Engineered CRISPR-Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing

Benjamin P. Kleinstiver, Alexander A. Sousa, Russell T. Walton, Y. Esther Tak, Jonathan Y. Hsu, Kendell Clement, Moira M. Welch, Joy E. Horng, Jose Malagon-Lopez, Irene Scarfo, Marcela V. Maus, Luca Pinello, Martin J. Aryee and J. Keith Joung

Broad use of CRISPR-Cas12a (formerly Cpf1) nucleases has been hindered by the requirement for an extended TTTV protospacer adjacent motif (PAM). To address this limitation, we engineered an enhanced Acidaminococcus sp. Cas12a variant (enAsCas12a) that has a substantially expanded targeting range, enabling targeting of many previously inaccessible PAMs. On average, enAsCas12a exhibits a twofold higher genome editing activity on sites with canonical TTTV PAMs compared to wild-type AsCas12a, and we successfully grafted a subset of mutations from enAsCas12a onto other previously described AsCas12a variants to enhance their activities. enAsCas12a improves the efficiency of multiplex gene editing, endogenous gene activation and C-to-T base editing, and we engineered a high-fidelity version of enAsCas12a (enAsCas12a-HF1) to reduce off-target effects. Both enAsCas12a and enAsCas12a-HF1 function in HEK293T and primary human T cells when delivered as ribonucleoprotein (RNP) complexes. Collectively, enAsCas12a provides an optimized version of Cas12a that should enable wider application of Cas12a enzymes for gene and epigenetic editing.

CRISPR–Cas nucleases are widely used for gene, epigenetic and base editing in human cells and other organisms. The study of alternative CRISPR nucleases beyond the commonly used Streptococcus pyogenes Cas9 (SpCas9), including Cas12a orthologs, has yielded additional enzymes with distinct and potentially advantageous properties. Cas12a nucleases, including AsCas12a and Lachnospiraceae bacterium ND2006 Cas12a (LbCas12a), recognize target sites with T-rich PAMs, require a single short ~40 nucleotide (nt) CRISPR RNA (crRNA) to program target specificity, and possess RNase activity that enables multiplex targeting through poly-crRNA transcript processing. Although Cas12a enzymes have shown utility for multiplex gene editing, gene activation and combinatorial library screens, one constraint is their requirement for a longer PAM of the form 5′-TTTV (where V is A, C, or G), which restricts targeting approximately sixfold relative to SpCas9. Although Cas12a orthologs from Francisella novicida (FnCas12a) and Moraxella bovoculi 237 (MbCas12a) were previously reported to recognize an increased number of PAMs in vitro, our own findings (Supplementary Fig. 1a,b and see Supplementary Results) and those previously reported by others have shown their activities to be less consistent and robust in human cells. Additionally, although two engineered AsCas12a variants (referred to as RVR and RR) were previously described that can recognize alternative TATV and TYCV PAMs, respectively, many PAMs remain inaccessible to Cas12a. Thus, additional variants with expanded targeting capabilities are needed to enable applications requiring high targeting density and flexibility.

We used structure-guided protein engineering to attempt to expand the PAM recognition of Cas12a nucleases, focusing on AsCas12a because it has been widely used for genome editing and structural information was available. To do so, we engineered ten variants bearing single amino acid substitutions to positively charged arginine residues that might be expected to alter or form novel PAM proximal DNA contacts (Supplementary Fig. 1c). Four of the 10 variants we tested in human cells (S170R, E174R, S542R and R548R) displayed higher gene editing activities on sites with canonical and non-canonical PAMs relative to wild-type AsCas12a (Supplementary Fig. 1d,e, respectively). Testing of additional variants harboring combinations of these four substitutions on an expanded number of targets showed that two variants (E174R/S542R and E174R/S542R/K548R) exhibited the highest editing activities on sites with non-canonical PAMs while still retaining robust activities on a canonical PAM site (Fig. 1a and Supplementary Fig. 1f).

To comprehensively profile the PAM preferences of these AsCas12a variants, we optimized an unbiased in vitro PAM determination assay (PAMDA) similar to other previously described methods (see Supplementary Results, Supplementary Fig. 2a–d and Supplementary Table 1). Using the PAMDA, we defined the PAM preferences of wild-type AsCas12a and variants with all possible single, double and triple combinations of the E174R, S542R and K548R substitutions. Plots of the mean PAMDA log10 values on all 256 4-nt PAM sequences revealed that, as expected, targeting
with wild-type AsCas12a was most efficient against TTTV PAMs, and that E174R/S542R and E174R/S542R/K548R showed the most expanded PAM preferences among the seven variants tested (Fig. 1b). Strikingly, the E174R/S542R/K548R variant could target many PAMs including TTYN (TTTN/TTCN), VTTT (ATTV/CTTV/GTTV), TRTV (ATTV/CTTV/GTTV), and TTV (ATTV/CTTV/GTTV) and others.

