RNA Micro Kit (Megan). cDNA was prepared using the Hifair® II 1st Strand cDNA Synthesis SuperMix, and real-time quantitative PCR was performed using the Hieff® qPCR SYBR Green Master Mix.

Next-generation sequencing and Data Analysis

NGS primers were designed to ensure PCR products of 140-200 bp in size. The primers incorporated 5′ overhang sequences (forward primer: 5′–ggagtgagtacggtgtgc–3′; reverse primer: 5′–gagttggatgctggatgg–3′). The above PCR products were subjected to a second round of PCR using primers with different barcode sequences. PCR products were mixed and sequenced on Illumina HiSeq platform. All targets information used in this study are detailed in the Supplementary Data 1.

For indel analysis, reads displaying mismatches on both ends of the reference sequence with a frequency under 1% were excluded. Allele quantification was performed by enumerating indel-containing reads which were normalized against the total read count using a custom Cas-analyzer-based script[58]. For base conversion analysis (A-to-G/C-to-T editing), the frequency of G/T substitutions at targeted A/C bases was calculated as a fraction of total aligned reads, using a custom BE-Analyzer-derived script[59].

Lentiviral Packaging and Stable Cell Line Construction

HEK293T cells were seeded in 6-well plates and transfected at ~80%. A total of 4.5 μg plasmid DNA per well was transfected, comprising 1.5 μg each of the lentiviral vector plasmid, psPAX2, and pMD2.G, using PEI for transfection. After 8 hours, the medium was replaced with fresh

complete medium. Viral supernatant was collected after 48 hours post transfection and filtered through a 0.22 μm filter. For stable cell line establishing, HEK293T cells in 6-well plates were transduced with 100 μL of viral supernatant. After 24 hours, the medium was replaced with fresh complete medium. Once the cells were confluent, they were passaged into 10 cm dishes and cultured with complete medium supplemented with 2 ug/mL puromycin. Cell viability was monitored under selection pressure, and the puromycin-containing medium was replaced if excessive cell death occurred a stable cell line was established after two passages.

Calculation of Probability Mass Function for Adjacent PAM Distances

The calculation method for adjacent PAM distances referred to published articles[35,39]. PAM sites for each Cas effector on both strands human genome (GRCh38) were scanned and sorted. Distances between adjacent PAMs were computed as: PAM_position_(i+1) – PAM_position_i, in base pairs. Probability Mass Function were constructed as: number of adjacent PAM pairs with distance X / total number of valid adjacent distances of this PAM.

GUIDE-seq assay

GUIDE-seq was carried out according to the established protocols[51]. Briefly, 1 × 10^6 HEK293T cells were electroporated with target plasmids and double-stranded oligodeoxynucleotide (dsODN), followed by 72 hours of culture before genomic DNA extraction. The isolated DNA was fragmented, subjected to end-repair and A-tailing, ligated to adapters, and amplified using dsODN-specific primers. Sequencing was performed on an MGI-seq 2000 platform (BGI) in paired-end mode. For data processing, the open-source GUIDE-seq software was employed with proper PAMs. Reads were analyzed using the GUIDE-seq computational pipeline (https://github.com/aryeelab/guideseq), followed by downstream analysis.

Animal Studies

6–10-week-old female SPF-grade C57BL/6 strain and ICR strain mice purchased from Laboratory Animal Center in Shanghai were housed at the Experimental Animal Center of East China Normal

University. All mice were housed in a specific pathogen-free facility on 12-h light/dark cycles with a temperature between 20°C and 22°C and humidity of 40-60%. All mice had ablibitum access to food and water in standard cages. All animal experiments were approved by the Animal Experiment Ethics Committee of East China Normal University (approval number: m20240305). evoCas12f mRNA was in vitro transcribed using the MEGAscript™ T7 AM1334 kit, and sgRNA was synthesized by GenScript. The sgRNA and mRNA were mixed at a final concentration of 100 ng/ul and centrifuged at 14800 g for 60 minutes at 4°C. The samples were injected into the cytoplasm of one-cell-stage embryos and then transferred into the uterus of surrogate mothers.

Statistical analysis and reproducibility

Data analysis was performed using GraphPad Prism software (version 9.3). Results are expressed as mean ± standard deviation. Statistical evaluations included at least three biologically independent experiments or samples. Comparisons between groups were analyzed using unpaired two-tailed Student's t-tests, with a P value < 0.05 considered statistically significant. No data were excluded from the analyses.

Data Availability

The NGS data generated in this study have been deposited in the NCBI Sequence Read Archive (SRA) database under the accession codes: PRJNA1310035 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1310035], PRJNA1310152 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1310152], PRJNA1310189 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1310189], PRJNA1310199 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1310199], PRJNA1310212 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1310212], PRJNA1310251 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1310251], PRJNA1372820 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1372820], PRJNA1372987 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1372987].ClinVar database (ClinVar (nih.gov)) was used to identify pathogenic SNVs that can be correct by base editing. The published

structure of Un1Cas12f1 (PDB ID: 7C7L) can be accessed at the RCSB Protein Data Bank (PDB) 7C7L: Cryo-EM structure of the Cas12f1-sgRNA-target DNA complex. There are no restrictions on data availability. Source data are provided with this paper.

References

1. Mali, P., Esvelt, K. M. & Church, G. M. Cas9 as a versatile tool for engineering biology. *Nat Methods* **10**, 957–963 (2013).

2. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).

3. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).

4. Shmakov, S. *et al.* Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Mol Cell* **60**, 385–397 (2015).

5. Makarova, K. S. *et al.* An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722–736 (2015).

6. Shmakov, S. *et al.* Diversity and evolution of class 2 CRISPR–Cas systems. *Nat Rev Microbiol* **15**, 169–182 (2017).

7. Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).

8. Zetsche, B. *et al.* Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell* **163**, 759–771 (2015).

9. Kleinstiver, B. P. *et al.* Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481–485 (2015).

10. Walton, R. T., Christie, K. A., Whittaker, M. N. & Kleinstiver, B. P. Unconstrained genome targeting with near-PAMless engineered CRISPR-Cas9 variants. *Science* **368**, 290–296 (2020).

11. Silverstein, R. A. *et al.* Custom CRISPR–Cas9 PAM variants via scalable engineering and machine learning. *Nature* https://doi.org/10.1038/s41586-025-09021-y (2025) doi:10.1038/s41586-025-09021-y.

