

# Enhanced Cas12a editing in mammalian cells and zebrafish

Pengpeng Liu<sup>1</sup>, Kevin Luk<sup>1</sup>, Masahiro Shin<sup>1</sup>, Feston Idrizi<sup>1</sup>, Samantha Kwok<sup>1</sup>, Benjamin Roscoe<sup>1</sup>, Esther Mintzer<sup>1</sup>, Sneha Suresh<sup>1</sup>, Kyle Morrison<sup>2</sup>, Josias B. Frazão<sup>1</sup>, Mehmet Fatih Bolukbasi<sup>1,3</sup>, Karthikeyan Ponnienselvan<sup>1</sup>, Jeremy Luban<sup>3,4,5</sup>, Lihua Julie Zhu<sup>1,4,6</sup>, Nathan D. Lawson<sup>1,5,\*</sup> and Scot A. Wolfe<sup>1,3,5,\*</sup>

<sup>1</sup>Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA, <sup>2</sup>Department of Chemistry and Biochemistry, Worcester Polytechnic Institute, Worcester, MA, USA, <sup>3</sup>Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA, <sup>4</sup>Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA, USA, <sup>5</sup>Li Weibo Institute for Rare Diseases Research, University of Massachusetts Medical School, Worcester, MA, USA and <sup>6</sup>Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA, USA

Received October 31, 2018; Revised March 04, 2019; Editorial Decision March 05, 2019; Accepted March 15, 2019

## ABSTRACT

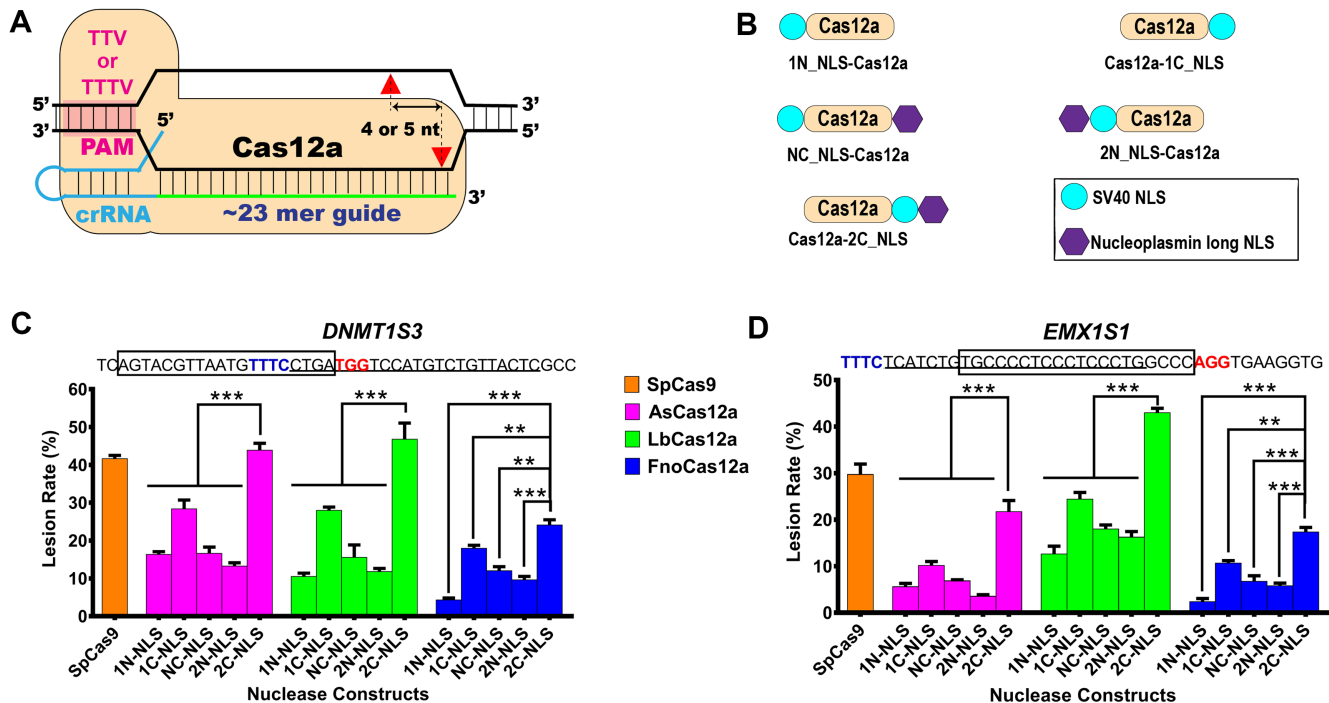
**Type V CRISPR–Cas12a systems provide an alternate nuclease platform to Cas9, with potential advantages for specific genome editing applications. Here we describe improvements to the Cas12a system that facilitate efficient targeted mutagenesis in mammalian cells and zebrafish embryos. We show that engineered variants of Cas12a with two different nuclear localization sequences (NLS) on the C terminus provide increased editing efficiency in mammalian cells. Additionally, we find that pre-crRNAs comprising a full-length direct repeat (full-DR-crRNA) sequence with specific stem-loop G-C base substitutions exhibit increased editing efficiencies compared with the standard mature crRNA framework. Finally, we demonstrate in zebrafish embryos that the improved LbCas12a and FnoCas12a nucleases in combination with these modified crRNAs display high mutagenesis efficiencies and low toxicity when delivered as ribonucleoprotein complexes at high concentration. Together, these results define a set of enhanced Cas12a components with broad utility in vertebrate systems.**

## INTRODUCTION

The breadth of Class 2 CRISPR–Cas single effector nucleases that have been identified and characterized continues to expand (1–4). Many of these newly discovered Class 2 systems have novel properties that differ from the ubiquitously employed Type II Cas9 system, making them particularly amenable to specific biological and therapeutic applications (5–7). In particular, the Type V Cas12a (Cpf1) DNA endonucleases have several unique attributes for genome editing applications (Figure 1A) (1). First, characterized Cas12a nucleases typically recognize a T-rich Protospacer adjacent motif (PAM) element at the 5' side of the protospacer, which facilitates targeting AT-rich genomic regions that can be challenging to target with Cas9-based nucleases (1,8,9). Second, unlike Cas9 nucleases, Cas12a nucleases are programmed with a single crRNA that does not include a tracrRNA (1). Thus, the shorter crRNA of Cas12a (~42 nt) compared with the sgRNA of Cas9 (~100 nt) is more amenable to the synthesis of chemically-modified guide RNAs that improve nuclease activity within mammalian cells (10,11). Third, Cas12a produces double-strand breaks with 5' overhangs that are distal from its PAM element (1,8), which are distinct from the blunt ends produced by SpCas9. The distal cleavage within the spacer region potentially permits Cas12a to continue cutting even after initial sequence alterations have occurred via imprecise DNA repair; this behavior is distinct from Cas9, where primary lesions usually prevent subsequent targeting. Consequently,

\*To whom correspondence should be addressed. Tel: +1 508 856 3953; Fax: +1 508 856 6200; Email: scot.wolfe@umassmed.edu  
Correspondence may also be addressed to Nathan D. Lawson. Tel: +1 508 856 1177; Fax: +1 508 856 6200; Email: nathan.lawson@umassmed.edu  
Present addresses:

Benjamin Roscoe, COGEN Therapeutics, Cambridge, MA, USA.  
Mehmet Fatih Bolukbasi, Exonics Therapeutics, Watertown, MA, USA.



**Figure 1.** Position and number of NLS improves genome editing by AsCas12a, LbCas12a and FnoCas12a. (A) General schematic of Cas12a (B). Schematic representation of a series of Cas12a constructs with different nuclear localization signals. Lesion rates determined by deep sequencing for SpCas9, AsCas12a, LbCas12a and FnoCas12a with different combinations of NLSs at the *DNMT1S3* (C) and *EMX1S1* (D) target sites, respectively. Boxed sequences represent SpCas9 targeting sites, red color labeled NGG PAM. Underlined sequences represent Cas12a targeting sites, blue color labeled TTTV PAM. Data are from three independent biological replicates performed on different days with expression constructs delivered by transient transfection in HEK293T cells (Supplementary Table S1). Error bars indicate  $\pm$ s.e.m. Statistical significance is determined by two-tailed Student's *t*-test: \*\*\*\* denotes  $P < 0.001$ , \*\*\* denotes  $P < 0.01$ , respectively (Supplementary Table S7).

Cas12a mutagenesis products are biased toward larger deletions than are typically produced by Cas9 (1,12). This mutagenesis behavior should increase the efficiency with which genomic features, such as transcriptional and splicing regulatory elements (13–15), can be selectively removed from the genome. Fourth, Cas12a proteins contain an active site for processing precursor crRNAs (pre-crRNA) arrays, which can be harnessed for multiplex genome editing from a single transcript (8,16–18). Finally, Cas12a displays higher genome editing precision than SpCas9 based on multiple unbiased genome-wide analyses (12,16,17). Thus, Cas12a nucleases could provide a valuable alternative to Cas9 for many genome editing applications.