Fig. 1 | Engineering and characterization of AsCas12a variants with expanded target range in human cells. a, Modification of endogenous sites in human cells by AsCas12a variants, assessed by T7E1 assay; mean shown for n ≥ 3. b, PAM preference profiles, assessed by PAMDA, for wild-type AsCas12a and all intermediate single and double substitution variants that comprise the E174R/S542R/K548R variant. The log10 rate constants ($k$) are the mean of four replicates, two each against two distinct spacer sequences (see Supplementary Fig. 2d). c, Mean activity plots for AsCas12a variants on sites with non-canonical PAMs, one for each PAM of the seven variants tested. d, Summary of the activities of wild-type AsCas12a and variants across sites in human cells encoding non-canonical PAMs, one for each PAM of the seven variants tested. e, Superimposition of the summaries of the human cell activities and PAMDA rate constants ($k$) for various targetable PAMs with enAsCas12a (E174R/S542R/K548R); mean and 95% confidence interval for human cell data are shown with black lines. Tier 1 PAMs exhibit greater than 20% mean targeting in human cells and a PAMDA $k$ greater than 0.01; tier 2 PAMs meet a modest threshold of greater than 10% mean targeting in cells and a PAMDA $k$ greater than 0.005 (see Supplementary Table 2).
Further characterization of the E174R/S542R and E174R/S542R/K548R variants in human cells showed robust editing activities on 60 endogenous target sites with VTTT and TTCN PAMs, and less effective modification of 15 target sites with VTTT PAMs (Fig. 1c and Supplementary Fig. 3a). Consistent with the PAMDA results (Fig. 1b), we observed efficient gene editing on 12 target sites bearing TATV PAMs with the E174R/S542R/K548R variant but not with E174R/S542R (Fig. 1c and Supplementary Fig. 3b). Both variants modified five sites with TTTT PAMs that were inefficiently edited with wild-type AsCas12a (Supplementary Fig. 3c). These data show that our variants enable robust editing of sites with non-canonical VTTT, TTTT and TTCN/TATV PAMs that cannot be modified efficiently by wild-type AsCas12a (summarized in Fig. 1d; see also Supplementary Figs. 1a and 3a–d).

Next, we examined the editing activities of the E174R/S542R/K548R variant on 97 other sites in human cells bearing 28 additional PAMs identified as targetable by the PAMDA. We observed efficient modification of 14 of 15 sites with TGTV PAMs (Supplementary Fig. 3e) and a range of editing activities across the other 82 sites harboring 25 additional PAMs (Supplementary Fig. 3f). Comparison of the mean PAMDA log10k and the mean human cell targeting values across the same PAMs showed a strong correlation for most PAMs (Supplementary Fig. 3g), suggesting that PAMs with a PAMDA log10k of –2.25 or higher were potentially targetable in human cells. PAMs accessible with E174R/S542R/K548R were binned into confidence tiers based on consistency between PAMDA and human cell experiments (Fig. 1e, Supplementary Fig. 3h and Supplementary Table 2; see Supplementary Results). Differences in activities across sites with the E174R/S542R/K548R variant could not be attributed to PAM or spacer sequence features (Supplementary Fig. 3i–n; see Supplementary Results). Taken together, our combined analyses illustrate that the E174R/S542R/K548R variant, henceforth referred to as enhanced AsCas12a (enAsCas12a), expands targeting range by approximately sevenfold (Fig. 1f).

While characterizing our different engineered AsCas12a variants, we noticed that certain substitutions were associated with increased editing activities in human cells (Fig. 1a and Supplementary Fig. 1d). To assess this improvement more comprehensively, we compared the gene modification activities of wild-type AsCas12a, E174R/S542R and enAsCas12a across 21 endogenous gene sites with canonical PAMs (Supplementary Fig. 4a). Compared to wild-type AsCas12a, both variants were on average nearly twofold more effective at modifying sites with TTTT PAMs (Fig. 2a).

Given the enhanced gene editing efficiencies observed with enAsCas12a, we speculated that this variant might also exhibit improved activity at lower temperatures, a property relevant for organisms that grow optimally at temperatures lower than 37°C. Previous studies showed that AsCas12a exhibited decreased activity at lower temperatures relative to LbCas12a18–22. In vitro cleavage reactions of AsCas12a, LbCas12a and enAsCas12a at 37, 32 and 25°C revealed that enAsCas12a is more active than AsCas12a at these temperatures, exhibiting activities more comparable to LbCas12a (Fig. 2b). Systematic examination of variants harboring all possible combinations of the E174R, S542R and K548R substitutions revealed that the improvements in cleavage efficiency with enAsCas12a at lower temperatures were largely attributable to E174R and to a lesser extent to S542R (Fig. 2b).

To determine whether the increased activity phenotype of enAsCas12a might be transferable to other AsCas12a variants, we added the E174R substitution to the previously described RVR and RR PAM recognition variants1 to create enRVR and enRR, respectively. Comparison of the activities of wild-type AsCas12a, enAsCas12a, RVR, enRVR, RR and enRR across 11 sites with TTTTN PAMs in human cells revealed that the original RVR and RR variants have similar or weaker activities compared to wild-type AsCas12a (Supplementary Fig. 4b). Furthermore, the enRVR and enRR variants generally showed more than twofold higher activities than RVR and RR, albeit with lower activities compared to enAsCas12a (Fig. 2c). Comparison of the editing activities of RVR and enRVR on 14 endogenous gene sites bearing TATN PAMs (Supplementary Fig. 4e) again showed an approximately twofold improvement with enRVR relative to the original RVR variant (Fig. 2d). Similarly, we examined the activities of enRR and RR in human cells on 29 sites bearing TYCN PAMs (Supplementary Fig. 4d) and found that enRR showed an average of 1.5-fold improved efficiency compared to RR (Fig. 2e). Importantly, the enRVR and enRR variants retained similar PAM preferences to their parental RVR and RR variants (Fig. 2f and Supplementary Fig. 4e). Collectively, enAsCas12, enRVR, and enRR exhibit improved activities and can target an expanded range of sequences compared to wild-type AsCas12a (Supplementary Fig. 4f and Supplementary Table 2).

