Nucleic Acids Research, 2020 **1**
doi: 10.1093/nar/gkaa110

Improved LbCas12a variants with altered PAM specificities further broaden the genome targeting range of Cas12a nucleases

Eszter Tóth[1], Éva Varga[1,2,3], Péter István Kulcsár[1,2,3], Virág Kocsis-Jutka[1,4], Sarah Laura Krausz[1,5], Antal Nyeste[1,3,4], Zsombor Welker[6], Krisztina Huszár[1,6], Zoltán Ligeti[1,2,3], András Tálas[1,5] and Ervin Welker[1,3,*]

[1]Institute of Enzymology, Research Centre of Natural Sciences of the Hungarian Academy of Sciences, Budapest H-1117, Hungary, [2]Doctoral School of Multidisciplinary Medical Science, University of Szeged, Szeged H-6726, Hungary, [3]Institute of Biochemistry, Biological Research Centre of the Hungarian Academy of Sciences, Szeged H-6726, Hungary, [4]ProteoScientia Kft, Cserhátszentiván, H-3066, Hungary, [5]School of Ph.D. Studies, Semmelweis University, Budapest, H-1085, Hungary and [6]Biospirál-2006 Kft., Szeged, H-6726, Hungary

Received June 25, 2019; Revised February 07, 2020; Editorial Decision February 10, 2020; Accepted February 24, 2020

ABSTRACT

The widespread use of Cas12a (formerly Cpf1) nucleases for genome engineering is limited by their requirement for a rather long TTTV protospacer adjacent motif (PAM) sequence. Here we have aimed to loosen these PAM constraints and have generated new PAM mutant variants of the four Cas12a orthologs that are active in mammalian and plant cells, by combining the mutations of their corresponding RR and RVR variants with altered PAM specificities. LbCas12a-RVRR showing the highest activity was selected for an in-depth characterization of its PAM preferences in mammalian cells, using a plasmid-based assay. The consensus PAM sequence of LbCas12a-RVRR resembles a TNTN motif, but also includes TACV, TTCV CTCV and CCCV. The D156R mutation in improved LbCas12a (impLbCas12a) was found to further increase the activity of that variant in a PAM-dependent manner. Due to the overlapping but still different PAM preferences of impLbCas12a and the recently reported enAsCas12a variant, they complement each other to provide increased efficiency for genome editing and transcriptome modulating applications.

INTRODUCTION

The identification and characterization of novel CRISPR–Cas systems expand the repertoire of effector proteins that can be used for gene, epigenome and base editing applications in eukaryotic cells (1,2). Beside the commonly used SpCas9, many other Cas effector nucleases with distinct properties have been successfully repurposed for genome engineering tasks (3–5). Among these several unique features highlight the Cas12a (formerly Cpf1) protein family that involves As- (*Acidaminococcus* sp.) and Lb- (*Lachnospiraceae bacterium*) Cas12a nucleases (3). These features include: (i) the utilization of a single, shorter crRNA; (ii) the recognition of T-rich PAM sequences; (iii) cleaving the target DNA at a PAM distal position and (iv) generating protruding DNA ends upon cleavage, as well as (v) having indiscriminate single stranded DNAse activity upon target binding (3,6). However, the most significant advantage of Cas12a nucleases is their ribonuclease activity which allows them to process their own crRNA from a longer precursor (3,7). This latter characteristic greatly simplifies their use for multiplex gene editing, transcription modulation and imaging, which tasks typically require simultaneous targeting of more sequences for efficient operation (8,9). In spite of these favourable features, their broader application is limited by the relatively rare occurrence of their long TTTV PAM sequence in mammalian genomes. The theoretically calculated frequency of the TTTV PAM motif in DNA sequences is 3/256, a number which is considerably smaller than the 16/256 frequency of the canonical NGG motif of SpCas9.

Searching for a robust Cas12a nuclease preferring a more frequent PAM sequence is still ongoing. Two Cas12a orthologues, namely Fn- (*Francisella novicida* U112) and Mb-Cas12a (*Moraxella bovoculi* 237), were reported to require NTTN PAM sequences *in vitro* (3); however, this finding has been confuted, and the PAM sequence was found to be stricter in mammalian cells (10,11). Another approach for loosening the PAM constrains of Cas12a orthologues is the generation of mutant variants with al-

*To whom correspondence should be addressed. Tel: +36 1 382 6610; Email: welker.ervin@ttk.hu

tered PAM preferences. Such mutant variants, AsCas12a-RVR and AsCas12a-RR were generated, and were shown to exhibit TATV and TYCV PAM preferences, respectively (12,13). It was shown that these mutants also preserved the canonical TTTV PAM preference of the AsCas12a-WT in mammalian cells (10,12,13), thus they are characterized by a PAM frequency of 6/256 (RVR) or 9/256 (RR), respectively. Mutants analogous to the other (Lb-, Fn-, Mb-) Cas12a orthologues were also generated, and were found to exhibit similarly loosened PAM constrains, but slightly different target preferences. In all orthologues, the RVR and RR mutations seem to alter the preferences primarily at the third and second position of the PAM sequence (10). When multiplexing with Cas12a, a single variant that can access all PAM sequences available for the WT and the RVR and RR mutants would be advantageous. Being inspired by this challenge, we have combined these mutations in order to create a more potent Cas12a nuclease variant.

MATERIALS AND METHODS

Materials

Restriction enzymes, Taq DNA polymerase (recombinant), Platinum Taq DNA polymerase and T4 ligase were purchased from Thermo Fischer Scientific. DNA oligonucleotides were acquired from Sigma-Aldrich. All DNA constructs were verified by Sanger sequencing (Microsynth AG). Plasmids were purified with the GenElute HP Plasmid Miniprep kit (Sigma-Aldrich). Q5 polymerase was purchased from New England BioLabs Inc. Dulbecco's modified Eagle's Medium, foetal bovine serum, penicillin and streptomycin were acquired from Thermo Fisher Scientific. 2 mm electroporation cuvettes were acquired from Cell Projects Ltd. and Bioruptor 0.5 ml Microtubes for DNA Shearing from Diagenode. Agencourt AMPure XP beads were purchased from Beckman Coulter Inc. T4 DNA ligase (for GUIDE-seq) and end-repair mix were acquired from Enzymatics Inc. KAPA universal qPCR Master Mix was purchased from KAPA Biosystems.

The following plasmids were gifts from Jia Chen (Addgene number: pCMV-dCpf1-BE #107685) (14), from Feng Zhang (Addgene number: pcDNA3.1-hLbCpf1 #69988, pcDNA3.1-hAsCpf1 #69982) (3) and from Ron Vale (Addgene number: pHRdSV40-scFv-GCN4-sfGFP-VP64-GB1-NLS #60904) (15).

Plasmid construction

Vectors were constructed using standard molecular biology techniques. For details see Supplemental File 1—Plasmid Construction and Supplementary Table S1. The sequences of DNA oligonucleotides utilized in this study are listed in Supplementary File 2. Deposited plasmids are summarized in Supplementary File 1—Supplementary Table S2.

Cell culturing

N2a (Neuro-2a mouse neuroblastoma cells, ATCC – CCL-131), HEK293 (293 H, Gibco), HEK293.EFGFP and HEK293.clover cells were grown at 37°C in a humidified atmosphere of 5% CO_2 content in high glucose Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% heat inactivated foetal bovine serum, 4 mM L-glutamine (Gibco), 100 units/ml penicillin and 100 μg/ml streptomycin.

HEK293.EGFP cells were generated, maintained and transfected as described in reference (16). HEK293.clover cells were generated as follows: a modified clover sequence was inserted after the ATG of human prion protein coding sequence with a self-cleaving donor plasmid by NHEJ-knock-in (17). The sequence of the inserted clover sequence is given in Supplementary File 1.

GFxFP assay

The modified GFxFP assay was carried out as described in reference (10,18). Spacer sequences and vectors used are listed in Supplementary File 2.

Cells were seeded onto 48-well plates a day before transfection at a density of 3×10^4 cells/well. The next day, at around 40% confluence, cells were transfected with plasmid constructs using Jetfect reagent (Biospiral-2006 Ltd.), briefly as follows: 250 ng total plasmid DNA (2 ng GFxFP plasmid, 124 ng crRNA and nuclease expression plasmid, and 124 ng mCherry expression plasmid to monitor the transfection efficiency) and 1 μl Jetfect were mixed in 50 μl serum free DMEM and the mixture was incubated for 30 min at room temperature prior being added to cells. Three parallel transfections were made from each sample. Cells were analysed by flow cytometry two days post-transfection. The EGFP signal was analysed in the mCherry positive population. Background fluorescence was determined by using a crRNA-less, inactive LbCas12a nuclease expression vector as negative control and this background value was subtracted from each sample. Three parallel transfections were made for each case. Activities of the Cas12a-WT nucleases on targets with TTTC PAM sequence are counted as 100%.