Like Cas9, Cas12a has been employed for targeted mutagenesis in fruit flies (18), mammalian cells (1,9,12,16), mouse embryos (19–21), zebrafish (22) and a variety of plant systems (26–28). Furthermore, Cas12a has been used successfully to restore dystrophin function via targeted gene correction in embryos of a mouse model of Duchenne muscular dystrophy (DMD) or by exon skipping via the generation of segmental deletions in an DMD-iPSC line (7). Additionally, Cas12a has been adapted to facilitate targeted cytosine base editing within the genome (23). Together, these results demonstrate that Cas12a-based systems have the potential to facilitate a broad variety of genome editing goals with both research and therapeutic applications.

Most studies have employed LbCas12a or AsCas12a in vertebrate systems because of their promising activity in cell

culture assays in initial reports (1,12). LbCas12a and AsCas12a prefer a TTTV PAM element (1,9), and the range of targetable sequences has been extended through modifications to residues involved in their PAM recognition (24). Notably, FnoCas12a prefers a more compact TTN PAM element (1,25,26), which provides an expanded targeting range relative to either LbCas12a or AsCas12a. Recently, FnoCas12a has been shown to be effective in mammalian systems (26) and plants (27). In many instances, the mutagenesis activity of Cas12a nucleases is modestly lower than observed for SpCas9 (12,19). Therefore, improvements in the cleavage activity of Cas12a-based nucleases could expand its utility in a variety of eukaryotic systems.

We have focused on improving the editing efficiency of Cas12a for targeted mutagenesis in mammalian cells and zebrafish embryos through alteration of its protein and crRNA components. The mutagenesis efficiency of AsCas12a, LbCas12a and FnoCas12a in mammalian systems was improved through the incorporation of an additional nuclear localization signal (NLS) sequence within the protein. The editing efficiency of Cas12a in mammalian cells and zebrafish embryos was further increased through extension of the 5' end of the crRNA and by altering the base composition of the hairpin within the pseudoknot. Together, these enhancements will increase the utility of Cas12a nucleases in a variety of systems for genome editing applications.

## MATERIALS AND METHODS

### Zebrafish husbandry

The use of Zebrafish was in accordance with established protocols (28) and in conformity with Institutional Animal Care and Use Committee guidelines of the University of Massachusetts Medical School. All Cas12a nuclease protein and crRNA injections were performed in the wild type Ekkwill 2 (EK2) zebrafish line.

### Plasmid constructs

Cas12a nuclease experiments for transient transfection in cell culture employed the following plasmids: All Cas12a crRNAs are expressed via a U6 promoter from a pBlue-script2 SK+ vector (Agilent). AsCas12a, LbCas12a and FnoCas12a were cloned from Addgene vectors (#69982, #69988 and #69976) that were a generous gift of the Zhang laboratory (1). All new Cas12a constructs were expressed via a CMV IE94 promoter from a pCS2-Dest gateway plasmid (29). The SpCas9 plasmid expression construct for transient transfections has been previously described (29). A corresponding pBluescript U6 SpCas9 sgRNA expression vector was constructed for these assays. All Cas12a and SpCas9 protein expression for protein purification utilized pET-21a protein expression plasmids (Novagen). LbCas12a-2xNLS and FnoCas12a-2xNLS expression constructs were constructed containing a 6xHis tag at the C-terminus for affinity purification. Our pET-21a 3xNLS-SpCas9 protein expression plasmid has been previously described (Wu *et al. Nature Medicine in press*). Representative protein sequences for the Cas12a constructs (both plasmid expression and protein expression constructs) and cloning vectors for the U6 expression vectors for the crRNA expression cassettes are shown in Supplementary SFigure 23. These plasmids have been deposited with Addgene for distribution to the community.

### Cell culture nuclease assays

Human Embryonic Kidney (HEK293T) and HeLa cells were cultured in high glucose DMEM with 10% FBS and 1% Penicillin/Streptomycin (Gibco) in a 37°C incubator with 5% CO<sub>2</sub>. K562 and Jurkat cells were cultured in RPMI 1640 medium with 10% FBS and 1% Penicillin/Streptomycin (Gibco) in a 37°C incubator with 5% CO<sub>2</sub>. These cells were authenticated by University of Arizona Genetics Core and tested for mycoplasma contamination at regular intervals. For transient transfection, we used early to mid-passage cells (passage number: 5–25). Approximately  $1.6 \times 10^5$  cells are transfected using Polyfect transfection reagent (Qiagen) in 24-well format according to the manufacturer's suggested protocol. Following plasmid amounts were used for transient transfections for single target sites: 50 ng of Cas12a nuclease expression vector, 50ng crRNA expression vector and 100 ng mCherry plasmid. Cas12a RNPs were delivered to HEK293T, Jurkat or K562 cells by nucleofection. LbCas12a or FnoCas12a protein were complexed with the desired crRNA generated by *in vitro* transcription (described below) at a ratio of 1:2.5 (16 pmol protein complexed to 40 pmol crRNA) in Neon

R buffer (Thermo Fisher Scientific) and incubated at RT for 15–20 min. For HEK293T cells, the Cas12a RNP complex was then mixed with  $1 \times 10^5$  cells in Neon R buffer at the desired concentration and electroporated using Neon<sup>®</sup> Transfection System 10 L Kit (Thermo Fisher Scientific) using the suggested electroporation parameters: Pulse voltage (1500 V), Pulse width (20 ms), Pulse number (2). For Jurkat and K562 cells, the Cas12a RNP complex was then mixed with  $2 \times 10^5$  cells in Neon R buffer at the desired concentration and electroporated using Neon<sup>®</sup> Transfection System 10 L Kit (Thermo Fisher Scientific) using the suggested electroporation parameters: Pulse voltage (1600 V), Pulse width (10 ms), Pulse number (3).

### AsCas12a immunohistochemistry

HEK293T cells are transfected in 6-well format via Polyfect transfection reagent (Qiagen) using the manufacturer's suggested protocol with 300 ng of each AsCas12a plasmid (or SpCas9–2xNLS) and 150 ng of crRNA (or sgRNA) expression plasmid on a cover slip. Forty eight hours following transfection, transfection media was removed, cells were washed with 1x PBS and fixed with 4% formaldehyde in 1xPBS for 15 min at room temperature. Following blocking (blocking solution: 2% BSA, 0.3% Triton X-100 within 1x PBS), samples were stained with mouse anti-hemagglutinin (Sigma, H9658, 1:500), and Alexa 488 donkey anti-mouse IgG (H+L; Invitrogen, A-21202, 1:2000), sequentially. VECTASHIELD Mounting Medium with DAPI (Vector Laboratories, H-1200) was used to stain the nuclei and to mount the samples on slide. Images were taken with Zeiss AxioPlan 2 IE Motorized Microscope System.

### LbCas12a and FnoCas12a protein purification

Protein purification for LbCas12a-2xNLS or FnoCas12a-2xNLS used a common protocol. The plasmid expressing LbCas12a-2xNLS (or FnoCas12a-2xNLS) was introduced into *Escherichia coli* Rosetta (DE3) pLysS cells (EMD Millipore) for protein overexpression. Cells were grown at 37°C to an OD<sub>600</sub> of ~0.2, then shifted to 18°C and induced for 16 h with IPTG (1 mM final concentration). Following induction, cells were pelleted by centrifugation and then resuspended with Nickel-NTA buffer (20 mM TRIS + 1 M NaCl + 20 mM imidazole + 1 mM TCEP, pH 7.5) supplemented with HALT Protease Inhibitor Cocktail, EDTA-Free (100X) [ThermoFisher] and frozen on dry ice. Cells were then thawed, an additional aliquot of HALT Protease Inhibitor Cocktail added followed by cell lysis with M-110s Microfluidizer (Microfluidics) following the manufacturer's instructions. The protein was purified with Ni-NTA resin and eluted with elution buffer (20 mM Tris, 500 mM NaCl, 250 mM Imidazole, 10% glycerol, pH 7.5). Cas12a protein was dialyzed overnight at 4°C in 20 mM HEPES, 500 mM NaCl, 1 mM EDTA, 10% glycerol, pH 7.5. Subsequently, Cas12a protein was step dialyzed from 500 mM NaCl to 200 mM NaCl (final dialysis buffer: 20 mM HEPES, 200 mM NaCl, 1 mM EDTA, 10% glycerol, pH 7.5). Next, the protein was purified by cation exchange chromatography (column = 5ml HiTrap-S, Buffer A = 20 mM HEPES pH 7.5 + 1 mM TCEP, Buffer B = 20 mM HEPES pH 7.5 + 1 M

NaCl + 1 mM TCEP, Flow rate = 5 ml/min, CV = column volume = 5 ml) followed by size-exclusion chromatography (SEC) on Superdex-200 (16/60) column (Isocratic size-exclusion running buffer = 20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP). The primary protein peak from the SEC was concentrated in an Ultra-15 Centrifugal Filters Ultracel -30K (Amicon) to a concentration of between 30 and 50  $\mu$ M. The purified protein quality was assessed by SDS-PAGE/Coomassie staining to be >95% pure.