We next sought to assess whether enAsCas12a could improve the efficiency of various Cas12a-based applications. One potential advantage of Cas12a enzymes is their RNase activity, which enables the processing of individual crRNAs from a poly-crRNA transcript27 and simplifies multiplex targeting in cells24–26. We compared the activities of AsCas12a, enAsCas12a and LbCas12a programmed with poly-crRNA arrays targeted to three endogenous genes in human cells (Fig. 3a–c). In all cases, we observed comparable or higher editing with enAsCas12a relative to AsCas12a and LbCas12a. We also designed multiplex arrays encoding two sets of proximally targeted crRNAs to generate small genomic deletions. Pairs of crRNAs were expressed from either poly-crRNA transcripts or pools of single crRNA plasmids, and we again observed comparable or improved deletion efficiencies with enAsCas12a relative to AsCas12a and LbCas12a (Fig. 3d).

Cas12a has also been used for epigenetic editing of endogenous human and plant genes by fusing DNase-inactive Cas12a (dCas12a) to heterologous effectors27–30. We found that a fusion of DNase-inactive enAsCas12a to the synthetic VPR activation domain (denAsCas12a–VPR) outperformed analogous dAsCas12a–, dLbCas12a– and dSpCas9–based VPR fusions (Supplementary Results and Supplementary Fig. 5a–f). Additional experiments comparing dAsCas12a–, denAsCas12a– and dLbCas12a–VPR targeted to the promoters of three endogenous human genes revealed that the most potent gene activation (range of 10- to 10,000-fold upregulation) was consistently achieved with the denAsCas12a–VPR fusion using pools of crRNAs targeted to sites with canonical (Fig. 3e) or non-canonical PAMs (Fig. 3f–g).

Cas12a enzymes have recently been adapted for base editing to induce targeted C-to-T alterations27. Base editors (BEs) consist of cytosine deaminases and uracil glycosylase inhibitor domains fused to nickase versions of Cas9 or DNase-inactive forms of LbCas12a27–30. Comparable DNase-inactive AsCas12a-BEs (AsBEs) have been reported as being minimally active27. To determine whether the enhanced activities of enAsCas12a could enable efficient base editing, we compared four different denAsCas12a base-editor fusions (enAsBE1.1–1.4; Supplementary Fig. 6a) to two analogous dAsCas12a constructs (AsBE1.1 and 1.4). Consistent with previous reports27, we observed minimal (<2%) C-to-T editing with AsBEs across all C5 for seven of eight sites in human cells with a maximum of 6% editing on the eighth site (Fig. 3h and Supplementary Table 3). However, enAsBE.fusions exhibited substantially improved C-to-T editing across the same eight sites (range of 2–34% editing; Fig. 3h and Supplementary Fig. 6b). Assessment of two analogous dLbCas12a base editors (LbBE1.1 and 1.4) revealed levels of C-to-T editing comparable to those of enAsBEs (range of 2–19% C-to-T editing; Fig. 3h,i). Similar to previous studies of SpCas9-BEs, GC motifs were edited less efficiently by Cas12a-BEs than AC, CC or TC motifs (Supplementary Fig. 6c). Additionally, C-to-T conversion was the predominant edit outcome with enAsBEs and LbBEs (Supplementary Fig. 6d), while levels of insertion and deletion
mutations (indels) for Cas12a-BEs were low presumably due to DNase inactivation of the nuclease (Supplementary Fig. 6c). To attempt to improve the specificity of enAsCas12a, we employed a strategy that we and others previously used to engineer high-fidelity variants of SpCas9 (refs. 35–37). Using structure-guided design, we created a series of AsCas12a and enAsCas12a variants with substitutions in amino acid residues expected to make non-specific contacts to DNA (see Supplementary Results and Supplementary Fig. 8a,b). Among the variants we examined, we found that enAsCas12a-N282A exhibited the greatest improvement in single mismatch intolerance while retaining on-target activity similar to enAsCas12a (Supplementary Fig. 8b). Comparison of the two nucleases using the PAMDA revealed nearly identical PAM preference profiles (Supplementary Fig. 8c–e), suggesting that the N282A substitution into enAsCas12a reduced both the number of off-target sites and the magnitude of GUIDE-seq read counts at off-target sites for three of the four crRNAs (Fig. 4b and Supplementary Fig. 7e). Additional GUIDE-seq experiments using 10 crRNAs targeted to sites with non-canonical PAMs again revealed that enAsCas12a-N282A reduced the number of off-target sites and GUIDE-seq read counts compared to enAsCas12a (Fig. 4b and Supplementary Fig. 7f), Based on these results, we conclude that the N282A substitution can be combined with enAsCas12a to generate an enhanced high-fidelity AsCas12a variant, which we refer to as enAsCas12a-HF1.