12. Yan, W. X. *et al.* Functionally diverse type V CRISPR-Cas systems. *Science* **363**, 88–91 (2019).

13. Tang, N. & Ji, Q. Miniature CRISPR-Cas12 Systems: Mechanisms, Engineering, and Genome Editing Applications. *ACS Chem Biol* **19**, 1399–1408 (2024).

14. Wu, W. Y. *et al.* The miniature CRISPR-Cas12m effector binds DNA to block transcription. *Molecular Cell* **82**, 4487-4502.e7 (2022).

15. Chen, W. *et al.* Cas12n nucleases, early evolutionary intermediates of type V CRISPR, comprise a distinct family of miniature genome editors. *Molecular Cell* **83**, 2768-2780.e6 (2023).

16. Harrington, L. B. *et al.* Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* **362**, 839–842 (2018).

17. Yin, S. *et al.* Enhanced genome editing to ameliorate a genetic metabolic liver disease through co-delivery of adeno-associated virus receptor. *Sci. China Life Sci.* **65**, 718–730 (2022).

18. Zheng, R. *et al.* Multiplex gene editing reduces oxalate production in primary hyperoxaluria type 1. *Zoological Research* **44**, 1–10 (2023).

19. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat Rev Drug Discov* **18**, 358–378 (2019).

20. Wang, J.-H., Gessler, D. J., Zhan, W., Gallagher, T. L. & Gao, G. Adeno-associated virus as a delivery vector for gene therapy of human diseases. *Signal Transduct Target Ther* **9**, 78 (2024).

21. Chew, W. L. *et al.* A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat Methods* **13**, 868–874 (2016).

22. Karvelis, T. *et al.* PAM recognition by miniature CRISPR–Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Research* **48**, 5016–5023 (2020).

23. Xin, C. *et al.* Comprehensive assessment of miniature CRISPR-Cas12f nucleases for gene disruption. *Nat Commun* **13**, 5623 (2022).

24. Xu, X. *et al.* Engineered miniature CRISPR-Cas system for mammalian genome regulation and editing. *Molecular Cell* **81**, 4333-4345.e4 (2021).

25. Kim, D. Y. *et al.* Efficient CRISPR editing with a hypercompact Cas12f1 and engineered guide RNAs delivered by adeno-associated virus. *Nat Biotechnol* **40**, 94–102 (2022).

26. Su, M. *et al.* Molecular basis and engineering of miniature Cas12f with C-rich PAM specificity. *Nat Chem Biol* **20**, 180–189 (2024).

27. Kong, X. *et al.* Engineered CRISPR-OsCas12f1 and RhCas12f1 with robust activities and expanded target range for genome editing. *Nat Commun* **14**, 2046 (2023).

28. Wu, Z. *et al.* Programmed genome editing by a miniature CRISPR-Cas12f nuclease. *Nat Chem Biol* **17**, 1132–1138 (2021).

29. Wang, Y. *et al.* Guide RNA engineering enables efficient CRISPR editing with a miniature Syntrophomonas palmitatica Cas12f1 nuclease. *Cell Reports* **40**, 111418 (2022).

30. Kim, D. Y. *et al.* Hypercompact adenine base editors based on a Cas12f variant guided by engineered RNA. *Nat Chem Biol* **18**, 1005–1013 (2022).

31. Zhang, S. *et al.* TadA reprogramming to generate potent miniature base editors with high precision. *Nat Commun* **14**, 413 (2023).

32. Wang, X. *et al.* Robust miniature Cas-based transcriptional modulation by engineering Un1Cas12f1 and tethering Sso7d. *Mol Ther* S1525-0016(24)00082–0 (2024) doi:10.1016/j.ymthe.2024.02.013.

33. Park, S.-J. *et al.* Robust genome editing activity and the applications of enhanced miniature CRISPR-

Cas12f1. *Nat Commun* **16**, 677 (2025).

34. Kleinstiver, B. P. *et al.* Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* **33**, 1293–1298 (2015).

35. Gao, L. *et al.* Engineered Cpf1 variants with altered PAM specificities. *Nat Biotechnol* **35**, 789–792 (2017).

36. Hu, J. H. *et al.* Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* **556**, 57–63 (2018).

37. Miller, S. M. *et al.* Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nat Biotechnol* **38**, 471–481 (2020).

38. Anders, C., Bargsten, K. & Jinek, M. Structural Plasticity of PAM Recognition by Engineered Variants of the RNA-Guided Endonuclease Cas9. *Mol Cell* **61**, 895–902 (2016).

39. Nishimasu, H. *et al.* Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* **361**, 1259–1262 (2018).

40. Hirano, H. *et al.* Structure and Engineering of Francisella novicida Cas9. *Cell* **164**, 950–961 (2016).

41. Doyon, J. B., Pattanayak, V., Meyer, C. B. & Liu, D. R. Directed Evolution and Substrate Specificity Profile of Homing Endonuclease I-SceI. *J. Am. Chem. Soc.* **128**, 2477–2484 (2006).

42. Zhang, Y., Wei, J., Wang, H. & Wang, Y. Characterization of NiCas12b for In Vivo Genome Editing. *Advanced Science* 2400469 (2024) doi:10.1002/advs.202400469.

43. Hino, T. *et al.* An AsCas12f-based compact genome-editing tool derived by deep mutational scanning and structural analysis. *Cell* S0092867423009637 (2023) doi:10.1016/j.cell.2023.08.031.

44. 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).

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. Chen, L. *et al.* Engineering a precise adenine base editor with minimal bystander editing. *Nat Chem Biol* **19**, 101–110 (2023).

47. Doman, J. L., Raguram, A., Newby, G. A. & Liu, D. R. Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors. *Nat Biotechnol* **38**, 620–628 (2020).

48. Yang, L. *et al.* Engineering APOBEC3A deaminase for highly accurate and efficient base editing. *Nat Chem Biol* **20**, 1176–1187 (2024).

49. Liao, H.-K. *et al.* In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation. *Cell* **171**, 1495-1507.e15 (2017).

50. Matharu, N. *et al.* CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. *Science* **363**, eaau0629 (2019).

51. 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).

52. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84–88 (2016).