Disruption assay

Disruption assay was carried out as described in reference (10). Spacer sequences and vectors used are listed in Supplementary File 2.

All EGFP disruption experiments were conducted on HEK293.EGFP or HEK293.clover cells. Cells were plated one day prior to transfection on 48-well plates at a density of ∼25 000–30 000 cells/well. Cells were co-transfected with two types of plasmids: Cas12a variant expression plasmid (137 ng) and crRNA and mCherry coding plasmid (97 ng) using 1 μl Jetfect reagent per well in 48-well plates. Transfected cells were analysed six days post-transfection by flow cytometry. For control an inactive LbCas12a expressing nuclease vector was used. Disruption activities of the nucleases were calculated as follows: 1 – (sample GFP % / average GFP % of negative controls). Transfection efficiency was calculated via mCherry expressing cells. Transfections were performed in triplicate.

Base editor assay

Base-editor assay was carried out as follows: HEK293 cells were seeded onto 48-well plates one day before transfection

at a density of 1×10^4 cells/well. The next day, at around 25% confluence, cells were transfected with plasmid constructs using Jetfect reagent, briefly as follows: 350 ng total plasmid DNA (145 ng crRNA and mCherry expression plasmid, plus 205 ng nuclease expression plasmid) and 1.5 μl Jetfect reagent were mixed in 50 μl serum free DMEM, and the mixture was incubated for 30 min at room temperature prior being added to cells. Three parallel transfections were made from each sample. Transfection efficiency was analysed by flow cytometry seven days post-transfection via mCherry fluorescence. Next, cells were centrifuged at 100 g for 10 min and genomic DNA was purified according to the Puregene DNA Purification protocol (Gentra systems). PCR primers used to amplify on- and off-target genomic sequences are listed in Supplementary File 2. PCR products were purified by the Nucleospin Gel Extract Kit (Macherey-Nagel) and sent for Sanger sequencing (Microsynth). Sequencing data was analysed by EditR (19). Spacer sequences and vectors used are listed in Supplementary File 2.

Flow cytometry

Flow cytometry analysis was carried out on an Attune Acoustic Focusing Cytometer (Applied Biosystems by Life Technologies). For data analysis the Attune Cytometric Software was used. In all experiments, a total of 10 000 viable single cells were acquired and were gated based on side and forward light-scatter parameters. Cells expressing EGFP and mCherry from a control plasmid were used to adjust the parameters for the identification of EGFP and mCherry positive cells in the samples. The EGFP signal was detected using a 488 nm diode laser for excitation and a 530/30 nm filter for emission. The mCherry signal was detected using a 561 nm diode laser for excitation and a 620/15 nm filter for emission.

Next generation sequencing

HEK293 cells were seeded onto 48-well plates one day before transfection at a density of 1×10^4 cells/well. The next day, at ∼25% confluence, cells were transfected with plasmid constructs using Jetfect reagent (Biospiral-2006. Ltd.), briefly as follows: 234 ng total plasmid DNA (97 ng crRNA and mCherry expression plasmid, plus 137 ng nuclease expression plasmid) and 1 μl Jetfect reagent were mixed in 50 μl serum free DMEM, and the mixture was incubated for 30 min at room temperature prior to being added to cells. Three parallel transfections were made from each sample. Transfection efficiency was analysed by flow cytometry 5 days post-transfection via mCherry fluorescence. Next, cells were centrifuged at 100 g for 10 min and genomic DNA was purified according to the Puregene DNA Purification protocol (Gentra systems). Amplicons for deep sequencing were generated in two rounds of PCR using Q5 high-fidelity polymerase to attach Illumina handles. The first step PCR primers used to amplify target genomic sequences are listed in reference (10). After being quantified with the Qubit dsDNA HS Assay kit (Invitrogen), PCR products were pooled for deep sequencing. Sequencing on an Illumina Miseq instrument was performed by ATGandCo Ltd. Indels were counted computationally among reads that matched at least 75% to the first 20 bp of the reference amplicon. Indels and mismatches were searched at ±60 bp around the cut site. For each sample, the indel frequency was determined as (number of reads with an indel)/(number of total reads). The following software were used: BBMap 38.08, samtools 1.8, BioPython 1.71, PySam 0.13. SRA accession: PRJNA590635. Spacer sequences and vectors used are listed in Supplementary File 2. NGS raw data are available in Supplementary File 3.

GUIDE-seq

GUIDE-seq experiments were performed with LbCas12a-WT and LbCas12a-RVRR for seven different crRNAs. Briefly, 3×10^6 HEK293.EGFP cells were transfected with 2.6 μg of LbCas12a variant expressing plasmid and 1.9 μg of mCherry and crRNA coding plasmid. 100 pmol of the dsODN containing phosphorothioate bonds at both ends [according to the original GUIDE-seq protocol (20)] was mixed with 100 μl home-made nucleofection solution and was added to the plasmids, and it was electroporated as described in Vriend *et al.* (21) using Nucleofector (Lonza) with A23 program and 2 mm electroporation cuvettes. Transfected cells were analysed ∼72 h post-transfection by flow cytometry. Cells were then centrifuged at 1000 g for 10 min, and genomic DNA was purified according to the Puregene DNA Purification protocol (Gentra systems). Genomic DNA was sheared with BioraptorPlus (Diagenode) to 550 bp in average. Sample libraries were assembled as previously described (20) and sequenced on an Illumina MiSeq instrument by ATGandCo Ltd. Data were analysed using the open-source guideseq software (version 1.1) (22). Consolidated reads were mapped to the human reference genome GrCh37 supplemented with the integrated EGFP sequence. Upon identification of the genomic regions integrating double-stranded oligodeoxynucleotides (dsODNs) in aligned data, off-target sites were retained when no more than seven mismatches compared to the on-target site were present and when they were absent in the background controls. Visualization of aligned off-target sites are provided as a color-coded sequence grid. Further details are available in Supplementary File 4. SRA accession: PRJNA590635. Spacer sequences and vectors used are listed in Supplementary File 2.

TIDE

Tracking of Indels by DEcomposition (TIDE) method (23) was applied for analysing mutations and determining their frequency in a cell population using different crRNAs and LbCas12a variants. From the isolated genomic DNA PCR was conducted with Q5 High-Fidelity DNA Polymerase in triplicates (for PCR primer details, see Supplementary File 4). Genomic PCR products were gel excised via NucleoSpin Gel and PCR Clean-up kit and the purified products were Sanger sequenced. Indel efficiencies were analysed by TIDE webtool (https://tide.nki.nl/) by comparing Cas12a treated and control samples.

Statistical analysis

Differences between LbCas12a variants were tested by Student's *t* test for paired-samples. Normality of the data and of the differences were tested by Shapiro–Wilk Normality Test. Statistical tests were performed on data including all parallel sample points, using the IBM SPSS version 20 software.

Upregulation, RT-PCR

N2a cells were seeded onto 6-well plates one day before transfection at a density of 2×10^5 cells/well. The next day, at ∼25% confluence, cells were transfected with plasmid constructs using Turbofect reagent (Thermo Fischer Scientific), briefly as follows: 4000 ng total plasmid DNA (1,200 ng crRNA and mCherry expression plasmid, 1200 ng scFv-BFP-VP64 and 1600 ng dead nuclease-SunTag plasmid) and 6 μl Turbofect reagent were mixed in 400 μl serum free DMEM, and the mixture was incubated for 30 min at room temperature prior to being added to the cells. Three parallel transfections were made from each sample. Cells were harvested 2 days post-transfection. A small proportion was used to measure transfection efficiency by flow cytometry utilising mCherry and BFP fluorescence. The rest of the cells were centrifuged at 100 g for 5 minutes, and total RNA was extracted using RNAzol-RT (Sigma-Aldrich). Genomic DNA contamination was digested by DNAseI (Thermo Fischer Scientific). After DNAseI inactivation, RNA was purified using Nucleospin RNA Clean-up XS (Macherey-Nagel). Reverse transcription was performed by SuperScript III Reverse transcriptase (Thermo Fischer Scientific) using 1300 ng purified RNA template. Real-Time PCR was performed using SYBR Green Master Mix (Thermo Fischer Scientific) on a StepOnePlus™ platform (Thermo Fischer Scientific). Results are expressed as fold change above no-target crRNA control cells after being normalized to HPRT1 expression using the ΔΔCt method. All related information is listed in Supplementary File 2.