To more thoroughly determine if the N282A substitution impacts on-target activity, we compared enAsCas12a and enAsCas12a-HF1 using several methods. We first performed in vitro cleavage assays to assess temperature tolerance, which revealed similar cleavage profiles among enAsCas12a, enAsCas12a-HF1 and LbCas12a at 37, 32 and 25 °C (Supplementary Fig. 9a). Next, we compared the on-target activities of both AsCas12a variants when delivered by plasmid DNase inactivation of the nuclease (Supplementary Fig. 6e). To more thoroughly determine if the N282A substitution impacts on-target activity, we compared enAsCas12a and enAsCas12a-HF1 using several methods. We first performed in vitro cleavage assays to assess temperature tolerance, which revealed similar cleavage profiles among enAsCas12a, enAsCas12a-HF1 and LbCas12a at 37, 32 and 25 °C (Supplementary Fig. 9a). Next, we compared the on-target activities of both AsCas12a variants when delivered by plasmid electroporation into human U2OS cells on sites with canonical and non-canonical PAMs (Supplementary Fig. 9b,c, respectively). These experiments revealed similar on-target activities for enAsCas12a and enAsCas12a-HF1 across six sites with TTTT PAMs, where the black line represents the mean of 21 sites (see also Supplementary Fig. 4a); NS, P > 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001 (Wilcoxon signed-rank, two-tailed; P values in Supplementary Table 8). b, Quantification of time-course in vitro cleavage reactions of Cas12a orthologs with LbCas12a and the same crRNAs10, or contained mismatches in target sites and the magnitude of GUIDE-seq read counts at off-target sites for three of the four crRNAs (Fig. 4a and Supplementary Fig. 7e). Additional GUIDE-seq experiments using 10 crRNAs targeted to sites with non-canonical PAMs again revealed that enAsCas12a-N282A reduced the number of off-target sites and GUIDE-seq read counts compared to enAsCas12a (Fig. 4b and Supplementary Fig. 7f). Based on these results, we conclude that the N282A substitution can be combined with enAsCas12a to generate an enhanced high-fidelity AsCas12a variant, which we refer to as enAsCas12a-HF1.

To more thoroughly determine if the N282A substitution impacts on-target activity, we compared enAsCas12a and enAsCas12a-HF1 using several methods. We first performed in vitro cleavage assays to assess temperature tolerance, which revealed similar cleavage profiles among enAsCas12a, enAsCas12a-HF1 and LbCas12a at 37, 32 and 25 °C (Supplementary Fig. 9a). Next, we compared the on-target activities of both AsCas12a variants when delivered by plasmid electroporation into human U2OS cells on sites with canonical and non-canonical PAMs (Supplementary Fig. 9b,c, respectively). These experiments revealed similar on-target activities for enAsCas12a and enAsCas12a-HF1 across six sites with TTTT PAMs, where both variants again exhibited approximately twofold improved editing compared to wild-type AsCas12a (Fig. 4c). Additionally, we observed comparable activities of enAsCas12a and enAsCas12a-HF1 on 17 target sites with various non-canonical PAMs (Fig. 4d). Taken together, these results show that the introduction of N282A does not abrogate the improved targeting range, temperature tolerance or higher gene editing activities of enAsCas12a-HF1.

Because the delivery of nucleases as RNP complexes offers advantages for research use and potentially therapeutic applications, we assessed whether our enhanced variants could be transfigured as
enAsCas12a-HF1 targeted to sites with canonical and non-canonical PAMs (see Supplementary Results and Supplementary Fig. 9d,e). We then examined the activities of the same RNP complexes when delivered to primary human T cells (Supplementary Fig. 9f) and found that enAsCas12a and enAsCas12a-HF1 both showed
Characterization and improvement of AsCas12a specificity and activity. **a,b**, Histograms illustrating the number of GUIDE-seq detected off-target sites for AsCas12a variants on sites with canonical TTTV PAMs (a; see Supplementary Fig. 7e) or non-canonical PAMs (b; see Supplementary Fig. 7f). **c–e**, Summary of on-target activities of wild-type, enAsCas12a and enAsCas12a-HF1 across sites encoding TTTV PAMs (c; n = 6) or enAsCas12a and enAsCas12a-HF1 on non-canonical PAMs (d; n = 17) (see Supplementary Fig. 9b,c, respectively). **e**, Assessment of the gene editing activities of AsCas12a, enAsCas12a and enAsCas12a-HF1 on-target sites harboring TTTV PAMs or non-canonical PAMs (n = 5 and 6, respectively) in primary human T cells when delivered as RNPs (see Supplementary Fig. 9f). For c–e, percentage modified assessed by T7E1 assay, with the mean shown by black bars and where each point is the mean of three independent experiments (see Supplementary Fig. 9b,c, respectively). na, not assessed.

The enhanced AsCas12a variants described herein substantially improve the targeting range, on-target activities and fidelity of Cas12a nucleases, which are properties that are important for multiplex gene editing, epigenetic editing, cytosine base editing and gene knockout in primary human T cells. Our in vitro PAMDA and human cell experiments suggest that enAsCas12a can target approximately 1 in every 6 bp of DNA, a roughly sevenfold improvement compared to most Cas12a orthologs. enAsCas12a also exhibits superior on-target activity relative to wild-type AsCas12a, increasing editing efficiencies by approximately twofold on sites with canonical TTTV PAMs in two cell lines and in primary human T cells. The enhanced enRVR and enRR variants also show improved activities compared to their parental variants. Our results provide an important proof-of-concept that the on-target potency of CRISPR enzymes can be augmented through engineering, a strategy that may be extensible to other CRISPR nucleases. Future structural studies will be helpful to characterize the roles of the substitutions in our AsCas12a variants (see Supplementary Discussion), and additional work may be required to determine whether the potency of enAsCas12a and enAsCas12a-HF1 RNPs are sufficient for therapeutic applications. In sum, the superior properties of the enhanced Cas12a enzymes described herein enable a wide range of applications that should encourage more widespread adoption of this class of nucleases.

