nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-021-00868-6

Check for updates

Programmed genome editing by a miniature CRISPR-Cas12f nuclease

Zhaowei Wu[1], Yifei Zhang[1,2], Haopeng Yu[3], Deng Pan[1,2], Yujue Wang[1,2], Yannan Wang[1,2], Fan Li[1], Chang Liu[1], Hao Nan[4], Weizhong Chen[1] and Quanjiang Ji[1,5]✉

The RNA-guided CRISPR-associated (Cas) nucleases are versatile tools for genome editing in various organisms. The large sizes of the commonly used Cas9 and Cas12a nucleases restrict their flexibility in therapeutic applications that use the cargo-size-limited adeno-associated virus delivery vehicle. More compact systems would thus offer more therapeutic options and functionality for this field. Here, we report a miniature class 2 type V-F CRISPR-Cas genome-editing system from *Acidibacillus sulfuroxidans* (AsCas12f1, 422 amino acids). AsCas12f1 is an RNA-guided endonuclease that recognizes 5′ T-rich protospacer adjacent motifs and creates staggered double-stranded breaks to target DNA. We show that AsCas12f1 functions as an effective genome-editing tool in both bacteria and human cells using various delivery methods, including plasmid, ribonucleoprotein and adeno-associated virus. The small size of AsCas12f1 offers advantages for cellular delivery, and characterizations of AsCas12f1 may facilitate engineering more compact genome-manipulation technologies.

Many archaea and bacteria adopt CRISPR-Cas (clustered regularly interspaced short palindromic repeat and CRISPR-associated proteins) systems for adaptive immunity against foreign genetic elements from invading bacteriophages and plasmids[1]. CRISPR-Cas systems are diverse in the numbers and configurations of Cas effectors involved in interference. They are currently categorized into two distinct classes (class 1 and 2) and six major types (I–VI)[1]. The RNA-guided Cas9 and Cas12a effectors (generally greater than 1,000 amino acids), from type II and V CRISPR systems, respectively, have been harnessed as widely used genome-editing tools[2–9] or as therapeutic medicines for correcting genetic deficiency in clinical use[10–12]. However, the large size of these Cas effectors poses difficulties for their efficient packing in cargo-size-limited delivery vehicles, such as adeno-associated virus (AAV)[13].

Recently, the small CRISPR effectors CasX (Cas12e, 986 amino acids) and CasΦ (Cas12j, 700–800 amino acids) were reported as functional genome editors in human cells[14,15]. The type V-F systems have smaller Cas effector proteins (400–700 amino acids), sharing sequence similarity to transposon-encoded TnpB nuclease[1]. Type V-F effectors were originally considered to be evolutionary intermediates from TnpB to fully fledged type V effectors and to be incapable of double-stranded DNA (dsDNA) targeting[1]. Harrington et al.[16] designated a subset of type V-F effectors into Cas14, and demonstrated the in vitro protospacer adjacent motif (PAM)-independent single-stranded DNA (ssDNA) targeting activity of Cas14a1 (also known as Un1Cas12f1). However, Cas14a1 failed to interfere with plasmid transformation in bacteria, leading to the conclusion that Cas14 may not target dsDNA[16]. Makarova et al.[1] updated the nomenclature and grouped these type V-F effectors into Cas12f1 (Cas14a and type V-U3), Cas12f2 (Cas14b) and Cas12f3 (Cas14c, type V-U2 and V-U4). Karvelis et al.[17] showed that Cas12f effectors were capable of in vitro dsDNA targeting in a PAM-dependent manner[18]. However, biochemical characterizations on cleaved products by tris borate EDTA–urea polyacrylamide gel electrophoresis (TBE-urea PAGE) and run-off sequencing revealed inconsistent patterns in dsDNA cleavage[17]. Moreover, only two Cas12f effectors, *Acidibacillus sulfuroxidans* Cas12f1 (AsCas12f1) and *Syntrophomonas palmitatica* Cas12f1 (SpCas12f1), mediated plasmid interference in *Escherichia coli* among all 10 tested Cas12f effectors[17]. Recently, the complex structure of Cas14a1/single-guide RNA (sgRNA)/dsDNA was determined by cryo-EM, revealing a unique asymmetric protein dimer configuration to bind one sgRNA and target dsDNA, which was fundamentally different from other CRISPR-Cas effectors[19,20]. These studies suggest additional, uncharacterized properties and application potentials of the Cas12f effectors, leading us to explore the possibility of using a type V-F effector protein as a genome-editing tool.

In this Article, we systemically characterize the biochemical properties of a miniature type V-F effector protein, AsCas12f1, selected from a previous collection of compact Cas12f nucleases[17]. We demonstrate that the AsCas12f1–crRNA–tracrRNA complex efficiently cleaves dsDNA targets when a 5′ T-rich PAM is recognized and its downstream vicinity matches with the guide sequence in CRISPR RNA (crRNA), creating asymmetric staggered DNA double-stranded breaks (DSBs). We show that AsCas12f1 serves as an effective genome-editing tool in *E. coli* and *Klebsiella pneumoniae*, as well as various human cells via different delivery methods, including plasmid, ribonucleoprotein (RNP) and AAV. Finally, we assess the editing fidelity of AsCas12f1 by using mismatched guides and circularization for in vitro reporting of cleavage effects by sequencing (CIRCLE-seq), respectively. Taken together, these data lay the groundwork for understanding the dsDNA cleaving mechanism by miniature Cas12f nucleases and establish the application potential of AsCas12f1 as a genome-editing tool.

Results

AsCas12f1 recognizes 5′ T-rich PAMs for dsDNA targeting. We started with AsCas12f1 originating from *A. sulfuroxidans*, with an effector size of 49 kDa (422 amino acids), which is less than half

[1]School of Physical Science and Technology, ShanghaiTech University, Shanghai, China. [2]University of Chinese Academy of Sciences, Beijing, China. [3]Department of Cell and Developmental Biology, John Innes Centre, Norwich Research Park, Norwich, UK. [4]College of Life Sciences, Northwest A&F University, Yangling, China. [5]Guangzhou Laboratory, Guangzhou, China. ✉e-mail: quanjiangji@shanghaitech.edu.cn

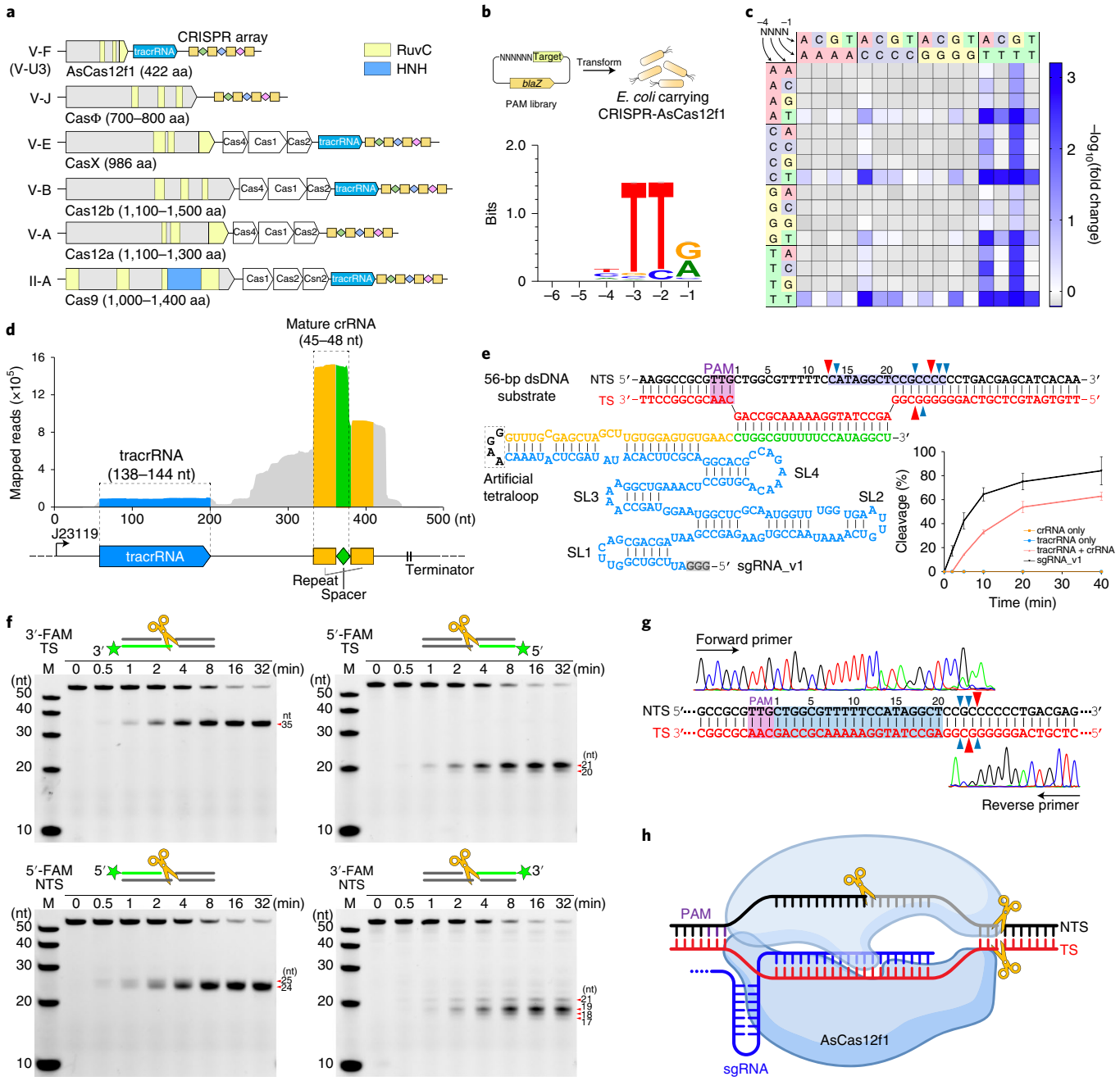


Fig. 1 | AsCas12f1, an RNA-guided nuclease, cleaves dsDNA in a PAM-dependent manner. a, Illustration of the genomic CRISPR-Cas loci of AsCas12f1 and other systems previously used for genome-editing applications. **b**, Graphical representation of the PAM depletion assay and WebLogo view for the PAM of AsCas12f1. **c**, PAM preference for all 256 four-nucleotide PAMs by AsCas12f1. **d**, Small RNA-seq results mapped onto the non-coding region of the CRISPR-AsCas12f1 system in *E. coli*. **e**, Map for the dsDNA cleavage pattern of AsCas12f1. Black and red sequences represent the NTS and TS, respectively. Blue and orange-to-green sequences indicate the tracrRNA and mature crRNA identified by small RNA-seq, respectively. The four nucleotides within the dashed box indicate the artificial tetraloop to connect tracrRNA and crRNA. Red and blue arrowheads indicate the major and minor cleavage sites, respectively. The purple shaded sequence is cut off by a dual-NTS cleavage event. The gray-shaded tri-guanosine is added to the beginning of sgRNA to maximize RNA production. The engineered sgRNA_v1 efficiently guided AsCas12f1 to target dsDNA substrates. Data represent the mean ± s.d. of three biological replicates. **f**, In vitro cleavage assay for mapping the cleavage sites. **g**, Run-off sequencing for the AsCas12f1-cleaved plasmid. **h**, Scheme of the AsCas12f1-mediated dsDNA cleavage.