53. Kleinstiver, B. P. *et al.* High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

54. Wu, Z. *et al.* Structure and engineering of miniature Acidibacillus sulfuroxidans Cas12f1. *Nat Catal* **6**, 695–709 (2023).

55. Shaw, W. M. *et al.* Inducible expression of large gRNA arrays for multiplexed CRISPRai applications. *Nat Commun* **13**, 4984 (2022).

56. Nissim, L., Perli, S. D., Fridkin, A., Perez-Pinera, P. & Lu, T. K. Multiplexed and programmable regulation of gene networks with an integrated RNA and CRISPR/Cas toolkit in human cells. *Mol Cell* **54**, 698–710 (2014).

57. Campa, C. C., Weisbach, N. R., Santinha, A. J., Incarnato, D. & Platt, R. J. Multiplexed genome engineering by Cas12a and CRISPR arrays encoded on single transcripts. *Nat Methods* **16**, 887–893 (2019).

58. Park, J., Lim, K., Kim, J.-S. & Bae, S. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* **33**, 286–288 (2017).

59. Hwang, G.-H. *et al.* Web-based design and analysis tools for CRISPR base editing. *BMC Bioinformatics* **19**, 542 (2018).

Acknowledgements

We thank Y. Zhang from the Flow Cytometry Core Facility of School of Life Sciences at ECNU and support from the ECNU Public Platform for innovation (011). This work was partially supported by grants from the National Key R&D Program of China (2023YFE0209200 and 2023YFC3403400 to D.L., 2022YFC3400200 to Y.G.), National Natural Science Foundation of China (32025023, 32230064 and 32311530111 to D.L., U24A20677 to L.W.), Innovation Program of Shanghai Municipal Education Commission (2025GDZKZD03 to D.L.), Shanghai Municipal Commission for Science and Technology (24J22800400 to D.L.).D.L. is a Shanghai Academy of Natural Sciences Exploration Scholar.

Author contributions

Y. H., L. W., L. Y., and D.L., conceived the rational engineering of Un1Cas12f1. Y. H., L. W., and D. C., performed the experiments of Un1Cas12f1 mutation library construction. Y. H., L. W., and J. M., performed the experiments of bacterial positive selection assay. Y. H., J. M., B. Y., performed the experiments of plasmid construction, cell culture, cell transfection, cell sorting, and genomic DNA preparation. Y. H. and J. M. performed the experiments of PCR, reverse transcription, qPCR and prepared the HTS libraries. L. W. , X. W., and Dan Zhang performed computational analysis. Y. H., C. D., and S. Y., performed the experiments of lentiviral vector production and the creation of stable cell lines. Y. H., J. M., Dexin Zhang and M. L., performed the animal experiments. G. S. performed the structural analysis. L. W., Y. H., L. Y., D. L., Y. G., B. D., Y. W., H. L., and Z. Z., designed the experiments and wrote the manuscript with the input from all the authors. D. L., L. Y., and L. W., supervised the study.

Competing interests Statement

D. L., L. W, Y. H., J. M, L. Y, and Dan Zhang have submitted patent applications (application number CN 202510015341.4, under review) based on the results reported in this study. This patent mainly relates to evoCas12f in this paper. The remaining authors declare no competing interests.

Figure legends

Fig. 1 Evolution of Un1Cas12f1 variants with broader targeting range. a, Schematic of the bacteria positive selection system. **b,** Scatter plots of the screening readout, each dot represents a WT or distinct mutant, highlighting enriched variants. The dashed lines indicate 100-fold enrichment. Q244R, L152S and D143S are pointed out with arrows. **c,** Heatmaps illustrating indel frequency ratios of each mutant to WT-Un1Cas12f1 at corresponding genomic loci. **d,** Indel frequencies by WT-Un1Cas12f1, Cas12f1-SSR at 14 genomic loci in HEK293T cells. **e,** Scatter plots of the screening readout, each dot represents a WT or distinct mutant, highlighting enriched variants. The dashed lines indicate 100-fold enrichment. N133S and E151V are pointed out with arrows. **f,** Heatmaps illustrating indel frequency ratios of each mutant to WT-Un1Cas12f1 at corresponding genomic loci. **g,** Indel frequencies by WT-Un1Cas12f1, evoCas12f at 24 genomic loci in HEK293T cells. In heatmaps, indel frequency ratios shown represent the mean of three biologically independent replicates. For **d** and **g**, dots represent individual values, error bars represent the mean ± SD from three independent experiments.

Fig. 2 Characterization of evoCas12f. a, Indel frequencies of WT-Un1Cas12f1 and evoCas12f at 40 endogenous sites with NNTR PAMs in HEK293T cells. **b,** WebLogos of the PAM sequences for WT-Un1Cas12f1 and evoCas12f. **c,** Indel frequencies of WT-Un1Cas12f1 and evoCas12f at 40 NTNR PAM genomic loci in HEK293T cells. **d,** Summary of indel frequencies by WT-Un1Cas12f1 and evoCas12f at NNTR PAM (40 sites), NTNR PAM (40 sites), and NNTR + NTNR PAM sites in HEK293T cells. **e,** Distribution of target sites of Cas12f family members in the human genome. There are 94,117,975, 240,147,795, 480,224,192, 237,466,111, and 1250,643,991 potential target sites of WT-Un1Cas12f1, enAsCas12f1, enOsCas12f1 and evoCas12f in the human genome (total n=2302,600,064), respectively. Plots show the probability mass function of the distance (in base pairs) to the nearest cleavage site. The red lines indicate

median. **f, g,** Indel frequencies summary for evoCas12f and enAsCas12a. For **a**, **c** and **f**, dots represent individual values, error bars represent the mean ± SD from three independent experiments. For **d** and **g**, each dot represents the mean of three independent experiments at the same genomic locus, error bars represent the mean ± SD. For **g**, *P* values were calculated using the two-tailed Student's t-test. ns, not significant.

