5624–5638 *Nucleic Acids Research, 2020, Vol. 48, No. 10* *Published online 24 April 2020*
doi: 10.1093/nar/gkaa272

Characterization of Cas12a nucleases reveals diverse PAM profiles between closely-related orthologs

Thomas Jacobsen[1], Fani Ttofali[1], Chunyu Liao[2], Srinivas Manchalu[2], Benjamin N. Gray[3] and Chase L. Beisel[1,2,4,*]

[1]Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC 27695, USA, [2]Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz-Centre for Infection Research (HZI), 97080 Würzburg, Germany, [3]Benson Hill, St. Louis, MO, 63132, USA and [4]Medical Faculty, University of Würzburg, 97080 Würzburg, Germany

Received August 14, 2019; Revised April 05, 2020; Editorial Decision April 06, 2020; Accepted April 07, 2020

ABSTRACT

CRISPR-Cas systems comprise diverse adaptive immune systems in prokaryotes whose RNA-directed nucleases have been co-opted for various technologies. Recent efforts have focused on expanding the number of known CRISPR-Cas subtypes to identify nucleases with novel properties. However, the functional diversity of nucleases within each subtype remains poorly explored. Here, we used cell-free transcription-translation systems and human cells to characterize six Cas12a single-effector nucleases from the V-A subtype, including nucleases sharing high sequence identity. While these nucleases readily utilized each other's guide RNAs, they exhibited distinct PAM profiles and apparent targeting activities that did not track based on phylogeny. In particular, two Cas12a nucleases encoded by *Prevotella ihumii* (PiCas12a) and *Prevotella disiens* (PdCas12a) shared over 95% amino-acid identity yet recognized distinct PAM profiles, with PiCas12a but not PdCas12a accommodating multiple G's in PAM positions -2 through -4 and T in position -1. Mutational analyses transitioning PiCas12a to PdCas12a resulted in PAM profiles distinct from either nuclease, allowing more flexible editing in human cells. Cas12a nucleases therefore can exhibit widely varying properties between otherwise related orthologs, suggesting selective pressure to diversify PAM recognition and supporting expansion of the CRISPR toolbox through ortholog mining and PAM engineering.

INTRODUCTION

Clustered regularly interspaced short palindromic repeats (CRISPR) and their CRISPR-associated (Cas) proteins comprise adaptive immune systems that protect bacteria and archaea from invading plasmids and bacteriophages (1–3). These systems rely on effector nucleases that are directed by CRISPR-encoded guide RNAs (gRNAs) to bind and cleave complementary nucleic acids often flanked by a short protospacer-adjacent motif (PAM) (4,5). The programmable nature of these nucleases lent to their direct use for genome-editing, gene regulation, and various other applications (6). These applications have been spurred in part by the ongoing discovery of Cas nucleases with distinct properties such as DNA or RNA targeting, varying recognized PAM profiles, different optimal temperatures, and reduced propensity for off-targeting (7–14).

The available set of Cas nucleases are part of a remarkably diverse assortment of CRISPR-Cas systems encompassing various proteins, mechanisms, and functions. This diversity is hypothesized to have emerged from the ongoing arms race between bacteria and invasive genetic elements such as phages (15,16). Attempts to capture this diversity are now reflected in a hierarchical classification scheme that groups systems into two classes, six types, and over 30 subtypes (7,8). Ongoing bioinformatics and biochemical characterizations have mainly focused on expanding the list of subtypes, with recent reports expanding Type V systems to nine subtypes and Type VI systems to five subtypes (7,17–19). However, emerging evidence suggests that incredible diversity lies within each subtype. For instance, characterization of ranging single-effector Cas9 nucleases within the Type II-A subtype have shown that these nucleases not only share limited sequence identity but also can recognize distinct PAM profiles, exhibit ranging propensities to accept mismatches between the guide and target, and do not recognize each other's processed crRNA:tracrRNA duplexes serving as the gRNAs (20–23). While these distinctions are normally observed for phylogenetically distinct nucleases, little is known about functional differences separating otherwise closely related nucleases.

*To whom correspondence should be addressed. Email: cbeisel@ncsu.edu

© The Author(s) 2020. Published by Oxford University Press on behalf of Nucleic Acids Research.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

A unique opportunity to explore the functional diversity between related Cas nucleases rests within the V-A subtype of CRISPR-Cas systems (24). This subtype is exemplified by Cas12a (also known as Cpf1) nucleases that exhibit unique properties compared to other known Cas nucleases. Specifically, these nucleases process gRNAs from a transcribed CRISPR array lacking accessory factors (e.g. tracrRNA), recognize T-rich PAMs located 5′ of the displaced strand of target DNA, utilize a RuvC endonucleolytic domain to nick both strands of target DNA, and can non-specifically cleave single-stranded DNA upon target recognition (24–26). In turn, these capabilities have led Cas12a to be harnessed for numerous applications in genome-editing, gene regulation, and nucleic acid sensing (24,27,28). Ongoing characterization of Cas12a nucleases has also revealed variability among these V-A effectors, such as the inability to use each other's gRNAs, a propensity to recognize a C or G at various PAM positions, and different temperature ranges in which these nucleases are active (11,14,24,29,30). Many of these characterization efforts have focused on sets of phylogenetically distinct Cas12a nucleases, with the assumption that phylogenetically similar nucleases exhibit similar properties.

Here, we characterized a set of six Cas12a nucleases, including nucleases exhibiting highly similar identity to each other or with well-established nucleases. We found that the nucleases were able to process and utilize each other's gRNAs for DNA targeting, although they diverged in their apparent DNA cleavage activities and PAM preferences independently of their phylogenic relationship. Further investigation of two Cas12a nucleases from *Prevotella ihumii* (PiCas12a) and *Prevotella disiens* (PdCas12a) revealed surprisingly different PAM profiles despite sharing 95.7% identity. Furthermore, mutating PiCas12a toward PdCas12a revealed PAM profiles distinct from those associated with either PiCas12a or PdCas12a. These findings demonstrate that otherwise closely related CRISPR nucleases and the intervening mutants can exhibit divergent properties, with implications for the evolution of CRISPR-Cas systems and expanding the set of nucleases available for CRISPR technologies.

METHODS

Strains, plasmids, oligonucleotides, and gBlocks

All strains, plasmids, oligos, and gBlocks used in this work can be found in Supplementary Table S1. All PCR amplifications were performed using the Q5 Hot Start High-Fidelity 2X Master Mix (NEB, Cat: M0494S).

The pET28b+ plasmids expressing the Cas12a nucleases were synthesized by GenScript. The PdCas12a bacterial expression plasmid was constructed by PCR-amplifying the PdCas12a from a previous construct (Addgene, Cat: 69990) and inserting it into the BamHI and NheI sites of the plasmid CB1095. All gRNAs used in this work were synthetized from IDT as custom gBlocks, which contain the constitutive J23119 promoter and a rho-independent terminator. The Q5 mutagenesis kit (NEB, Cat: E0554S) was used to generate the PiCas12a variants, the catalytically-dead Cas12a nucleases, and the GFP reporter plasmids following the manufacturer's instructions.

To construct the mammalian-expressed HkCas12a and PiCas12a plasmids, the nucleases and an N-terminal nuclear localization tag were PCR-amplified from their respective pET28b+ plasmids and inserted into the plasmid CB1067 using NEBuilder HiFi DNA Assembly Master Mix (NEB, Cat: E2621L). The empty gRNA expression plasmids were constructed using a previously built plasmid expressing an empty *Acidaminococcus* sp. BV3L6 Cas12a gRNA (Addgene, Cat: 78956). We used the Q5 mutagenesis kit to convert the direct repeat of AsCas12a to that of Hk and PiCas12a. The plasmids containing the empty Hk and PiCas12a gRNAs were then digested with BsmBI (NEB, Cat: R0580S) and ligated with 5′ phosphorylated and annealed oligos containing a target sequence-of-interest.

GFP reporter assay using a cell-free transcription-translation (TXTL) system

The materials and methods to perform this assay was described in detail elsewhere (31,32). Briefly, we used the commercially available cell-free TXTL system developed from an all-*Escherichia coli* lysate (Arbor Biosciences, Cat: 507096) (33) to rapidly express Cas12a from a plasmid and targeting or non-targeting gRNAs from custom gBlocks. For this assay, we used GFP reporter plasmids containing a target sequence flanked by potential PAM sequences. The target site was located immediately upstream of the −35 element of the P70a promoter driving GFP expression. GFP fluorescence was measured with a Synergy H1 plate reader (BioTek) using excitation and emission wavelengths of 488 nM and 553 nM, respectively. The reactions were incubated for 16 h at 29°C and the resulting fluorescence data were analyzed using end-point and time-course analyses. The reported production of GFP was calculated using a linear standard calibration curve developed from recombinant GFP as we have performed previously (29,32). For the plate reader used for our experiments, the raw fluorescence values were divided by the conversion factor 9212.6 l/μmol.

crRNA generation in TXTL and Northern blotting analysis

The crRNA was generated by adding 2.56 nM plasmid encoding Cas12a, 0.21nM plasmid encoding T7 RNA polymerase, and 6.25 μM GamS protein in TXTL. After incubation for 2 h at 29°C, 2 μl of a gBlock encoding the repeat-spacer was spiked in the TXTL mix. After incubation for an additional 14 h at 29°C, total RNA was extracted using the Direct-zol RNA MiniPrep kit (Zymo Research, Cat: R2051) and treated with Turbo DNase (Life Technologies, AM2238). Northern blotting analysis was conducted as previously described (34). As part of the blotting analysis, the oligodeoxyribonucleotide CLTJNB (ACCGCGAAAG GTTTTGCACTCGAC) was end-labeled with γ-^{32}P-ATP and used for hybridization.

TXTL-based PAM screen

For PAM determination, we utilized our previously developed TXTL-based PAM screen, which had been used to successfully recapitulate the known PAMs for NmCas9, FnCas12a, and AsCas12a, and determine the PAMs for previously unreported Cas12a nucleases (29,31,32). We first

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

constructed a plasmid containing a 5N-randomized PAM library flanked by a sequence targeted by the targeting gRNA using methods described previously (31,32). Briefly, the PAM library was generated by PCR-amplifying CB847 and gBlock TJ460 with primer pairs CSMpr1308/1309 and CSMpr1310/1311, respectively, followed by assembling the two amplicons with NEBuilder HiFi DNA Assembly Master Mix. Following generation of the PAM library, we performed a high-throughput PAM determination screen (31,32) using the same protocol as the GFP reporter assay in TXTL, though the GFP reporter plasmid was replaced with the PAM library plasmid. After incubating the samples for 16 h at 29°C, the uncleaved 5N-randomized PAM library from the targeting and non-targeting reactions for each Cas12a were PCR-amplified and prepared for next-generation sequencing (NGS). The NGS data, including the raw data and post-processing reads, were deposited in the NCBI gene expression omnibus (accession #GSE130377). The code used for the processing and analyses of the NGS data can be found in the following public repository: https://bitbucket.org/csmaxwell/crispr-txtl-pam-counting-script/. The methods used to generate the PAM wheels are described in detail elsewhere (35). The resulting Krona plots used to generate the PAM wheels can be found in Data S1.