the size of most Cas9 and Cas12 effectors (Fig. 1a)[2,5,14,15,18]. First, to identify the PAM specificity of the CRISPR-AsCas12f1 system, we transformed a plasmid library containing randomized PAM sequences into the *E. coli* constitutively expressing a minimal CRISPR-AsCas12f1 system (Fig. 1b). Plasmids containing valid PAMs for AsCas12f1 were depleted during the process, and high-throughput sequencing of the remaining plasmids revealed that AsCas12f1 generally recognizes PAMs as 5′-NTTR (where R represents A or G; Fig. 1b). To comprehensively show the profile of PAM preference by AsCas12f1, $-\log_{10}$(fold change) values of all 256 four-nucleotide PAMs in the PAM depletion assay were plotted in Fig. 1c. Targeting with AsCas12f1 was most efficient against

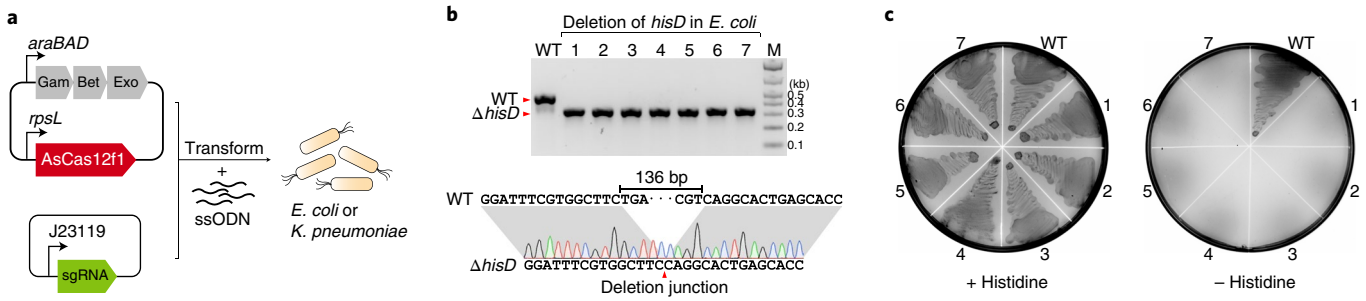


Fig. 2 | AsCas12f1-mediated genome editing in bacteria. a, Scheme of AsCas12f1-mediated genome editing in bacteria. **b**, Deletion of *hisD* in *E. coli* using AsCas12f1. WT, wild type. **c**, Deletion of *hisD* disrupts the de novo biosynthesis of histidine and thereby causes auxotrophy.

5′-NTTR PAMs. By contrast, 5′-NTTY (where Y is C or T) PAMs were less effective, and the −4 nucleotide affected the targeting activity following the order T > C > A > G (Fig. 1c and Supplementary Fig. 1a,b). In addition, 5′-CTCA and 5′-TTCA were also efficient PAMs for AsCas12f1. These findings are slightly different from the 5′-TTN PAMs determined by in vitro DNA cleavage for AsCas12f1 (ref. [17]).

AsCas12f1 is guided by tracrRNA and crRNA. To identify the essential RNA components in the CRISPR-AsCas12f1 system, small RNA-seq was performed using *E. coli* expressing a minimal CRISPR-AsCas12f1 system. As shown in Fig. 1d, an ~45–48-nt mature crRNA and an ~138–144-nt putative trans-activating CRISPR RNA (tracrRNA) were identified. The mature crRNA contains a 28-nt direct repeat and a following 17–20-nt guide sequence. Through RNA secondary structure prediction, a putative tracrRNA was identified with four stem loops, and its 3′ end is mostly complementary to the direct repeat of the mature crRNA, possibly forming a tracrRNA–crRNA duplex (Fig. 1e). Like other tracrRNA-containing CRISPR-Cas systems (all type II systems and some type V systems), a non-Cas RNase III-like ribonuclease most likely is responsible for crRNA maturation in the native host as well as in many heterologous hosts, such as *E. coli* and human cells[2,14,18]. Through analytical size-exclusion chromatography (SEC), we found that AsCas12f1 can assemble with tracrRNA and crRNA to form an RNP complex in vitro (Supplementary Fig. 2a). Moreover, AsCas12f1 was unable to exclusively process its cognate pre-crRNA in vitro (Supplementary Fig. 2b), which is consistent with the finding in Cas14a1 (ref. [16]). For engineering purposes, the tracrRNA and the mature crRNA were fused into a sgRNA (sgRNA_v1) by adding a connecting tetraloop, which enhanced the dsDNA targeting activity by AsCas12f1 (Fig. 1e).

Characterization of AsCas12f1-mediated dsDNA cleavage. To uncover the biochemical properties of AsCas12f1, we first determined the effect of environmental factors on its nuclease activity. Generally, the nuclease activity of AsCas12f1 is magnesium-dependent (the optimal concentration ranges from 5 to 20 mM), NaCl-sensitive (less than 50 mM) and thermophilic (optimal temperature from 45 to 60 °C) (Supplementary Fig. 3a–e). The optimal spacer length for the maximal in vitro dsDNA cleavage activity was 17 bp (Supplementary Fig. 4a,b). Furthermore, by truncating the sgRNA_v1, a shortened version, sgRNA_v2, was obtained with similar in vitro and in vivo performance compared to sgRNA_v1 (Extended Data Fig. 1a–d).

To reveal the dsDNA recognition and cleavage pattern of AsCas12f1 in vitro, four 56-nt dsDNA substrates with identical sequence and a single FAM label at individual ends were used in a time-dependent cleavage assay, and the cleavage products were analyzed by denaturing TBE-urea PAGE. As shown in Fig. 1e,f, the AsCas12f1-derived DNA break site on the target strand (TS) was predominantly located 3 nt downstream from the spacer. For the non-target strand (NTS), AsCas12f1 created two DNA breaks. The left one was mainly located 12 nt downstream from the PAM, while the right one was centered at 5 ± 2 nt downstream from the spacer. Interestingly, the cleavage pattern was instantly formed rather than generating a time-dependent ladder-like pattern, indicating that a 10- to 15-nt displaced NTS fragment was directly cut off from the dsDNA substrate instead of trimming one ssDNA end, nucleotide by nucleotide, via its exonuclease activity. Similar cutoff event on NTS were also observed in Cas12a nucleases with shorter cutoff lengths[21]. In addition, the run-off sequencing of the cleavage product of a plasmid carrying the same target also showed consistent results (Fig. 1g). We found that the break site on the TS and NTS slightly shifted when the guide length was less than 18 nt (Supplementary Fig. 4c,d). A similar shift was also observed in Cas12a nucleases[21]. Taken together, AsCas12f1 asymmetrically cleaves dsDNA by creating an approximately 11-nt-long 5′ TS overhang at the PAM-proximal segment, and a near blunt end for the PAM-distal segment (Fig. 1e,h).

Through sequence alignment with other Cas12 family proteins and structural modeling, four putative active-site residues were identified in the conserved RuvC domain of AsCas12f1 (Extended Data Fig. 2a,b). D225A and E324A completely abolished the nuclease activity, whereas R383A and D401A retained minor nickase activity for NTS, creating a DNA break 18 nt downstream from the PAM (Extended Data Fig. 2c).

In addition to the dsDNA cleaving activity, most Cas12 nucleases possess ssDNA-cleaving capacity[14,15,22–25]. AsCas12f1 also cleaved the ssDNA substrate in a PAM-independent manner, and the DNA break site was identical to that in dsDNA cleavage. AsCas12f1 also trimmed the PAM-distal free ssDNA end from 3′ to 5′ nucleotide by nucleotide, generating a time-dependent ladder-like pattern (Extended Data Fig. 3a–f). Furthermore, AsCas12f1 possessed target-activated collateral ssDNA cleavage activity (Supplementary Fig. 5a–c). Similar activities were also observed in Cas14a1 (ref. [16]).