Initial studies with FnoCas12a-2xNLS protein in zebrafish were produced from a FnoCas12a-2C-NLS-TEV-MBP-6xHis expression construct, where MBP was cleaved following the Hexa-His purification. TEV cleavage of MBP incorporated two extra steps: 1. Determine total protein concentration following Hexa-His purification (described above) by measuring absorbance at 280 nm blanked against elution buffer. Add 1:10 (w/w) of TEV protease. 2. Dialyze FnoCas12a-MBP eluate + TEV protease overnight against 20 mM HEPES pH 7.5, 150 mM NaCl, 1 mM TCEP, and 10% glycerol. The final cleaved FnoCas12a product was purified by cation exchange chromatography (Column = 5 ml HiTrap-S, Buffer A = 20 mM HEPES pH 7.5 + 1 mM TCEP, Buffer B = 20 mM HEPES pH 7.5 + 1 M NaCl + 1 mM TCEP, Flow rate = 5 ml/min, CV = column volume = 5 ml) followed by size-exclusion chromatography (SEC) on Superdex-200 (16/60) column (Isocratic size-exclusion running buffer = 20 mM HEPES pH 7.5, 300 mM NaCl, 1 mM TCEP). The primary protein peak from the SEC was concentrated in an Ultra-15 Centrifugal Filters Ultracel -30K (Amicon) to a concentration of between 30 and 50  $\mu$ M. The purified protein quality was assessed by SDS-PAGE/Coomassie staining to be >95% pure.

### ***In vitro* transcription of crRNAs**

For crRNA generation, DNA templates for crRNA transcription were PCR assembled by annealing and extending a common scaffold oligonucleotide and a target sequence specific oligonucleotide (Invitrogen), which contains sequence complementary to the common scaffold oligonucleotide on the 3' end (Supplementary Table S8). To generate each crRNA expression template, a 60  $\mu$ l PCR containing 1X Phusion Buffer, 200  $\mu$ M common scaffold oligonucleotide and target sequence specific oligonucleotide, and 1.2 U of Phusion DNA polymerase (NEB Cat#M0530L) was performed using the following thermocycling program: initial incubation 98°C for 15 s, 40 cycles of 98°C for 10 s, 60°C for 15 s, 72°C for 5 s, final extension at 72°C for 1 min, and hold at 4°C. Production of each double stranded crRNA DNA template was validated by 3% TBE agarose gel with ethidium bromide, and then crRNA DNA templates were purified by ethanol precipitation using 3 M sodium acetate and 2  $\mu$ g glycogen as a carrier.

For crRNA generation, IVT was performed using the MEGAscript<sup>®</sup> T7 Kit (Ambion Cat#AM1334). Half reactions (10  $\mu$ l) using 800 ng of crRNA DNA template were incubated for a minimum of 5 h at 37°C. To eliminate the DNA template following crRNA production by IVT, reactions were treated with DNase TURBO (ThermoFisher) for 40 min at 37°C. Production of each crRNA was validated by 3% TBE agarose gel with ethidium bromide. Following

validation, crRNAs were purified by ethanol precipitation with 3 M sodium acetate and resuspended in 30  $\mu$ l of nuclease free water. Finally, crRNA quality was again validated by 3% TBE agarose gel and quantified by A260.

### **Injection of zebrafish embryos with RNP**

We performed microinjection of Cas12a protein and crRNAs into one-cell-stage zebrafish embryos according to standard methods (30). Briefly, in order to assemble RNPs, LbCas12a, or FnoCas12a protein was complexed with the corresponding crRNA at 2.5-fold excess (Supplementary Table S8) in nuclease free water with 1:10 phenol red and incubated for 20 minutes at room temperature. Injection solutions were held on ice until utilization. Injections of 1 and 2 nl RNP solution were performed into the blastomere of ~60 one-cell-stage zebrafish embryos. Heat shock treated embryos were incubated at 34°C for 4 h immediately following injection and then shifted to 28.5°C (22). Embryos without heat shock were incubated at 28.5°C. All experiments were performed in biological triplicate by three individuals.

At 24 hpf, the morphology of injected embryos was assessed. The embryos were assigned to three morphology classes: normal, abnormal and dead (Supplementary Table S6). The distributions of these morphological classes were used to assess the statistical significance of the impact of different crRNA frameworks or different Cas12a RNP doses on zebrafish development. R, a system for statistical computation and graphics, was used for this analysis (31). Nuclease toxicity, measured as proportion of abnormal and dead animals, was first arcsin transformed to homogenize the variance. Levene's test indicates that the assumption of homogeneity of variances was met for both Heatshock+ and Heatshock- experiments. Three-way analysis of variance (ANOVA) with Completely Randomized Design was performed to test whether there are main effects of RNP dose, target sites and crRNA framework, and whether there is a significant dose-dependent guide frameworks impact (Supplementary Table S7).

### **Preparation of zebrafish genomic DNA for analysis of editing efficiency**

For analysis of RNP activity and resulting insertions and deletions for each nuclease treatment group (or uninjected controls), pooled genomic DNA samples ( $n = 20$ ) were harvested at 24 hpf from embryos displaying normal developmental morphology using the DNeasy Blood & Tissue kit (QIAGEN Cat#69506).

For initial analysis, pooled genomic samples were PCR amplified for Sanger sequencing (Supplementary Table S8). For each genomic DNA sample, a 25  $\mu$ l reaction containing 1X Phire Buffer, 100  $\mu$ M forward and reverse primer, and 0.5  $\mu$ l of Phire Hot Start II DNA Polymerase (ThermoFisher Cat#F122L) was performed using the following thermocycling program: initial denaturation 98°C for 3 minutes, 31 cycles of 98°C for 20 s, 62°C 20 s, 72°C for 30 s, final extension at 72°C for 5 min, and hold at 12°C. DNA amplicon production was validated by 3% TAE agarose gel with ethidium bromide. DNA amplicons were purified by Zymo

DNA clean and concentrator 5 column (Zymo D4014). Sanger sequencing was performed by a commercial service (Genewiz) and chromatograms were analyzed for indel rate and composition using TIDE Analysis (32) or Synthego ICE (<https://ice.synthego.com/#/>).

### Target site indel frequency analysis in mammalian cells or zebrafish embryos by deep sequencing

Library construction for deep sequencing is modified from our previous report (29). For analysis of mammalian cell culture experiments, cells were harvested 72 h after transfection (or nucleofection) and genomic DNA extracted with GenElute Mammalian Genomic DNA Miniprep Kit (Sigma). For analysis of zebrafish embryos the genomic DNA preparation is described above. Briefly, regions flanking each target site were PCR amplified using locus-specific primers (Supplementary Table S8) bearing tails complementary to the Truseq adapters as described previously (29). 25–50 ng input genomic DNA is PCR amplified with Phusion High Fidelity DNA Polymerase (New England Biolabs): (98°C, 15 s; 67°C 25 s; 72°C 18 s) × 30 cycles. 1 µl of each PCR reaction was amplified with barcoded primers to reconstitute the TruSeq adaptors using the Phusion High Fidelity DNA Polymerase (New England Biolabs): (98°C, 15 s; 61°C, 25 s; 72°C, 18 s) × 9 cycles. Equal amounts of the products were pooled and gel purified. The purified library was deep sequenced using a paired-end 150 bp Illumina MiSeq run. For off-target analysis, we choose 13 potential DNMT1S3 off target sites which were identified and described previously (12,16). We used the same approach to build and sequence the Illumina deep sequencing library as for Cas12a target sites described above.

MiSeq data analysis for editing at target sites or off-target sites was performed using a suite of Unix-based software tools. First, the quality of paired-end sequencing reads (R1 and R2 fastq files) was assessed using FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Raw paired-end reads were combined using paired end read merger (PEAR) (33) to generate single merged high-quality full-length reads. Reads were then filtered by quality (using Filter FASTQC (34)) to remove those with a mean PHRED quality score under 30 and a minimum per base score under 24. Each group of reads was then aligned to a corresponding reference sequence using BWA (version 0.7.5) and SAMtools (version 0.1.19). To determine indel frequency, size and distribution, all edited reads from each experimental replicate were combined and aligned, as described above. Indel types and frequencies were then cataloged in a text output format at each base using bam-readcount (<https://github.com/genome/bam-readcount>). For each treatment group, the average background indel frequencies (based on indel type, position and frequency) of the triplicate negative control group were subtracted to obtain the nuclease-dependent indel frequencies.