Indel formation in HEK293T cells

The target sites and their PAM sequences in the *DNMT1* gene in HEK293T cells can be found in Supplementary Table S2. The indel formation assays were conducted as described previously (29). Briefly, 2×10^5 HEK293T cells were seeded in 12-well plates 24h prior to performing transient transfections. For each reaction, we transiently transfected 160 ng of the gRNA and 640 ng of the Cas12a plasmid using jetPRIME (Polypus Transfection, Cat: 114-07). After a 20-h incubation at 37°C, the growth media was replaced with fresh media in each well. The cells were then incubated for an additional 52 h at 37°C prior to the isolation of genomic DNA.

Tracking of indels by decomposition (TIDE) analysis

Genomic DNA from transfected HEK293T cells were isolated using the GeneJET Genomic DNA Purification Kit (ThermoFisher Scientific, Cat: K0721) following the manufacturer's instructions. The genomic DNA was PCR-amplified using primer pairs TJ719/720 or TJ699/722. After validating amplification on a 1% agarose gel, the amplicon was prepared for Sanger sequencing with the primer closest to the expected cleavage site. The chromatograms obtained from each sequencing reaction were analyzed using TIDE analysis (36). Each gRNA tested was analyzed against a non-PAM (GAAAT) negative control targeting a site in the *DNMT1* gene (Supplementary Table S2).

Statistical analyses

For a subset of the data, we performed statistical analyses to determine statistical significance between multiple datasets. The end-point fluorescence measurements were inputted in MATLAB's ttest2 function assuming unequal variances, which used a two-tailed *t*-test and a 95% confidence interval to discern statistical differences between the two samples tested.

RESULTS

A phylogenetically diverse set of Cas12a nucleases exhibit ranging effective activities in TXTL

A phylogenetically diverse set of Cas12a nucleases was selected from publicly available sequences for characterization. The ensuing set came from Firmicutes bacterium ADurb.Bin193 (Adurb193Cas12a), archaeon ADurb.Bin336 (Adurb336Cas12a), *Francisella novicida* (Fn3Cas12a), *Prevotella ihumii* (PiCas12a), *Prevotella disiens* (PdCas12a) and *Helcococcus kunzii* ATCC 51366 (HkCas12a). These nucleases tended to share low identity with each other at the amino-acid level, with the exception of the PiCas12a and PdCas12a nucleases, which shared ∼95.7% amino-acid identity. Additionally, the Fn3Cas12a nuclease characterized here shares 91.4% amino acid identity with the well-characterized *Francisella novicida* U112 Cas12a (FnCas12a) (24). The other tested nucleases showed approximately 30 to 40% amino-acid identity with each other and with the well-characterized Cas12a nucleases from FnCas12a, *Lachnospiraceae bacterium* ND2006 (LbCas12a), and *Acidaminococcus* sp. BV3L6 (AsCas12a) in pairwise comparisons (Figure 1A). Nonetheless, the direct repeat sequences associated with these Cas12a nucleases were relatively well-conserved, consistent with previously characterized Cas12a nucleases (Figure 1B) (24,37).

We first assessed the ability of the six tested nucleases to target DNA using an all-*E. coli* cell-free TXTL assay we previously employed to rapidly characterize CRISPR-Cas systems (31,32). As part of the assay, DNA constructs each encoding the nuclease, a targeting or non-targeting gRNA, and a targeted GFP reporter were added to the TXTL mix, and GFP fluorescence was monitored over time using a plate reader (Figure 1C). Loss of GFP production reflects expression of the CRISPR components, formation of the active nuclease:gRNA complex, cleavage of the GFP reporter plasmid, and subsequent plateauing of GFP levels. For these experiments, each Cas12a was expressed with a gRNA with a processed repeat, a 24-nt guide, and a downstream rho-independent terminator. The target sequence was flanked by a 5′ TTTC PAM recognized by all known Cas12a nucleases (24,37). The resulting fluorescence time-courses from the assays confirmed that all nucleases could actively target the DNA encoding the GFP reporter (Figure 1D). We do note variability in the time to cessation of GFP expression—a reflection of the time to express the nuclease, form the ribonucleoprotein complex, and finally bind and cleave the target DNA—with the fold-reduction in GFP levels (comparing targeting and non-targeting gRNAs) after a 16 h reaction at 29°C ranging between 1.9-fold (for Adurb336Cas12a) and 4.5-fold (for PiCas12a). Given the challenge of distinguishing these factors as part of TXTL, we refer to the 'apparent targeting activity' when making comparisons.

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

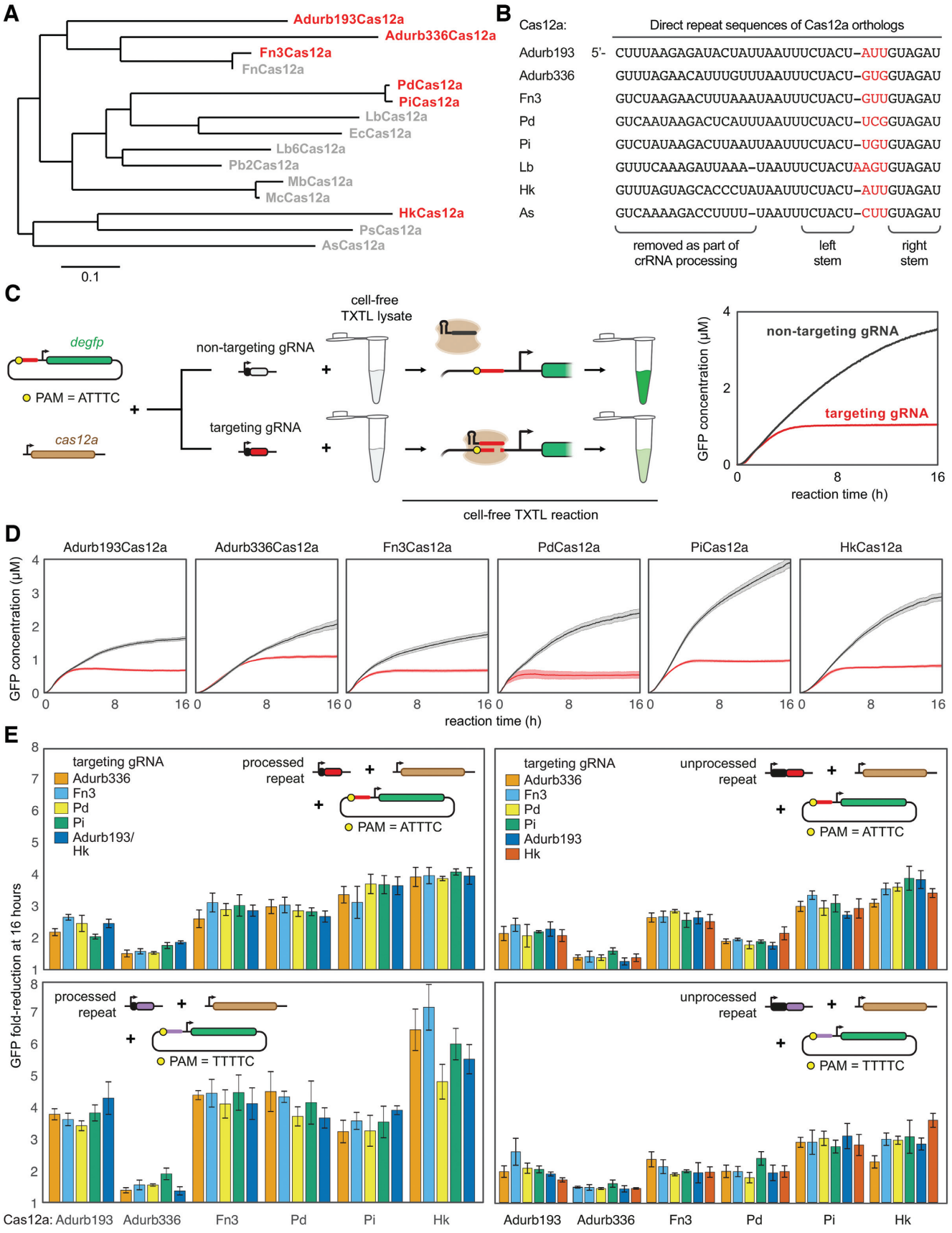


Figure 1. Phylogenetically diverse Cas12a nucleases exhibit varying apparent targeting activities and can process and utilize each other's gRNAs. (**A**) Phylogenetic tree of the Cas12a nucleases analyzed in this work (red) and other Cas12a nucleases that are commonly used or were reported in our previous

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

Each Cas12a nuclease can process and utilize the gRNA for the other nucleases

Due to the phylogenetic diversity exhibited between these Cas12a nucleases, we asked whether each nuclease could utilize each other's gRNAs. Similar to prior work on characterizing Cas12a nucleases (24,37), the gRNAs associated with the Cas12a nucleases reported here share the same left and right stem sequences, with variability in the sequences removed following gRNA maturation (Figure 1B). The length and sequence of the connecting loop between the two stems also varied, though the total length of the full repeat was well-conserved (35–36 nts) across these nucleases.

Cas12a binds a canonical hairpin formed near the 3′ end of the repeat and cleaves upstream of this hairpin through an endoribonuclease domain within the nuclease (Figure 1C) (25). Correspondingly, gRNAs have been expressed with the full-length repeat (∼36 nts) or a processed repeat (∼19 nts) (24,25). We therefore used two sets of gRNAs as part of the GFP reporter assay: one set containing the processed versions of the repeat, and another set containing the unprocessed gRNAs. We note that the processed repeat is identical for Adurb193Cas12a and HkCas12a. To account for a potential impact of the target sequence, we targeted two different sites flanked by a 5′ TTTC PAM that were present in the GFP reporter plasmid (Figure 1E). We then performed the GFP reporter assay using each nuclease paired with each gRNA and target site. Based on the resulting fluorescence time-courses, all Cas12a nucleases tested were able to utilize each other's gRNAs (Figure 1E, Supplementary Figure S1), with only minor differences in apparent targeting activity. These findings can be explained by the repeat sequences differing only in the trimmed portion and the hairpin loop of the repeats, which were both shown to accommodate mutations without disrupting the activity of FnCas12a (24). They also match that of previous work showing that AsCas12a and LbCas12a were able to utilize gRNAs associated with phylogenetically diverse Cas12a nucleases when the sequence length of the loop region was conserved (30).