AsCas12f1-mediated genome editing in bacteria. Inspired by the PAM-dependent plasmid interference activity by AsCas12f1 in *E. coli*, we first sought to explore its programmability for genome editing in bacteria. Given that most bacteria lack non-homologous end joining (NHEJ) machines to repair a DSB in emergence, the majority of bacteria die from the genomic break[26]. We thus introduced lambda-red recombinases to assist homology-directed repair (HDR) and supplemented synthesized single-stranded donor oligonucleotides (ssODNs) as repair templates (Fig. 2a). As shown in Fig. 2b, a 136-bp fragment of *hisD* was deleted in *E. coli*, disrupting the de novo biosynthesis of histidine and creating an auxotroph phenotype (Fig. 2c). Similarly, a 236-bp fragment of *cyaA* was deleted in *E. coli*, which slows down the aerobic growth of *E. coli*

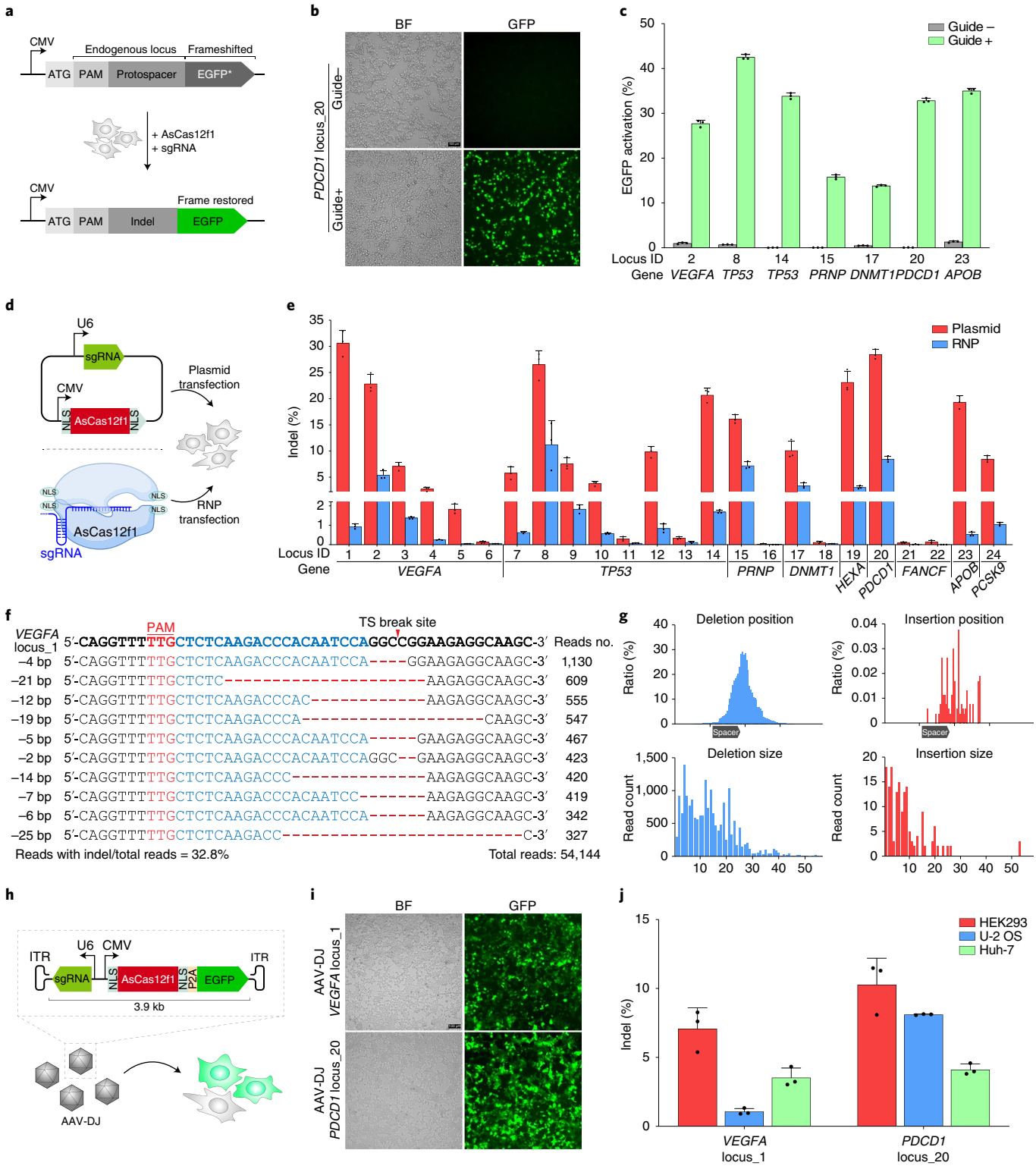


Fig. 3 | AsCas12f1 facilitates genome editing in human cells. a, Scheme of AsCas12f1-mediated EGFP activation in human cells. **b**, Transfection of plasmids carrying AsCas12f1 and sgRNA activated EGFP expression. BF, bright field. **c**, EGFP activation efficiencies on seven target sites determined by flow cytometry. Data represent mean ± s.d. of three biological replicates. **d**, Experimental workflow of AsCas12f1-mediated genome disruption in HEK293 cells using plasmid and RNP for delivery. **e**, Indel efficiencies on 24 endogenous loci in HEK293 cells, determined by NGS. Data represent mean ± s.d. of three biological replicates. **f**, Sequencing of locus_1 in *VEGFA* after AsCas12f1-mediated gene disruption. The top 10 most frequent NHEJ products are displayed. **g**, Size and position information of indels generated by AsCas12f1. **h**, Scheme of AAV-derived AsCas12f1 for human cell editing. **i**, The EGFP reporter indicated that AAV-DJ efficiently delivered AsCas12f1 into HEK293 cells. **j**, AAV-derived indel efficiencies on two endogenous loci in HEK293, U2-OS and Huh-7 cells determined by NGS. Data represent mean ± s.d. of three biological replicates.

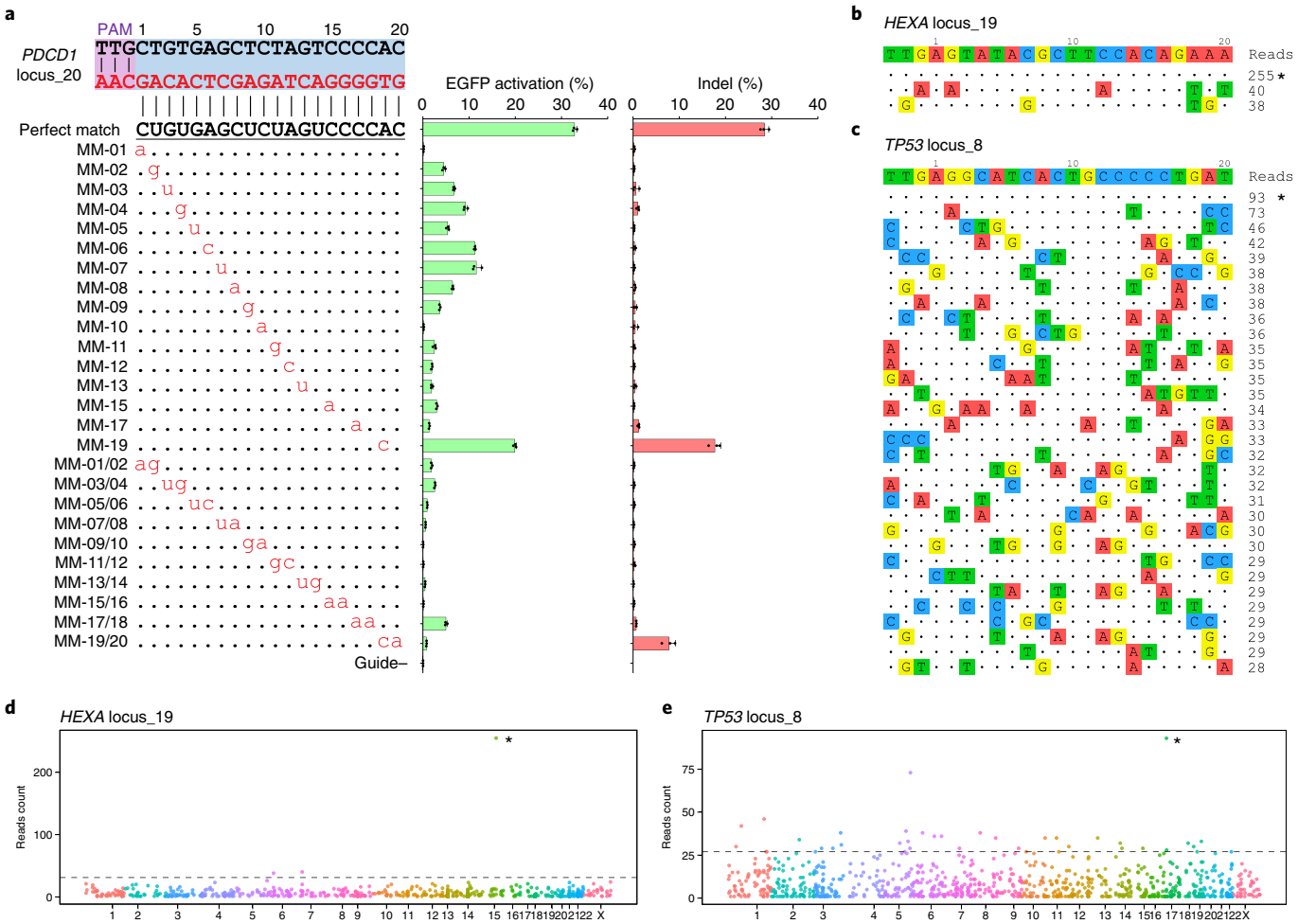


Fig. 4 | Mismatch tolerance and genome-wide editing fidelity of AsCas12f1. a, Impacts of mismatches on AsCas12f1-mediated EGFP activation activity and indel activity. Data represent the mean ± s.d. of three biological replicates. **b**,**c**, Sequence alignments of CIRCLE-seq-detected genome-wide off-target sites for AsCas12f1 on *HEXA* locus_19 (**b**) and *TP53* locus_8 (**c**). The on-target sites are indicated by asterisks, and off-target sites are listed by their CIRCLE-seq read counts. Mismatches to the on-target sequence are indicated by colored nucleotides. **d**,**e**, Manhattan plots for all CIRCLE-seq-detected loci with no more than six mismatches to the on-target site for *HEXA* locus_19 (**d**) and *TP53* locus_8 (**e**). Each detected locus is represented as a single dot and mapped on the complete human genome. The dot height indicates the CIRCLE-seq read count. On-target sites are indicated by asterisks. Dashed lines indicate the CIRCLE-seq threshold.

in glucose-rich medium (Supplementary Fig. 6a,b)[27]. In addition, this system was also effective in *K. pneumoniae*, a notorious human pathogen with increasing concern given its multidrug-resistant nature[28]. Gene disruptions were achieved on both metabolic and drug-resistant genes by deleting fragments ranging from 100 to 299 bp (Supplementary Fig. 7a–e). Moreover, single nucleotide substitution in *galK* was also accomplished, which converts a regular codon into a stop codon and eliminates the PAM site for AsCas12f1 (Supplementary Fig. 7f). These results demonstrate the potential of AsCas12f1 for in vivo genome editing.