Fig. 3 Generating Un1Cas12f1-ABE via RNA-Protein Recruitment Modules. a, Schematic view of the miniature ABE based on the MS2-MCP system. **b,** Schematic view of the Un1Cas12f1-ABE constructs. **c,** Comparison of A-to-G editing frequencies at 2 TTTR PAM sites by four different Un1Cas12f1-ABE constructs. **d,** Schematic of the MS2 tag insertion. The tracrRNA is shown in yellow, the crRNA in pink, and MS2 tags in blue, with insertion positions indicated by dashed boxes. **e,** A-to-G editing frequencies with different MS2 tag insertion positions. Dots represent individual values, error bars represent the mean ± SD from three independent experiments.**f,** Comparison of A-to-G editing frequencies at 2 TTTR PAM genomic loci by different numbers and types of MS2 tags. **g,** Introducing mutations into TadA8e reduces bystander editing of Un1Cas12f1-ABE**.** In heatmaps, editing frequencies shown represent the mean of three biologically independent replicates.

Fig 4. Generating evoCas12f-derived tools based on the MS2-MCP strategy. a, A-to-G editing frequencies of Un1Cas12f1-ABE and evoCas12f-ABE at non-canonical PAM targets. **b,** Average A-to-G editing frequencies of Un1Cas12f1-ABE and evoCas12f-ABE at non-canonical PAM targets. **c,** A-to-G conversion efficiencies by Un1Cas12f1-ABE and evoCas12f-ABE at endogenous targets. PAM sequences are indicated in red. **d,** Comparison of A-to-G conversion efficiencies at 6 TTTR PAM targets by evoCas12f-ABE, dCasMINI-ABE, TaRGET-ABE-C3.1, miniABE and eCas12f-ABE. **e,** Summary of A-to-G conversion efficiencies of evoCas12f-ABE

at each adenine position. Values represent average editing frequencies of three independent experiments from 17 endogenous sites. **f,** Editing window of evoCas12f-ABE, dCasMINI-ABE, TaRGET-ABE-C3.1, miniABE and eCas12f-ABE. **g,** Average on-target editing versus average off-target editing for Cas12f related ABEs. The y-axis reflects the average on-target A-to-G editing, the x-axis reflects the mean cumulative off-target A-to-G editing in orthogonal R-loops assay, editing efficiencies shown represent the mean of three biologically independent replicates. **h,i,** Comparison of C-to-T editing frequencies at non-canonical PAM and TTTR PAM targets by evoCas12f-CBE and miniCBE. **j,** Schematic diagram of the mechanism of the GFP reporter system. **k,** GFP activation efficiencies of evoCas12fa, SminiCRa, dCasMINI-VPR and AsCas12f-HKRA-MPH. For **a**, **d** and **k**, dots represent individual values, error bars represent the mean ± SD from three independent experiments. For **b** and **i**, each dot represents the mean of three independent experiments at the same genomic locus, error bars represent the mean ± SD.For **c** and **h**, values and error bars represent mean ± SD. from three independent experiments.

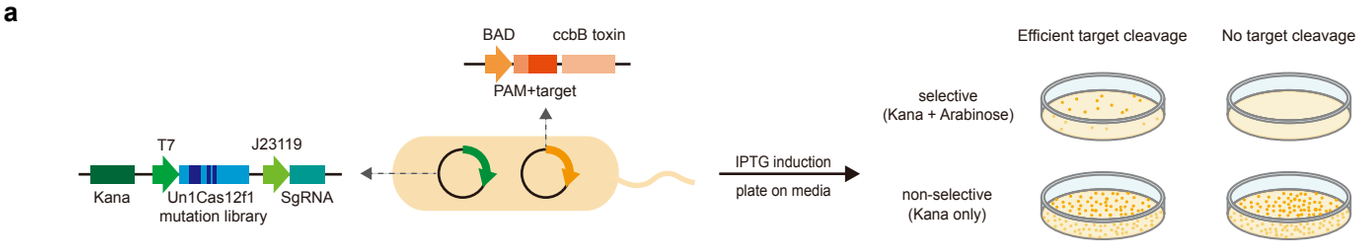
Fig. 5 | Applications of evoCas12f and derived base editors. a, Frequencies of genetic mutations in ClinVar that could, in principle, be corrected by evoCas12f-ABE and evoCas12f-CBE. evoCas12f-ABE targeting A_3-A_4 ,evoCas12f-CBE targeting C_3-C_4 and C_{16}-C_{18}. **b**, Design of sgRNAs targeting pathogenic SNVs and their correction by base editing. sgRNAs were designed to target the pathogenic SNVs *ACVRL1* c.1436G>A, *LAMA2* c.7071G>A , *GAA* c.896T>C and *NOTCH3* c.1363T>C. These pathogenic SNVs were rescued using either evoCas12f-ABE or evoCas12f-CBE. The histogram shows the correction efficiencies, with the targeted pathogenic SNVs highlighted in red. **c**, Illustration of exon 1 and the relative positions of targeting sgRNA. **d**, Indel frequencies of Un1Cas12f1 and evoCas12f at 8 *Tyr* targets in NIH/3T3 cells **e**, In vivo editing frequencies of F0 generation mice, each dot represents one F0 generation mouse. **f**, Phenotype of F0 generation pups: 3-day-old mice on the left and 7-day-old mice on the right. For **b** and **d**, dots represent individual values, error bars represent the mean ± SD from three independent experiments.

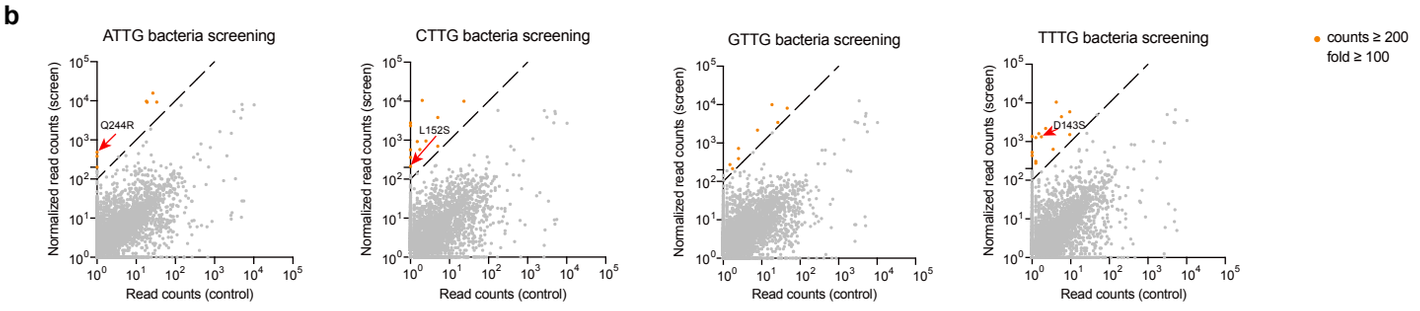
Fig. 6 Specificity analysis and optimization of evoCas12f. a, Sensitivity of Un1Cas12f1, evoCas12f, and CasMINI using sgRNAs with one or two bp mismatches. **b,** Sequence alignments of off-target sites for Un1Cas12f1 and evoCas12f at the TTTG-12 and ATTA-4 targets. Mismatched bases are highlighted in different colors. The number of reads is listed to the right of the corresponding sequence. **c, d,** Indel frequencies of variants with potential non-specific DNA-protein contacts disrupted. For **a**, **c** and **d**, dots represent individual values, error bars represent the mean ± SD from three independent experiments.