The Cas12a nucleases may be utilizing each other's gRNAs without processing the 5′ end of the repeat. To explore this possibility directly, we performed Northern blotting analysis on gRNAs with the full-length repeat expressed with each of the Cas12a nucleases in TXTL. We specifically used three different repeats exhibiting the greatest sequence diversity (Supplementary Figure S2A), and we used a probe that hybridizes to the 24-nt guide present in all gRNAs. The analysis revealed that the 5′ end of the repeat was efficiently processed off of each repeat (Supplementary Figure S2B). Interestingly, we had difficulty detecting gRNAs paired with Adurb336Cas12a, suggesting that the gRNAs were destabilized in the presence of this nuclease. Overall, we conclude that the tested Cas12a nucleases can process and utilize each other's gRNAs, indicating a commonly shared property across the nucleases.

PAM determination reveals distinct recognition profiles

We next were interested in deciphering the PAM profiles recognized by these Cas12a nucleases and how their profiles compared. To this end, we performed a TXTL-based PAM screen that we developed previously (Figure 2A) (31,32). As part of the assay, separate TXTL reactions contained DNA encoding a Cas12a nuclease, a targeting or non-targeting gRNA, and a target sequence flanked by a 5N library of potential PAM sequences. The reactions containing each Cas12a resulted in the cleavage of library members containing a recognizable PAM. After a 16 h incubation at 29°C, the uncleaved library members from the samples with the targeting or non-targeting gRNA were PCR-amplified and subjected to NGS. The NGS data were analyzed for depletion of PAM sequences in the presence of the targeting gRNA, where greater depletion yielded more defined PAM profiles. We refer to 'apparent cleavage activity' when comparing activities between nucleases, as the assay relies on DNA cleavage but would also be affected by the expression, folding, gRNA binding, and DNA target binding for each nuclease.

The PAM profiles deciphered for each Cas12a (Figure 2B, Supplementary Figure S3A) showed that all six nucleases recognized the canonical TTTV (V = A/C/G) motif commonly associated with Cas12a nucleases (24,37). However, there were notable deviations from this motif often unique to each nuclease. For instance, the PAM profiles for Adurb336Cas12a, PdCas12a and Adurb193Cas12a all closely resembled those for LbCas12a and AsCas12a, with the consensus TTTV and the ability to accommodate a C at the −2 position. Separately, the PAM profile for Fn3Cas12a closely resembled that of its close ortholog FnCas12a, with the consensus YTV (Y = C/T) and no strong bias at the -4 position (24). We do note that some T-rich sequences (e.g. TTTC) were less prominent in the PAM wheel for Fn3Cas12a, although this can be attributed to the absence of specific sequences within the tested PAM library (Supplementary Table S3). Finally, the PAM profiles for HkCas12a and PiCas12a recognized PAM profiles distinct from the other nucleases. HkCas12a recognized motifs with a mix-

work (32) (gray). (**B**) Direct repeat sequences of the Cas12a nucleases represented in the phylogenetic tree. Red text indicates non-conserved bases of the gRNA. Note that Fn3Cas12a shares the same gRNA as FnCas12a. (**C**) Representative figure showing the GFP reporter assay in TXTL. A GFP reporter plasmid, Cas12a plasmid, and a targeting or non-targeting gRNA are added to a TXTL reaction. At the beginning of the reactions, each reaction expresses GFP, Cas12a, and the gRNA. The reaction containing the non-targeting gRNA will continue to express GFP as the gRNA will be unable to direct the Cas12a to cleave the GFP reporter plasmid, while the reaction containing the targeting gRNA will halt GFP production cleavage of the remaining reporter plasmid. (**D**) Time-courses for the different Cas12a nucleases in the GFP reporter assay using a targeting (red) or non-targeting (gray) gRNA. In all cases, the target was flanked by a canonical 5′ TTTC PAM and was immediately upstream of the −35 element of the promoter driving GFP expression. Each solid line and shaded region represent the mean and standard deviation from three separate TXTL reactions. (**E**) Fold-reduction in GFP levels for each Cas12a nuclease targeting two different protospacers flanked by ATTTC PAMs. For both target sequences, each Cas12a target and cleave the DNA using each other's full-length (bottom row) and mature (top row) gRNAs. Note that the Adurb193 and HkCas12a share the same processed gRNA sequence. The fold-reduction was calculated using the GFP fluorescence data from the 16 h time-point from the reactions containing the targeting non-targeting gRNA. The error bars represent the standard deviation from three separate TXTL reactions. See Supplementary Figure S1 for the associated time-courses.

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

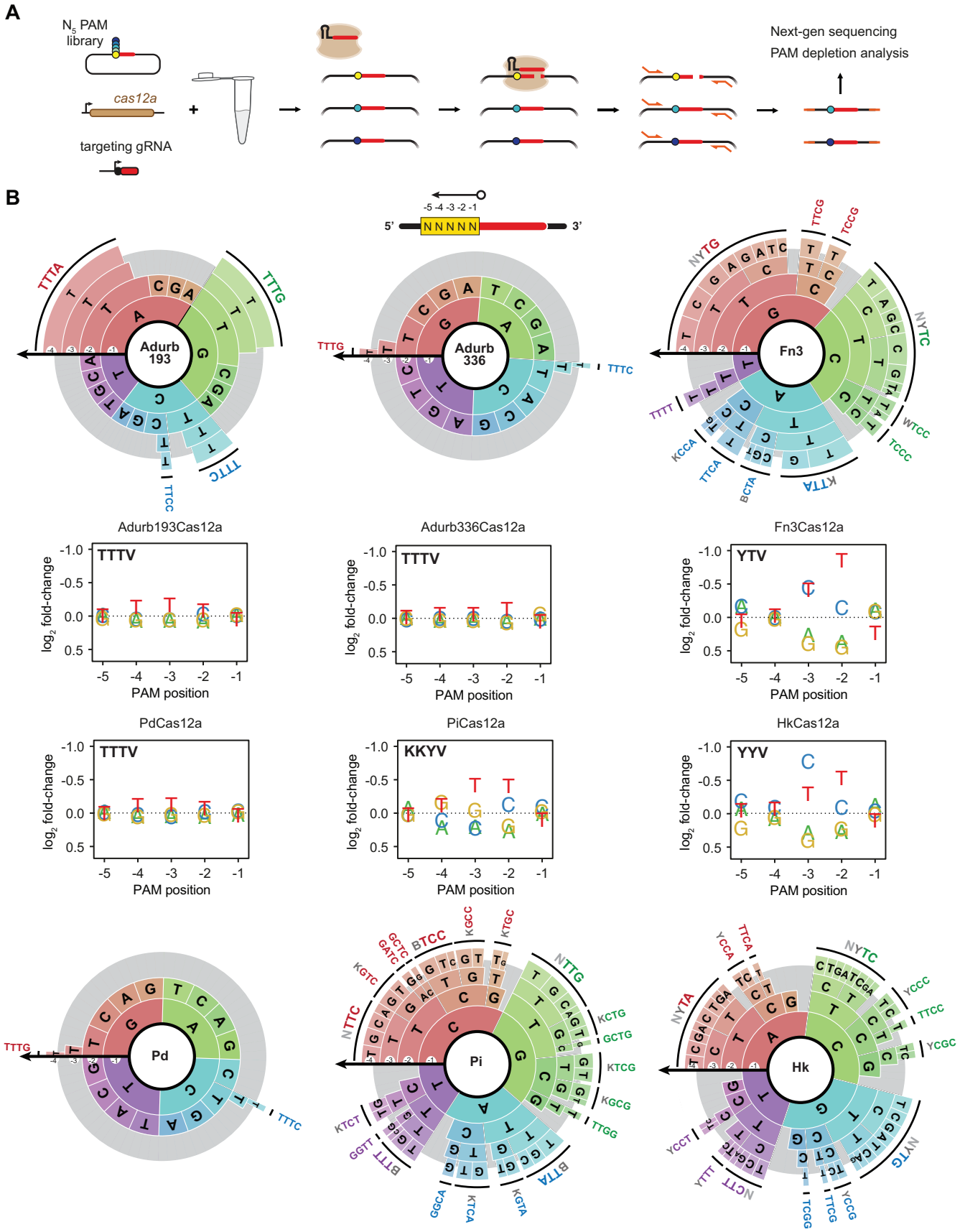


Figure 2. The characterized Cas12a nucleases are associated with various apparent cleavage activities and PAM sequences. (**A**) Representative figure showing the TXTL-based PAM screen. A 5N PAM library, Cas12a, and a targeting or non-targeting gRNA are added to a TXTL reaction. After a 16 h

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

ture of C or T, with bias toward C at the −3 position and T at the −2 position. There was also evidence that HkCas12a could weakly recognize a G at the −2 position for some sequences. Separately, PiCas12a generally recognized a KKYV (K = G/T) motif and, to a lesser degree, a TTGS (S = C/G) motif. In addition, the PAM profiles for HkCas12a and PiCas12a indicated that both nucleases could accommodate a T at the −1 position when paired with a subset of sequences (e.g. NYTT for HkCas12a, KTYT for PiCas12a). Most importantly, the different PAM profiles did not track with the phylogenic relationships between the nucleases, as underscored by the contrasting PAM profiles of PiCas12a and PdCas12a despite sharing 95.7% protein sequence identity. This insight is further supported by accounting for a larger set of Cas12a nucleases we previously subjected to the TXTL-based PAM assay (32) (Supplementary Figure S3B), revealing little correlation between PAM profiles and phylogenetic relationships.

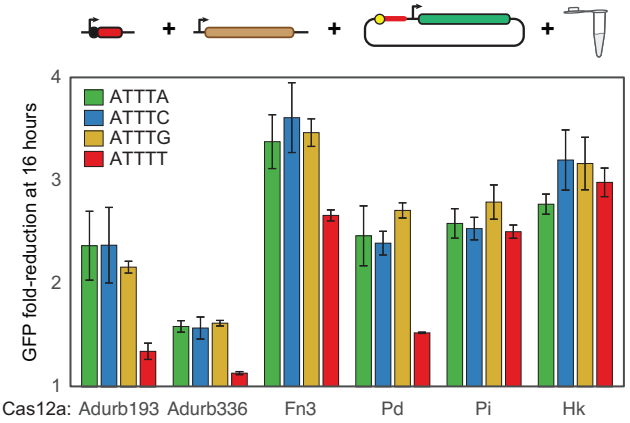


Figure 3. Some of the Cas12a nucleases can accommodate a T at the –1 position within the canonical TTTV PAM. Fold-reduction of GFP expression for each Cas12a and motif indicated. The fold-reduction was calculated using the GFP fluorescence data from the 16 h time-point from the TXTL reactions containing the targeting and non-targeting gRNA. The error bars represent the standard deviation from three separate TXTL reactions. See Supplementary Figure S5 for the associated time-courses.