### Comparison of the activity of our LbCas12a-2xNLS and NEB LbCas12a protein

LbCas12a protein was purchased from NEB (EnGen Lba Cas12a, Catalog # 0653S). For activity comparison in

mammalian cells a DRf-crRNA targeting the AAVS1 locus (Supplementary Table S8) was *in vitro* transcribed as described above. Different amounts (5pmol, 10pmol, 20pmol, 40pmol and 80pmol) of our LbCas12a or NEB LbCas12a Protein was mixed with the desired amount crRNAs at 2.5-fold excess to form each RNP complex. After 15–20 min incubation at room temperature, the RNP complex was mixed with  $1 \times 10^5$  cells in Neon R buffer and electroporated using Neon® Transfection System (Thermo Fisher Scientific) using the suggested electroporation parameters: Pulse voltage (1500v), Pulse width (20ms), Pulse number (2). To compare the activity of our LbCas12a with NEB commercial LbCas12a protein in zebrafish embryos, we prepared RNP complex using different amount of our LbCas12a or NEB LbCas12a protein (4, 8, 16 and 24 fmol) complexed with the Albino crRNA at 2.5-fold excess, performed microinjection of Cas12a protein and crRNAs into one-cell-stage zebrafish embryos as previous described. LbCas12a induced indel rates in mammalian cells or zebrafish were determined by deep sequencing as described above.

## RESULTS

### Improved nuclear localization increases Cas12a activity in mammalian cells

NLS sequences control the nuclear import of proteins (35). Previous studies with SpCas9 have reported inefficient nuclear localization mediated by a single SV40 NLS (36). While AsCas12a and LbCas12a have been employed for effective genome editing in a variety of mammalian systems, in many prior studies these proteins were modified to contain only a single NLS appended to the C-terminus (Cas12a-1C\_NLS, Figure 1B) (1,9,12,16). However, in the case of AsCas12a we find that a single NLS leads to inefficient nuclear localization in HEK293T cells (Supplementary Figure S1). To improve the efficiency of nuclear localization, we examined a number of different combinations of N- and C-terminal NLSs, and found that the fusion of two different of NLSs (SV40 and Nucleoplasmin long NLS (35)) onto the C-terminus facilitated the most efficient nuclear localization of Cas12a (Supplementary Figure S1). Next, we examined the influence of the different NLS combinations on AsCas12a, LbCas12a and FnoCas12a nuclease activity at two previously defined active target sites in human cells (*DNMT1S3* and *EMX1S1*; Figure 1C, D and Supplementary Table S1) (1). At both genomic target sites, we observed significantly improved lesion frequency associated with Cas12a proteins bearing two C-terminal NLSs (Figure 1C and D). Together, these results suggest that improving the nuclear localization of Cas12a proteins can significantly increase lesion frequencies at their respective target sites.

Consistent with previous studies of the mutational end products produced by AsCas12a and LbCas12a editing (1,12), we found that FnoCas12a primarily produced deletions at the *DNMT1S3* and *EMX1S1* target sites (Supplementary Figures S2A–D and 3A–D). We observed that the types of deletion products produced for the three different Cas12a orthologs were generally similar, although the rank order of their frequency varied (Supplementary Figures S2A–L and S3A–L). Consistent with previous obser-

vations (12), the profile of Cas12a lesions, which are primarily deletions, were distinct from those induced by SpCas9, which produced smaller deletions and more insertions than Cas12a (Supplementary Figures S2M–P and S3M–P). Notably, many deletions mediated by all three Cas12as have evidence of short microhomologies flanking the deleted sequence segments (e.g. 15 base deletion for FnoCas12a in Supplementary Figure S2D), as previously reported (12).

### Optimization of Cas12a crRNA direct repeat length

The majority of Cas12a studies in mammalian cells have utilized mature (processed) crRNAs to target the nuclease to a particular locus (1,12,16,19). However, in the native CRISPR system, the pre-crRNA array is processed by Cas12a through cleavage within the direct repeat (DR) at four nucleotides (AAUU; the repeat recognition sequence, [RRS]) upstream of the stem-loop to generate the mature crRNA species (DRt-crRNA, Figure 2A) (8). We examined the impact of including sequence components from the unprocessed pre-crRNA on the nuclease activity of Cas12a in mammalian cells. We designed four different crRNAs with different compositions of the Direct Repeats flanking the core crRNA: truncated Direct-Repeat crRNA (DRt-crRNA), full-length Direct-Repeat crRNA (DRf-crRNA), full-length Direct-Repeat crRNA with a 3' truncated Direct-Repeat (DRf-crRNA-DRt), and full-length Direct-Repeat crRNA with a 3' full-length Direct-Repeat (DRf-crRNA-DRf) (Figure 2A). We tested these four different crRNA compositions with LbCas12a and FnoCas12a using the native DR sequence for each nuclease at eleven endogenous target sites in the human genome. Expression vectors for Cas12a and the crRNA were introduced via transient transfection assays in HEK293T cells. We found that the DRf-crRNA structure provided the highest editing rates for both LbCas12a and FnoCas12a across 11 genomic target sites, whereas crRNA constructs encoding 3' terminal DR components displayed lower activity than the mature DRt-crRNA (Figure 2B and C). Consistent with prior studies (1,12), we found that LbCas12a has higher activity than FnoCas12a at overlapping target sequences, but notably FnoCas12a displayed good editing efficiency when employing the DRf-crRNA framework (median rate 26.3% versus 34.2% for LbCas12a; Figure 2C).

To examine the requirement for crRNA processing by Cas12a for the improved activity of the DRf-crRNA constructs, we tested the impact on nuclease activity of either RNase-dead Cas12a or mutations within the RRS of the crRNA that reduce processing (8). Lesion rates were significantly reduced with RNase-dead LbCas12a or FnoCas12a when used in conjunction with DRf-crRNAs (Figure 2D–G). Likewise, DRf-crRNAs bearing RRS mutations showed decreased lesion rates when used with wild type Cas12a proteins. By contrast, loss of RNA processing only modestly reduced lesion rates associated with DRt-crRNA (Figure 2D–G). These data demonstrate that the incorporation of the full length DR at the 5' end of the Cas12a crRNA can significantly improve its nuclease activity when expressed via a U6 promoter, and that this enhanced activity is dependent on the potential of Cas12a to process the DR with its RNase cleavage domain.

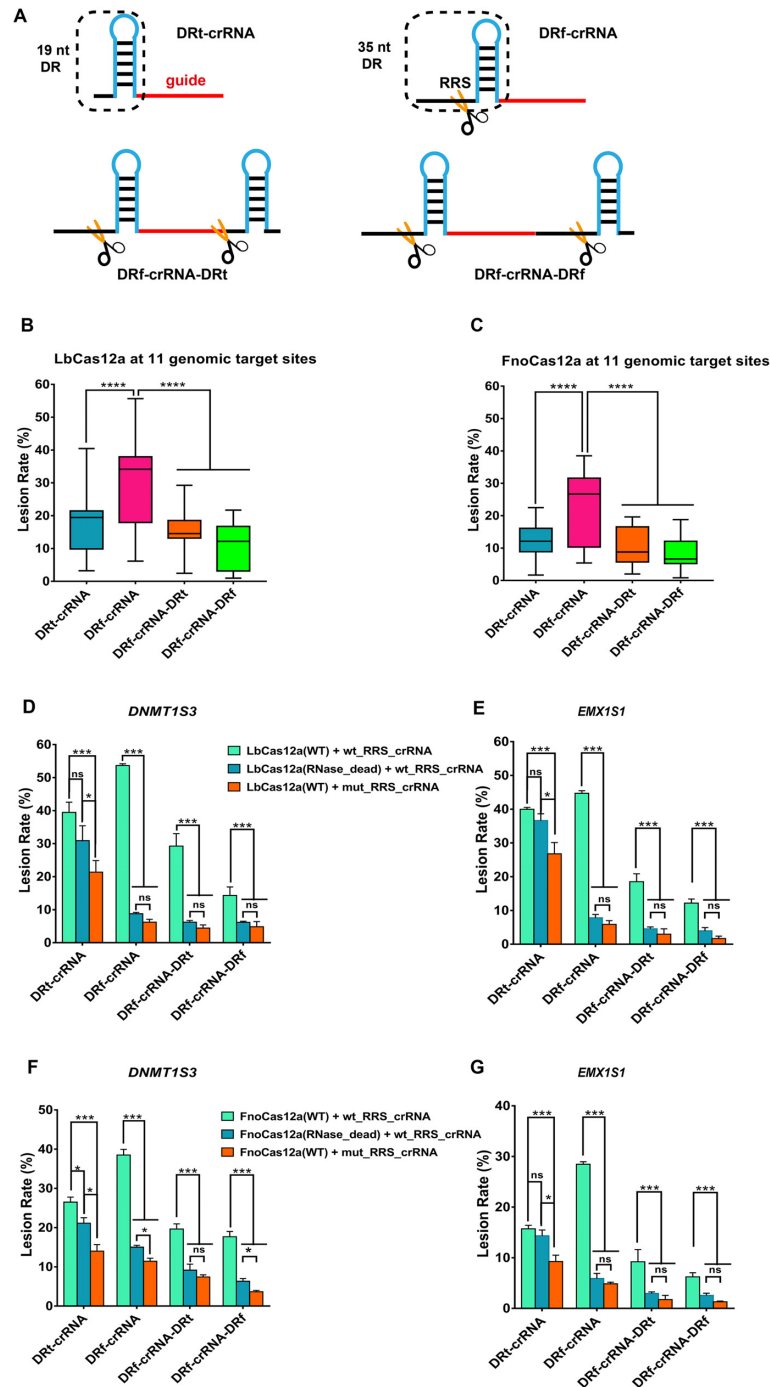
### Optimization of Cas12a crRNA hairpin sequence