RESULTS AND DISCUSSION

The four mutations of the RVR and RR variants of AsCas12a [S542R, K548V, N552R and K607R, (12)] are combined resulting in the AsCas12a-RVRR variants. The four mutations at the analogous positions of the RVR and RR variants of Lb-, Fn- and MbCas12a have also been introduced into the sequences of these three nucleases creating the corresponding RVRR mutants (Figure 1A and B) (10,12). Their activities and PAM preferences were compared in the plasmid-based GFxFP assay described previously (18,24). We prefer this mammalian system for determining PAM preferences over bacterial or *in vitro* selections, on the one hand, since these nucleases have been reported to exhibit different PAM preferences when tested *in vitro* and in mammalian cells (10). On the other hand, testing mammalian genomic targets with various PAMs to reveal the PAM sequence requirements of RNA-guided nucleases is burdened with the uncertainties of the varying cleavage efficiencies of the different target sequences involved. By contrast, in the GFxFP system the PAM preferences are tested on the same spacer sequence, and thus, the differences in the cleavage activities of nucleases can be more unambiguously assigned to the PAM sequences examined. Since the PAM preferences of the Cas12a nucleases show some sequence dependence (10), in this study we have explored at least four different targets to establish PAM preferences that are more generally valid.

Figure 1C shows a comparison of the activities of the RVRR variants of all four Cas12a nucleases relative to the corresponding WT, RVR and RR mutants on targets with their optimal PAM sequences: TTTC to the WT, TATC to the RVR and TCTC/TTCC/TCCC to the RR variants (3,10,12,13,25,26). Introducing the RVRR mutations to As- and FnCas12a was found to abolish their activity on the PAM sequences examined. In contrast, Mb- and LbCas12a-RVRR mutants were found to exhibit considerable activities, the LbCas12a-RVRR variant being compatible with most of these PAM sequences, while exhibiting lower activity on TCTC PAM (Figure 1C, Supplementary Figures S1–S4). LbCas12a has further advantages including that it exhibits the highest activity among these four Cas12a nucleases in certain experimental systems (10,18,25,27,28). Furthermore, it remains active at lower temperatures which makes it the choice of Cas12a nuclease for gene editing and transcription activations in plant and ectotherm animals (28–32). Based on these characteristics, we have decided to subject LbCas12a-RVRR to a deeper investigation.

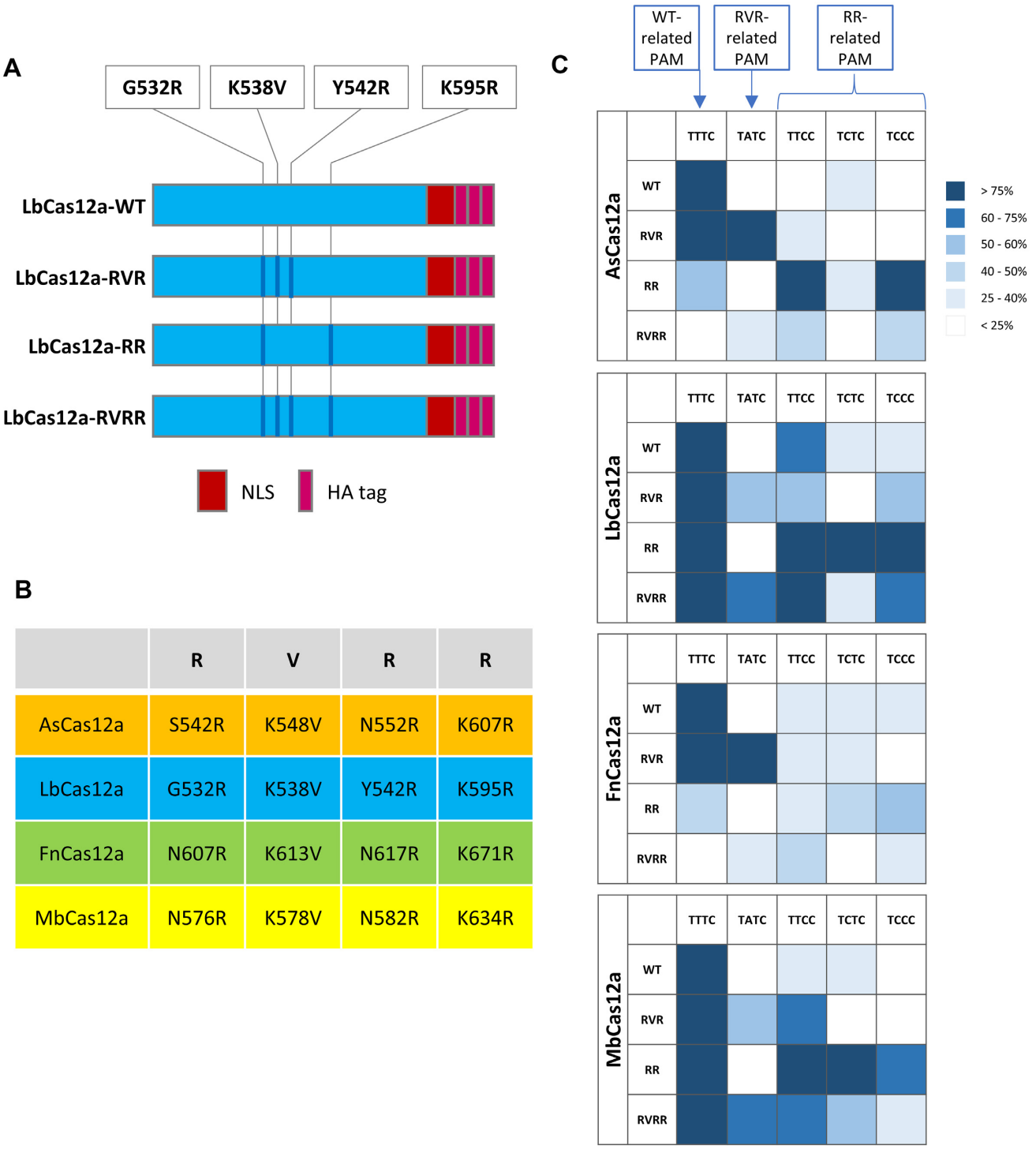
First, we have analysed the preferences of LbCas12a-RVRR at all combinations of PAM positions –3 and –1 on four targets (Figure 2A and B, Supplementary Figure S5). For clarity, we discuss the results according to four panels keeping position –3 fix and varying the four possible nucleotides at position –1.

TTTN PAMs: LbCas12a-WT shows robust activity on its optimal PAM sequences (TTTV, where V is A, C or G) and lower but considerable activity on TTTT, which is consistent with former reports on LbCas12a disfavouring the T nucleotide at position -1 (3,10,25,33). The RVRR variant works with all the four nucleotides including 'T' at the –1 position (TTTN PAM) and shows highest activity with a –1 'A' nucleotide (TTTA) (Figure 2A, first panel, Supplementary Figure S5).

TATN PAMs: Activity profiling of targets with TATN PAM sequences has revealed that the RVR mutant demonstrates robust activity on TATV sequences, while the RVRR variant shows activity on all the four sequences (TATN PAM), although its activity is higher with -1 R nucleotides (TATA and TATG PAMs) (Figure 2A, second panel, Supplementary Figure S5).

TCTN PAMs: As it was previously reported, –3 C in PAM sequence is disfavoured by the WT and the RVR variant; however, LbCas12a-RR demonstrates robust activity with TCTN PAMs (10). The RVRR variant shows moderate activity on TCTV sequences, while its activity is negligible with a –1 'T' (Figure 2A, third panel, Supplementary Figure S5).

TGTN PAMs: The analysis of the effect of –3 G in PAM sequence has revealed an unexpectedly high activity of the RVRR variant of LbCas12a [reaching the highest activity with M at position -1 (TGTA or TGTC)], whereas the other variants, i.e. WT, RVR and RR were found to show decreased or negligible cleavage efficiencies. No considerable



B

	R	V	R	R
AsCas12a	S542R	K548V	N552R	K607R
LbCas12a	G532R	K538V	Y542R	K595R
FnCas12a	N607R	K613V	N617R	K671R
MbCas12a	N576R	K578V	N582R	K634R

Figure 1. Activity of Cas12a variants on targets with candidate PAM sequences measured in the GFxFP assay. (**A**) Schematics depicting the main features of the WT and three mutant variants of Cas12a used: each protein sequence is flanked by a nuclear localization signal (NLS) and is preceded by a 3xHA tag. The RVRR variant combines all the mutations included in RVR and RR Cas12a. (**B**) Mutations used for RVR and RR variants are listed for all four Cas12a orthologues in the table (orange – AsCas12a, blue – LbCas12a, green – FnCas12a, yellow – MbCas12a). (**C**) Average activity of WT As-, Lb-, Fn- and MbCas12a orthologues and their mutant variants (RVR, RR, RVRR) on four targets with TTTC, TATC, TTCC, TCTC and TCCC PAM sequences, measured in the GFxFP assay. See also Supplementary Figure S1–4.