TXTL results suggest PdCas12a exhibits different capacities for transcriptional silencing and DNA cleavage

While PdCas12a exhibited comparable efficiencies of GFP silencing to the other nucleases (Figure 1D), PdCas12a exhibited noticeably weak apparent cleavage activity in the PAM screen (Figure 2B) even though the target sequence was the same for both experiments. In contrast, the two other nucleases exhibiting weak apparent cleavage activity in the PAM screen (Adurb193Cas12a, Adurb336Cas12a) also exhibited similarly weak apparent targeting activity in the GFP reporter assay. One notable difference between the assays is that the target site in the GFP reporter assay is immediately upstream of the −35 element, and this proximity was previously shown to induce transcriptional silencing with a catalytically-dead nuclease (32); by contrast, the PAM screen absolutely requires DNA cleavage. To further interrogate a potential role of transcriptional silencing, we targeted the same sequence used in the PAM screen flanked by the same PAM in two different positions in the GFP reporter plasmid: the same site directly upstream of the promoter and a new site ∼100 bp upstream of the promoter (Supplementary Figure S4A). While PdCas12a was able to silence GFP when targeting a sequence directly upstream of the promoter, GFP was not efficiently silenced when PdCas12a targeted the same sequence located at a site further upstream (Supplementary Figure S4B). In contrast, targeting both locations in the GFP reporter led to comparable gene silencing with PiCas12a. We also evaluated catalytically-dead versions of PiCas12a and PdCas12a harboring mutations in the RuvC domain (D946A, E1035A, and D1279A for PiCas12a, D943A for PdCas12a), where equivalent mutations preserved potent silencing by FnCas12a (34). However, these variants did not exhibit any silencing activity when targeting immediately upstream of the promoter (Supplementary Figure S4C), potentially due to the mutations being disruptive or GFP repression occurring entirely through cleavage and benefiting from targeting a transcriptional activity region. Setting aside the inconclusive results with dPiCas12a and dPdCas12a, these data suggest that PdCas12a may more readily drive transcriptional silencing than DNA cleavage, although further investigation is needed.

TXTL confirms variable bias against T at the −1 PAM position

We next aimed to assess the variable PAM recognition across the set of Cas12a nucleases. One of the more notable insights was the lack of bias against T at the -1 position of the PAM for HkCas12a and PiCas12a for some sequences (Figure 2B, Supplementary Figure S3A). We therefore directly interrogated the impact of this PAM position for all six nucleases using the GFP reporter assay (Figure 3, Supplementary Figure S5). As part of the assay, we used a 5′ TTTN PAM flanking the target sequence in the GFP reporter plasmid. We observed that all tested nucleases were able to recognize the TTTV motif, albeit at various apparent targeting activities, with marginal differences based on the identity of the −1 nucleotide. By contrast, four of the nucleases (Adurb336Cas12a, Fn3Cas12a, PdCas12a, Adurb193Cas12a) exhibited significantly reduced targeting with T versus V at the −1 PAM position ($P = <0.0001$–0.0002 comparing TTTT versus TTTV, $n = 3$). The two exceptions were PiCas12a and HkCas12a, which yielded GFP silencing that was statistically indistinguishable between T

incubation at 29°C, the library members containing a recognizable PAMs are cleaved, while the non-PAMs are left remaining in the reaction. The non-PAM members are PCR amplified and prepared for NGS. (**B**) Fold-change plots and PAM wheels representing the depleted motifs from the 5N PAM library for each Cas12a tested. For the fold-change plots, note the inverted y-axis. See Supplementary Figure S3A for versions of the fold-change plots rescaled to maximize separation between nts. As part of the screen for Fn3Cas12a, 72 out of the 1024 PAM sequences were absent from both libraries or from the targeted library, preventing their inclusion in the PAM wheel. See Supplementary Table S3 for the complete list of these sequences.

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

and V at this PAM position ($P = 0.15$ and 0.65 comparing TTTT versus TTTV for PiCas12a and HkCas12a respectively, $n = 3$ for TTTT and $n = 9$ for TTTV). These findings confirm that PiCas12 and HkCas12a can readily accommodate T at the –1 position for some motifs, adding to the list of Cas12a nucleases with this distinct capability (37). Given that PiCas12a and HkCas12a are phylogenetically distinct and separated by other Cas12a nucleases that were biased against T at the –1 PAM position, this finding provides further support that PAMs do not necessarily track with phylogenetic relationships.

HkCas12a recognizes C-rich PAMs in TXTL and in human cells

From our PAM screen, HkCas12a exhibited an ability to efficiently recognize C-rich motifs not associated with other well-characterized Cas12a nucleases (37), with a bias toward a C at the −3 position and a T at the −2 position (Figure 2B). To interrogate this unique recognition and bias further, we applied the GFP reporter assay to assess HkCas12a's ability to recognize CCCC, TCCC, CTCC, and CCTC as PAMs in TXTL (Figure 4A, Supplementary Figure S6). HkCas12a exhibited similar apparent targeting activities for all four PAM sequences. We did notice that CTCC and CCTC yielded the lowest and highest apparent targeting activities, respectively, matching the biases observed in the PAM screen at these two positions (Figure 2B). These differences were statistically significant when comparing TCCC to CCTC ($P = 0.0021$, $n = 3$) but not to CTCC ($P = 0.0627$, $n = 3$), while the error for CCCC was too large to establish statistical significance.

We next asked if similar sequences could be recognized by HkCas12a as part of genome-editing in human cells. Our results from the PAM screen and GFP reporter assay suggested that this nuclease did not explicitly require T in the PAM, while T at the −3 position could reduce targeting activity and T at the −2 position could enhance targeting activity. To assess this directly, we designed gRNAs targeting different sites in the *DNMT1* gene in HEK293T cells (Figure 4B, Supplementary Table S2) following our prior success generating indels in this gene and cell line with AsCas12a (29). Sites were selected to capture a range of PAM sequences, including GTTTC, ATTTC, GCCCC, ACCCC, GCCTG, ACCTC, GCTCA, ACTCA, TTCCC and ATCCA. We transiently transfected a plasmid expressing HkCas12a and a plasmid expressing processed gRNAs. After 72 hours post-transfection, we analyzed the frequency of indels formed by TIDE analysis (36) using the non-PAM TAGCT as a reference (Figure 4C). HkCas12a generated indels for all targets that were all significantly different from a control with the non-PAM GAAAT ($P = 0.0009$–0.0046, $n = 3$), with the target associated with the GTTTC motif resulting in the highest editing efficiency. When considering the average indel frequency for each pair of targets flanked by the same motif, the CTCV motif was associated with the lowest frequency, in line with the observed bias at the -3 position for a C over a T. However, the two indel frequencies for each pair often diverged, suggesting a contribution of the target sequence as we and others observed previously (29,38). Accordingly, one site flanked by the CTCV motif (GTCA) yielded an indel frequency that was statistically indistinguishable from individual targets flanked by the CCCV motif (GCCCC; $P = 1.0$, $n = 3$) and the CCTV motif (GCCTG; $P = 0.26$, $n = 3$). We did not observe the bias toward a T at the −2 position. Finally, we assessed editing at the TTTV-flanking sites by AsCas12a (Supplementary Figure S7), which showed that AsCas12a yielded higher frequencies of editing compared to HkCas12a. In total, we found that the HkCas12a can recognize C-rich PAMs in TXTL and in human cells, with some influence the exact PAM sequence on targeting activity.

Mutating PiCas12a toward PdCas12a reveals distinct PAM profiles in TXTL

We next turned our attention to PiCas12a, whose recognition of PAMs containing multiple G's (Figure 2B) distinguishes it from other previously characterized Cas12a nucleases (24,37). We first tested PiCas12a's ability to recognize the canonical ATTTC motif, as well as more unique G-containing sequences (ATGTC, CGGTC, AGGCG, AGTGC) uncovered from the PAM screen (Figure 2B) as part of the GFP reporter assay (Figure 5B, Supplementary Figure S9A). PiCas12a was able to recognize unique sequences containing a single G in the −3 position (ATGTC) as well as multiple G's in the −4/−3 (CGGTC) and −4/−2 positions (AGTGC) (Figure 5B). The only exception was AGGCG, which was not recognized in the GFP reporter assay but was strongly depleted in the PAM screen. In contrast, PdCas12a only strongly recognized ATTTC and at best modestly recognized ATGTC.

The disparate PAM profiles distinguishing PiCas12a and PdCas12a raised the question: which of the 57 mutations are responsible for these differences? This question is particularly intriguing given that most of these mutations (41/57) fall within the RuvC and Nuc domains that have not been implicated in PAM recognition, while only five mutations fall within the PAM-interacting domain (Figure 5A). Previous work showed that the PAM specificity of AsCas12a could be altered by mutating specific residues located within or between the recognition (REC) domain and the PAM-interacting domain (39–41). We therefore aligned the protein sequences of AsCas12a, PiCas12a and PdCas12a to identify associated residues that altered the PAM preferences of AsCas12a and varied between PiCas12a and PdCas12a (Document S1). This led us to identify five residues in PiCas12a: K163, N599, S600, F604 and T628 (Figure 5A, Document S1). The equivalent residues in AsCas12a (K164, N547, K548, N552, Q571) directly interact with the PAM (N547, K548), can interact when mutated in the RVR variant (N552), or are in the immediate vicinity of the PAM (K164, Q571) (Supplementary Figure S8) (41,42). We converted each residue or a combination of the residues in PiCas12a to match that of PdCas12a to create the PiCas12a variants with K163E, N599D, S600N, F604Y, T628A, DNYA (containing the four mutations N599D/S600N/F604Y/T628A), and EDNYA (containing all five mutations). Although the corresponding residues could also be mutated in PdCas12a, we focused on PiCas12a as the starting point given this nuclease's unique PAM profile.