Online content
Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability, and associated accession codes are available at https://doi.org/10.1038/s41587-018-0011-0.
References

1. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. Cell 163, 759–771 (2015).
2. Kim, H. K. et al. In vivo high-throughput profiling of CRISPR-Cpf1 activity. Nat. Methods 14, 153–159 (2017).
3. Gao, L. et al. Engineered Cpf1 variants with altered PAM specificities. Nat. Biotechnol. 35, 789–792 (2017).
4. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. Cell 157, 1262–1278 (2014).
5. Dominguez, A. A., Lim, W. A. & Qi, L. S. Beyond editing: repurposing CRISPR-Cas9 for precision genome regulation and interrogation. Nat. Rev. Mol. Cell Biol. 17, 5–15 (2016).
6. Wright, A. V., & Nuñez, J. K. & Doudna, J. A. Biology and applications of CRISPR systems: harnessing nature’s toolbox for genome engineering. Cell 164, 29–44 (2016).
7. Komor, A. C., Badran, A. H. & Liu, D. R. CRISPR-based technologies for the manipulation of eukaryotic genomes. Cell 168, 20–36 (2017).
8. Shmakov, S. et al. Discovery and functional characterization of diverse class 2 CRISPR-Cas systems. Mol. Cell 60, 385–397 (2015).
9. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR-Cas systems. Nat. Rev. Microbiol. 15, 169–182 (2017).
10. Kleinstiver, B. P. et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat. Biotechnol. 34, 869–874 (2016).
11. Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A. & Charpentier, E. The CRISPR-Cas9 system employs a novel mechanism of DNA cleavage. eLife 6, e26863 (2017).
12. Zetsche, B. et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat. Biotechnol. 35, 31–34 (2017).
13. Tak, Y. E. et al. Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors. Nat. Methods 14, 1163–1166 (2017).
14. Zhang, G., Wang, H., Li, Y., Tran, M. H. & Farzaneh, M. CRISPR nucleases excise CRISPR RNAs from mRNA transscripts in mammalian cells. Nat. Chem. Biol. 13, 839–841 (2017).
15. Chow, R. D., Wang, G., Codina, A., Ye, L. & Chen, S. Mapping in vivo genetic interactions through Cpf1 crRNA array screening. Preprint at https://www.biorxiv.org/content/early/2017/06/21/153486 (2017).
16. Kim, D. et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. Nat. Biotechnol. 34, 863–868 (2016).
17. Yamano, T. et al. Crystal structure of Cpf1 in complex with guide RNA and target DNA. Cell 165, 949–962 (2016).
18. Gao, P., Yang, H., Rajashankar, K. R., Huang, Z. & Patel, D. J. Type V CRISPR-Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. Cell Res. 26, 901–913 (2016).
19. Karvevila, T. et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. Genome Biol. 16, 253 (2015).
20. Moreno-Mateos, M. A. et al. CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. Nat. Commun. 8, 2024 (2017).
21. Tang, X. et al. A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. Nat. Plants 3, 17018 (2017).
22. Kim, H. et al. CRISPR/Cpf1-mediated DNA-free plant genome editing. Nat. Commun. 8, 14406 (2017).
23. Liu, Y. et al. Engineering cell signaling using tunable CRISPR-Cpf1-based transcription factors. Nat. Commun. 8, 2095 (2017).
24. Zhang, X. et al. Gene activation in human cells using CRISPR/Cpf1-p300 and CRISPR/Cpf1-SunTag systems. Protein Cell 9, 380–383 (2018).
25. Zhang, X. et al. Multiplex gene regulation by CRISPR-dCas9p1. Cell Discov. 3, 17018 (2017).
26. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. Nat. Methods 12, 326–328 (2015).
27. Li, X. et al. Base editing with a Cpf1–citrine deaminase fusion. Nat. Biotechnol. 36, 324–327 (2018).
28. Komor, A. C., Kim, Y. B., Packer, M. S., Zuris, J. A. & Liu, D. R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature 533, 420–424 (2016).
29. Nishida, K. et al. Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science 353, aa8729 (2016).
30. Komor, A. C., Badran, A. H. & Liu, D. R. Editing the genome without double-stranded DNA breaks. ACS Chem. Biol. 13, 383–388 (2018).
Methods

Plasmids and oligonucleotides. New plasmids described in this study have been deposited with the non-profit plasmid repository Addgene (http://www.addgene.org/crispr-cas). Descriptions and sequences of plasmids can be found in Supplementary Table 5 and the Supplementary Information, respectively. The target site sequences for crRNAs and oligonucleotide sequences are available in Supplementary Tables 6 and 7, respectively. Human expression plasmids for wild-type AsCas12a, Lb Cas12a, FrCas12a and MbCas12a (SQT1659, SQT1665, AAS1472 and AAS2134, respectively) were generated by sub-cloning the nucleosome open-reading frames of the plasmids pY004 and pY010 (Addgene plasmid 47327; a gift from A. Schier) to generate BPK3079 and RTW645, respectively. All Cas12a variants, activator constructs and base-editor fusions were generated via standard molecular cloning and isolation as assembly. Human cell expression plasmids for Cas12a crRNAs were generated by annealing and ligating oligonucleotides corresponding to spacer sequence duplexes into BsmBI-digested BPK3079, BPK3082 (ref. 10), BPK4446 and BPK4449 for U6 promoter-driven transcription of As, Lb, MbCas12a, respectively. Substrate plasmids for in vitro cleavage reactions were generated by cloning target site sequences into the EcoRI and SplI sites of p11-lact-ct1. Plasmids for in vitro transcription of Cas12a crRNAs were generated by annealing strands corresponding to spacer sequence duplexes into Bsal-digested MSP3491 and MSP3495 for T7 promoter-driven transcription of As and LbCas12a crRNAs, respectively.