AsCas12f1 facilitates genome editing in human cells. Motivated by the efficient genome editing via AsCas12f1 in bacteria, we sought to explore whether AsCas12f1 can be harnessed as a genome-editing tool in human cells. An enhanced green fluorescent protein (EGFP) reporter assay was designed by introducing two plasmids into HEK293 cells. One carried an inactivated EGFP cassette, which was frameshifted by inserting a short endogenous fragment sequence at the beginning of the EGFP coding sequence (CDS), and the other plasmid contained the expression cassettes of the N- and C-terminal nuclear localization sequence (NLS)-fused AsCas12f1, as well as the sgRNA targeting to the endogenous site on the first plasmid (Fig. 3a)[29]. Insertion or deletion (Indel) generated on the target site may restore the frameshifted EGFP cassette. As shown in Fig. 3b,c and Extended Data Fig. 4, EGFP was activated by targeting seven plasmid carrying endogenous loci with efficiencies ranging from 13.8% to 42.5%. Because these target loci originated from the human genome, an endogenous gene disruption assay was performed in HEK293 cells by delivering either a plasmid expressing NLS-AsCas12f1-NLS and its sgRNA, or an RNP (Fig. 3d). As shown in Fig. 3e, plasmid-derived CRISPR-AsCas12f1 induced indels (>1%) in 17 targets among 24 tested loci with a maximal indel efficiency of 32.8%. The plasmid-based gene disruption was also effective in other cell lines, such as HepG2 and Huh-7 (Supplementary Fig. 8a,b). Moreover, the RNP-based delivery system was also effective for gene editing, but had a lower overall editing efficiency than that of the plasmid-based delivery system (Fig. 3e). Next-generation sequencing (NGS) of the target loci revealed that AsCas12f1 introduced more deletions than insertions (Fig. 3f,g). The deletion position is centered at the TS break site observed in vitro (Fig. 3g).

To our knowledge, AsCas12f1 is currently the smallest RNA-guided Cas nuclease functioning in human cell editing. The

extremely small size of AsCas12f1 (422 amino acids) enables packaging of both a hU6-driven sgRNA cassette and a cytomegalovirus (CMV)-driven NLS-AsCas12f1-NLS-P2A-EGFP expression cassette into a single AAV vector within a cargo size of 3.9 kb (3.1 kb without EGFP; Fig. 3h). Using AAV serotype DJ (ref. [30]), we targeted locus_1 of the human vascular endothelial growth factor A gene (*VEGFA*) and locus_20 of the programmed cell death 1 gene (*PDCD1*) in HEK293, U-2 OS and Huh-7 cells. As shown in Fig. 3i,j, AAV-DJ efficiently transduced the CRISPR-AsCas12f1 system into human cells and successfully induced indels in the *VEGFA* and *PDCD1* genes within all three cell lines with a maximum efficiency of 11.5%.

Finally, we evaluated the mismatch tolerance of AsCas12f1-mediated gene editing by tiling 1- or 2-nt mismatches across the guide sequence in sgRNAs (Fig. 4a). Owing to the high sensitivity of the EGFP activation assay, an internal seed sequence was identified near the middle of the target sequence. This finding was similar to the previous observations in the mismatched ssDNA cleavage by Cas14a1 (ref. [16]). However, 1- or 2-nt mismatches substantially reduced the indel activity of AsCas12f1 on the endogenous locus, indicating that AsCas12f1 exhibited sufficient editing fidelity (Fig. 4a). To measure the genome-wide cleavage activity of AsCas12f1 directly, we applied CIRCLE-seq[31], a highly sensitive in vitro method for identifying off-target cleavage events by AsCas12f1. As shown in Fig. 4b–e, CIRCLE-seq analysis identified 2 and 31 off-target sites for locus_19 of the hexosaminidase subunit alpha gene (*HEXA*) and locus_8 of the tumor protein 53 gene (*TP53*) cleaved by AsCas12f1, respectively.

Discussion

In this study, we demonstrate that AsCas12f1 is a programmable RNA-guided dsDNA nuclease and a bona fide genetic manipulation tool functioning in both bacteria and human cells. Intriguingly, Karvelis et al.[17] reported inconsistent dsDNA cleavage patterns on Cas14a1 (Un1Cas12f1) digested products with two biochemical approaches. TBE-urea PAGE analysis indicated 8- to 13-nt-long 5′ staggered ends crossing the protospacer, and run-off sequencing revealed 2- to 4-nt-long 5′ staggered ends downstream of the protospacer. The authors attributed this contradiction to the difference in substrate size[17]. However, cleavage pattern analyses by either TBE-urea PAGE using 5′ ^{32}P-labeled short DNA substrates or run-off sequencing can only reveal the break sites of two single-strand ends among the four single-strand ends generated by a single DSB event. In this study, we utilized four dsDNA substrates with identical sequence but singly labeled with FAM at each DNA end for a time-dependent cleavage assay, revealing a distinct dsDNA cleavage activity whereby AsCas12f1 applies one cut to the TS and two further cuts to the NTS. A 10- to 15-nt-long displaced NTS fragment is directly cut off from the dsDNA substrate, generating cleavage products with asymmetric staggered ends. The structure of this cleavage product may facilitate NHEJ-based gene insertion by introducing a double-stranded donor oligonucleotide (dsODN) carrying sticky ends complementary to the AsCas12f1-mediated DSBs, allowing oriented DNA insertion, particularly in non-dividing cells whose HDR is temporarily dormant.

NGS revealed another feature, that the AsCas12f1-induced indels mainly occurred at the distal end of the protospacer, and a reasonable portion of the indels kept the target site intact by only changing the downstream sequence. Unlike the Cas9-mediated gene editing process, in which most indels generated by NHEJ disrupt the target site, AsCas12f1-derived edits can be iteratively targeted by the same guide RNA until the target site is eliminated by HDR or long-sequence deletion. Such a compact system and its minimal PAM requirement offer particular advantages, with both easier cellular delivery using cargo-size-limited vectors and a wider range of targetable genomic loci, and may serve as a new addition to the CRISPR-Cas toolbox and bring new solutions for some challenges in the field of genome editing.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41589-021-00868-6.

Received: 22 June 2021; Accepted: 23 July 2021;

References

1. Makarova, K. S. et al. Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants. *Nat. Rev. Microbiol.* **18**, 67–83 (2020).
2. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
3. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
4. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
5. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759–771 (2015).
6. Shan, Q. et al. Targeted genome modification of crop plants using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 686–688 (2013).
7. Wang, H., La Russa, M. & Qi, L. S. CRISPR/Cas9 in genome editing and beyond. *Annu. Rev. Biochem.* **85**, 227–264 (2016).
8. Ran, F. A. et al. In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature* **520**, 186–191 (2015).
9. Anzalone, A. V., Koblan, L. W. & Liu, D. R. Genome editing with CRISPR-Cas nucleases, base editors, transposases and prime editors. *Nat. Biotechnol.* **38**, 824–844 (2020).
10. Amoasii, L. et al. Gene editing restores dystrophin expression in a canine model of Duchenne muscular dystrophy. *Science* **362**, 86–91 (2018).
11. Pickar-Oliver, A. & Gersbach, C. A. The next generation of CRISPR-Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490–507 (2019).
12. Nelson, C. E. et al. Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. *Nat. Methods* **25**, 427–432 (2019).
13. 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).
14. Liu, J. J. et al. CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature* **566**, 218–223 (2019).
15. Pausch, P. et al. CRISPR-CasΦ from huge phages is a hypercompact genome editor. *Science* **369**, 333–337 (2020).
16. Harrington, L. B. et al. Programmed DNA destruction by miniature CRISPR-Cas14 enzymes. *Science* **362**, 839–842 (2018).
17. Karvelis, T. et al. PAM recognition by miniature CRISPR-Cas12f nucleases triggers programmable double-stranded DNA target cleavage. *Nucleic Acids Res.* **48**, 5016–5023 (2020).
18. Strecker, J. et al. Engineering of CRISPR-Cas12b for human genome editing. *Nat. Commun.* **10**, 212 (2019).
19. Takeda, S. N. et al. Structure of the miniature type V-F CRISPR-Cas effector enzyme. *Mol. Cell* **81**, 558–570 (2021).
20. Xiao, R., Li, Z., Wang, S., Han, R. & Chang, L. Structural basis for substrate recognition and cleavage by the dimerization-dependent CRISPR-Cas12f nuclease. *Nucleic Acids Res.* **49**, 4120–4128 (2021).
21. Lei, C. et al. The CCTL (Cpf1-assisted Cutting and Taq DNA ligase-assisted Ligation) method for efficient editing of large DNA constructs in vitro. *Nucleic Acids Res.* **45**, e74 (2017).
22. Chen, J. S. et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* **360**, 436–439 (2018).
23. Teng, F. et al. CDetection: CRISPR-Cas12b-based DNA detection with sub-attomolar sensitivity and single-base specificity. *Genome Biol.* **20**, 132 (2019).
24. Harrington, L. B. et al. A scoutRNA is required for some type V CRISPR-Cas systems. *Mol. Cell* **79**, 416–424 (2020).
25. Yan, W. X. et al. Functionally diverse type V CRISPR-Cas systems. *Science* **363**, 88–91 (2019).
26. Wu, Z. et al. Strategies for developing CRISPR-based gene editing methods in bacteria. *Small Methods* **4**, 1900560 (2020).
27. Perrenoud, A. & Sauer, U. Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr and Mlc on glucose catabolism in *Escherichia coli*. *J. Bacteriol.* **187**, 3171–3179 (2005).