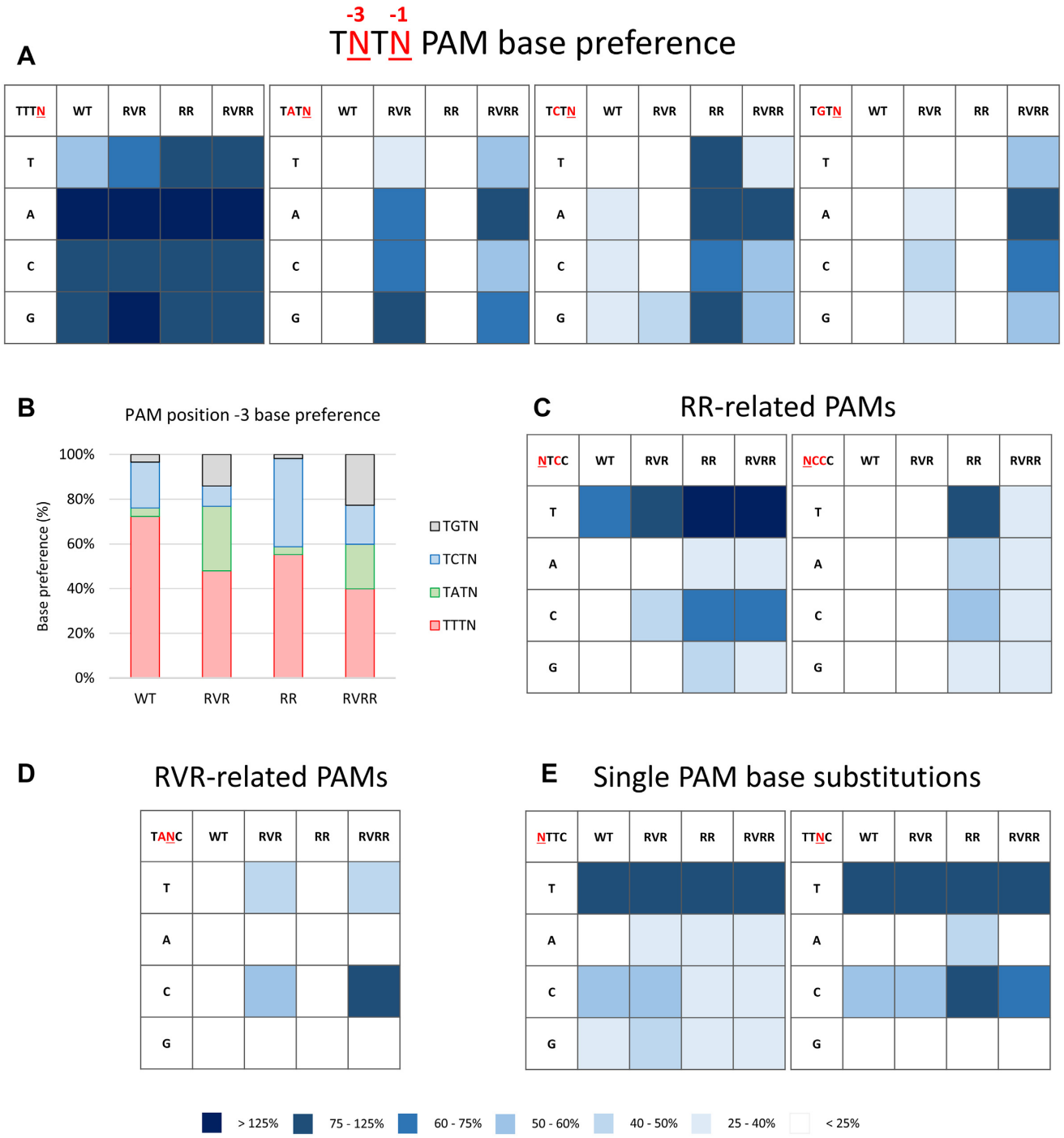


Figure 2. Activity of LbCas12a variants on targets with candidate PAM sequences measured in the GFxFP assay. (**A**) Average activity of LbCas12a-WT and its mutant variants (RVR, RR, RVRR) on four targets with TNTN PAM sequences measured in the GFxFP assay (N: A, G, C or T). See also Supplementary Figure S5. (**B**) PAM position -3 base preference of LbCas12a variants. See also Supplementary Figure S5. (**C**) Average activity of LbCas12a-WT and its mutant variants (RVR, RR, RVRR) on four targets with additional RR-related (NTCC and NCCC) PAM sequences measured in GFxFP assay (N: A, G, C or T). See also Supplementary Figures S6 and S7. (**D**) Average activity of LbCas12a-WT and its mutant variants (RVR, RR, RVRR) on four targets with additional RVR-related (TANC) PAM sequences measured in the GFxFP assay (N: A, G, C or T). See also Supplementary Figure S8. (**E**) Average activity of LbCas12a-WT and its mutant variants (RVR, RR, RVRR) on four targets with additional single PAM base substitutions (NTTC and TTNC PAM sequences) measured in the GFxFP assay (N: A, G, C or T). See also Supplementary Figures S9 and S10.

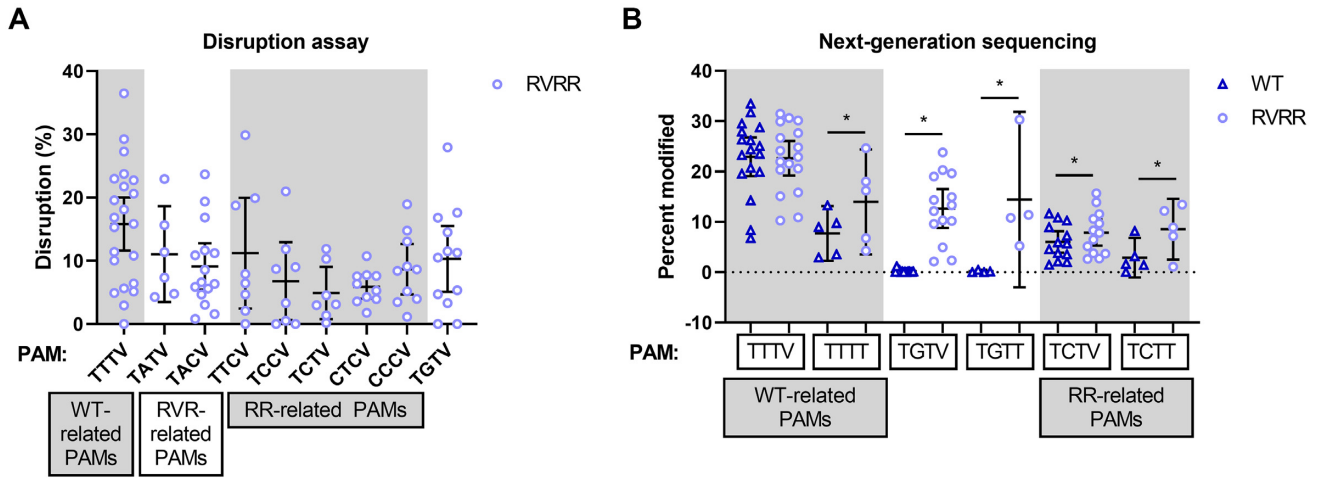


Figure 3. Modifications mediated by the LbCas12a-RVRR variant on targets with candidate PAM sequences. (**A**) Disruption mediated by the LbCas12a-RVRR variant on targets with TTTV (WT-related PAMs), TATV, TACV (RVR-related PAMs), TTCV, TCCV, TCTV, CTCV, CCCV (RR-related PAMs), or TGTV PAM sequences in HEK293.EGFP or HEK293.clover cells (V: A, G or C). The means and interquartile ranges are shown; data points are plotted as open circles representing the mean of biologically independent triplicates. See also Supplementary Figures S11–S19. (**B**) Indels mediated by the LbCas12a-WT and LbCas12a-RVRR variant on human endogenous targets with TTTN (WT-related PAMs), TGTN or TCTN (RR-related PAMs) PAM sequences in HEK293 cells (N: A, G, C or T). The means and interquartile ranges are shown; data points are plotted as open circles representing the mean of biologically independent triplicates. See also Supplementary Figures S20–S22.

activity of either the WT or the RVR or RR mutants of LbCas12a on targets with a 'G' nucleotide at PAM position -3 have been reported in earlier studies (Figure 2A, fourth panel, Supplementary Figure S5) (12,13).