Cell culture conditions and isolation of primary human T cells. Human U2OS (from the American Type Culture Collection, ATCC) cells were cultured in DMEM supplemented with 1% heat-inactivated FBS (HI-FBS), 1% penicillin/streptomycin, 1% GlutaMAX, 1% non-essential amino acids, 0.05 mM 2-mercaptoethanol and 0.5% DMSO (ATCC). HEK293 cells and HEK293T cells were cultured in high glucose DMEM supplemented with 10% HI-FBS, 1% penicillin/streptomycin, 1% GlutaMAX, 1% non-essential amino acids, 0.05 mM 2-mercaptoethanol (Millipore Sigma), and 10% DMSO (ATCC). U2OS, 293, 293T and BHK cells were cultured in Eagle Medium supplemented with 10% heat-inactivated FBS (HI-FBS), 1% penicillin/streptomycin and 2 mM GlutaMAX. HEK293 cells (Invitrogen) and HEK293T cells were cultured in DMEM supplemented with 10% HI-FBS, 1% penicillin/streptomycin, 1% GlutaMAX, 1% non-essential amino acids, 0.05 mM 2-mercaptoethanol and 10% DMSO (ATCC). HEK293 cells were treated with 200 μM EDTA, 0.05% SDS, 1.4 mg ml–1 Proteinase K (New England Biolabs; NEB), and 12.5 mM DTT for 12–20 h at 55 °C. To extract gDNA, the lysate was combined with 165 μl paramagnetic beads, mixed thoroughly, incubated for 5 min, separated on a magnetic plate and washed three times with 70% EtOH, allowed to dry for 5 min and eluted in 65 μl elution buffer (1 mM Tris–HCl pH 8.0). Genomic loci were amplified by PCR with Phusion Hot Start Flex DNA Polymerase (NEB) using approximately 100 ng of gDNA as a template and the primers listed in Supplementary Table 7. Following analysis on a QIAxcel capillary electrophoresis machine (Qiagen), PCR products were purified with purification columns (Additional file 1).

For nuclease experiments, the percentage modification of endogenous human target sites was determined by T7 Endonuclease I (T7EI) assays, similar to as previously described18. The T7EI assay was selected to quantify relative differences in activities between Cas12a nucleases because it has previously been shown that the T7EI assay is effective at detecting indels greater than 1 nt (refs. 21, 22), consistent with indel profiles that are commonly observed with Cas12a nucleases12. Briefly, 200 ng of purified PCR products were denatured, annealed and digested with 10 U T7EI (NEB) at 37 °C for 25 min. Digests were purified with paramagnetic beads and analyzed using a QIAxcel to estimate target site modification.

For base editing experiments, target deep sequencing was performed essentially as previously described19. Dual-indexed Tru-seq libraries were generated from purified and pooled PCR products using a KAPA HT Library Preparation Kit (KAPA Biosystems) and sequenced on an Illumina MiSeq Sequencer. Samples were sequenced to an average of 57,833 reads (minimum of 8,534 reads) per replicate and an average of 173,499 (minimum of 70,022) per triplicate condition. Negative control transfections included Cas12a and U6-null plasmids.

Gene activation experiments. For experiments with crRNAs or sgRNAs targeting the VEGFA promoter, 1.6 × 10⁷ HEK293 cells per well were seeded in 24-well plates roughly 24 h before transfection with plasmids encoding Cas12a or Cas9 activators and pools of crRNAs or sgRNAs (750 ng and 250 ng, respectively), 1.5 μl TransIT-L1 (Mirus) and Opti-MEM to a total volume of 50 μl. The cell culture media were changed 22 h post-transfection, and aliquots of the media supernatant were collected 44 h post-transfection to determine VEGFA concentration using an Human VEGF Quantitative ELISA Kit (R&D Systems). For experiments targeting the AR, HR or NPY1R promoters, 8.6 × 10⁵ HEK293 cells per well were seeded in 12-well plates roughly 24 h before transfection with 750 ng Cas12a activator expression plasmid, 250 ng crRNA plasmid pools, 3 μl TransIT-L1 (Mirus) and 100 μl Opti-MEM. Total RNA was extracted from the transfected cells 72 h post-transfection using the NucleoSpin RNA Plus Kit (Clontech). Complementary DNA synthesis was performed using a High-Capacity Reverse Transcription Kit (Applied Biosystems) and diluted cDNA was amplified by PCR using primers specific to purified RNA, and 3 μl of 1:20 diluted cDNA was amplified by RT-qPCR using Fast SYBR Green Master Mix (Termo Fisher Scientific) and the primers listed in Supplementary Table 7. RT-qPCR reactions were performed on a LightCycler480 (Roche) with the following cycling program: initial denaturation at 95 °C for 20 s followed by 45 cycles of 95 °C for 3 s and 60 °C for 30 s. If sample amplification did not reach the detection threshold after 35 cycles, cycles to threshold (Ct) values were considered as 35 due to Ct fluctuations typical of transcripts expressed at very low levels. Gene expression levels over negative controls experiments (Cas12a activator and empty crRNA plasmids) were normalized to the expression of HPRT1.