28. Wyres, K. L. & Holt, K. E. *Klebsiella pneumoniae* as a key trafficker of drug resistance genes from environmental to clinically important bacteria. *Curr. Opin. Microbiol.* **45**, 131–139 (2018).
29. Hu, Z. et al. A compact Cas9 ortholog from *Staphylococcus auricularis* (SauriCas9) expands the DNA targeting scope. *PLoS Biol.* **18**, e3000686 (2020).
30. Grimm, D. et al. In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J. Virol.* **82**, 5887–5911 (2008).
31. Tsai, S. Q. et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat. Methods* **14**, 607–614 (2017).

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

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021

Methods

Plasmid construction. *Escherichia coli*-codon-optimized AsCas12f1, human-codon-optimized AsCas12f1 and the non-coding region of the CRISPR-AsCas12f1 locus from *A. sulfuroxidans* were ordered from Sangon Biotech. To express the CRISPR-AsCas12f1 system in bacteria, a p15a-based plasmid was constructed to carry a *rpsL*-driven AsCas12f1 and a J23119-driven CRISPR non-coding region with a single repeat–spacer–repeat unit, amenable to stuffer–spacer exchange by Golden Gate assembly (*Bsa*I-restriction sites). For PAM depletion, a plasmid library was constructed as previously described[32], which contained six randomized nucleotides located at the immediate 5′ upstream of a target sequence. For in vitro dsDNA cleavage assay, pUC19-based plasmids carrying different PAMs were constructed via circular polymerase extension cloning (CPEC)[33]. For heterologous expression of AsCas12f1 in *E. coli*, a pET28a-based plasmid was constructed with an N-terminal hexa-histidine tag, a SUMO tag and an HRV3c recognition site. Two NLSs were inserted onto the N and C termini to express NLS-tagged AsCas12f1. Mutations in the RuvC active sites were created by CPEC. For bacterial genome editing, a p15a-based *rpsL*-driven AsCas12f1-expressing plasmid was constructed with *araBAD*-driven lambda-red recombinase cassette in two versions, with different antibiotic resistance markers (the carbenicillin-resistant marker, *blaZ*, for *E. coli* and the apramycin-resistant marker, *apmR*, for *K. pneumoniae*).The sgRNA_v1 was separately expressed on another kanamycin-resistant pSGKP-based plasmid, which is amenable to stuffer–spacer exchange by Golden Gate assembly (*Bsa*I-restriction sites). For human genome editing, a CMV-driven AsCas12f1 (human-codon-optimized, N- and C-terminal NLS tagged)-expressing plasmid was constructed with a U6-driven sgRNA_v1 cassette (containing *Bsa*I sites for spacer insertion). For EGFP activation, a short endogenous fragment was inserted into a CMV-driven EGFP-expressing plasmid at the beginning of the EGFP CDS, causing a frameshift of the EGFP open reading frame. For AAV packing, a CMV-driven NLS-AsCas12f1-NLS-P2A-EGFP cassette and a U6-driven sgRNA_v1 cassette were assembled into an AAV2-based plasmid. Detailed map and sequence information for the plasmids used in this study is available in Supplementary Table 1.

PAM depletion assay and analysis. The PAM depletion assay was carried out as previously described[14]. Briefly, 100 ng of the PAM library plasmid was electroporated into the *E. coli* carrying the CRISPR-AsCas12f1 system with a spacer targeting the PAM library. Approximately 100,000 colonies were collected to ensure appropriate coverage of all possible PAMs. Plasmids were extracted using a SPARKeasy Superpure Mini Plasmid Kit (Shandong Sparkjade Biotechnology Co.). The amplicon sequencing library was generated by a two-step amplification method. The polymerase chain reaction (PCR) products were purified by 1.5× VAHTS DNA Clean Beads (Vazyme Biotech Co.) and subjected to Illumina HiSeq sequencing (PE150). Sequencing reads were mapped to the reference plasmid sequence and PAM sequences were extracted. The abundance of each PAM was counted from the aligned reads and normalized to the total reads. The fold change of PAM depletion was calculated by the ratio compared to the abundance in the control group (using the same plasmid but without targeting spacer). PAMs depleted over 10-fold were used to generate a WebLogo.

Small RNA-seq and analysis. The total RNA of the *E. coli* carrying the CRISPR-AsCas12f1 system with a non-targeting spacer was extracted using an EasyPure RNA Kit (TransGen Biotech Co.). RNA (15 μg) was treated with 2 U of DNase I (NEB) for 30 min at 37 °C and subsequently purified by phenol/chloroform extraction. The DNA-free RNA was then treated with 20 U of T4 polynucleotide kinase (NEB) for 6 h at 37 °C for 2′-3′-dephosphorylation. The reaction was then supplemented with 1 mM adenosine triphosphate (ATP) and incubated at 37 °C for another 1 h for 5′-phosphorylation. The T4 polynucleotide kinase-treated RNA was subsequently purified by phenol/chloroform extraction and treated with 5 U of RppH (NEB) to hydrolyze 5′-pyrophosphate for 1 h at 37 °C. The resulting RNA was purified by phenol/chloroform extraction. The quality was checked by agarose gel electrophoresis. Pre-treated RNA (5 μg) was used to prepare the small RNA library using a VAHTS Small RNA Library Prep Kit for Illumina (Vazyme) according to the manufacturer's instructions, with modifications in library amplification using two index primers containing two different 8-nt indices. The small RNA library was subjected to Illumina HiSeq sequencing (PE150). The sequencing reads were mapped to the reference sequence by HISAT2 (ref. [34]) after the removal of adaptors and low-quality reads. Total mapped unique reads were extracted and sorted using SAMtools[35] and the coverage depth of each site was calculated.

AsCas12f1 expression and purification. The *E. coli* BL21(DE3) cells containing the pET28a-based AsCas12f1-expressing plasmid were incubated at 37 °C while shaking vigorously until the optical density at 600 nm (OD_{600}) reached 0.6. The expression of AsCas12f1 was induced by supplementing IPTG to a final concentration of 0.25 mM, then incubation was continued at 16 °C overnight. Cells were collected by centrifugation and resuspended in buffer A (20 mM Tris-HCl, pH 7.5, 1 M NaCl, 15 mM imidazole and 1 mM DTT), then disrupted by a French press and clarified by centrifugation. The clear lysate was loaded onto a 5-ml HisTrap Ni-NTA column (GE healthcare/Cytiva) pre-equilibrated with the same buffer. Unbound proteins were washed away with 10 column volumes (CVs) of buffer A with 50 mM imidazole, and the His-tagged AsCas12f1 protein was eluted with 5 CV buffer A with 250 mM imidazole. The eluted protein was buffer-exchanged by centrifugation to remove imidazole, and subjected to tag removal by HRV3c with incubation at 4 °C for overnight. The digested product was concentrated to 2 ml and then loaded onto a HiLoad 16/600 Superdex 200-pg column (GE healthcare/Cytiva) pre-equilibrated with buffer A containing no imidazole. Peak fractions were merged and concentrated to ~1.2 mg ml^{-1} (~25 μM). The whole procedure of protein purification was carried out at 4 °C. The resulting protein can be kept at 4 °C for more than a week without detectable activity loss. For long-term storage, the protein was supplemented with 10% glycerol, then flash-frozen in liquid nitrogen and stored at −80 °C.

sgRNA preparation. The DNA templates for sgRNA transcription were ordered from Sangon Biotech Co. The sgRNA was synthesized in vitro using a HiScribe T7 High Yield RNA Synthesis Kit (NEB). Generally, 20 μl of reaction was prepared containing 1 μg of templates and incubated at 37 °C overnight. The sgRNA was treated with 4 U of DNase I (NEB) for 15 min and subjected to phenol/chloroform extraction and ethanol precipitation. The sgRNA pellet was then dissolved in RNase-free water and diluted to a concentration of ~1,000 ng μl^{-1}. The sgRNA templates used in this study are provided in Supplementary Table 2.

Analytical size-exclusion chromatography. To analyze the formation of RNP complex, 5 nmoles of AsCas12f1 were incubated with 10 nmoles of tracrRNA, and 10 nmoles of crRNA in 250 μl of buffer B (20 mM Tris-HCl, pH 7.5, 600 mM NaCl, 5 mM $MgCl_2$ and 1 mM DTT) at 45 °C for 10 min. The sample was then injected onto a 25-ml Superose 6 10/300 GL column (GE Healthcare/Cytiva) pre-equilibrated with the same buffer for SEC. The OD_{280} and OD_{260} values were recorded simultaneously.

In vitro DNA cleavage assay for the dsDNA substrate. For 56-bp 6-FAM-labeled dsDNA cleavage, 20 nM substrate was incubated with 500 nM AsCas12f1 and 500 nM sgRNA in 1× reaction buffer (10 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$ and 50 mM NaCl) at 45 °C, then 10-μl aliquots were removed at time intervals and quenched with 10 μl of formamide stop buffer (90% formamide, 0.075% xylene cyanol FF, 0.075% bromophenol blue and 50 mM EDTA). The cleavage products were denatured by incubating at 95 °C for 10 min and then analyzed by 20% TBE-urea-PAGE gel electrophoresis. For run-off sequencing, 5 nM pUC19 plasmid was incubated with 250 nM AsCas12f and 250 nM sgRNA in 1× reaction buffer at 45 °C for 60 min. The reaction was quenched with 6× gel loading dye (NEB). The digest product was gel-purified and subjected to Sanger sequencing. For linearized pUC19 cleavage, 5 nM linear plasmid was incubated with 250 nM AsCas12f1 and 250 nM sgRNA in 1× reaction buffer at 45 °C, then10-μl aliquots were removed at time intervals and quenched with 6× gel loading dye (NEB). The cleavage products were analyzed by 1% agarose gel electrophoresis. Sequence information for the substrates used in this study is provided in Supplementary Data 1.