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

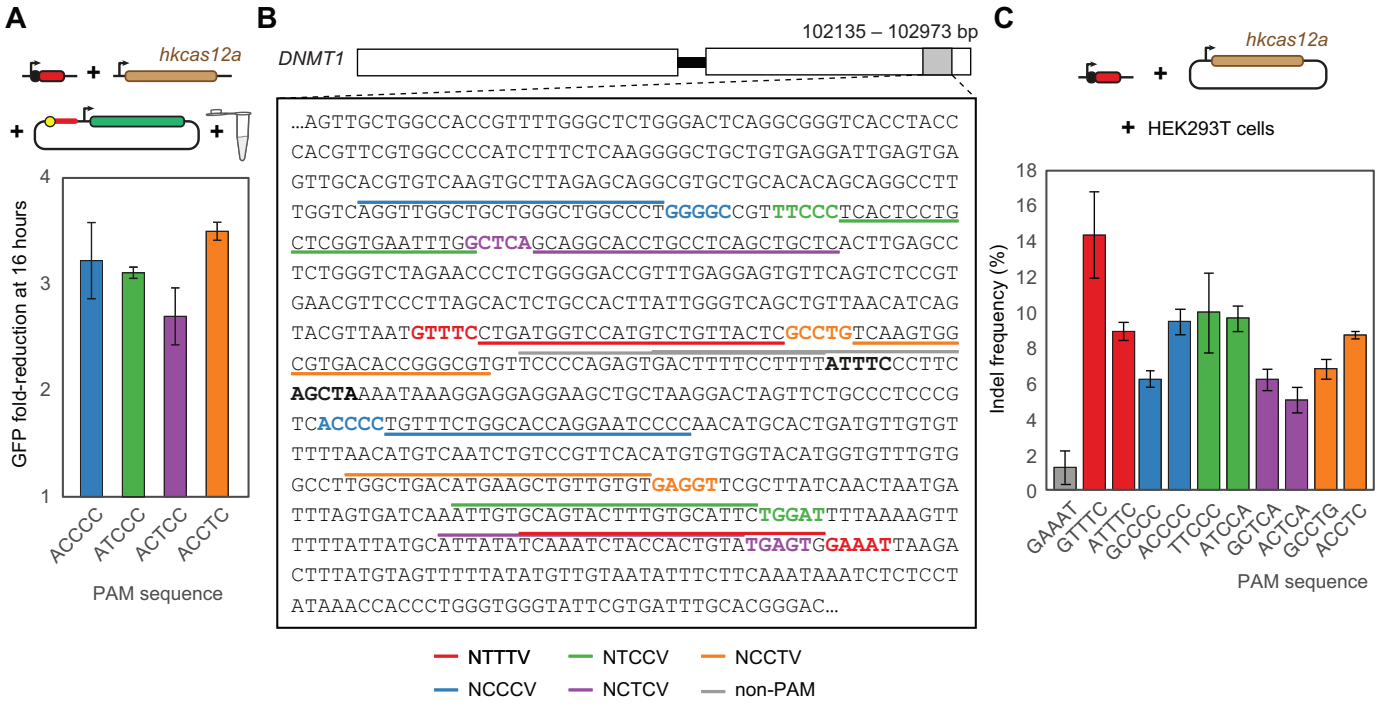


Figure 4. HkCas12a recognizes C-rich PAM sequences in TXTL and human cells. (**A**) Fold-reduction of GFP expression for HkCas12a targeting sequences containing various C-rich PAMs. The fold-reduction was calculated using the GFP fluorescence data from the 16 h time-point from the reactions containing the targeting non-targeting gRNA. The error bars represent the standard deviation from three separate TXTL reactions. See Supplementary Figure S6 for the associated time-courses. (**B**) A section of the *DNMT1* gene containing the targeted sequences (highlighted) and their associated PAMs (bold) used for the indel formation assays. The sequences highlighted in red and gray represent the targets associated with the canonical (NTTTC) and unrecognized (TAGCT, GAAAT) PAM sequences, respectively. The other colors represent targets associated with non-canonical motifs that were tested in the GFP reporter assay. Note that some of the target sequences and their associated PAMs are located in the bottom strand. (**C**) Indel frequencies for each *DNMT1* target site tested in HEK293T cells quantified using TIDE (37). The indel frequency data for each target sequence were compared to that of a target sequence flanked by a non-PAM (TAGCT). Error bars represent the standard deviation from three independent transfections.

We performed the GFP reporter assay to assess PAM recognition by these PiCas12a variants as well as PdCas12a using the same PAM sequences tested with PiCas12a (Figure 5B, Supplementary Figure S9A). The PiCas12a variants each recognized at least one of the non-canonical sequences recognized by PiCas12a. In many cases, each mutation reduced or eliminated recognition of a non-canonical PAM, such as impaired recognition of AGTGC by the K163E, S600N, DNYA and EDNYA variants, or impaired recognition of ATGTC by the K163E and T628A variants. However, the mutations also altered PAM recognition in ways that did not follow the expected progression from PiCas12a to PdCas12a. Specifically, some mutations enhanced recognition of motifs poorly recognized by either PiCas12a or PdCas12a. Specifically, the K163E and F604Y variants exhibited enhanced recognition of AGGCG, while the S600N, F604Y, T628A and DNYA variants exhibited enhanced recognition of CGGTC compared to ATTTC. Another example was the non-additive effect of the mutations. In particular, the DNYA variant did not recognize the AGGCG motif, while the addition of the K163E mutation that exhibited enhanced recognition of AGGCG yielded the EDNYA variant that was also unable to recognize this sequence. We also tested two recognized sequences in the PAM screen (AGGCC, AGGTC), although neither were recognized by any of the variants in the GFP reporter assay (Supplementary Figure S9B).

We next sought to uncover the PAM profile of a subset of the PiCas12a variants investigated in TXTL. We were particularly interested in the F604Y, DNYA, and EDNYA variants, as the F604Y variant exhibited increased apparent targeting activity and recognition of AGGCG in comparison to PiCas12a, while the DNYA and EDNYA variants exhibited activities approaching that of PdCas12a (Figure 5B). Thus, we performed the PAM screen on these three variants to uncover their full range of recognized motifs (Figure 6). The PAM profiles revealed that the F604Y variant maintained recognition of G-rich motifs similar to PiCas12a, although the variant exhibited stronger recognition of G at the −3 and −4 positions, stronger recognition of C over T at the −2 position, and a more overt bias against A at the -4 position. The F604Y variant also recognized fewer sequences with T at the −1 position. The DNYA variant exhibited a reduced ability to recognize G-rich motifs, although it could readily recognize T at the −1 position within a TTYT motif. Finally, the EDNYA variant showed a lack of non-canonical PAM recognition, although the GFP reporter assays showed recognition of CGGTC (Figure 5B). Interestingly, the EDNYA variant exhibited poor apparent cleavage activity in the PAM screen similar to PdCas12a (Figure 2B), suggesting that the K163E mutation alone or in

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

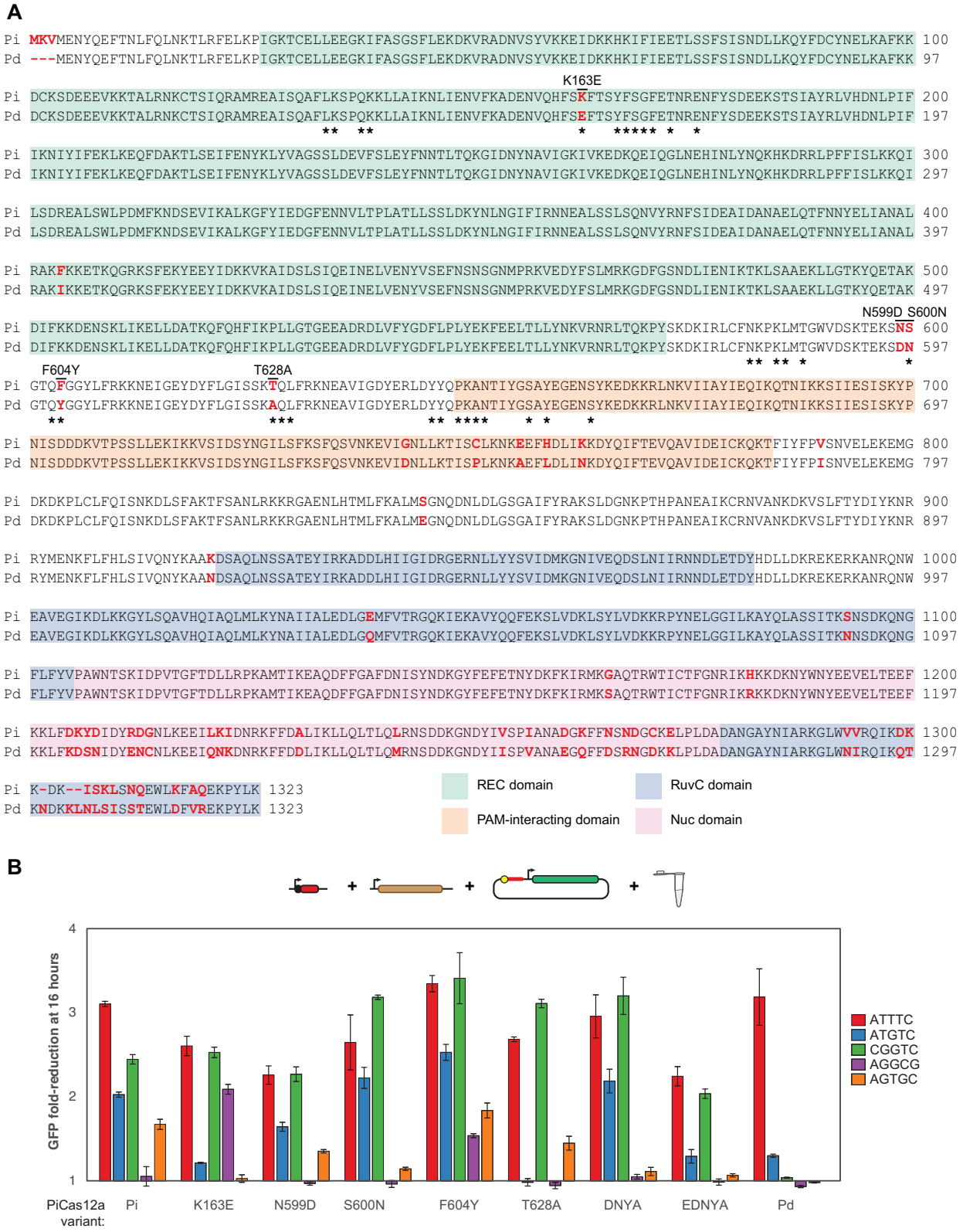


Figure 5. Altering residues in the PAM domain between PdCas12a and PiCas12a modifies PAM recognition of PiCas12a. (**A**) Sequence alignment of Pd and PiCas12a. These nucleases share 95.7% amino-acid identity, with most of the variations stemming from the 3′ end of the sequence. The black/bold/underlined sequences indicate matching sequences, while regular black text indicates unmatched residues. The red/bolded sequence indicate the residues investigated in this work. The asterisks indicate residues shown to alter PAM specificity when aligned with AsCas12a (39,40) (see Document S1 for alignment with AsCas12a and other Cas12a nucleases). The sequences were aligned using MUSCLE with default settings. (**B**) Fold-reduction of GFP expression for each PiCas12a variant and motif. The fold-reduction was calculated using the GFP fluorescence data from the 16 h time-point from the reactions containing the targeting non-targeting gRNA. Error bars represent the standard deviation from three separate TXTL reactions. See Supplementary Figure S9A for the associated time-courses.

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

the context of the other mutations accounts for the reduced apparent activity. Through these multiple experiments, we conclude that mutating specific residues in PiCas12a found in PdCas12a and associated with altered PAM recognition in AsCas12a leads to changes in the recognized PAM profile, in some cases deviating from either PiCas12a or PdCas12a.