We have also examined some motifs related to TTCC/CTCC/TCCC and TAYC PAM sequences on which RR or RVR LbCas12a variants, respectively, exhibit some activities (10,12,13). These screens have revealed three additional PAM motifs on which the RVRR variant shows robust activity, namely TACC, TTCC and CTCC PAMS (Figure 2C and D, Supplementary Figures S6–S8). Inspired by the observation that the RVRR mutant is active on PAM sequences, on which the former mutants are not, we have examined their preferences at positions -2 and -4 on NTTC and TTNC PAM motifs. However, no newer PAM preferences were revealed in these screens, although the TTTC and TTCC motives of RVRR were confirmed (Figure 2E, Supplementary Figures S9 and S10).

Altogether, the RVRR variant seems to unite the PAM preferences of both the WT and the RVR and RR variants examined, except for the TCTT PAM of the LbCas12a RR mutant. It demonstrates activity on 15 of the 16 possible combinations of TNTN PAM sequences, although it preserves a slightly higher preference for a –3 'T' nucleotide (Figure 2B). Thus, the PAM of the LbCas12a-RVRR variant resembles the TNTN consensus PAM sequence with a permissiveness for –2 'C' in some combinations.

The ultimate aim of the development of new variants is to be applied on genomic target sequences. Here, we aimed to confirm the above described PAM preferences of the RVRR variant on genomic targets, although this approach is plagued by the high variability of the LbCas12a nuclease activities on diverse target sequences. EGFP disruption has been shown to be a versatile and effective approach to test the activity of the SpCas9 nuclease, utilizing a copy of EGFP sequence integrated into the genome (34,35). However, the rarity of the 'TTT' motif strongly limits its applicability for Cas12a nucleases, since only two targets are found in the entire EGFP sequence (10,11). To overcome this limitation, we have generated a HEK293 cell line containing an integrated copy of clover, that is codon optimized to harbour a great number of TTT motifs for the Cas12a nucleases. We have determined the disruption activity of the LbCas12a-RVRR mutant on 98 targets. As shown in Figure 3A, we have categorized the targets according to the PAM motifs, most of them differing in the –2 and/or –3 positions, to form groups that are related to the preferences of the WT or that of the RR or RVR mutants. TGTV PAMs are new to LbCas12a variants therefore they are categorised as a new, distinct PAM group.

In line with the observation that Cas12a nucleases are more 'pickier' in their target selection than SpCas9 is (18,36), LbCas12a-RVRR was found to show a relatively weak activity on several targets in almost all PAM groups. The highest activity is evident with the TTTV PAM, and a moderately higher activity was detected with TAYV, TGTV TTCV, CCCV and TCCV compared to the CTCV and TCTV sequences (Figure 3A, Supplementary Figures S11–S19), a finding consistent with the results obtained in the GFxFP assay.

To make a meaningful comparison of the activity of the LbCas12a-WT and the LbCas12a-RVRR mutant we give a side by side presentation of their indel percentages revealed by next-generation sequencing separately on 17 targets with TTTV PAMs that are optimal for the WT protein, as well as on 5 targets with TTTT PAM sequence on which the WT protein shows decreased activity (Figure 3B). The RVRR mutations do not reduce the activity of LbCas12a on targets with TTTV PAM, but significantly increase its activity on targets with TTTT PAM, consistent with the PAM preferences discerned in the GFxFP assay (Figures 2A and 3B, Supplementary Figure S20). We have also compared their

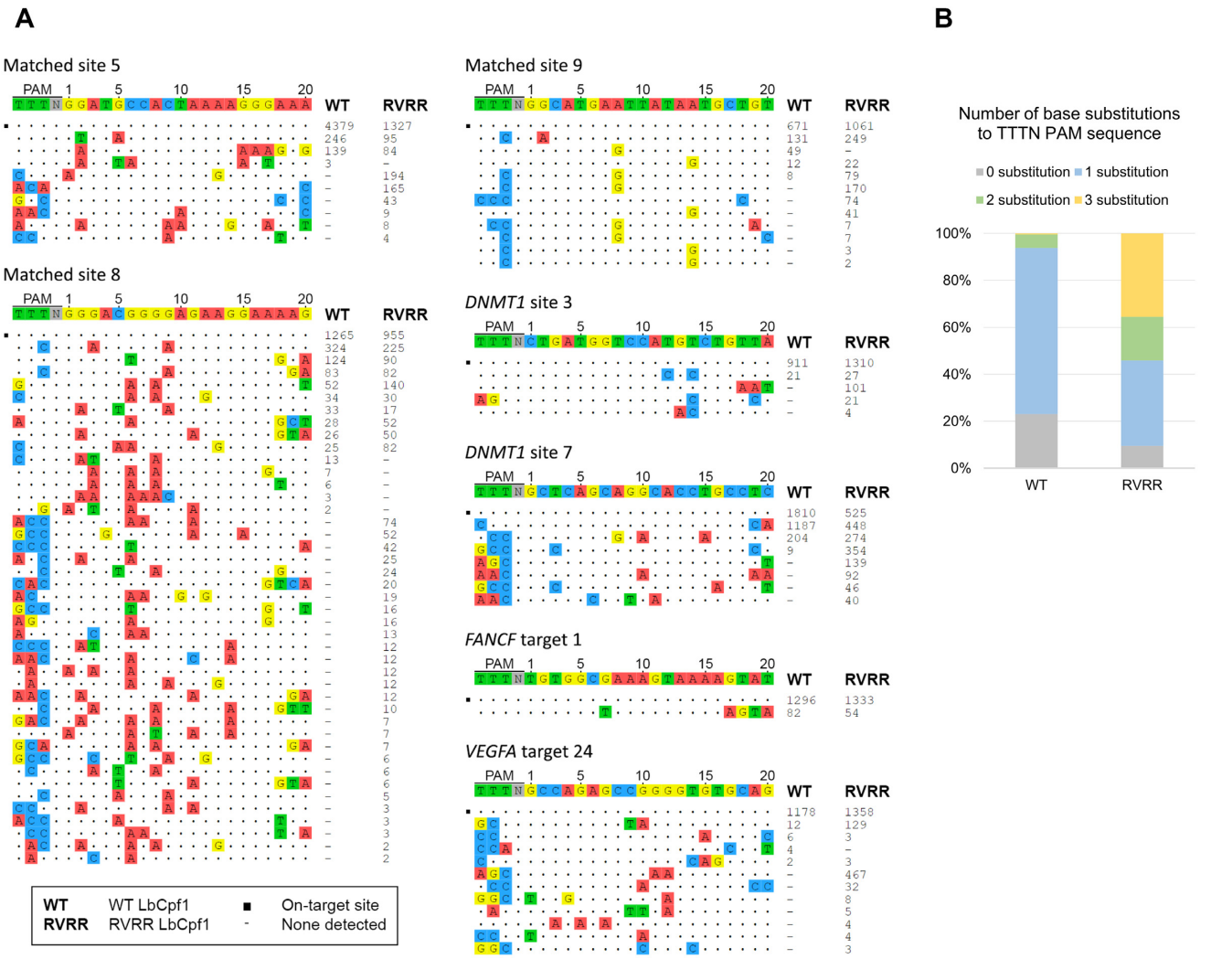


Figure 4. Genome-wide specificities of the LbCas12a-WT and the LbCas12a-RVRR variant. (**A**) Off-target sites for the LbCas12a-WT and the LbCas12a-RVRR variant with seven different crRNAs determined using GUIDE-seq in HEK293 cells. Mismatches compared to the on-target sites are highlighted in colour, and GUIDE-seq read counts are shown on the right of the on- and off-target sequences representing a measure of cleavage efficiency at a given site. See also Supplementary File 4. (**B**) Distribution of the number of base substitutions to the WT-related TTTN PAM sequence of LbCas12a-WT and LbCas12a-RVRR off-targets.

activities on 18–18 targets with TCTN or TGTN PAMs. As expected, the mutations increased the indel inducing activities significantly on targets with TGTV and TGTT PAMs and slightly on targets with TCTV and TCTT PAMs (Figure 3B, Supplementary Figures S21 and S22).

We were also curious to see the fidelity of this variant compared to the WT nuclease, thus conducted GUIDE-seq genome wide off-target analyses (20) on 7 endogenous targets with TTTV PAM sequence in HEK293 cells. Five of seven targets had been tested previously, and had been found to have off-target sites with LbCas12a-WT (27). Here we revealed that the RVRR variant cleaves at more off-target sites than the WT nuclease does, except for the *FANCF* target 1 (Figure 4A, Supplementary Figure S23A). While the LbCas12a-WT cleaves off-target sequences typically having the canonical TTTV PAM or with single T to C substitutions in its PAM sequence, the RVRR variant cleaves several additional off-target sites having PAM sequences differing in up to all three thymidines (Figure 4A and B, Supplementary Figure S23). This finding is in line with the more permissive PAM requirements of this mutant and suggests that the increased off-target propensity is primarily the consequence of the wider sequence space available to this variant. Nevertheless, if we restrict the comparison to the off-target sites of LbCas12a-WT, a slight increase in the off- versus on-target reads for LbCas12a-RVRR is discernible (Supplementary Figure S23B).