In vitro DNA cleavage assay for the ssDNA substrate. For 56-nt 6-FAM-labeled ssDNA cleavage, 20 nM substrate was incubated with 500 nM AsCas12f1 and 500 nM sgRNA in 1× reaction buffer at 45 °C, then 10-μl aliquots were removed at time intervals and quenched with 10 μl of formamide stop buffer. The cleavage products were denatured and analyzed by 20% TBE-urea-PAGE gel electrophoresis. For M13 ssDNA cleavage, 5 nM M13 ssDNA substrate was incubated with 100 nM AsCas12f1, 100 nM sgRNA, and 50 nM of targetable dsDNA or ssDNA activator in 1× reaction buffer at 45 °C, then 10-μl aliquots were removed at time intervals and quenched for agarose gel electrophoresis. Sequence information for the substrates used in this study is provided in Supplementary Data 1.

In vitro pre-crRNA processing assay. The templates for tracrRNA and pre-crRNA (containing a single repeat–spacer–repeat unit) were ordered from Sangon Biotech Co. RNAs were synthesized in vitro by using a HiScribe T7 High Yield RNA Synthesis Kit and purified by phenol/chloroform extraction. To determine whether AsCas12f1 exerts the pre-crRNA processing, 100 nM pre-crRNA and 100 nM tracrRNA were incubated with 250 nM AsCas12f1 in 1× reaction buffer at 45 °C, then 10-μl aliquots were removed at time intervals and quenched with formamide stop buffer. The denatured cleavage products were analyzed by 10% TBE-urea-PAGE gel electrophoresis. The RNA templates used in this study are provided in Supplementary Table 2.

Bacterial genome targeting and editing. The bacterial genome targeting assay was carried out as previously described[36]. Briefly, the *E. coli* carrying the p15a-based AsCas12f-expressing plasmid was electroporated with the sgRNA-transcription plasmid containing a genome-targeting spacer. After incubation at 37 °C overnight, the colony forming units were counted to calculate the transformation efficiency (CFU per μg plasmid). The bacterial genome editing was performed as previously described[37,38]. The *E. coli* carrying the p15a-based AsCas12f-expressing plasmid with lambda-red recombinase was electroporated with the sgRNA-transcription plasmid containing a genome-targeting spacer along with 1 nmole of 90-nt ssDNA repair template. After incubation at 37 °C overnight, gene-editing events in the

resulting colonies were verified by PCR amplification and Sanger sequencing of the region surrounding the target locus. The targets and primers used in this study are provided in Supplementary Table 3 and Supplementary Data 1.

Cell culture and transfection. The HEK293, HepG2, Huh-7 and U-2 OS cells were maintained in DMEM medium (Corning) with 10% FBS (Biological Industrial) at 37 °C in 5% CO_2. A day before transfection, cells were trypsinized and seeded to each well of 24-well plates in appropriate concentrations (1.5×10^5 cells per well for HEK293 and 1.0×10^5 cells per well for HepG2, Huh-7 and U-2 OS). For plasmid-based genome editing, 500 ng of plasmid was transfected using Lipofectamine 3000 (Life Technologies) according to the manufacturer's instructions. For RNP transfection, 300 pmole of purified NLS-tagged AsCas12f1 was incubated with 300 pmole of sgRNA in 50 µl of Opti-MEM supplemented with 5 mM $MgCl_2$ at 45 °C for 10 min. The ribonucleoprotein was transfected using Lipofectamine 3000 (Life Technologies).

EGFP activation assay. The EGFP activation assay was performed as previously described, with modifications[29]. Briefly, 600 ng of EGFP plasmid and 400 ng of AsCas12f1-expressing plasmid were co-transfected into HEK293 cells at ~60% confluency. After 24 h of incubation at 37 °C, the cells were trypsinized and analyzed by flow cytometry using an LSRFortessa instrument (BD Biosciences). The gating strategy is shown in Extended Data Fig. 4.

AAV production and cell transduction. The AAV production was carried out by PackGene Biotech Co., giving a final physical titration of 10^{13} GC ml^{-1} (GC, genome copy number). HEK293, Huh-7 and U-2 OS cells were seeded at 0.5×10^5 per well to each well of 24-well plates, respectively. AAV was added to the cells at a multiplicity of infection of 10^5 GC per cell. Cells were collected after incubation at 37 °C for the appropriate days (three days for HEK293, four days for U-2 OS and six days for Huh-7) to evaluate the endogenous gene disruption.

Evaluation of endogenous gene disruption. Genomic DNA was extracted at 48 h post-transfection using an Ezup Column Animal Genomic DNA Purification Kit (Sangon). Approximately 500-bp regions surrounding the target loci were amplified using a Phanta Max Super-Fidelity DNA Polymerase Mastermix (Vazyme). PCR products were gel-purified and re-annealed. The indels were detected using T7 endonuclease I (Vazyme) digestion. For NGS, ~200-bp regions surrounding the target loci were amplified using primers with dual 8-nt barcodes. The PCR products were merged and gel-purified. The amplicon-seq library was prepared by using a VAHTS Universal DNA Library Prep Kit for Illumina (Vazyme) according to the manufacturer's instructions. The library was subjected to Illumina NovaSeq sequencing (PE150) by Genewise Co. Sequencing data were demultiplexed according to the barcodes, and indels were analyzed using Cas-Analyzer[39]. The targets and primers used in this study are provided in Supplementary Table 3 and Supplementary Data 1.

CIRCLE-seq for identifying genome-wide off-target sites. CIRCLE-seq was carried out as previously described with modifications[40]. Purified genomic DNA from HEK293 cells was sheared to 200–500-bp fragments using NEBNext dsDNA Fragementase (NEB). Sheared genomic DNA was end-repaired, dA-tailed and ligated to a hairpin adaptor using a VAHTS Universal DNA Library Prep Kit for Illumina (Vazyme) according to the manufacturer's instructions. The adaptor-ligated DNA was treated with lambda exonuclease (NEB) and *E. coli* exonuclease I (NEB), then with USER enzyme (NEB) and T4 polynucleotide kinase (NEB). DNA was circularized with T4 DNA ligase (NEB) at a concentration of 5 ng $µl^{-1}$, then treated with Plasmid-Safe ATP-dependent DNase (Epicentre), then 150 ng of circular genomic DNA library was cleaved by 100 nM AsCas12f1 and 100 nM sgRNA in 1× reaction buffer at 37 °C for 1 h. The cleaved products were purified using VAHTS DNA Clean Beads (Vazyme), and end-repaired, dA-tailed and ligated to NEBNext adaptor (NEB) using a VAHTS Universal DNA Library Prep Kit for Illumina (Vazyme). The USER-treated library was amplified by PCR using Phanta Max Super-Fidelity DNA Polymerase Mastermix (Vazyme), then subjected to Illumina Novaseq sequencing (PE150). Data analysis was performed using the CIRCLE-seq analysis pipeline with minor modifications to fit the target sequence of AsCas12f1. The experimentally determined and in silicon predicted off-target sites are listed in Supplementary Data 2.

Statistics and reproducibility. All statistical analyses were performed using GraphPad Prism (v.8.2.1). The exact replication numbers are indicated in the figure legends. The experimental findings in all of the figures were successfully reproduced at least three times independently.

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

Data availability

NGS data are available at the NCBI Sequence Read Archive (SRA) under accession codes SRX9277304, SRX9277305, SRX9277306, SRX10828545, SRX10828546, SRX10828547 and SRX10839449. Plasmids can be accessed in Addgene under accession codes 171609, 171610, 171611, 171612, 171613 and 171614. Source data are provided with this paper.

Code availability

The custom Python scripts for calculating sequence frequencies in the PAM depletion assay are available at GitHub (https://github.com/atlasbioinfo/midseq). The computational pipeline of CIRCLE-seq to identify genome-wide off-target sites for AsCas12f1 is available at GitHub (https://github.com/tsailabSJ/circleseq).

References

32. Karvelis, T. et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol.* **16**, 253 (2015).
33. Quan, J. & Tian, J. Circular polymerase extension cloning for high-throughput cloning of complex and combinatorial DNA libraries. *Nat. Protoc.* **6**, 242–251 (2011).
34. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915 (2019).
35. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
36. Zhang, Y. et al. Catalytic-state structure and engineering of *Streptococcus thermophilus* Cas9. *Nat. Catal.* **3**, 813–823 (2020).
37. Chen, W. et al. CRISPR/Cas9-based genome editing in *Pseudomonas aeruginosa* and cytidine deaminase-mediated base editing in *Pseudomonas* species. *Iscience* **6**, 222–231 (2018).
38. Wang, Y. et al. CRISPR-Cas9 and CRISPR-assisted cytidine deaminase enable precise and efficient genome editing in *Klebsiella pneumoniae*. *Appl. Environ. Microbiol.* **84**, e01834 (2018).
39. 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).
40. Lazzarotto, C. R. et al. Defining CRISPR-Cas9 genome-wide nuclease activities with CIRCLE-seq. *Nat. Protoc.* **13**, 2615–2642 (2018).

Acknowledgements

This work was supported by grants 21922705 (Q.J.), 91753127 (Q.J.) and 2207783 (Q.J.) from the National Natural Science Foundation of China; 19QA1406000 (Q.J.) from the Shanghai Committee of Science and Technology, China; 2017YFA0506800 (Q.J.) from the National Key R&D Program of China; EKPG21-18 (Q.J.) from the Emergency Key Program of Guangzhou Laboratory; and 2019M651627 (Z.W.) from the China Postdoctoral Science Foundation. We thank C. He at The University of Chicago for helpful discussions. We thank R. Liu, X. Huang and J. Li at ShanghaiTech University for kindly providing HEK293 cells and valuable advice regarding cell culture experiments. We thank G. Luo and Z. Ren at Sun Yat-sen University for helpful advice in RNA-seq experimental design. We thank J. Li, Y. Xiong and X. Ren from the Molecular and Cell Biology Core Facility (MCBCF) at the School of Life Science and Technology, ShanghaiTech University, for providing technical support.