PiCas12a and the F604Y variant recognize distinct non-canonical PAMs in TXTL and in human cells and are sensitive to the -5 PAM position

Deeper interrogation of the PAM wheels for PiCas12a and the F604Y variant revealed biases at the −5 position as well as within the variable V at the −1 position of the PAM within the G-rich non-canonical motifs (Data S1). We thus used the GFP reporter assay to test the ability of PiCas12a and the F604Y variant to recognize individual PAMs within the CGGYV and NGGYG motifs (Figure 7A, Supplementary Figure S10). We found that PiCas12a recognized motifs within CGGYV with some bias against A at the −1 position. The bias at the −5 position was notably strong for the NGGYG motifs, with C preferred over G and no recognition of A or T. In contrast, the F604Y variant exhibited less bias at the −1 position within the CGGYV motif, and it exhibited stronger recognition of G and some recognition of A or T at the −5 position within the NGGYG motif. These results show that the −1 and −5 positions factor into recognition of these non-canonical PAMs.

To further interrogate non-canonical PAM recognition by PiCas12a and the F604Y variant, we next assessed indel formation in human cells. We selected targets within *DNMT1* flanked by various G-containing sequences, including GTTGC, GTTGA, ATGTC, CGTTG, TGTTC, GGTCA, AGTCC and CGGTG (Figure 7C) to initially test with PiCas12a. We then selected a subset of these targets to test with the F604Y variant (GTTGC, GTTGA and CGGTG) to specifically evaluate PAMs that were preferentially recognized by this variant over PiCas12a in the PAM screen and GFP reporter assay (Figure 5B, Supplementary Figure S9A). We performed this assay following the experimental set up used for the indel formation assay with HkCas12a (Figure 4), and we tested many of these same sites with AsCas12a as a basis of comparison (Supplementary Figure S7).

Both nucleases yielded significant indel formation across all tested PAMs ($P \leq 0.00001$–0.020, $n = 3$; $P = 0.020$ for ATGTC-1) compared to the non-PAM control (GAAAT) with only one exception: PiCas12a with ATGTC-2 ($P = 0.11$, $n = 3$). However, there were noticeable differences in the frequency of indel formation between the nucleases and targets (Figure 7D). PiCas12a yielded higher indel frequencies at the two TTTV-flanked targets than the tested targets flanked by non-canonical PAMs, with only the CGTTC-flanked target exhibiting similar frequencies (13% compared to 15% and 10% for the two TTTV-flanked targets). In contrast, the F604Y variant yielded similar frequencies of editing between one TTTV-flanked target site (11%) and 3/4 of the tested targets flanked by non-canonical PAMs (10% for ATGTC-2, 10% for CGGTG-1, 14% for CGGTG-2). Furthermore, when directly comparing PiCas12a and the F604Y variant at the same targets, the F604Y variant consistently yielded higher frequencies of indel formation, paralleling the results from the GFP reporter assays in TXTL. The targets flanked by ATGTC-1 and ATGTC-2 in particular exhibited enhanced editing for the F604Y variant over PiCas12a (7% versus 3%, 10% versus 2%). We do note that the indel frequencies sometimes varied between targets flanked by the same general motif as observed for HkCas12a (Figure 4C), where a greatly expanded set of sequences would need to be tested to assess the relative contribution of the target sequence and the specific PAM sequence. Similarly, we primarily evaluated PAM sequences that exhibited enhanced recognition by the F604Y variant, although an expanded target set could also probe PAM sequences such as AGGCG that appeared to be uniquely recognized by this variant. When evaluating indel formation with AsCas12a, the overall frequencies of editing with a canonical TTTV PAM were higher than those for PiCas12a or the F604Y variant (compare Figure 7D and Supplementary Figure S7). However, when targeting sites flanked by the G-containing sequences, AsCas12a exhibited low indel frequencies that were either modest but significant ($P = 0.027$, $n = 3$ for GGTCA) or not significant ($P = 0.18$–0.90, $n = 3$ for the others) compared to the non-PAM control. Overall, these data showed that the PAMs for PiCas12a and the F604Y variant that were identified with TXTL can mediate gene editing in human cells, with the F604Y variant being associated with higher editing efficiencies across all tested target sequences. Probing an expanded set of target sites would help establish the contribution of the target sequence and provide a validated list of PAM sequences that support editing.

DISCUSSION

Here, we characterized six phylogenetically diverse Cas12a nucleases, including Adurb336, Fn3, Pi, Pd, Adurb193 and HkCas12a. While a subset of these nucleases was associated with high amino-acid identity, their PAM profiles generally did not track with phylogeny (Figures 1 and 2). Specifically, Hk and PiCas12a were associated with distinct PAM profiles that varied greatly compared to other closely related orthologs, such as their ability to tolerate a T in the −1 PAM position (Figure 3, Supplementary Figure S5). HkCas12a recognized a C-rich PAM, with a bias toward a C in the −3 PAM position and toward a T in the −2 position (Figures 2B and 4). PiCas12a recognized PAMs containing multiple G's in its PAM, with strong biases at the −5 position, while the F604Y mutant exhibited both enhanced effective cleavage activity as well as more flexible PAM recognition. During the preparation of this manuscript, another group reported the characterization of HkCas12a, which also reported recognition of C-rich PAM sequences (37), where their data also suggested similar biases, although it was not explicitly stated. These nucleases highlight the remarkable diversity that can exist within CRISPR subtypes and suggest that other nucleases sharing strong identity could also recognize different PAM sequences, with the potential of further expanding the toolbox of CRISPR nucleases to collectively cover all possible PAM sequences. The

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

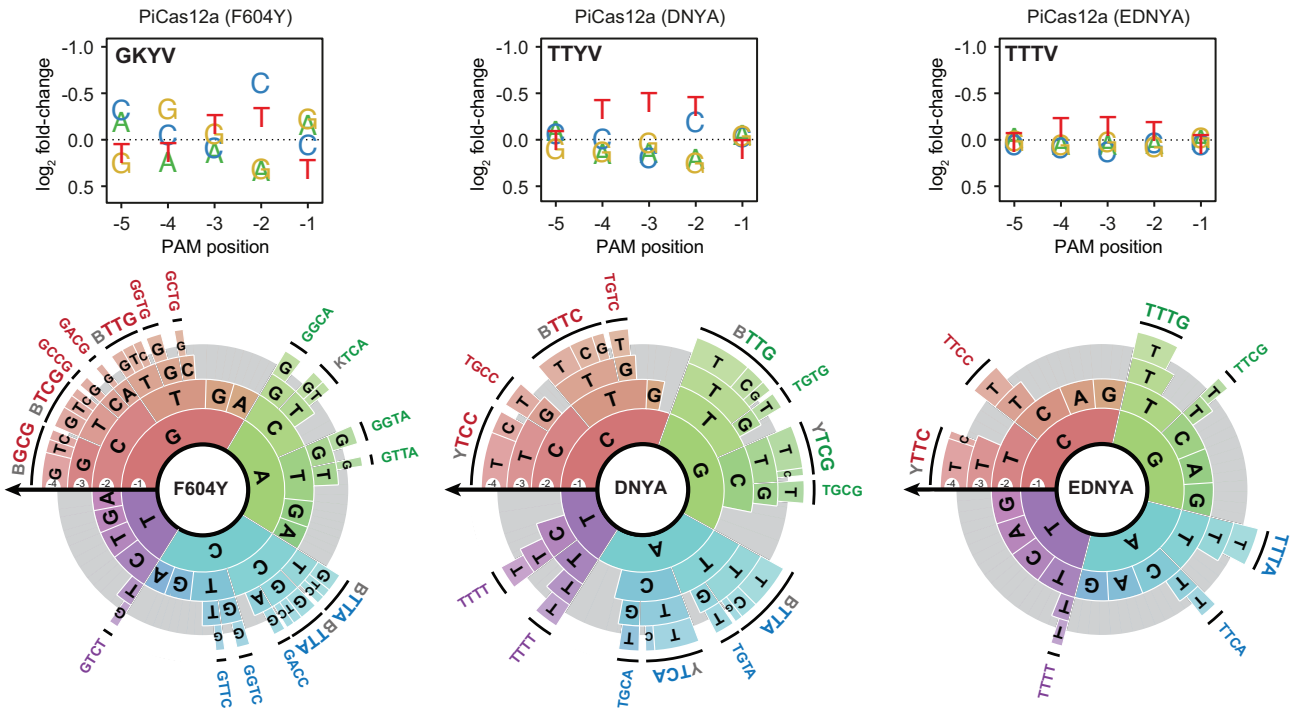


Figure 6. Mutating residues in PiCas12a to match those in PdCas12a yields distinct PAM profiles and apparent cleavage activities. Fold-change plots and PAM wheels of the F604Y, DNYA and EDNYA PiCas12a variants. The F604Y variant was associated with a more promiscuous PAM compared to the wild-type nuclease, while the DNYA and EDNYA variants were associated with a reduced ability to recognize non-canonical motifs. See Supplementary Figure S3A for versions of the fold-change plots rescaled to maximize separation between nts.

PAM screens also revealed variability in the apparent cleavage activities that further distinguished the nucleases, and probing other features not explicitly evaluated in this study (e.g. temperature dependence, location and length of the 5′ staggered cleavage product) could reveal additional differences. It would also be important to evaluate the propensity for off-targeting using well-established methodologies (43) given the potential of altered PAM recognition to impact targeting across a genome.

We also found that PiCas12a and PdCas12a exhibited strikingly different apparent targeting activities and PAM recognition profiles despite their shared amino-acid identity. To gain insights into the DNA targeting properties that differentiate these orthologs, we mutated a combination of five residues that were previously shown to alter PAM recognition or were proximal to these residues (39,40) to match that of PdCas12a (Figure 5A, Document S1). The PiCas12a variants investigated in this work includes the K163E, N599D, S600N, F604Y and T628A variants, as well as the combinatorial DNYA and EDNYA variants. While many of these variants were associated with reduced non-canonical PAM recognition (e.g. reduced recognition of ATGTC from K163E variants), the effects from mutating these residues were not predictable (Figure 5B). In fact, some of these mutations led to broadening of the PAM profile recognized by wild-type PiCas12a (e.g. CGGTC recognition of S600N, F604Y, T628A and DNYA variants). These mutational analyses suggest that while mutations can alter the PAM profile of CRISPR nucleases, there exists a complex and combinatorial relationship between the residues mutated and their effect on PAM recognition. While we investigated five specific residues based on prior work altering PAM recognition in AsCas12a (39,40), other residues could influence PiCas12a's unique PAM profile. One notable example was the EDNYA variant sharing PdCas12a's limited apparent cleavage activity while maintaining its ability to recognize a CGGTG PAM sequence.