Regarding the general shortage of TTTV PAM motifs in appropriate distances from targeted genomic positions, the expanded target space offered by the RVRR mutant is especially advantageous for using it in combination with base editors (14,37). We have replaced the inactive LbCas12a-WT with the inactive RVRR mutant in the dLbCas12a-WT-BE construct [(dCpf1-BE) (14)] and have compared their activities of C to T editing on six endogenous targets. As expected, the two base editors exhibited similar efficiencies

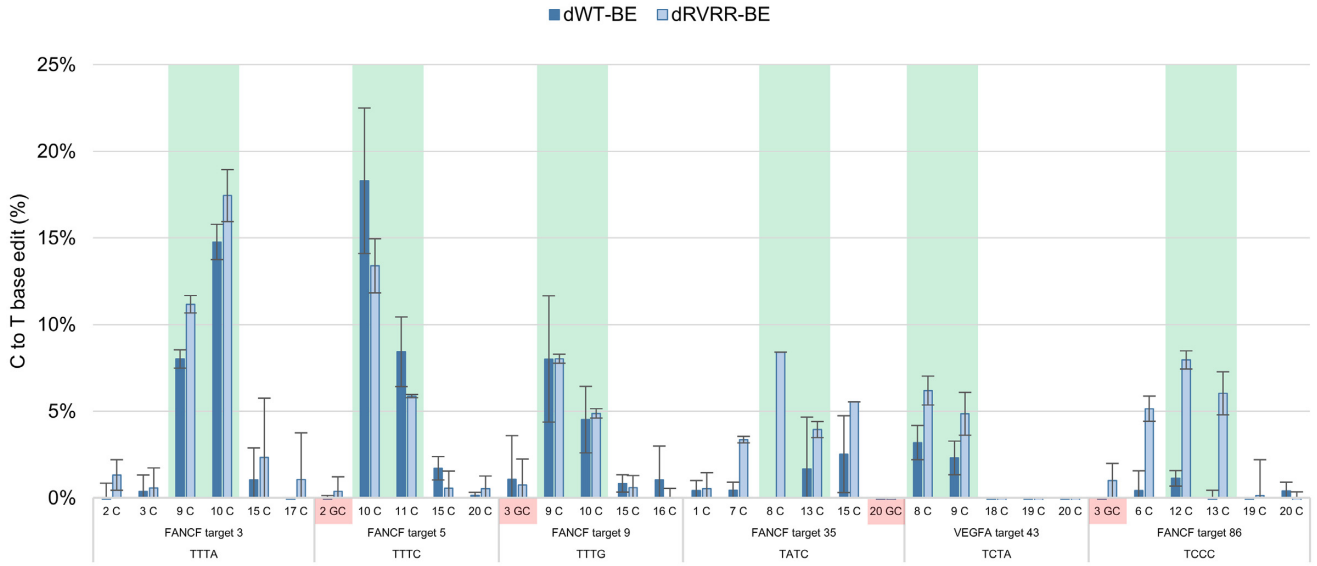


Figure 5. Comparison of C to T base editing efficiency of dLbCas12a-WT-BE and dLbCas12a-RVRR-BE in HEK293 cells. C to T base editing efficiency of dLbCas12a-WT-BE and dLbCas12a-RVRR-BE on six human endogenous targets with TTTV (WT-related PAMs), TATC (RVR-related PAMs), or TCTV and TCCC (RR-related PAMs) PAM sequences (V: A, G or C). C to T base editing efficiencies were analysed by EditR (19). The position of all candidate cytosines are shown. GC with red background: a guanine precedes the candidate cytosine. Green background shows the editing window determined by Li *et al.* (14). Bars correspond to the average of 3 parallel samples. See also Supplementary Figure S24.

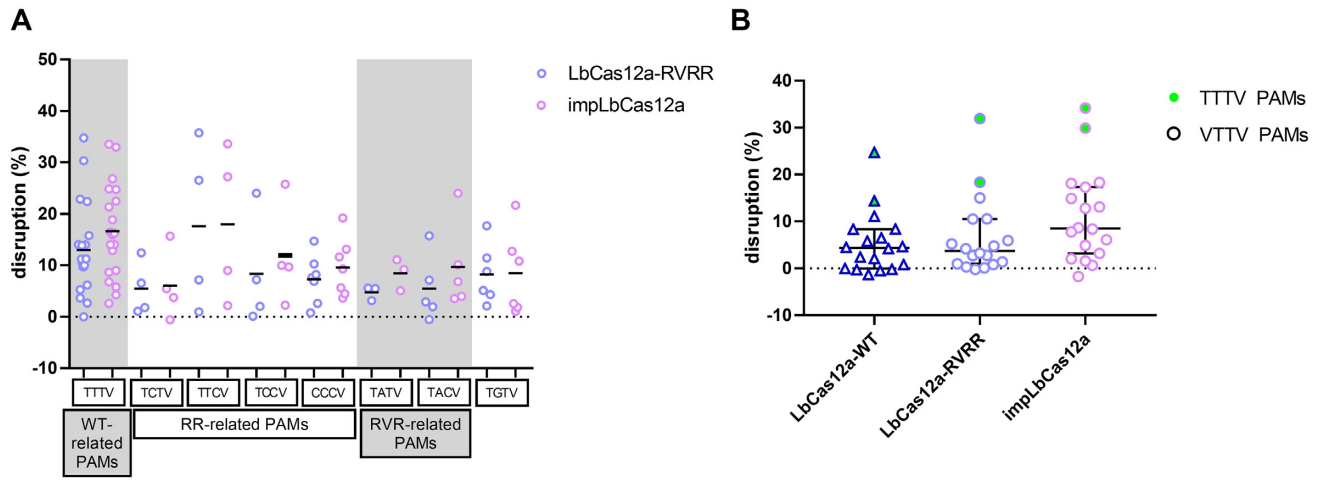


Figure 6. Effect of the D156R mutation on the activity of the LbCas12a-RVRR variant Disruption mediated by the LbCas12a-RVRR and the impLbCas12a variants on targets (**A**) with TTTV (WT-related PAMs), TCTV, TTCV, TCCV, CCCV (RR-related PAMs), TATV, TACV (RVR-related PAMs) and TGTV PAM sequences, (**B**) NTTV PAM sequences with in HEK293.clover cells (V: A, G or C; N: A, G, C or T). Green dots mark the two available targets with TTTV PAM sequences. The means and interquartile ranges are shown; data points are plotted as open circles representing the mean of biologically independent triplicates. See also Supplementary Figures S25 and S26.

on the targets with TTTV PAM sequences, while the RVRR variant showed higher editing on targets with TNTN PAMs (Figure 5). The C to T edits detected were mostly confined to the 8 to 13 editing window described for the dLbCas12a-WT-BE (14); however, we have noticed some exceptions (Figure 5, FANCF target 35 7C and FANCF target 86 6C). These results suggest that the dLbCas12a-RVRR-BE variant preserves the editing activity of the WT variant, but it is compatible with a broadened target range. We have also checked whether the fidelity of base editing mediated by dLbCas12a is altered by RVRR mutations. We have selected 21 predicted off-target sites (38) of the six targets shown in figure 5 containing up to four mismatches (Supplementary File 2), and analysed the rates of off-target base editing. Compared to the dLbCas12a-WT-BE, we have detected no increase in off-target editing for the RVRR variant on these sequences (Supplementary Figure S24). These results are in line with those of the GUIDE-seq analysis, showing minute differences between the fidelities of the WT and the RVRR nucleases on targets with canonical TTTV PAM sequences.