There are three A:U base pairs within the stem-loop region of the DR (37–41). Mutations within the hairpin stem that disrupt basepairing abolish Cas12a nuclease activity *in vitro*, whereas mutations within the stem that preserve base pairing (e.g. A:U to G:C) do not (1). However the impact of the A:U to G:C substitutions (G–C swap) on editing in mammalian cells has not been investigated. We hypothesized that increasing the thermal stability of the stem would potentially increase the fraction of properly folded crRNA for loading into Cas12a and thereby nuclease activity. To test this hypothesis, we designed seven different G–C swapped stem-loop sequences in the DRf-crRNA backbone (Figure 3A) and tested their activities on six different endogenous sites in HEK293T cells. We found that simultaneous G–C swaps at positions 1 and 3 (DRf-GC@13-crRNA) showed significantly increased editing efficiencies for both LbCas12a and FnoCas12a compared with mature crRNAs (DRt-crRNA) or unmodified full-length Direct Repeat crRNAs (DRf-crRNA) (Figure 3B and C, Supplementary Figures S4 and S5). We observed similarly increased editing rates at three sites in HeLa cells using crRNAs with these modifications (Supplementary Figure S6A and B). These results illustrate that these improvements in activity are not a cell-type specific effect. We also evaluated the nuclease activity of LbCas12a and FnoCas12a protein complexed with different crRNA frameworks at two target sites when delivered as ribonucleoprotein complexes (Cas12a RNPs) by electroporation into HEK293T cells, Jurkat cells and K562 cells. LbCas12a or FnoCas12a loaded with either DRf-crRNA or DRf-GC@13-crRNA showed significantly higher activities than DRt-crRNA (Supplementary Figure S6C and D). These results demonstrate that across different delivery methods and different cell types, these crRNA modifications can dramatically improve the Cas12a nuclease activity.

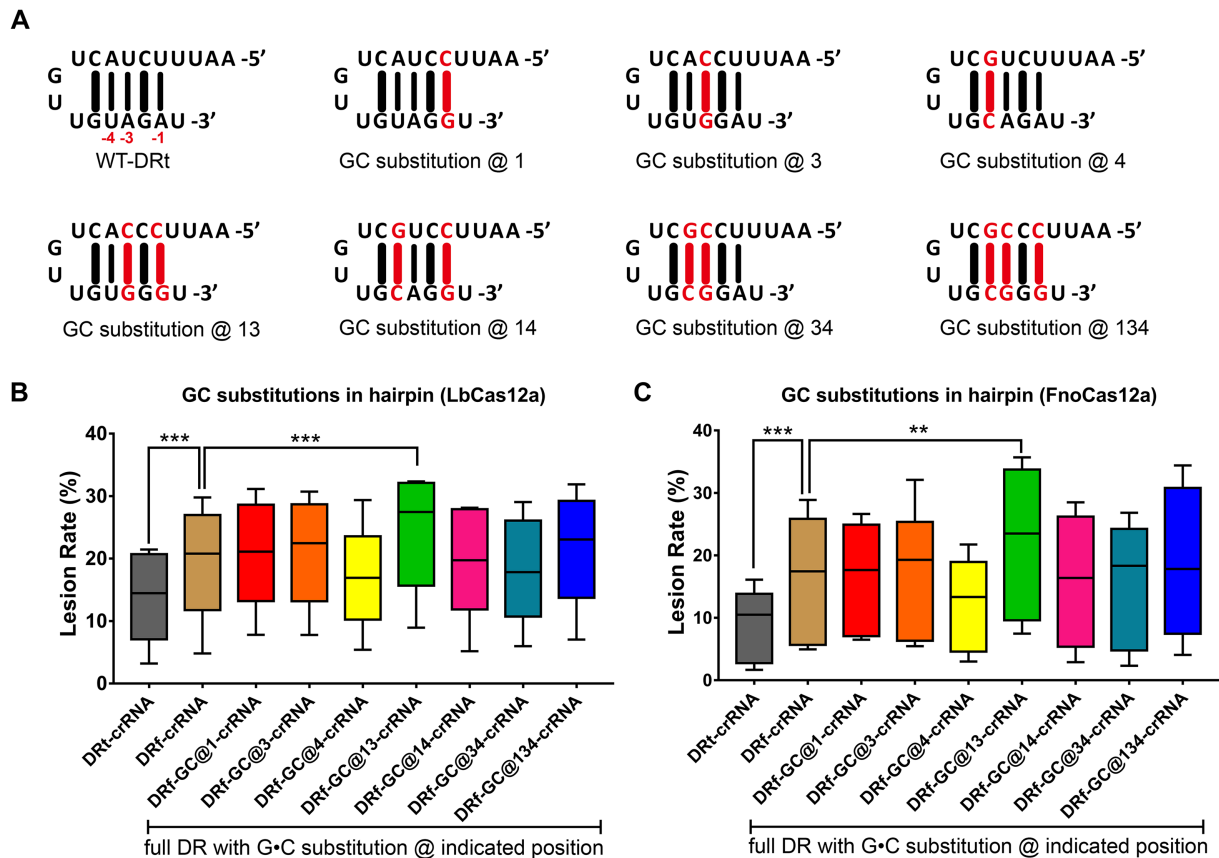
To compare the specificity of LbCas12a programmed with different crRNA frameworks, we carried out targeted deep sequencing analyses to detect the indels for 13 potential off-target sites for the DNMT1S3 crRNA, which has been previously characterized by genome-wide deep sequencing approaches (12,16). These data indicate that either crRNA fused full-length direct repeat (DRf-cr) or simultaneous G–C swaps at positions 1 and 3 (DRf-GC@13-crRNA) did not show significantly higher activities than DRt-crRNA on these off-target sites (Supplementary Figure S7, Supplementary Table S4). Thus, these crRNA modifications do not substantially alter the specificity of Cas12a.

### Genome editing in zebrafish embryos

Only a single previous study has successfully utilized LbCas12a and AsCas12a in zebrafish embryos for targeted genome editing (22). This work demonstrated activity at four independent loci, but found that editing was only possible when using RNPs rather than mRNA encoded protein. Furthermore, optimal editing was achieved using a transient heat shock of embryos at 34°C (4 h) following Cas12a RNP injection (22). Therefore, we sought to determine if the crRNA improvements described above were applicable in this model system.



**Figure 2.** Employing a full-length Direct-Repeat crRNA (DRf-crRNA) enhances editing efficiency. (A) General schematic of four different crRNAs: 19nt-truncated Direct-Repeat crRNA (DRT-crRNA), 35nt-full-length Direct-Repeat crRNA (DRf-crRNA), full-length Direct-Repeat crRNA with 3' truncated Direct-Repeat (DRf-crRNA-DRt), full-length Direct-Repeat crRNA with 3' full-length Direct-Repeat (DRf-crRNA-DRf). The 19-nt and 35-nt Direct Repeats are denoted with the dotted box, the guide sequences are marked in red. Scissors represents the Cas12a RNase domain, which processes the crRNA at the RRS. The nuclease activities of each crRNA type are determined by deep sequencing for LbCas12a (B) and FnoCas12a (C) at 11 endogenous target sites on genome. Each box represents the 25th and 75th percentile and median is indicated by a line. Whiskers in the box plots are defined by the Tukey method. Statistical significance is determined by one-way analysis of variance (ANOVA), "\*\*\*\*" denotes  $P < 0.0001$  (Supplementary Table S7). Deep sequencing data are from three independent biological replicates performed on different days with expression constructs delivered by transient transfection in HEK293T cells (Supplementary Table S2). Error bars indicate  $\pm$  s.e.m. (D–G). The ability of Cas12a to process the Full-length Direct Repeat (DRf) is important for enhanced nuclease activity of this crRNA. Evaluation of the editing activity of Wild-type (WT) or RNase-dead Cas12a with different crRNA constructs (wild-type crRNA sequence or crRNA containing a mutation within the RRS sequence of crRNA) for LbCas12a (D, E) and FnoCas12a (F, G) at the DNMT1S3 (D, F) and EMX1S1 (E, G) target sites. Lesion rates determined by deep sequencing. Data are from three independent biological replicates performed on different days with expression constructs delivered by transient transfection in HEK293T cells (Supplementary Table S2). Error bars indicate  $\pm$  s.e.m. Statistical significance is determined by two-tailed Student's *t*-test: "\*\*\*\*" denote  $P < 0.001$ , "\*\*\*" denote  $P < 0.01$ , "\*\*" denote  $P < 0.05$ , respectively (Supplementary Table S7).