Prior crystal structures of AsCas12a (PDB: 5B43) and the RVR variant (PDB: 5XH6) provide some hints at the structural impact of the residues mutated in PiCas12a (41,42). The residues in AsCas12a equivalent to those mutated in this work directly interact with the PAM (N599, S600, F604) or are located in the vicinity (K163, T628) in one or both crystal structures. Mutating the residues that directly interact with the PAM would be expected to readily impact PAM recognition, particularly when changing the chain length or chemical properties of the residue (e.g. the S600N mutation). Separately, mutating the residues in the vicinity of the PAM could have created new interactions. As support, the N552R mutation in AsCas12a (equivalent to F604) switched the side chain from having no interaction with the PAM to forming hydrogen-bonding interactions with the base at the −3 position of the target strand (41). By extension, K163 or T628 could be involved in PAM interactions in either PiCas12a or PdCas12a. The combination of these mutations could also have a synergistic effect, reflecting the varying impact of combining mutations. For instance, introducing K163E into PiCas12a enhances recognition of AGGCG, while introducing K163E into DNYA does not regain recognition of AGGCG and instead results

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

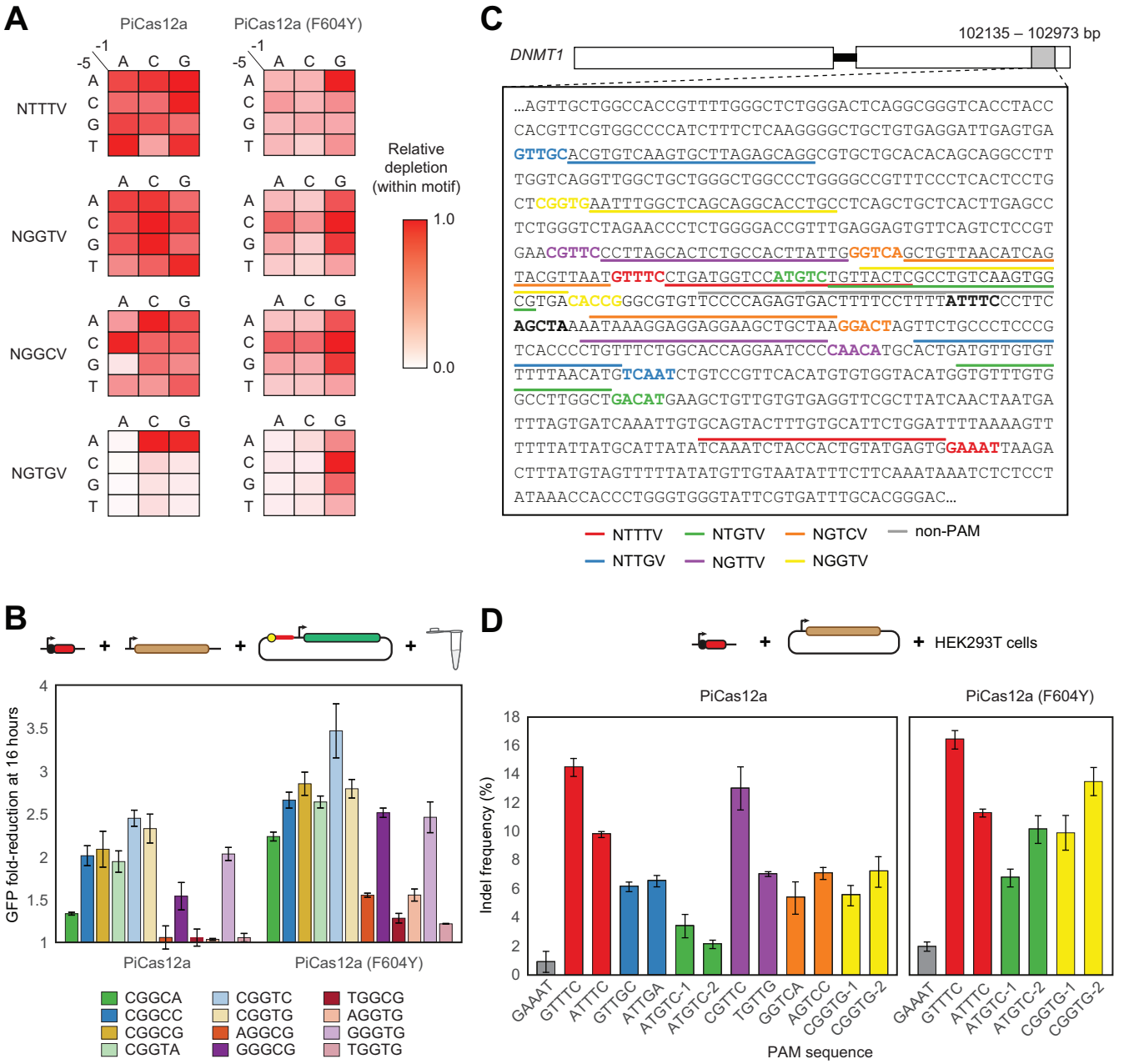


Figure 7. PiCas12a and the F604Y variant can recognize various non-canonical motifs in TXTL and in human cells. (**A**) Comparing biases at the −5 and −1 positions of the PAM for PiCas12a and the F604Y variant. Comparisons are based on the results from the PAM screen for each nuclease. (**B**) Fold-reduction of GFP expression in TXTL for each nuclease and different G-rich PAMs. The fold-reduction was calculated using the GFP fluorescence data from the 16 h time-point from the reactions containing the targeting non-targeting gRNA. The error bars represent the standard deviation from three separate TXTL reactions. See Supplementary Figure S10 for the associated time-courses. (**C**) A section of the *DNMT1* gene containing the targeted sequences (highlighted) and their associated PAMs (bold) used for the indel formation assays. The sequences highlighted in red and gray represent the targets associated with the canonical NTTTV motif and two non-PAM sequences (TAGCT, GAAAT) motifs, respectively. The other colors represent targets associated with non-canonical motifs that were tested in the GFP reporter assay. Note that some of the target sequences and their associated PAMs are located in the bottom strand. (**D**) Indel frequencies for each *DNMT1* target site tested in HEK293T cells quantified using TIDE (36). The indel frequency data for each target sequence were compared to that of a target sequence flanked by a non-PAM (TAGCT). Error bars represent the standard deviation from three independent transfections.

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

in lower apparent targeting and cleavage activity. An exhaustive mutational analysis along with detailed biochemical and structural studies could provide insights into PAM engineering of Cas12a nucleases, which in turn may lead to CRISPR nucleases associated with diverse PAM profiles.

The properties distinguishing Pi and PdCas12a suggest a specific selective pressure that led to diversified PAM recognition by PiCas12a. This mode of diversification is distinct from traditional modes such as spacer acquisition and horizontal gene transfer (44) and is in line with phages escaping CRISPR immunity by mutating the PAM sequence flanking existing protospacers (16). PAM diversification could also be driven by anti-CRISPR proteins (Acrs) (15), where multiple Acrs were discovered for Type V-A CRISPR systems (45,46) and one has been shown to inhibit Cas12a by acetylating a residue responsible for PAM recognition (47). The only major constraint on PAM diversification appears to be avoiding self-recognition through the presence of a PAM within the 3′ end of the repeat. Interestingly, all of the Cas12 nucleases tested in this work shared the same six nucleotides at the 3′ end of the repeat (GTAGAT) (Figure 1B), suggesting pressure to maintain this sequence while still leaving ample sequence space that could be recognized as non-canonical PAMs.

Through this work, we encountered a few discrepancies between the results of the PAM screen, the GFP reporter assay, and the indel formation assay in HEK293T cells. For instance, PiCas12a recognized AGGCG in the PAM screen and but not the GFP reporter assay, the EDNYA variant of PiCas12a recognized CGGTC in the GFP reporter assay but not in the PAM screen, and both PiCas12a and the F604Y variant recognized AGGTC and AGGCC in the PAM screen but not the GFP reporter assay. These results were relatively surprising given that the two assays shared the same target sequence. Aside from difference in the flanking sequences, one notable discrepancy was the proximity of the target site to the promoter in the GFP reporter assay. This proximity would be expected to result in transcriptional repression, suggesting differences in DNA binding versus cleavage, although further work is needed to clarify the contribution of each mechanism. Beyond discrepancies within the TXTL assays, we also observed some discrepancies between the TXTL results and the indel formation assays in human cells. For instance, HkCas12a exhibited a bias for a T at the −2 position of the PAM in TXTL, although this bias was not observed in human cells. Overall, these discrepancies underscore the need for validation experiments when performing PAM determination assays, particularly in cellular or acellular systems in which the nucleases will be employed (4).

DATA AVAILABILITY

The NGS data, including the raw data and post-processing reads, were deposited in the NCBI gene expression omnibus (accession #GSE130377). The code used for the processing and analyses of the NGS data can be found in the following public repository: https://bitbucket.org/csmaxwell/crispr-txtl-pam-counting-script/.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Matthew Begemann for his assistance with Cas12a identification and for helpful feedback on the manuscript, Gloria Yi for her assistance with the molecular cloning, and Tatjana Achmedov for assistance conducting the Northern blotting analyses. The PdCas12a expression plasmid pY018 (pcDNA3.1-hPdCpf1) was a gift from Feng Zhang. The empty gRNA expression plasmid (pU6-As-crRNA) was a gift from Jin-Soo Kim.

Author contribution: T.J. and F.T. performed the molecular cloning, PAM screens, GFP reporter assays, and indel formation assays. C.L. performed the northern blotting analyses. T.J. and C.L.B analyzed the data. S.M. assessed the PiCas12a mutants in relation to published structural data. B.N.G. identified the Cas12a nucleases. T.J. and C.L.B. wrote the paper, with critical feedback from B.N.G.

FUNDING

National Institutes of Health [1R35GM119561]; National Science Foundation [MCB-1413044]. Funding for open access charge: NIH grant [1R35GM119561].

Conflict of interest statement. C.L.B. is a co-founder and scientific advisory board member of Locus Biosciences and submitted provisional patent applications on CRISPR technologies. B.N.G. is a current employee of Benson Hill, which has proprietary CRISPR technologies on which patent applications have been filed.