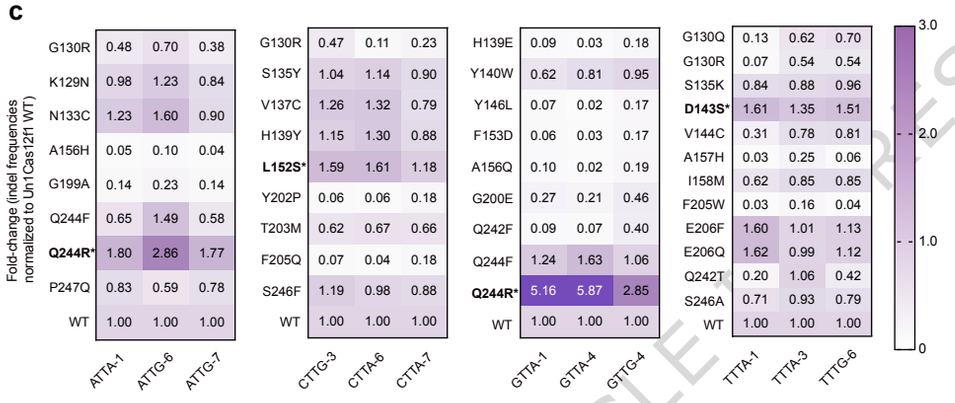
Editor's Summary:

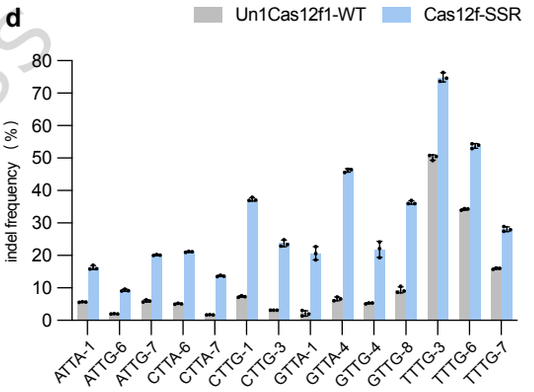
Compact Un1Cas12f1 enzyme suits AAV delivery but is limited by narrow PAMs and modest activity. Here, authors engineer evoCas12f to recognize broader NTNR/NYTR PAMs and boost editing efficiency up to 91%, enabling efficient multiplex editing, base editing, and gene activation in cells and mice.

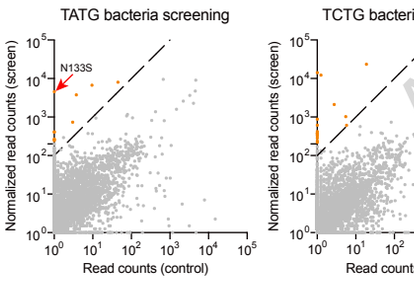
Peer review information: *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