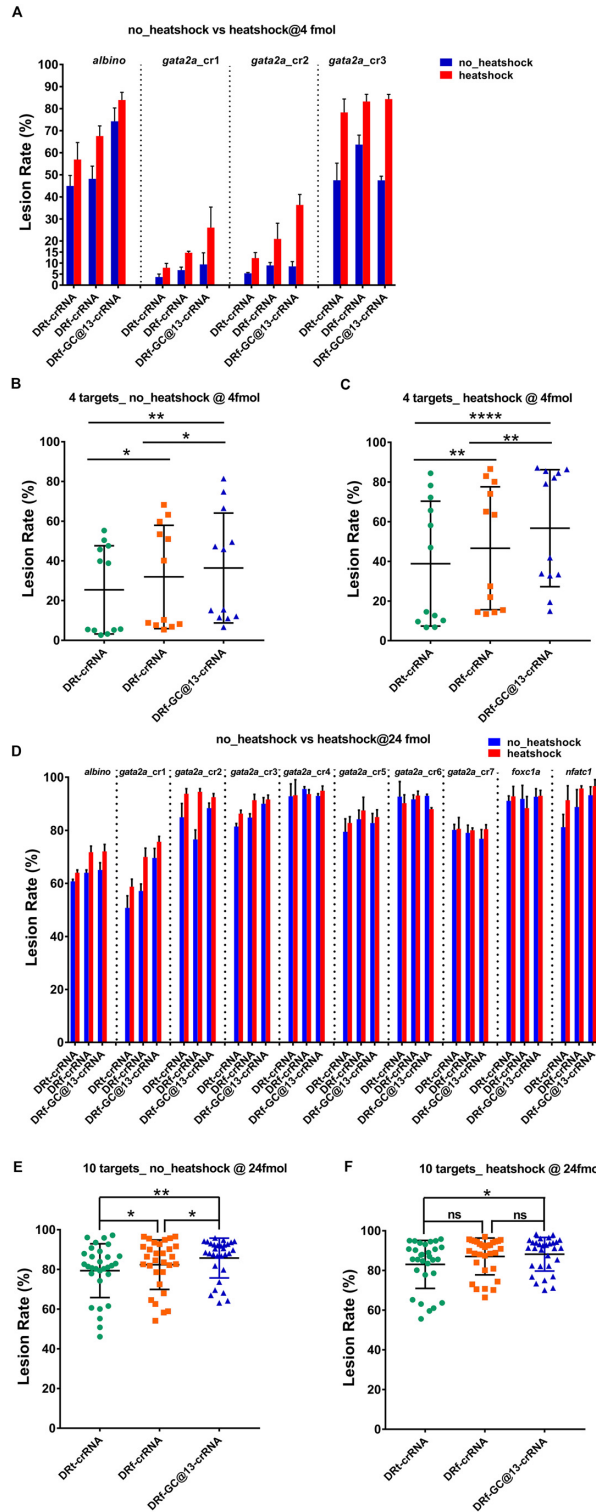
**Figure 3.** G-C swaps at specific position in the stem of the direct repeat increase editing efficiency. (A) General schematic of G-C swaps (indicated in red) at different position in the stem of the direct repeat hairpin. The activities of 9 different G-C swap crRNAs are determined by deep sequencing for LbCas12a (B) and FnoCas12a (C) at six different endogenous target sites on genome. Each box represents the 25th and 75th percentile and the middle line is the median. Whiskers in the box plots are defined by the Tukey method. Deep sequencing data are from three independent biological replicates performed on different days with expression constructs delivered by transient transfection in HEK293T cells (Supplementary Table S3). Error bars indicate  $\pm$  s.e.m. Statistical significance is determined by one-way analysis of variance (ANOVA), \*\*\* and \*\*\*\* denote  $P < 0.01$  and  $< 0.001$  respectively (Supplementary Table S7).

Since our Cas12a proteins differ in the composition of the C-terminal NLSs from those that were previously described (22), we performed an initial set of injections using 4 fmol LbCas12a RNP targeting four different sites at two genomic loci. In these injections, we evaluated three different crRNA frameworks (DRt-crRNA, DRf-crRNA, and DRf-GC@13-crRNA), which were synthesized by *in vitro* transcription. Following RNP injections, we performed deep sequencing on PCR amplicons spanning genomic target sites from treated embryos at 24 h post fertilization (hpf) to determine lesion rates. The DRf-crRNA significantly increased editing rates in embryos maintained at 28.5°C compared to DRt-crRNA (Figure 4A and B) and this effect was further enhanced with heat shock at 34°C (Figure 4A and C). At all four target sites, we also observed significant increases in lesion rates with RNPs containing the DRf-GC@13-crRNA in comparison to DRf-crRNA (Figure 4A–C). Importantly, there were no significant increases in toxicity associated with the higher editing rates achieved with either the DRf-crRNA or DRf-GC@13-crRNA frameworks (Supplementary Figure S8A and B). For one of the least active target sites (*gata2a\_cr2*), we performed a dose response analysis spanning four Lb-

Cas12a:crRNA concentrations (4, 8, 16 and 24 fmol), which revealed increasing mutagenesis as a function of nuclease concentration (Supplementary Figure S9A). At the highest dose (24 fmol), lesion rates approached 100% for the heat shocked embryos for all crRNA configurations (Supplementary Figure S9B and C) without substantial increases in toxicity with higher LbCas12a RNP dose (Supplementary Figure S9D).

Based on the LbCas12a:crRNA titration data, we examined the ability of 24 fmol LbCas12a RNPs programmed with the three different crRNA frameworks to edit ten target sites spanning four different genomic loci including some intronic sequences, which are easier to target with Cas12a than SpCas9 (22). Deep sequencing analysis of locus-specific editing rates revealed that both DRf-crRNA and DRf-GC@13-crRNA produced modestly higher mutagenesis levels than DRt-crRNA at these ten different target sites with or without heat shock (Figure 4). Consistent with the Moreno-Mateos study (22), we also observed that heat shock of the embryos at 34°C post injection modestly increased LbCas12a-mediated gene editing in aggregate at the ten target sites for each crRNA composition (Supplementary Figure S10A). Notably, lesion rates using these op-





**Figure 4.** Genome Editing by LbCas12a in zebrafish embryos. (A) Activity profiles of truncated Direct-Repeat crRNA (DRt-crRNA), full-length Direct-Repeat crRNA (DRf-crRNA) and G–C swapped crRNA (DRf-GC@13-crRNA) at four genomic sites in zebrafish embryos with and without heat shock using 4fmol RNP. Lesion rates are determined by deep sequencing. An aggregate analysis of the editing data (four target sites times three replicates) shows a significant increase in genome editing efficiency in fish embryos without heatshock (B) or with heatshock (C) for the DRf-crRNA and the DRf-GC@13-crRNA relative to the DRt-crRNA. (D) Activity profiles of three different crRNA frameworks at ten genomic sites in zebrafish embryos with and without heat shock treated with 24 fmol LbCas12a RNP. In the aggregate analysis across all ten target sites the DRf-GC@13-crRNA also provides a significant increase in genome editing efficiency in fish embryos without heatshock (E) or with heatshock (F). Deep sequencing data are from zebrafish embryos from three independent injections by three different individuals (Supplementary Table S5). For the bar charts, error bars indicate  $\pm$  s.e.m. For each dot plot the three lines represent 75th, 50th and 25th percentile, respectively. Statistical significance is determined by one-way analysis of variance (ANOVA), ‘\*’, ‘\*\*’, ‘\*\*\*’, ‘\*\*\*\*’ denotes  $P$  values of  $<0.05$ ,  $<0.01$  and  $<0.0001$ , respectively (Supplementary Table S7).

timized conditions exceeded 75% across all ten target sites. Accordingly, embryos injected with LbCas12a:crRNA targeting *albino* displayed a complete loss of pigmentation consistent with inactivation of both alleles (Supplementary Figure S10B). Analysis of the indels produced at these target sites revealed that the mutagenic end-products in the zebrafish embryos are microhomology-driven, similar to the observations in mammalian cell culture (Supplementary Figures S11–S16) (12). Interestingly, the spectrum of deletions that are produced within the embryos changed in the heat shock treated cohort, with a trend toward the reduction of insertions and creation of larger deletions in the genomes of heat shocked animals (Supplementary Figures S11–S16 and S21).