REFERENCES

1. Barrangou,R., Fremaux,C., Deveau,H., Richards,M., Boyaval,P., Moineau,S., Romero,D. and Horvath,P. (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, **315**, 1709–1712.
2. Sorek,R., Lawrence,C.M. and Wiedenheft,B. (2013) CRISPR-mediated adaptive immune systems in bacteria and archaea. *Annu. Rev. Biochem.*, **82**, 237–266.
3. Barrangou,R. and Marraffini,L.A. (2014) CRISPR-Cas systems: prokaryotes upgrade to adaptive immunity. *Mol. Cell*, **54**, 234–244.
4. Leenay,R.T. and Beisel,C.L. (2017) Deciphering, communicating, and engineering the CRISPR PAM. *J. Mol. Biol.*, **429**, 177–191.
5. Mojica,F.J., Diez-Villasenor,C., Garcia-Martinez,J. and Almendros,C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**, 733–740.
6. Adli,M. (2018) The CRISPR tool kit for genome editing and beyond. *Nat. Commun.*, **9**, 1911.
7. Koonin,E.V., Makarova,K.S. and Zhang,F. (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr. Opin. Microbiol.*, **37**, 67–78.
8. Makarova,K.S., Wolf,Y.I. and Koonin,E.V. (2018) Classification and nomenclature of CRISPR-Cas systems: where from here? *Cris. J.*, **1**, 325–336.
9. Mougiakos,I., Mohanraju,P., Bosma,E.F., Vrouwe,V., Finger Bou,M., Naduthodi,M.I.S., Gussak,A., Brinkman,R.B.L., van Kranenburg,R. and van der Oost,J. (2017) Characterizing a thermostable Cas9 for bacterial genome editing and silencing. *Nat. Commun.*, **8**, 1647.
10. Harrington,L.B., Paez-Espino,D., Staahl,B.T., Chen,J.S., Ma,E., Kyrpides,N.C. and Doudna,J.A. (2017) A thermostable Cas9 with increased lifetime in human plasma. *Nat. Commun.*, **8**, 1424.

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021

11. Moreno-Mateos,M.A., Fernandez,J.P., Rouet,R., Vejnar,C.E., Lane,M.A., Mis,E., Khokha,M.K., Doudna,J.A. and Giraldez,A.J. (2017) CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nat. Commun.*, **8**, 2024.
12. Kleinstiver,B.P., Tsai,S.Q., Prew,M.S., Nguyen,N.T., Welch,M.M., Lopez,J.M., McCaw,Z.R., Aryee,M.J. and Joung,J.K. (2016) Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat. Biotechnol.*, **34**, 869–874.
13. Kim,D., Kim,J., Hur,J.K., Been,K.W., Yoon,S.H. and Kim,J.S. (2016) Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat. Biotechnol.*, **34**, 863–868.
14. Malzahn,A.A., Tang,X., Lee,K., Ren,Q., Sretenovic,S., Zhang,Y., Chen,H., Kang,M., Bao,Y., Zheng,X. *et al.* (2019) Application of CRISPR-Cas12a temperature sensitivity for improved genome editing in rice, maize, and Arabidopsis. *BMC Biol.*, **17**, 9.
15. Bondy-Denomy,J., Pawluk,A., Maxwell,K.L. and Davidson,A.R. (2013) Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature*, **493**, 429–432.
16. Deveau,H., Barrangou,R., Garneau,J.E., Labonté,J., Fremaux,C., Boyaval,P., Romero,D.A., Horvath,P. and Moineau,S. (2008) Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *J. Bacteriol.*, **190**, 1390–1400.
17. Konermann,S., Lotfy,P., Brideau,N.J., Oki,J., Shokhirev,M.N. and Hsu,P.D. (2018) Transcriptome engineering with RNA-targeting Type VI-D CRISPR effectors. *Cell*, **173**, 665–676.
18. Yan,W.X., Chong,S., Zhang,H., Makarova,K.S., Koonin,E.V., Cheng,D.R. and Scott,D.A. (2018) Cas13d is a compact RNA-targeting Type VI CRISPR effector positively modulated by a WYL-domain-containing accessory protein. *Mol. Cell*, **70**, 327–339.
19. Yan,W.X., Hunnewell,P., Alfonse,L.E., Carte,J.M., Keston-Smith,E., Sothiselvam,S., Garrity,A.J., Chong,S., Makarova,K.S., Koonin,E.V. *et al.* (2019) Functionally diverse type V CRISPR-Cas systems. *Science*, **363**, 88–91.
20. Chatterjee,P., Jakimo,N. and Jacobson,J.M. (2018) Minimal PAM specificity of a highly similar SpCas9 ortholog. *Sci. Adv.*, **4**, eaau0766.
21. Fonfara,I., Le Rhun,A., Chylinski,K., Makarova,K.S., Lécrivain,A.L., Bzdrenga,J., Koonin,E.V. and Charpentier,E. (2014) Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res.*, **42**, 2577–2590.
22. Briner,A.E., Donohoue,P.D., Gomaa,A.A., Selle,K., Slorach,E.M., Nye,C.H., Haurwitz,R.E., Beisel,C.L., May,A.P. and Barrangou,R. (2014) Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol. Cell*, **56**, 333–339.
23. Komor,A.C., Badran,A.H. and Liu,D.R. (2017) CRISPR-based technologies for the manipulation of eukaryotic genomes. *Cell*, **168**, 20–36.
24. Zetsche,B., Gootenberg,J.S., Abudayyeh,O.O., Regev,A., Koonin,E.V., Zhang,F., Slaymaker,I.M., Makarova,K.S., Essletzbichler,P., Volz,S.E. *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, **163**, 759–771.
25. Fonfara,I., Richter,H., Bratovič,M., Le Rhun,A. and Charpentier,E. (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, **532**, 517–521.
26. Chen,J.S., Ma,E., Harrington,L.B., Da Costa,M., Tian,X., Palefsky,J.M. and Doudna,J.A. (2018) CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science*, **6245**, eaar6245.
27. Tak,Y.E., Kleinstiver,B.P., Nuñez,J.K., Hsu,J.Y., Horng,J.E., Gong,J., Weissman,J.S. and Joung,J.K. (2017) Inducible and multiplex gene regulation using CRISPR–Cpf1-based transcription factors. *Nat. Methods*, **14**, 1163–1166.
28. Gootenberg,J.S., Abudayyeh,O.O., Kellner,M.J., Joung,J., Collins,J.J. and Zhang,F. (2018) Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science*, **360**, 439–444.
29. Jacobsen,T., Liao,C. and Beisel,C.L. (2019) The *Acidaminococcus* sp. Cas12a nuclease recognizes GTTV and GCTV as non-canonical PAMs. *FEMS Microbiol. Lett.*, **336**, fnz085.
30. Li,B., Zhao,W., Luo,X., Zhang,X., Li,C., Zeng,C. and Dong,Y. (2017) Engineering CRISPR–Cpf1 crRNAs and mRNAs to maximize genome editing efficiency. *Nat. Biomed. Eng.*, **1**, 66.
31. Maxwell,C.S., Jacobsen,T., Marshall,R., Noireaux,V. and Beisel,C.L. (2018) A detailed cell-free transcription-translation-based assay to decipher CRISPR protospacer-adjacent motifs. *Methods*, **143**, 48–57.
32. Marshall,R., Maxwell,C.S., Collins,S.P., Jacobsen,T., Luo,M.L., Begemann,M.B., Gray,B.N., January,E., Singer,A., He,Y. *et al.* (2018) Rapid and scalable characterization of CRISPR technologies using an *E. coli* cell-free transcription-translation system. *Mol. Cell*, **69**, 146–157.
33. Garamella,J., Marshall,R., Rustad,M. and Noireaux,V. (2016) The all *E. coli* TX-TL toolbox 2.0: A platform for cell-free synthetic biology. *ACS Synth. Biol.*, **5**, 344–355.
34. Liao,C., Ttofali,F., Slotkowski,R.A., Denny,S.R., Cecil,T.D., Leenay,R.T., Keung,A.J. and Beisel,C.L. (2019) Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. *Nat. Commun.*, **10**, 2948.
35. Leenay,R.T., Maksimchuk,K.R., Slotkowski,R.A., Agrawal,R.N., Gomaa,A.A., Briner,A.E., Barrangou,R. and Beisel,C.L. (2016) Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol. Cell*, **62**, 137–147.
36. Brinkman,E.K., Chen,T., Amendola,M. and Van Steensel,B. (2014) Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.*, **42**, e168.
37. Teng,F., Li,J., Cui,T., Xu,K., Guo,L., Gao,Q., Feng,G., Chen,C., Han,D., Zhou,Q. *et al.* (2019) Enhanced mammalian genome editing by new Cas12a orthologs with optimized crRNA scaffolds. *Genome Biol.*, **20**, 15.
38. Esvelt,K.M., Mali,P., Braff,J.L., Moosburner,M., Yaung,S.J. and Church,G.M. (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat. Methods*, **10**, 1116–1121.
39. Gao,L., Cox,D.B.T., Yan,W.X., Manteiga,J.C., Schneider,M.W., Yamano,T., Nishimasu,H., Nureki,O., Crosetto,N. and Zhang,F. (2017) Engineered Cpf1 variants with altered PAM specificities. *Nat. Biotechnol.*, **35**, 789–792.
40. Kleinstiver,B.P., Sousa,A.A., Walton,R.T., Tak,Y.E., Hsu,J.Y., Clement,K., Welch,M.M., Horng,J.E., Malagon-Lopez,J., Scarfò,I. *et al.* (2019) Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat. Biotechnol.*, **37**, 276–282
41. Yamano,T., Zetsche,B., Ishitani,R., Zhang,F., Nishimasu,H. and Nureki,O. (2017) Structural basis for the canonical and non-canonical PAM recognition by CRISPR-Cpf1. *Mol. Cell*, **67**, 633–645.
42. Yamano,T., Nishimasu,H., Zetsche,B., Hirano,H., Ian,M., Li,Y., Fedorova,I., Nakane,T., Makarova,K.S., Koonin,E.V. *et al.* (2016) Crystal structure of Cpf1 in complex with guide RNA and target DNA. *Cell*, **165**, 949–962.
43. Martin,F., Sánchez-Hernández,S., Gutiérrez-Guerrero,A., Pinedo-Gomez,J. and Benabdellah,K. (2016) Biased and unbiased methods for the detection of off-target cleavage by CRISPR/Cas9: An overview. *Int. J. Mol. Sci.*, **17**, 1507.
44. Westra,E.R., Dowling,A.J., Broniewski,J.M. and van Houte,S. (2016) Evolution and ecology of CRISPR. *Annu. Rev. Ecol. Evol. Syst.*, **47**, 307–331.
45. Watters,K.E., Fellmann,C., Bai,H.B., Ren,S.M. and Doudna,J.A. (2018) Systematic discovery of natural CRISPR-Cas12a inhibitors. *Science*, **362**, 236–239.
46. Marino,N.D., Zhang,J.Y., Borges,A.L., Sousa,A.A., Leon,L.M., Rauch,B.J., Walton,R.T., Berry,J.D., Joung,J.K., Kleinstiver,B.P. *et al.* (2018) Discovery of widespread type I and type V CRISPR-Cas inhibitors. *Science*, **362**, 240–242.
47. Dong,L., Guan,X., Li,N., Zhang,F., Zhu,Y., Ren,K., Yu,L., Zhou,F., Han,Z., Gao,N. *et al.* (2019) An anti-CRISPR protein disables type V Cas12a by acetylation. *Nat. Struct. Mol. Biol.*, **26**, 308–314.

Downloaded from https://academic.oup.com/nar/article/48/10/5624/5824607 by guest on 04 August 2021