GUIDE-seq. GUIDE-seq experiments were performed as previously described. Briefly, U2OS cells were electroporated as described above but including 100 pmol of the double-stranded oligodeoxynucleotide (dODN) GUIDE-seq tag. Restriction-fragment length polymorphism (RFLP) assays (performed as previously described) and T7EI assays (as described above) were performed to assess GUIDE-seq tag integration and on-target modification percentages, respectively. GUIDE-seq libraries were sequenced using an Illumina MiSeq sequencer and data was analyzed using GUIDE-seq v1.1 (ref. 14) with an NNNN PAM, a 75-bp window, and allowing up to nine mismatches before downstream data processing (Supplementary Table 4). High-confidence, cell-type-specific, single-nucleotide polymorphisms were identified using SAMTools.

Expression and purification of Cas12a proteins. For in vitro experiments, plasmids encoding Cas12a-NLS (nuclear localization signal)-6xHis fusion proteins were transformed into Rosetta 2 (DE3) E. coli and single colonies were cultured.
inoculated into 25 ml LB medium cultures containing 50 mg l⁻¹ kanamycin and 25 mg l⁻¹ chloramphenicol (Kan/Cm) before growth at 25 °C for 16 h. Starter cultures were then diluted 1:100 into 150 ml LB medium containing Kan/Cm and grown at 37 °C. The absorbance, A₆₀₀, reached 0.1 (equivalent to 0.010 optical density) before shaking at 37 °C for 2 h. Cell pellets from 50 ml of culture were harvested by centrifugation at 1,200 g for 15 min and suspended in 1 ml lysis buffer v1 containing 20 mM Hepes pH 7.5, 100 mM KCl, 5 mM MgCl₂, 5% glycerol, 1 mM DTT, Sigmafast protease inhibitor (Sigma-Aldrich) and 0.1% Triton X-100. The cell suspension was localized into a 1 ml AFA fiber milliTUBE ( Covaris) and was lysed using an E220Revolution focused ultrasonicator (Covaris) according to the following conditions: peak intensity power, 150 W; 200 cycles per burst; duty factor, 10%; and treatment for 20 min at 5 °C. The cell lysate was centrifuged for 20 min at 21,000 g and 4 °C, and the supernatant was mixed with an equal volume of binding buffer v1 ( lysis buffer v1 with 10% glycerol and 0.1% Triton X-100) and applied to 100 μl of HisPur Ni-NTA Resin (Thermo Fisher Scientific) that was pre-equilibrated in binding buffer v1 and rocked at 4 °C for 4 h. The protein-bound resin was washed three times with 1 ml wash buffer v2 (20 mM Hepes pH 7.5, 100 mM KCl, 5 mM MgCl₂, 10% glycerol and 0.1% imidazole) and 0.1% Triton X-100) and then once with 1 ml binding buffer v1. Three sequential elutions were performed with 500 μl elution buffer v2 (20 mM Hepes pH 7.5, 100 mM KCl, 5 mM MgCl₂, 10% glycerol and 0.1% imidazole) and visualized by SDS–PAGE and coomassie staining. Selections were pooled and dialyzed using Spectra/ Por 4 Standard Cellulose Dialysis Tubing (Spectrum Chemical Manufacturing Corp) in three sequential 1:500 buffer exchanges, the first two into dialysis buffer (300 mM NaCl, 10 mM Tris-HCl pH 7.4, 0.1 mM EDTA and 1 mM DTT), and the last into dialysis buffer containing 50% glycerol. Proteins were then concentrated with Amicon Ultra-0.5 ml Centrifugal Filter Units (Millipore Sigma), digested with an equal volume of dialysis buffer with 80% glycerol to final storage conditions of 1X dialysis buffer with 50% glycerol and stored at −20 °C.

For experiments in human cells, starter cultures were grown as described above. For all 16 h, cultures were diluted 1:5024 into 100 μl of HCT116 (Stabilomed, Germany) and were used at 80% confluence. The DNA template was degraded by the addition of 1 μg RQ1 DNase and digestion over time was then fit to an exponential decay model (y = kt + c) to represent the profile of uncleavable substrates. The depletion of each PAM untreated control) and then normalized to account for the increased fractional representation of uncleavable substrates (by selecting the five PAMs with the highest average counts across all time points to represent the profile of uncleavable substrates). The depletion of each PAM over time was then fit to an exponential decay model (y(t) = A e⁻ᵏᵗ), where y(t) is the normalized PAM count, t is the time (minutes), k is the rate constant and A is a constant), by linear least-squares regression.

Targeting range calculations. The targeting ranges of wild-type and variant AsCas12a nucleases were assessed on various annotated genomic elements using GENCODE's Release 27 GTF file. Complete occurrences of targetable 4-nt PAMs were enumerated within regions encompassing 1 kilobase (kb) upstream of all transcription start sites (TSSs), within the first exon of all genes, and within all annotated micro RNAs (miRNAs). Parameter value(s) for each element in the GTF file were: exon1, feature-type exon, exon_number 1, gene_type protein_coding; TSS, feature-type transcript, gene_type gene, gene_type miRNA. For each element, PAM counts were normalized by length and were visualized through a boxplot. The PAM identification and enumeration script will be made available upon request. Targetable PAMs for Cas12a nucleases included: TTTV for wild-type AsCas12a; TTYN, RTTC, CTTV, TATM, TCCC, TACA (tier 1), and RTTS, TATA, TGTG, ANCC, CVCC, TGCC, GTCC, TTAG, TTAG (tier 2) PAMs for enAsCas12a (see Fig. 1c and Supplementary Fig. 3b); TATV for AsCas12a-RVR; and TYCV for AsCas12a-RIV; and TTYV for AsCas12a-RV (ref. 11).