While we had been finalizing these analyses, the generation of AsCas12a variants with similarly loosened PAM constrains have been reported in an impressive and detailed study (39). It has been reported that the E174R mutation

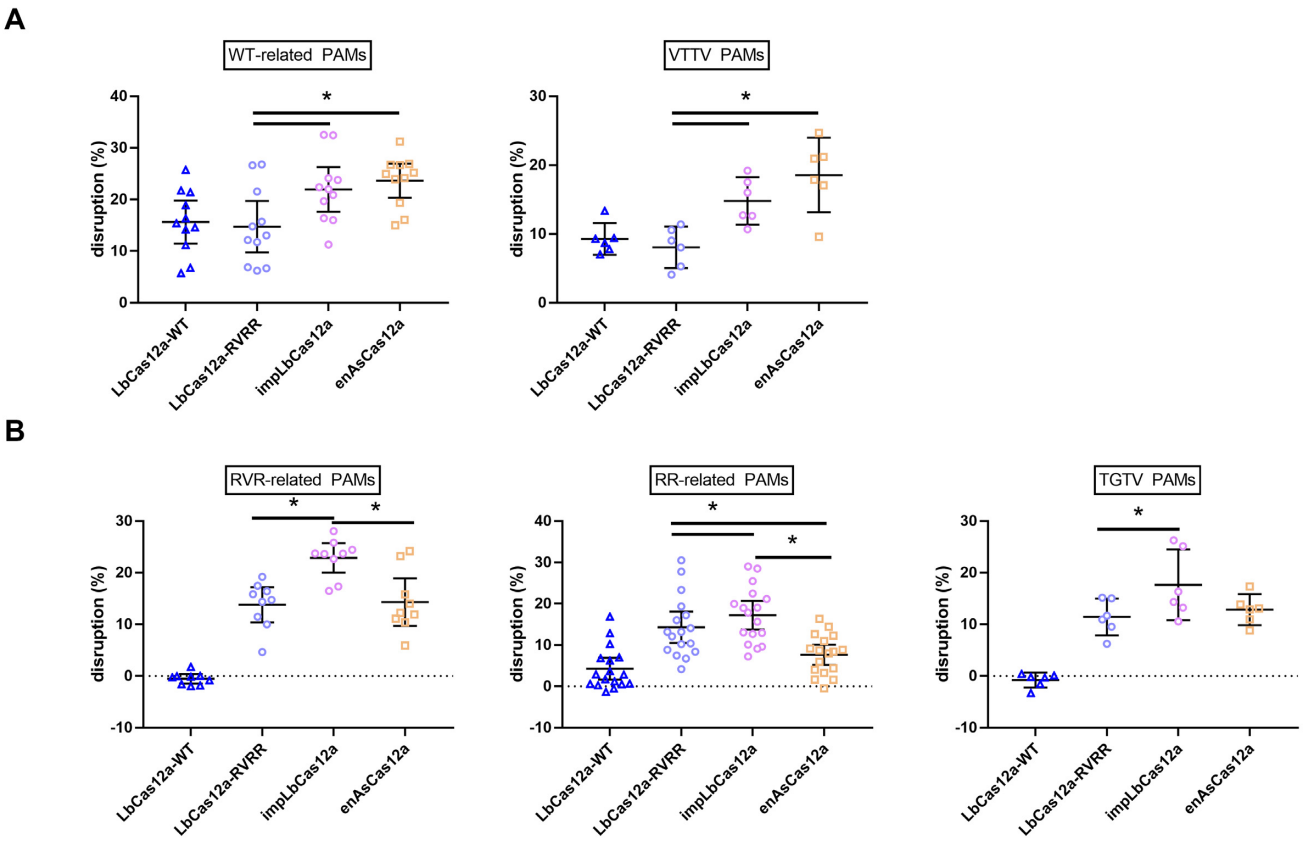


Figure 7. Comparison of the activities of Lb- and AsCas12a variants on targets with various PAM sequences in HEK293.GFP and HEK293.clover cells. Disruption mediated by the LbCas12a-WT, LbCas12a-RVRR, impLbCas12a and enAsCas12a nucleases on targets (**A**) with TTTV (WT-related PAMs) and VTTV PAM sequences, (**B**) with TAYV (RVR-related PAMs), YYYV (RR-related PAMs) and TGTV PAM sequences in HEK293.clover and HEK293.GFP cells (V: A, G or C). Means and interquartile ranges are shown; data points are plotted as open circles representing the mean of biologically independent triplicates. See also Supplementary Figures S27 and S28.

considerably enhances the activity of the AsCas12a variants to about two-fold of the WT's activity on targets with TTTV PAM sequences (39). We have introduced the analogous mutation (D156R) to LbCas12a-RVRR to create an improved LbCas12a variant (termed impLbCas12a, possessing overall the mutations D156R, G532R, K538V, Y542R and K595R), and we have tested whether this mutation in LbCas12a also results in an increased activity. Although an overall enhancing effect may be discernible in average (1.3-fold) primarily due to its enhanced activity on TTTV, TCCV, CCCV, TATC and TACV PAMs (the highest increase being 1.9-fold for TACV and 1.7-fold for TATV), it lags behind the degree observed in the case of AsCas12a on targets with TTTV PAM sequences (Figure 6A and Supplementary Figures S25 and S26).

We have also realized that this mutation makes the AsCas12a variants more permissive for NTTV PAM sequences (39), and we have tested the impLbCas12a mutant on targets with NTTV PAMs in the disruption assay. The D156R mutation of impLbCas12a seems to further broaden the PAM preferences of the variant to include NTTV PAM motifs (Figure 6B).

Next, we have compared the activity of enAsCas12a with that of the LbCas12a-RVRR and the impLbCas12a variants in disruption assays on 49 GFP and clover targets. We have focused either on those targets on which LbCas12a-RVRR exhibited higher activity (*i.e.* targets with TGTV PAMs or RR- and RVR-related PAMs) in our previous experiments (Figure 3A) or on those targets that we assumed to be preferred by enAsCas12a (*i.e.* targets with VTTV or TTTV PAMs). This latter PAM-altered variant was selected to tolerate NTTN PAMs and had an enhanced activity on TTTV PAMs (39). In line with our former experiments, and as it is apparent from Figure 7, RVRR mutations have been found to exert no increasing effect on targets with WT-related (TTTV) and VTTV PAMs, while greatly increasing the activity of LbCas12a on targets with TGTV, RR- and RVR-related PAMs. Addition of the D156R mutation does not seem to decrease the activity of the RVRR variant on any PAMs examined; however, it considerably increases its activity on targets with VTTV and RVR-related PAMs, offering a more versatile and efficient choice of Cas nucleases. Thus, these experiments have confirmed that the effects of the D156R mutation, namely loosening the PAM preferences of the Cas12a nuclease at the -4 position and enhancing its activity in a PAM- and target-dependent manner, seem to be transferable between As- and LbCas12a. Comparing the activities of the As- and LbCas12a variants, we have found that impLbCas12a does not surpass the activity of enAsCas12a on targets with NTTV PAMs

(Figure 7A and Supplementary Figure S27), but exceeds it on targets with RR- and RVR-related PAMs (Figure 7B and Supplementary Figure S28). Since impLbCas12a and enAsCas12a have partially overlapping but still different target and PAM preferences (10), they complement each other, and jointly they considerably increase the target space available for Cas12a nucleases.

The broadened target space of the LbCas12a variant developed here allows using multiple targets with different PAM sequences, which makes it easier to find targets in appropriate positions for transcriptional activation of genes. Thus, we have tested the inactive impLbCas12a variant in fusion with 10 consecutive SunTags to recruit transcription factors (40) to the promoter of the *SPRN* locus. We have used two sets of targets, one with canonical TTTV PAMs and one with alternative PAMs (Supplementary Figure S29A and B). As expected, the inactive impLbCas12a-SunTag variant leveraged the activity of the WT counterpart with both sets of targets (Supplementary Figure S29).

In summary, the PAM preference of the RVRR mutants of LbCas12a nuclease unite the PAM motifs on which the WT, RVR and RR variants demonstrate activities, and also involve new PAM motifs (TGTN) on which no LbCas12a variants have been demonstrated to exert substantial activities so far. Their target space is considerably broadened, thus the theoretical probability of finding a suitable PAM for the RVRR variants in DNA sequences increases to about three-fold compared to the RR variant, which, until now, has possessed the broadest PAM space among the LbCas12a nucleases. Our results suggest that the RVRR mutants offer superior alternatives to the WT, RVR or RR variants of LbCas12a in many applications, such as multiplex gene editing, transcriptome modulation, epigenetic editing and C to T base editing. We have shown that impLbCas12a is a useful alternative for enAsCas12a, owing to their different activities on various PAM sequences. Further advantages of the LbCas12a variants include their less pronounced temperature dependence compared to the AsCas12a nucleases, and high activity at lower temperatures (29,31,32). Thus, our findings suggest that the application of impLbCas12a is more advantageous than that of enAsCas12a for ectotherm animals and plants.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Ildikó Szűcsné Pulinka, Judit Szűcs, Dávid Fetter, Gábor Erdős, Orsolya Oravecz, Balázs Bohár for their excellent laboratory assistance. We thank Dora Bokor, PharmD, for proofreading the manuscript.

Author contributions: E.T. and E.W. conceived and designed the experiments, E.T., E.V., V.K.J., P.I.K., S.L.K., A.NY., Z. W., K.H., Z.L. and A.T. performed all experiments and analysed the data. E.T. and E.W. wrote the manuscript with input from all the authors.