Given our promising results with FnoCas12a editing in mammalian cells, we examined its efficacy in zebrafish embryos. Because FnoCas12a recognizes a TTN PAM element (1,25,26), it should be able to target a wider spectrum of sequences within the genome than LbCas12a (TTTV PAM). Based on the results with LbCas12a, we employed the most active crRNA configuration (DRf-GC@13-crRNA) at a 24 fmol RNP dose to evaluate FnoCas12a activity at the same ten target sites. Lesions were observed at all ten sites with a median rate of >50%, with heat shock modestly increasing the indel rates and shifting the distribution of products toward deletions (Supplementary Figures S17–21). At five of the ten target loci FnoCas12a was just as effective as LbCas12a in producing targeted mutagenesis, suggesting that FnoCas12a can provide an important alternative nuclease for zebrafish mutagenesis when an appropriate PAM is not available for SpCas9 or LbCas12a. Collectively, these results demonstrate that delivery of pre-assembled Cas12a-crRNA RNP complexes provides a robust genome editing system in zebrafish.

## DISCUSSION

Type V CRISPR–Cas12a systems are a more recent addition to the genome editing toolbox (1,3) that are still being optimized for editing in vertebrate systems. Herein we describe improvements to the Cas12a nuclease components that improve editing rates in mammalian cells and zebrafish embryos. Suzuki and colleagues compared the editing activities, and nuclear localization of SpCas9 fused to different types of NLSs, which revealed that a bipartite NLS was superior to the standard SV40 NLS (36). Similar to these results for SpCas9, we find that the degree of nuclear localization and editing rates for three different Cas12as are improved by two C-terminus NLSs, one of which (Nucleoplasmin NLS) is bipartite in nature (42). It is possible that both the number and the composition of the NLSs impact nuclease activity, as a comparison of the editing efficiency of our LbCas12a with that of NEB LbCas12a, which contains two SV40 NLSs, revealed improved activity in mammalian cells (*AAVSI*) and zebrafish embryos (*slc45a2*; Supplementary Figure S22).

We also examined the impact of different crRNA compositions on the editing rate of Cas12a. We found that the inclusion of the full-length Direct Repeat improved the editing rates for Cas12a in mammalian cells and zebrafish embryos. This enhancement of activity was irrespective of the

delivered form of the nuclease (transient plasmid transfection or Cas12a-crRNA RNP using *in vitro* transcribed crRNAs). Moreno-Mateos and colleagues also examined the use of the full-length Direct Repeat crRNA (DRf-crRNA) to program LbCas12a and AsCas12a for editing in zebrafish embryos (22). While there was little comparative analysis between the DRt-crRNA and the DRf-crRNA in that study, their data suggested that Cas12a programmed with the full-length DR may have increased activity in zebrafish embryos, which is consistent with our analysis. While our study was being completed, a similar conclusion about the enhancement in activity for extended crRNAs in mammalian cells was reached in another study (43). Unique to our analysis, we demonstrate that crRNA processing by Cas12a is an important component for the enhancement in editing rates achieved with the full length DR, as mutation of the RRS or mutation of the RNase domain within Cas12a dramatically reduced the activity of Cas12a loaded with a full length DR crRNA.

Prior studies have examined the impact of chemical modifications within the crRNA or changes to the native crRNA sequence on Cas12a editing rates *in vitro* and in mammalian cells (1,8,43–45). We incorporated isosteric A:U to G:C base substitutions within the hairpin stem in an attempt to increase the thermal stability of the pseudoknot that forms the core RNA structure recognized by Cas12a (37–41). Interestingly, Cas12a makes few, if any, base specific contacts with the core of the hairpin stem of the pseudoknot (38,39) suggesting that it may be tolerant to more radical changes within this region that preserve the structure but alter its stability. However, sequence changes that extend or reduce the length of the hairpin stem (8,44), or create mismatches within the stem (1) have been shown to disrupt Cas12a activity. Note that our base substitutions were utilized in the context of the full-length direct repeat crRNAs produced by transcription. Their inclusion in the context of truncated synthetic crRNAs may not be beneficial.

Importantly, improvements in Cas12a activity associated with these crRNA modifications in mammalian cells translated to zebrafish embryos. Interestingly, the improvements in activity for the modified crRNAs were only apparent at the lower concentration injections for Cas12a (4 and 8 fmol RNP, Supplementary Figure S9). At higher concentrations (24 fmoles), we observed saturation of the editing rates, such that even heat shock only had a minor impact on the bulk edit rates in the injected population for the target sequences (Figure 4C and D). The 24 fmol dose is substantially higher than the highest LbCas12a injection dose (10 fmol) employed by Moreno-Mateos *et al.* (22), and yet that we find this dose is still well tolerated by the zebrafish, with >70% developing normally and no statistically significant increase in mortality (Supplementary Figure S8). Finally, we find that the utilization of heat shock on the embryos shifts the distribution of indels towards deletions with larger segments being removed (Supplementary Figure S21). Thus, the utilization of heatshock with Cas12a editing in zebrafish embryos can not only increase editing rates, but may also increase the production of sequence modifications that are more likely to be disruptive to gene function (dele-

tion of critical elements for protein function or nearby splice sites).

The demonstration that FnoCas12a, which utilizes a TTN PAM, has robust activity should also provide utility for zebrafish genome editing given its broader targeting range than LbCas12a. Although the FnoCas12a system had overall lower editing rates than LbCas12a in both mammalian cells and zebrafish embryos in our experiments, for 50% of the target sites in zebrafish it achieved similar editing efficiency to LbCas12a. Given that FnoCas12a can potentially target more than four times as many sequences within the zebrafish genome, this nuclease could be particularly useful for the targeted knock in of DNA sequences, since the efficiency of knock-in is a function of the proximity of the nuclease target site to the desired sequence alteration (46).

The extremely high editing rates and the low toxicity that is observed for LbCas12a suggests that this system may be amenable to F0 reverse genetic screens in zebrafish. F0 screens in zebrafish have been employed successfully with SpCas9 RNP injections (47,48). However, the efficiency of LbCas12a editing, its low toxicity and the prevalence for this nuclease to create deletions within the target sequence may provide an interesting alternative to SpCas9. In addition, since LbCas12a has more target sites within the zebrafish coding sequence than SpCas9 (22), it or other Cas12a variants that are being characterized (1,25) may maximize the rate of inactivation of a single locus in the context of multiplexed guides (48). We anticipate that the modifications to the Cas12a system described herein that increase editing rates for both plasmid expression systems and RNPs should have broad utility for editing in a variety of eukaryotic systems.

## DATA AVAILABILITY

The next-generation sequencing data have been deposited in the NCBI Sequence Read Archive database under the BioProject accession code PRJNA498760. All other relevant data are available from corresponding authors upon reasonable request.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

We thank E. Kittler and the UMass Medical School Deep Sequencing Core for their assistance with the Illumina sequencing and E. Sontheimer for insightful discussions. All new reagents described in this work are being deposited with the nonprofit plasmid-distribution service Addgene.

## FUNDING

National Institutes of Health (NIH) [R01HL093766 to S.A.W., N.D.L., R01AI117839 to S.A.W., J.L., R35HL140017 to N.D.L., U01HG007910 to M.G., J.L., 1R01GM115911 to S.A.W., E.S., 1UG3TR002668 to S.A.W., E.S., A.K. J.W.]. Funding for open access charge: NIH [R01AI117839].

**Conflict of interest statement.** The authors have filed patent applications related to genome engineering technologies. S.A.W. is a consultant for Acworth Pharmaceuticals.