Statistics. Statistical significance between data sets was calculated using Wilcoxon signed-rank or Mann–Whitney tests using GraphPad Prism version 7.0c (see the results of Tests in Supplementary Table 8). A multiple comparisons adjustment was performed using the Bonferroni correction (34 total tests; P < 0.00147). P values are reported using GraphPad style: not significant (NS), P > 0.05; *P < 0.05; **P < 0.01; ***P < 0.001; ****P < 0.0001.

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

Code availability. The custom Python script for PAMaDA data analysis and the PAM identification and enumeration script will be made available upon request.

Data availability. Data sets from GUIDE-seq and high-throughput sequencing experiments (for PAMaDA and base editing experiments) have been deposited with the National Center for Biotechnology Information Sequence Read Archive under BioProject ID PRJNA508751.

References.

39. Rohland, N. & Reich, D. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22, 939–946 (2012).
40. Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. Nat. Biotechnol. 30, 460–465 (2012).
41. Vouillot, L., Thélise, A. & Pollet, N. Comparison of T7E1 and surveyor mismatch cleavage assays to detect mutations triggered by engineered nucleases. G3 (Bethesda) 5, 407–415 (2015).
42. Sentmanat, M. E., Peters, S. T., Florian, C. P., Connelly, J. P. & Pruett-Miller, S. M. A survey of validation strategies for CRISPR–Cas9 editing. Sci. Rep. 8, 888 (2018).
43. Kim, H. K. et al. Deep learning improves prediction of CRISPR–Cpf1 guide RNA activity. Nat. Biotechnol. 36, 239–241 (2018).
44. Clement, K. et al. Analysis and comparison of genome editing using CRISPResso2. Preprint at https://www.biorxiv.org/content/early/2018/08/15/392217 (2018).
45. Kleinstiver, B. P. et al. Engineered CRISPR–Cas9 nucleases with altered PAM specificities. Nature 523, 481–485 (2015).
46. Tsai, S. Q., Topkar, V. V., Joung, J. K. & Aryee, M. J. Open-source guideseq software for analysis of GUIDE-seq data. Nat. Biotechnol. 34, 483 (2016).
47. Studier, F. W. Stable expression clones and auto-induction for protein production in E. coli. Methods Mol. Biol. 1091, 17–32 (2014).
Reporting Summary

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

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
- [ ] An indication of 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 statistics 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
- [ ] Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI)

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection

Nucleic acids samples were visualized using Qiaxcel ScreenGel Software (v1.4 and v1.5). Sequencing data was acquired using an Illumina MiSeq machine.

Data analysis

Nucleic acids samples were analyzed using Qiaxcel ScreenGel Software (v1.4 and v1.5). Data visualization was performed using GraphPad Prism 7.0c. GUIDE-seq data was analyzed with guideseq software v1.1 (https://github.com/aryeelab/guideseq), and amplicon sequencing data was analyzed with CRISPResso (https://github.com/lucapinello/CRISPResso). Custom scripts to analyze PAM depletion data and perform targeting range calculations are available upon request.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. 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

Data sets from GUIDE-seq and high-throughput sequencing experiments have been deposited with the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) under BioProject ID PRJNA508751 (http://www.ncbi.nlm.nih.gov/bioproject/508751). Otherwise, data sets generated and analyzed as part of this study will be available upon request upon publication.

Field-specific reporting
Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences
- Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

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

| Sample size | No sample size calculation was performed. Final data sets are comprised of at least triplicate independent samples for all conditions for each data set. |
|--------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Data exclusions | No data were excluded.                                                                                                                        |
| Replication  | Experimental data are expressed as the mean and s.e.m. of at least three independent replicates, with replicate experiments performed on separate days when possible. |
| Randomization | No randomization was performed in our study as appropriate control samples were used to establish baseline/reference activities.               |
| Blinding     | Blinding was not performed during data analysis. Raw data not already included in the manuscript will be made available upon request.            |

Materials & experimental systems

Policy information about availability of materials

- n/a Involved in the study
- ☒ Unique materials
- ☒ Antibodies
- ☒ Eukaryotic cell lines
- ☒ Research animals
- ☒ Human research participants

Eukaryotic cell lines

Policy information about cell lines

| Cell line source(s) | Human U2OS (from Toni Cathomen, Freiburg), HEK293 cells (Invitrogen), HEK293T cells (ATCC), and primary human T cells (MGH Blood Transfusion Service) |
|---------------------|---------------------------------------------------------------------------------------------------------------------------------------------|
| Authentication      | Cell line identities were confirmed by STR profiling (ATCC)                                                                                   |
| Mycoplasma contamination | Media supernatant was analyzed biweekly for the presence of Mycoplasma and all tests were negative for contamination.                        |
| Commonly misidentified lines (See ICLAC register) | No commonly misidentified cell lines were used in this study.                                                                              |
Method-specific reporting

- ChIP-seq
- Flow cytometry
- Magnetic resonance imaging

n/a

Involved in the study