FUNDING

National Research, Development and Innovation Office [K128188 to E.W., PD125331 to E.T and 2018-1.1.1-MKI-2018-00270]; Ministry of National Economy [VEKOP-2.1.7-17-2016-00383]; E.V. and Z.L. were supported by EFOP [3.6.3-VEKOP-16-2017-00009] grant. Funding for open access charge: National Research, Development and Innovation Office [K128188].

Conflict of interest statement. The authors are inventors on patent applications covering the LbCas12a variants described in this work.

REFERENCES

1. Bae,T., Hur,J.W., Kim,D. and Hur,J.K. (2019) Recent trends in CRISPR–Cas system: genome, epigenome, and transcriptome editing and CRISPR delivery systems. *Genes Genomics*, **41**, 871–877.
2. Adli,M. (2018) The CRISPR tool kit for genome editing and beyond. *Nat. Commun.*, **9**, 1911.
3. Zetsche,B., Gootenberg,J.S., Abudayyeh,O.O., Slaymaker,I.M., Makarova,K.S., Essletzbichler,P., Volz,S.E., Joung,J., van der Oost,J., Regev,A. *et al.* (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR–Cas system. *Cell*, **163**, 759–771.
4. Ran,F.A., Cong,L., Yan,W.X., Scott,D.A., Gootenberg,J.S., Kriz,A.J., Zetsche,B., Shalem,O., Wu,X., Makarova,K.S. *et al.* (2015) In vivo genome editing using Staphylococcus aureus Cas9. *Nature*, **520**, 186–191.
5. Hou,Z., Zhang,Y., Propson,N.E., Howden,S.E., Chu,L.F., Sontheimer,E.J. and Thomson,J.A. (2013) Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc. Natl. Acad. Sci. U.S.A.*, **110**, 15644–15649.
6. 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*, **360**, 436–439.
7. Fonfara,I., Richter,H., Bratovic,M., Le Rhun,A. and Charpentier,E. (2016) The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, **532**, 517–521.
8. Zetsche,B., Heidenreich,M., Mohanraju,P., Fedorova,I., Kneppers,J., DeGennaro,E.M., Winblad,N., Choudhury,S.R., Abudayyeh,O.O., Gootenberg,J.S. *et al.* (2016) Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat. Biotechnol.*, **35**, 31–34.
9. Sun,H., Li,F., Liu,J., Yang,F., Zeng,Z., Lv,X., Tu,M., Liu,Y., Ge,X., Liu,C. *et al.* (2018) A single multiplex crRNA Array for FnCpf1-Mediated human genome editing. *Mol. Ther.*, **26**, 2070–2076.
10. Tóth,E., Czene,B.C., Kulcsár,P.I., Krausz,S.L., Tálas,A., Nyeste,A., Varga,É., Huszár,K., Weinhardt,N., Ligeti,Z. *et al.* (2018) Mb- and FnCpf1 nucleases are active in mammalian cells: activities and PAM preferences of four wild-type Cpf1 nucleases and of their altered PAM specificity variants. *Nucleic Acids Res.*, **46**, 10272–10285.
11. Tu,M., Lin,L., Cheng,Y., He,X., Sun,H., Xie,H., Fu,J., Liu,C., Li,J., Chen,D. *et al.* (2017) A new lease of life': FnCpf1 possesses DNA cleavage activity for genome editing in human cells. *Nucleic Acids Res.*, **45**, 11295–11304.
12. 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.
13. Nishimasu,H., Yamano,T., Gao,L., Zhang,F., Ishitani,R. and Nureki,O. (2017) Structural basis for the Altered PAM recognition by engineered CRISPR-Cpf1. *Mol. Cell*, **67**, 139–147.
14. Li,X., Wang,Y., Liu,Y., Yang,B., Wang,X., Wei,J., Lu,Z., Zhang,Y., Wu,J., Huang,X. *et al.* (2018) Base editing with a Cpf1-cytidine deaminase fusion. *Nat. Biotechnol.*, **36**, 324–327.
15. Tanenbaum,M.E., Gilbert,L.A., Qi,L.S., Weissman,J.S. and Vale,R.D. (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell*, **159**, 635–646.
16. Kulcsár,P.I., Tálas,A., Huszár,K., Ligeti,Z., Tóth,E., Weinhardt,N., Fodor,E. and Welker,E. (2017) Crossing enhanced and high fidelity

SpCas9 nucleases to optimize specificity and cleavage. *Genome Biol.*, **18**, 1–17.
17. Tálas,A., Kulcsár,P.I., Weinhardt,N., Borsy,A., Tóth,E., Szebényi,K., Krausz,S.L., Huszár,K., Vida,I., Sturm,Á. *et al.* (2017) A convenientmethod to pre-screen candidate guide RNAs for CRISPR/Cas9 gene editing by NHEJ-mediated integration of a 'self-cleaving' GFP-expression plasmid. *DNA Res.*, **24**, 609–621.
18. Tóth,E., Weinhardt,N., Bencsura,P., Huszár,K., Kulcsár,P.I., Tálas,A., Fodor,E. and Welker,E. (2016) Cpf1 nucleases demonstrate robust activity to induce DNA modification by exploiting homology directed repair pathways in mammalian cells. *Biol. Direct*, **11**, 1–14.
19. Kluesner,M.G., Nedveck,D.A., Lahr,W.S., Garbe,J.R., Abrahante,J.E., Webber,B.R. and Moriarity,B.S. (2018) EditR: A method to quantify base editing from sanger sequencing. *Cris. J.*, **1**, 239–250.
20. Tsai,S.Q., Zheng,Z., Nguyen,N.T., Liebers,M., Topkar,V. V., Thapar,V., Wyvekens,N., Khayter,C., Iafrate,A.J., Le,L.P. *et al.* (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR–Cas nucleases. *Nat. Biotechnol.*, **33**, 187–197.
21. Vriend,L.E.M., Jasin,M. and Krawczyk,P.M. (2014) Assaying break and nick-induced homologous recombination in mammalian cells using the DR-GFP reporter and cas9 nucleases. *Methods Enzymol.*, **546**, 175–191.
22. Tsai,S.Q., Topkar,V. V, Joung,J.K. and Aryee,M.J. (2016) Open-source guideseq software for analysis of GUIDE-seq data. *Nat. Biotechnol.*, **34**, 483.
23. 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.
24. Mashiko,D., Fujihara,Y., Satouh,Y., Miyata,H., Isotani,A. and Ikawa,M. (2013) Generation of mutant mice by pronuclear injection of circular plasmid expressing Cas9 and single guided RNA. *Sci. Rep.*, **3**, 3355.
25. 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.
26. Dong,Ren, K., Qiu,X., Zheng,J., Guo,M., Guan,X., Liu,H., Li,N., Zhang,B., Yang,D. *et al.* (2016) The crystal structure of Cpf1 in complex with CRISPR RNA. *Nature*, **532**, 522–526.
27. 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.*, **532**, 522–526.
28. Tang,X., Lowder,L.G., Zhang,T., Malzahn,A.A., Zheng,X., Voytas,D.F., Zhong,Z., Chen,Y., Ren,Q., Li,Q. *et al.* (2017) A CRISPR–Cpf1 system for efficient genome editing and transcriptional repression in plants. *Nat. Plants*, **3**, 17018.
29. 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.
30. Kim,H., Kim,S.T., Ryu,J., Kang,B.C., Kim,J.S. and Kim,S.G. (2017) CRISPR/Cpf1-mediated DNA-free plant genome editing. *Nat. Commun.*, **8**, 14406.
31. Fernandez,J.P., Vejnar,C.E., Giraldez,A.J., Rouet,R. and Moreno-Mateos,M.A. (2018) Optimized CRISPR-Cpf1 system for genome editing in zebrafish. *Methods*, **150**, 11–18.
32. 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.
33. Kim,H.K., Song,M., Lee,J., Menon,A. V, Jung,S., Kang,Y.M., Choi,J.W., Woo,E., Koh,H.C., Nam,J.W. *et al.* (2016) In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat. Methods*, **14**, 153–159.
34. Fu,Y., Foden,J.A., Khayter,C., Maeder,M.L., Reyon,D., Joung,J.K. and Sander,J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR–Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–826.
35. Reyon,D., Tsai,S.Q., Khayter,C., Foden,J.A., Sander,J.D. and Joung,J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.*, **30**, 460–465.
36. 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.
37. Yang,B., Yang,L. and Chen,J. (2019) Development and application of base editors. *Cris. J.*, **2**, 91–104.
38. Bae,S., Park,J. and Kim,J.S. (2014) Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics*, **30**, 1473–1475.
39. 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.
40. Zhang,X., Wang,W., Shan,L., Han,L., Ma,S., Zhang,Y., Hao,B., Lin,Y. and Rong,Z. (2018) Gene activation in human cells using CRISPR/Cpf1-p300 and CRISPR/Cpf1-SunTag systems. *Protein Cell*, **9**, 380–383.