## REFERENCES

- 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.
- Shmakov, S., Abudayyeh, O.O., Makarova, K.S., Wolf, Y.I., Gootenberg, J.S., Semenova, E., Minakhin, L., Joung, J., Konermann, S., Severinov, K. *et al.* (2015) Discovery and functional characterization of diverse class 2 CRISPR–Cas systems. *Mol. Cell*, **60**, 385–397.
- Shmakov, S., Smargon, A., Scott, D., Cox, D., Pyzocha, N., Yan, W., Abudayyeh, O.O., Gootenberg, J.S., Makarova, K.S., Wolf, Y.I. *et al.* (2017) Diversity and evolution of class 2 CRISPR–Cas systems. *Nat. Rev. Microbiol.*, **15**, 169–182.
- Koonin, E.V., Makarova, K.S. and Zhang, F. (2017) Diversity, classification and evolution of CRISPR–Cas systems. *Curr. Opin. Microbiol.*, **37**, 67–78.
- East-Seletsky, A., O’Connell, M.R., Burstein, D., Knott, G.J. and Doudna, J.A. (2017) RNA targeting by functionally orthogonal type VI-A CRISPR–Cas enzymes. *Mol. Cell*, **66**, 373–383.
- Gootenberg, J.S., Abudayyeh, O.O., Lee, J.W., Essletzbichler, P., Dy, A.J., Joung, J., Verdine, V., Donghia, N., Daringer, N.M., Freije, C.A. *et al.* (2017) Nucleic acid detection with CRISPR–Cas13a/C2c2. *Science*, **356**, 438–442.
- Zhang, Y., Long, C., Li, H., McAnally, J.R., Baskin, K.K., Shelton, J.M., Bassel-Duby, R. and Olson, E.N. (2017) CRISPR–Cpf1 correction of muscular dystrophy mutations in human cardiomyocytes and mice. *Sci. Adv.*, **3**, e1602814.
- 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.
- 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.* (2017) In vivo high-throughput profiling of CRISPR–Cpf1 activity. *Nat. Methods*, **14**, 153–159.
- Hendel, A., Bak, R.O., Clark, J.T., Kennedy, A.B., Ryan, D.E., Roy, S., Steinfeld, I., Lunstad, B.D., Kaiser, R.J., Wilkens, A.B. *et al.* (2015) Chemically modified guide RNAs enhance CRISPR–Cas genome editing in human primary cells. *Nat. Biotechnol.*, **33**, 985–989.
- Rahdar, M., McMahon, M.A., Prakash, T.P., Swayze, E.E., Bennett, C.F. and Cleveland, D.W. (2015) Synthetic CRISPR RNA–Cas9-guided genome editing in human cells. *Proc. Natl. Acad. Sci. U.S.A.*, **112**, E7110–E7117.
- 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.
- Canver, M.C., Lessard, S., Pinello, L., Wu, Y., Ilboudo, Y., Stern, E.N., Needleman, A.J., Galacteros, F., Brugnara, C., Kutlar, A. *et al.* (2017) Variant-aware saturating mutagenesis using multiple Cas9 nucleases identifies regulatory elements at trait-associated loci. *Nat. Genet.*, **49**, 625–634.
- Canver, M.C., Smith, E.C., Sher, F., Pinello, L., Sanjana, N.E., Shalem, O., Chen, D.D., Schupp, P.G., Vinjamur, D.S., Garcia, S.P. *et al.* (2015) BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature*, **527**, 192–197.
- Vierstra, J., Reik, A., Chang, K.-H., Stehling-Sun, S., Zhou, Y., Hinkley, S.J., Paschon, D.E., Zhang, L., Psatha, N., Bendana, Y.R. *et al.* (2015) Functional footprinting of regulatory DNA. *Nat. Methods*, **12**, 927–930.
- 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.
- Yan, W.X., Mirzazadeh, R., Garnerone, S., Scott, D., Schneider, M.W., Kallas, T., Custodio, J., Wernersson, E., Li, Y., Gao, L. *et al.* (2017) BLISS is a versatile and quantitative method for genome-wide profiling of DNA double-strand breaks. *Nat. Commun.*, **8**, 15058.

18. Port, F. and Bullock, S.L. (2016) Augmenting CRISPR applications in *Drosophila* with tRNA-flanked sgRNAs. *Nat. Methods*, **13**, 852–854.
19. Hur, J.K., Kim, K., Been, K.W., Baek, G., Ye, S., Hur, J.W., Ryu, S.-M., Lee, Y.S. and Kim, J.-S. (2016) Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nat. Biotechnol.*, **34**, 807–808.
20. Kim, Y., Cheong, S.-A., Lee, J.G., Lee, S.-W., Lee, M.S., Baek, I.-J. and Sung, Y.H. (2016) Generation of knockout mice by Cpf1-mediated gene targeting. *Nat. Biotechnol.*, **34**, 808–810.
21. Watkins-Chow, D.E., Varshney, G.K., Garrett, L.J., Chen, Z., Jimenez, E.A., Rivas, C., Bishop, K.S., Sood, R., Harper, U.L., Pavan, W.J. *et al.* (2017) Highly efficient Cpf1-Mediated gene targeting in mice following high concentration pronuclear injection. *G3 (Bethesda)*, **7**, 719–722.
22. 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.
23. 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.
24. 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.
25. 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.
26. 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.
27. Endo, A., Masafumi, M., Kaya, H. and Toki, S. (2016) Efficient targeted mutagenesis of rice and tobacco genomes using Cpf1 from *Francisella novicida*. *Sci. Rep.*, **6**, 38169.
28. Westerfield, M. (2007) *The Zebrafish Book*. 5th edn. University of Oregon Press, Eugene.
29. Bolukbasi, M.F., Gupta, A., Oikemus, S., Derr, A.G., Garber, M., Brodsky, M.H., Zhu, L.J. and Wolfe, S.A. (2015) DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat. Methods*, **12**, 1150–1156.
30. Holder, N. and Xu, Q. (1999) Microinjection of DNA, RNA, and Protein into the Fertilized Zebrafish Egg for Analysis of Gene Function. In: *Molecular Embryology, Molecular Embryology: Methods and Protocols*. Humana Press, New Jersey, Vol. **97**, pp. 487–490.
31. Ihaka, R. and Gentleman, R. (1996) R: a language for data analysis and graphics. *J. Comput. Graph. Stat.*, **5**, 299–314.
32. 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.
33. Zhang, J., Kobert, K., Flouri, T. and Stamatakis, A. (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, **30**, 614–620.
34. Blankenberg, D., Gordon, A., Kuster Von, G., Coraor, N., Taylor, J., Nekrutenko, A. and Team, T.G. (2010) Manipulation of FASTQ data with Galaxy. *Bioinformatics*, **26**, 1783–1785.
35. Makkerh, J.P., Dingwall, C. and Laskey, R.A. (1996) Comparative mutagenesis of nuclear localization signals reveals the importance of neutral and acidic amino acids. *Curr. Biol.*, **6**, 1025–1027.
36. Suzuki, K., Tsunekawa, Y., Hernandez-Benitez, R., Wu, J., Zhu, J., Kim, E.J., Hatanaka, F., Yamamoto, M., Araoka, T., Li, Z. *et al.* (2016) In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature*, **540**, 144–149.
37. Dong, D., 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.
38. Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I.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.
39. Gao, P., Yang, H., Rajashankar, K.R., Huang, Z. and Patel, D.J. (2016) Type V CRISPR–Cas Cpf1 endonuclease employs a unique mechanism for crRNA-mediated target DNA recognition. *Cell Res.*, **26**, 901–913.
40. Swarts, D.C., van der Oost, J. and Jinek, M. (2017) Structural basis for guide RNA processing and Seed-Dependent DNA targeting by CRISPR–Cas12a. *Mol. Cell*, **66**, 221–233.
41. Stella, S., Alcón, P. and Montoya, G. (2017) Structure of the Cpf1 endonuclease R-loop complex after target DNA cleavage. *Nature*, **546**, 559–563.
42. Kosugi, S., Hasebe, M., Matsumura, N., Takashima, H., Miyamoto-Sato, E., Tomita, M. and Yanagawa, H. (2009) Six classes of nuclear localization signals specific to different binding grooves of importin alpha. *J. Biol. Chem.*, **284**, 478–485.
43. Park, H.M., Liu, H., Wu, J., Chong, A., Mackley, V., Fellmann, C., Rao, A., Jiang, F., Chu, H., Murthy, N. *et al.* (2018) Extension of the crRNA enhances Cpf1 gene editing in vitro and in vivo. *Nat. Commun.*, **9**, 3313.
44. 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**, 0066.
45. Yin, H., Song, C.-Q., Suresh, S., Kwan, S.-Y., Wu, Q., Walsh, S., Ding, J., Bogorad, R.L., Zhu, L.J., Wolfe, S.A. *et al.* (2018) Partial DNA-guided Cas9 enables genome editing with reduced off-target activity. *Nat. Chem. Biol.*, **14**, 311–316.
46. Paquet, D., Kwart, D., Chen, A., Sproul, A., Jacob, S., Teo, S., Olsen, K.M., Gregg, A., Noggle, S. and Tessier-Lavigne, M. (2016) Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature*, **533**, 125–129.
47. Shah, A.N., Davey, C.F., Whitebirch, A.C., Miller, A.C. and Moens, C.B. (2015) Rapid reverse genetic screening using CRISPR in zebrafish. *Nat. Methods*, **12**, 535–540.
48. Wu, R.S., Lam, I.I., Clay, H., Duong, D.N., Deo, R.C. and Coughlin, S.R. (2018) A rapid method for directed gene knockout for screening in G0 Zebrafish. *Dev. Cell*, **46**, 112–125.