Author contributions

Z.W., Y.Z. and Q.J. conceived the initial study. Z.W. and Yujue Wang performed the PAM depletion assay. Z.W. and H.Y. conducted the small RNA-seq and computational analyses. Z.W., Y.Z., Yannan Wang, F.L., D.P. and W.C. performed plasmid construction, protein purification and biochemical experiments. Z.W., Yujue Wang and Yannan Wang performed the genome editing in bacteria. Z.W., H.Y., D.P. and Yujue Wang performed the genome editing in human cells and NGS data analyses. Z.W., H.N. and D.P. designed the AAV vector and performed the AAV-based genome editing. Z.W. and C.L. analyzed the sequence homology among Cas12 nucleases and built the structure model. Z.W., H.Y. and D.P. performed the CIRCLE-seq. Z.W. and Q.J. prepared the figures and wrote the manuscript. The manuscript was reviewed and approved by all authors.

Competing interests

Q.J., Z.W., Y.Z., Yujue Wang and Yannan Wang have filed a patent application (PCT/CN2020/132759) related to this work through ShanghaiTech University. The remaining authors declare no competing interests.

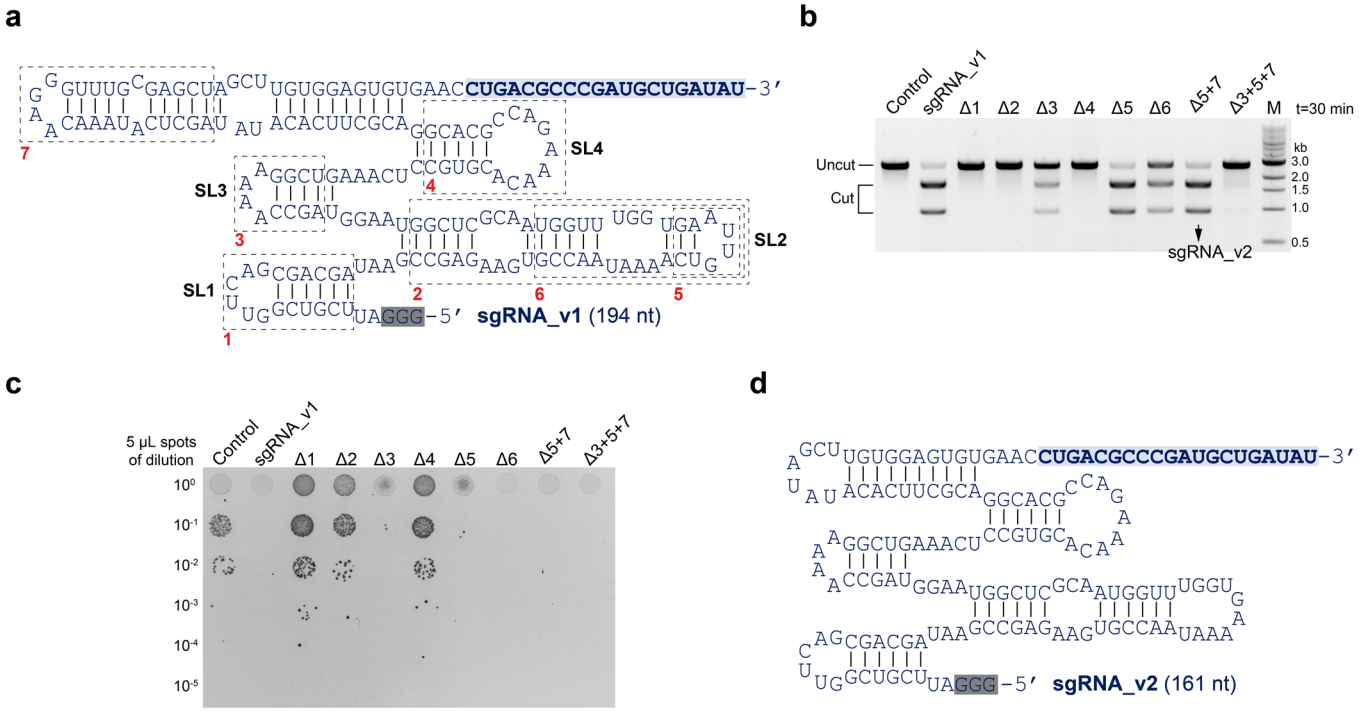
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41589-021-00868-6.

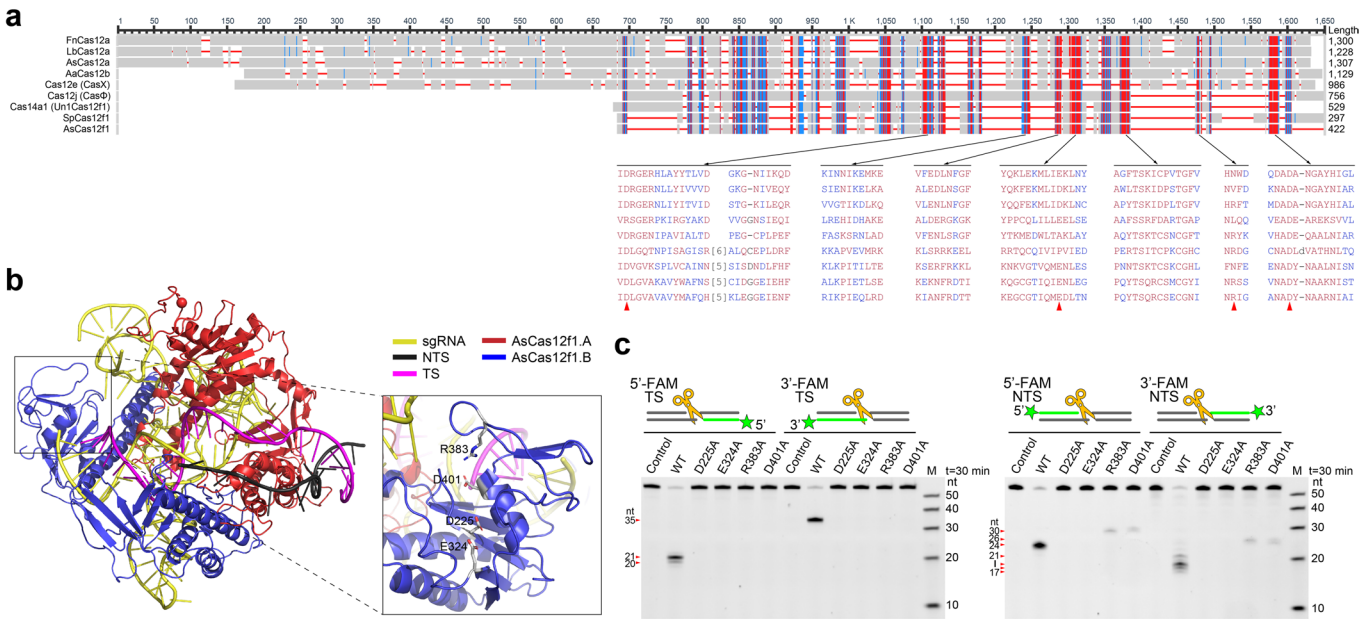
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41589-021-00868-6.

Correspondence and requests for materials should be addressed to Q.J.

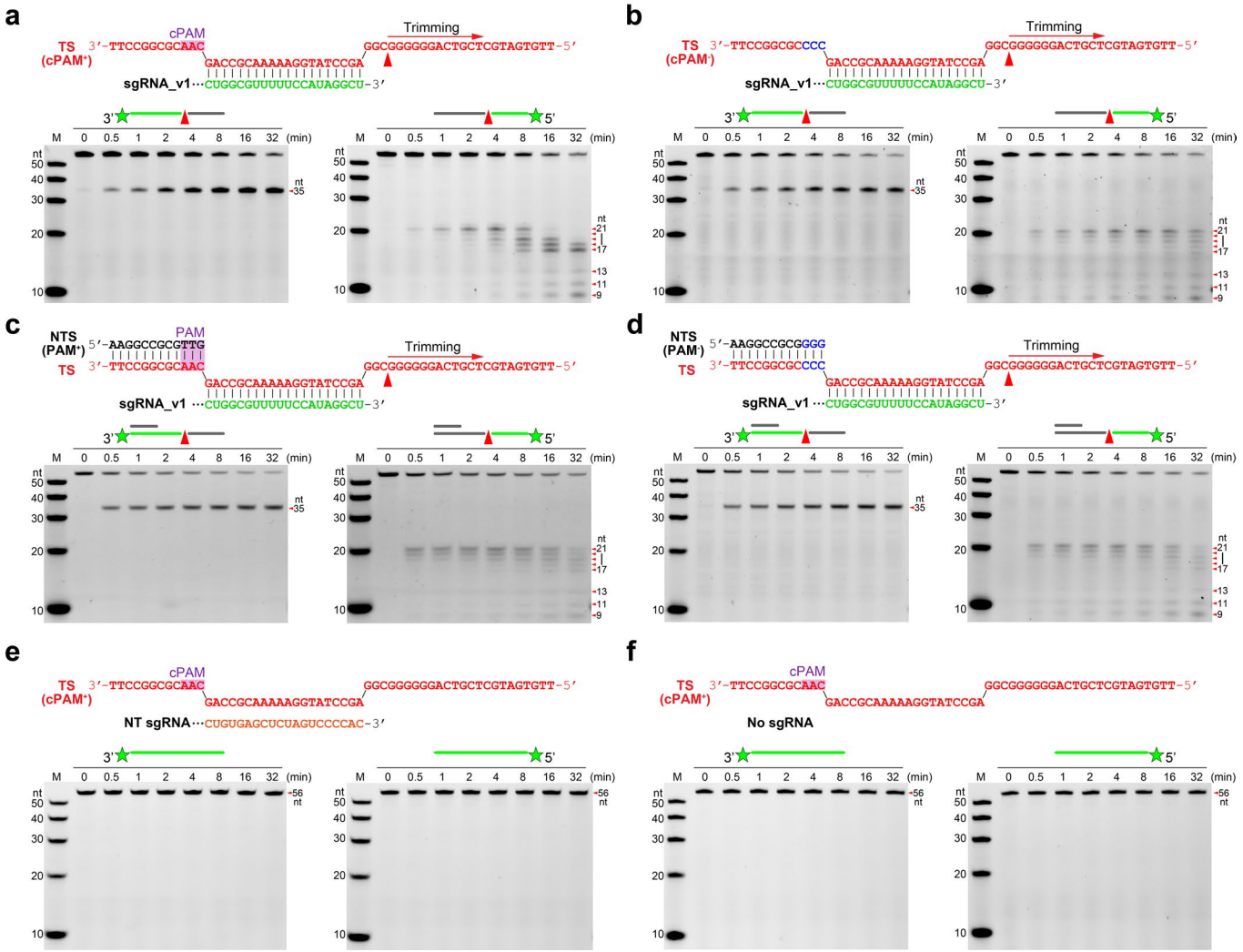
Reprints and permissions information is available at www.nature.com/reprints.