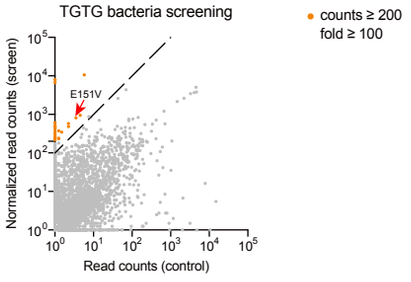


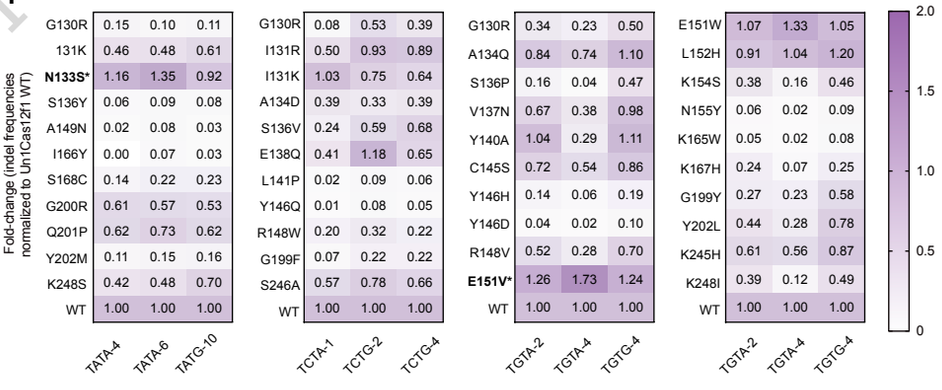


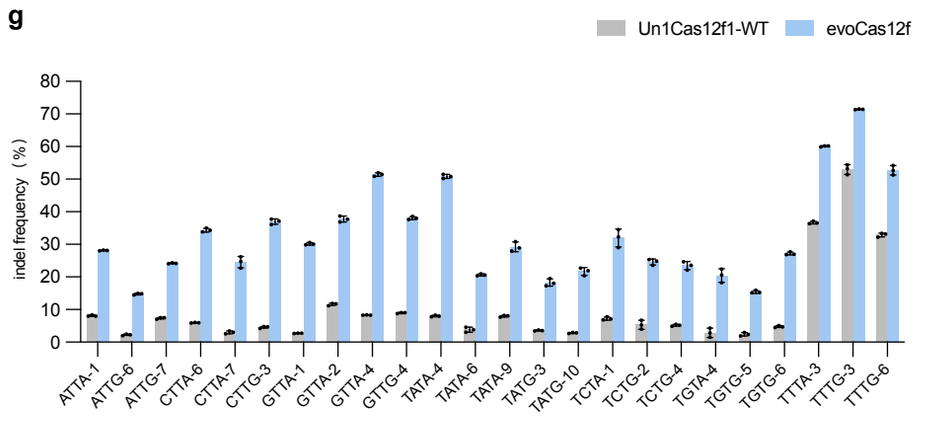


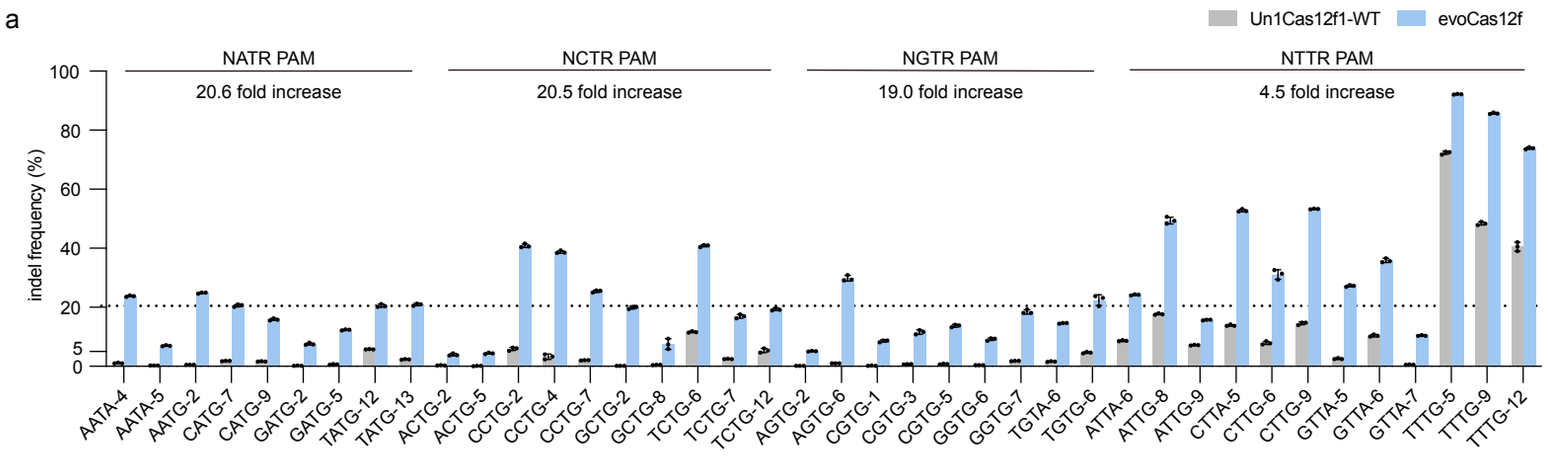


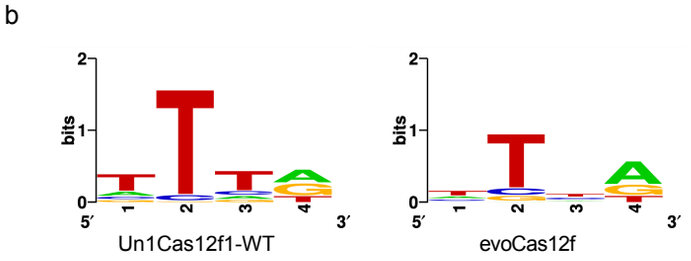


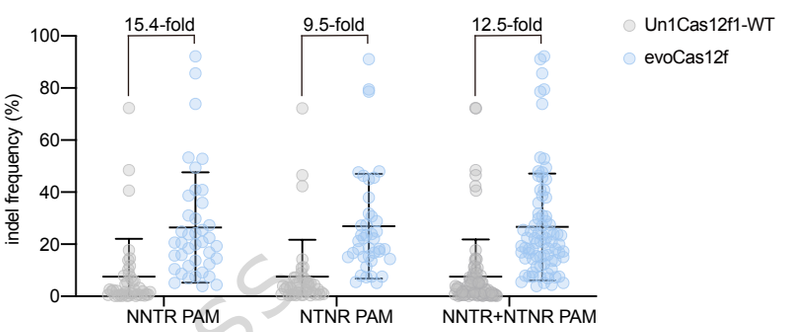


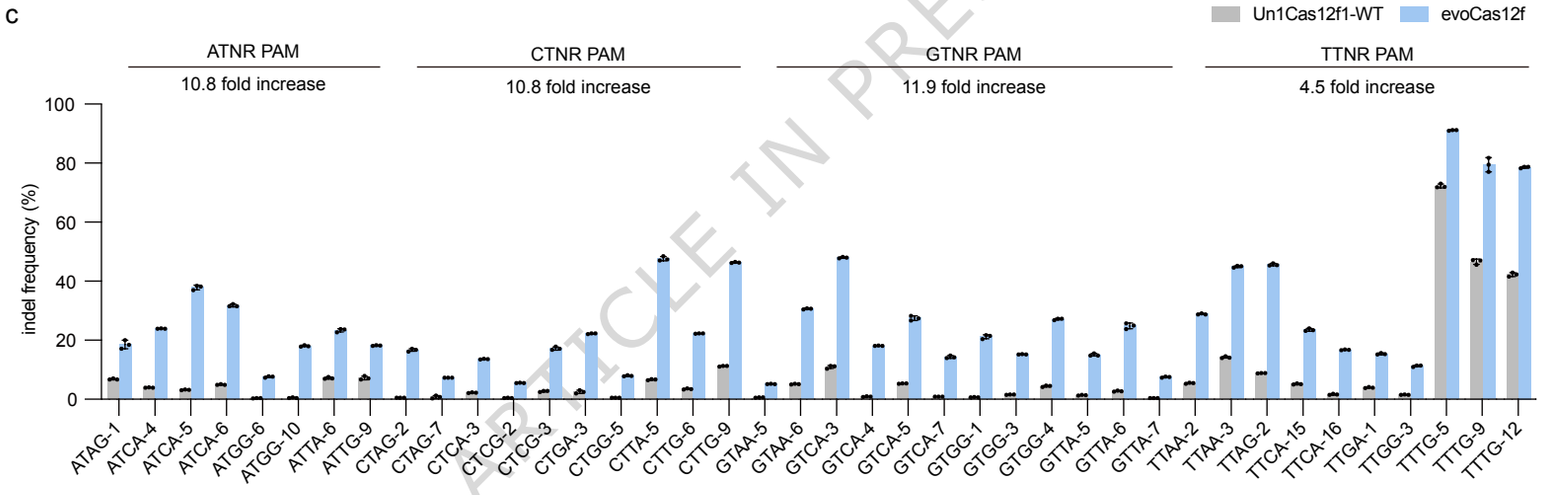


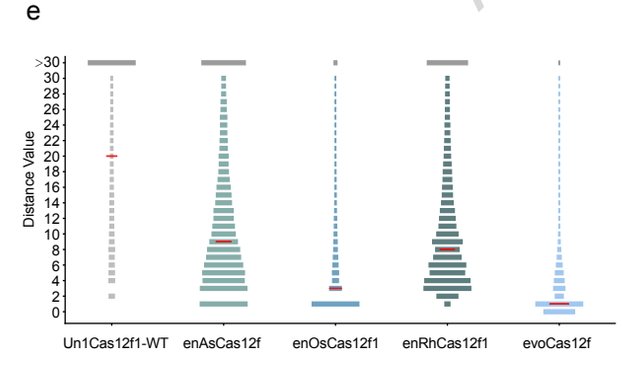


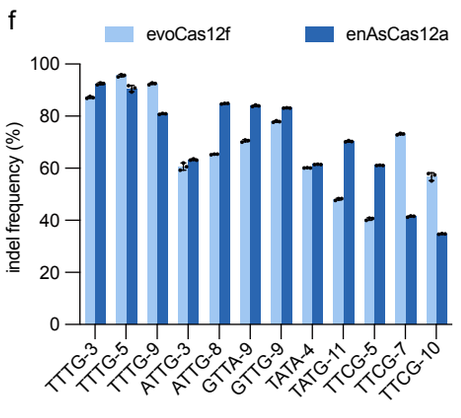


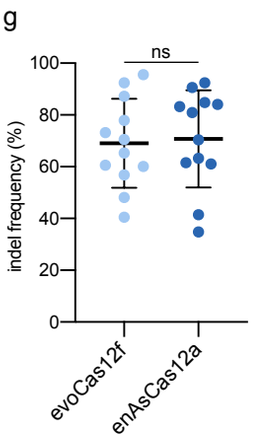


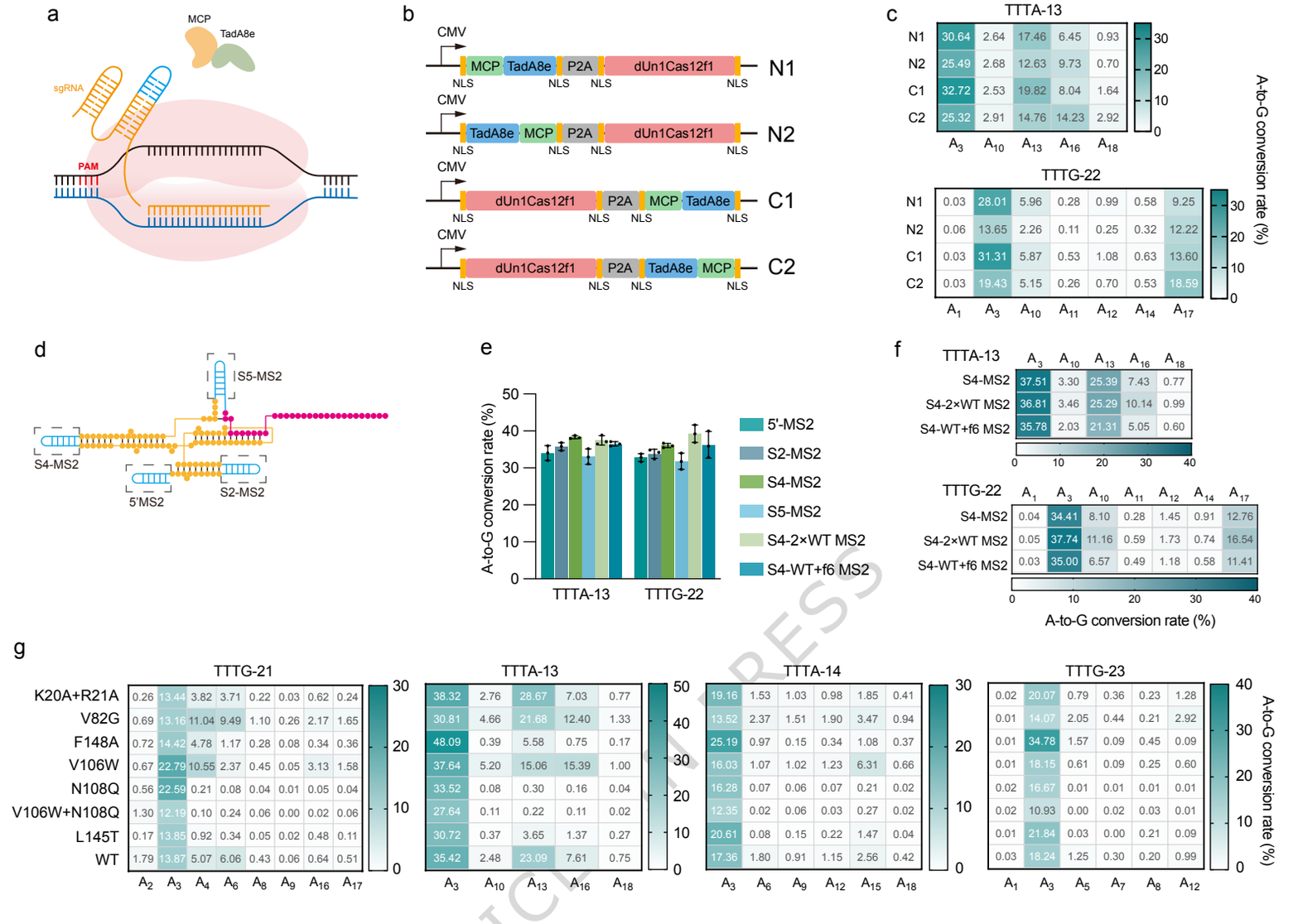


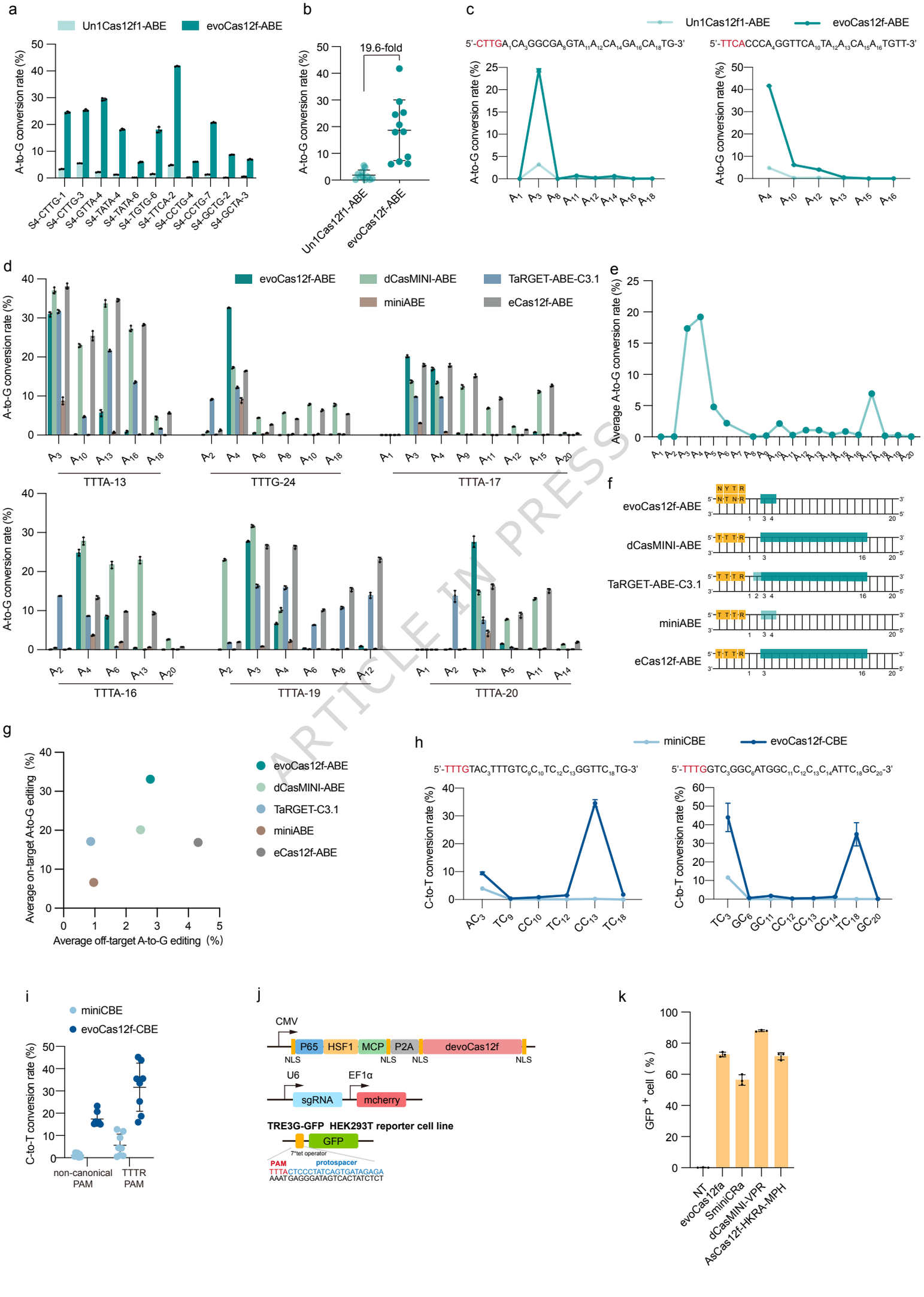











a
Un1Cas12f1-ABE
evoCas12f-ABE
A-to-G conversion rate (%)