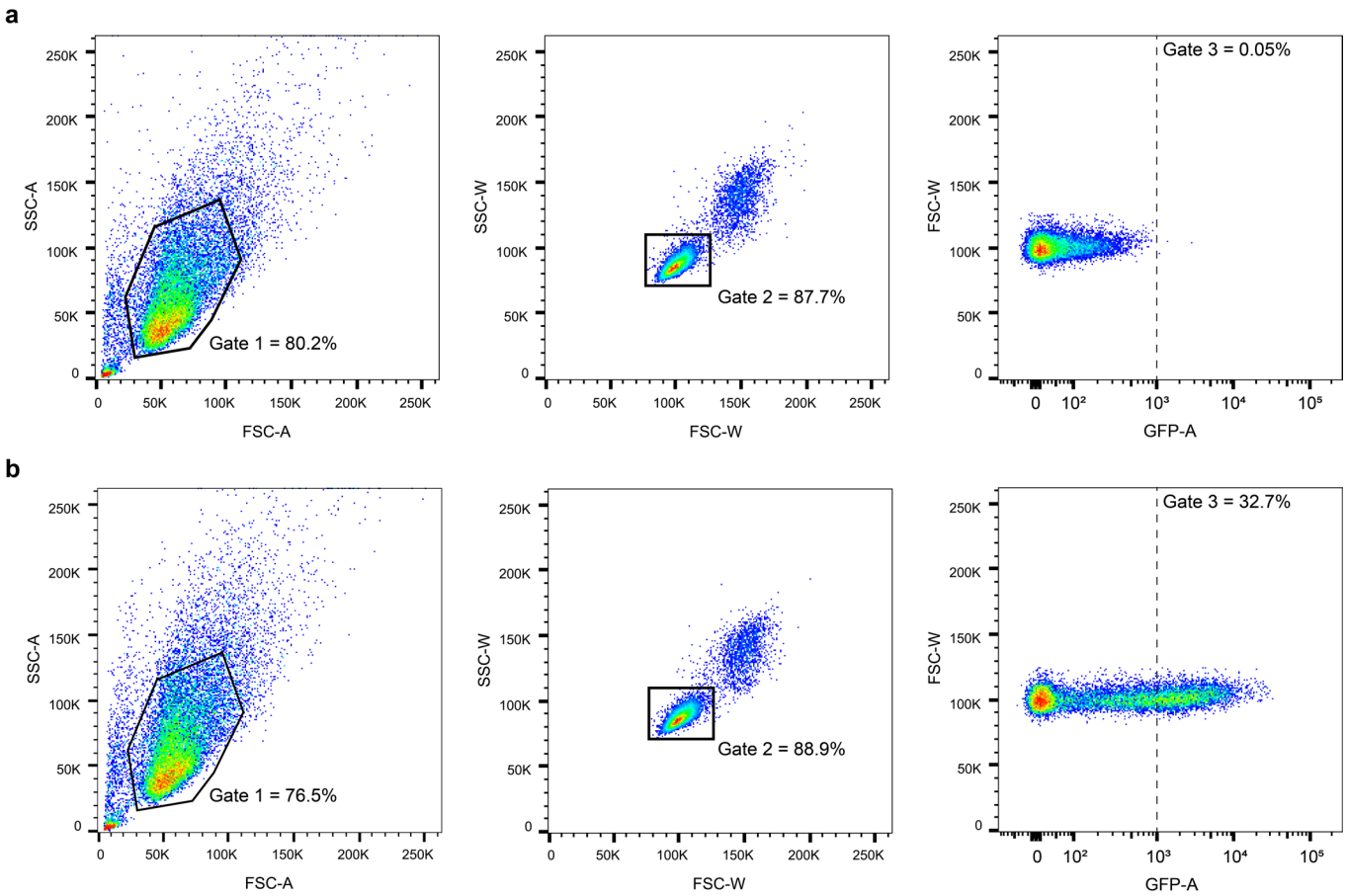
Extended Data Fig. 1 | Optimization of the sgRNA for AsCas12f1. a, The predicted secondary structure of sgRNA_v1 and the truncation information. The gray shaded tri-guanosine is added to the beginning of sgRNA to maximize the RNA production. **b**, in vitro dsDNA cleavage assay for different truncations of sgRNA. **c**, Bacterial genome targeting using different truncations of sgRNA. **d**, The predicted secondary structure of sgRNA_v2.



Extended Data Fig. 2 | Sequence and structural analysis of AsCas12f1 and its mutations in the RuvC active site. a, Multiple sequence alignment of Cas12 nucleases. **b**, Structure model of AsCas12f1 was built using the cryo-EM structure of Cas14a1/sgRNA/dsDNA (PDB ID: 7C7L) as a template using SWISS-MODEL web server with its default parameters. **c**, D225A and E324A completely abolished the nuclease activity of AsCas12f1, whereas R383A and D401A retained minor NTS nickase activity at 18-nt downstream the PAM.



Extended Data Fig. 3 | AsCas12f1 cleaves ssDNA in a PAM-independent manner. a, Cleavage pattern for complementary-PAM (cPAM)-containing ssDNA. **b**, Cleavage pattern for non-cPAM-containing ssDNA. **c**, Cleavage pattern for PAM-containing staggered substrate. **d**, Cleavage pattern for non-PAM-containing staggered substrate. **e**, Non-targeting sgRNA cannot mediate targeted cleavage on short ssDNA substrate. **f**, AsCas12f1 cannot non-specifically cleave short ssDNA substrate.



Extended Data Fig. 4 | Flow cytometry gating strategies. a, **b**, Gates used for the analysis of flow cytometric data shown in Fig. 3c. Gate 1 is used to remove the debris from the cell population, gate 2 specifies single cells from cell aggregates, and gate 3 identifies the cells with GFP fluorescence. Gates were set on the control sample (**a**) and was then kept constant for the experimental group (**b**).

nature research

Corresponding author(s): Quanjiang Ji

Last updated by author(s): Jul 13, 2021

Reporting Summary

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

Statistics

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

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

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

Software and code

Policy information about availability of computer code

Data collection	BD FACSDiva (Version 8.0.1) for flow cytometry data collection.
Data analysis	Hisat (version 2.2.1); SAMtools (version 1.11); Cas-Analyzer (http://www.rgenome.net/cas-analyzer/); CIRCLE-seq Pipeline (version 1.1); SWISS-MODEL (https://swissmodel.expasy.org/); Flowjo (version 10); midSeq (version 1.0)

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

Data

Policy information about availability of data

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

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

NGS data are available at the NCBI Sequence Read Archive (SRA) under accession codes SRX9277304, SRX9277305, SRX9277306, SRX10828545, SRX10828546, SRX10828547, and SRX10839449. Plasmids can be accessed in Addgene under accession codes 171609, 171610, 171611, 171612, 171613, and 171614. Source data are provided with the paper.

Field-specific reporting

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

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

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

Life sciences study design

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

Sample size	No statistical methods were used to determine the sample size, because the sample size was not a factor of the analysis for the experiment. All sample size were determined from preliminary test experiments.
Data exclusions	No data were excluded from analysis
Replication	All biochemical experiments were independently reproduced at least three times. All attempts at replication were successful.
Randomization	No randomization was applied to this study. No animals or human subjects were involved. The sample size for analysis was small. All experiments provided quantitative results. Uncontrollable conditions of the experiment did not affect the results, because controls were always included in assays to avoid biased results.
Blinding	Blinding was not performed because the data we analyzed (DNA cleavage, genome editing efficiency, etc.) are not subject to biased interpretation.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293, HepG2, Huh-7, U-2 OS from ATCC
Authentication	No cell lines were authenticated
Mycoplasma contamination	Cells were not tested for mycoplasma contamination
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines.

Flow Cytometry

Plots

Confirm that:

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

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

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

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

Methodology

Sample preparation	600 ng of EGFP plasmid and 400 ng of AsCas12f1-expressing plasmid were co-transfected into HEK293 cells in approximately 60% confluency. After 24-hrs incubation at 37°C, the cells were trypsinized for flow cytometry.
Instrument	LSRFortessa (BD Biosciences) for data collection.
Software	BD FACSDiva (Version 8.0.1) for collection. FlowJo (version 10) for analysis.
Cell population abundance	A minimum of 20,000 cells/sample were acquired for flow cytometry.
Gating strategy	HEK293 cells were identified by forward and side scatter, followed by doublet exclusion. Cells were examined for GFP positivity. cells transfected with control plasmid (without targeting spacer) were used as a negative control to draw boundaries between GFP-positive and GFP-negative cells.

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