S4-CTTG-1
S4-CTTG-3
S4-GTTA-4
S4-TATA-4
S4-TATA-6
S4-TGTG-6
S4-TTCA-2
S4-CCTG-4
S4-CCTG-7
S4-GCTG-2
S4-GCTA-3
b
19.6-fold
A-to-G conversion rate (%)
Un1Cas12f1-ABE
evoCas12f-ABE
c
Un1Cas12f1-ABE
evoCas12f-ABE
5'-CTTGA1CA3GGCGA8GTA11A12CA14GA16CA18TG-3'
5'-TTCACCCA4GGTTCA10TA12A13CA15A16TGTT-3'
A-to-G conversion rate (%)
d
evoCas12f-ABE
dCasMINI-ABE
TaRGET-ABE-C3.1
miniABE
eCas12f-ABE
A-to-G conversion rate (%)
TTTA-13
TTTG-24
TTTA-17
TTTA-16
TTTA-19
TTTA-20
e
Average A-to-G conversion rate (%)
f
evoCas12f-ABE
dCasMINI-ABE
TaRGET-ABE-C3.1
miniABE
eCas12f-ABE
g
Average on-target A-to-G editing (%)
Average off-target A-to-G editing (%)
evoCas12f-ABE
dCasMINI-ABE
TaRGET-C3.1
miniABE
eCas12f-ABE
h
miniCBE
evoCas12f-CBE
5'-TTTGTAC3TTTGTC9C10TC12C13GGTTC18TG-3'
5'-TTTGGTC3GGC6ATGGC11C12C13C14ATTC18GC20-3'
C-to-T conversion rate (%)
i
miniCBE
evoCas12f-CBE
C-to-T conversion rate (%)
non-canonical PAM
TTTR PAM
j
CMV
P65
HSF1
MCP
P2A
devoCas12f
NLS
U6
EF1α
sgRNA
mcherry
TRE3G-GFP HEK293T reporter cell line
GFP
7*tet operator
PAM
protospacer
TTTACTCCCTATCAGTGATAGAGA
AAATGAGGGATAGTCACTATCTCT
k
GFP+ cell (%)
NT
evoCas12fa
SminiCRa
dCasMINI-VPR
AsCas12f-HKRA-MPH