Enhanced Cas12a editing in mammalian cells and zebrafish

Pengpeng Liu
University of Massachusetts Medical School

Et al.

Let us know how access to this document benefits you.
Follow this and additional works at: https://escholarship.umassmed.edu/oapubs

Part of the Biochemistry, Biophysics, and Structural Biology Commons, Cell and Developmental Biology Commons, Embryonic Structures Commons, Genetic Phenomena Commons, Genetics and Genomics Commons, and the Nucleic Acids, Nucleotides, and Nucleosides Commons

Repository Citation

Creative Commons License
This work is licensed under a Creative Commons Attribution-Noncommercial 4.0 License
This material is brought to you by eScholarship@UMassChan. It has been accepted for inclusion in Open Access Publications by UMass Chan Authors by an authorized administrator of eScholarship@UMassChan. For more information, please contact Lisa.Palmer@umassmed.edu.
Enhanced Cas12a editing in mammalian cells and zebrafish

Pengpeng Liu¹, Kevin Luk¹, Masahiro Shin¹, Feston Idrizi¹, Samantha Kwok¹, Benjamin Roscoe¹, Esther Mintzer¹, Sneha Suresh¹, Kyle Morrison², Josias B. Frazão¹, Mehmet Fatih Bolukbasi¹,³, Karthikeyan Ponnienselvan¹, Jeremy Luban³,⁴,⁵, Lihua Julie Zhu¹,⁴,⁶, Nathan D. Lawson¹,⁵,* and Scot A. Wolfe¹,³,⁵,*

¹Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA, ²Department of Chemistry and Biochemistry, Worcester Polytechnic Institute, Worcester, MA, USA, ³Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA, ⁴Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, MA, USA, ⁵Li Weibo Institute for Rare Diseases Research, University of Massachusetts Medical School, Worcester, MA, USA and ⁶Program 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,
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 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 by 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 Figure 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% CO2. 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% CO2. 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 × 10^5 cells are transfected using Polyfect transfection reagent (Qiagen) 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 × 10^5 cells in Neon R buffer at the desired concentration and electroporated using Neon® 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 (DE3) pLysS cells (EMD Milipore) for protein overexpression. Cells were grown at 37°C to an OD600 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 dialed 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 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 × 10^5 cells in Neon R buffer at the desired concentration and electroporated using Neon® 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 × 10^5 cells in Neon R buffer at the desired concentration and electroporated using Neon® Transfection System 10 L Kit (Thermo Fisher Scientific) using the suggested electroporation parameters: Pulse voltage (1600 V), Pulse width (10 ms), Pulse number (3).
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 Ultrace -30K (Amicon) to a concentration of between 30 and 50 µM. 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 Ultrace -30K (Amicon) to a concentration of between 30 and 50 µM. 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 complementarity to the common scaffold oligonucleotide on the 3’ end (Supplementary Table S8). To generate each crRNA expression template, a 60 µl PCR containing 1X Phusion Buffer, 200 µM 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 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 computations 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 Heat shock+ and Heat shock– 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 µl reaction containing 1X Phire Buffer, 100 µM forward and reverse primer, and 0.5 µl of Phire Hot Start II DNA Polymerase (Thermofisher Cat#KF122L) was performed using the following thermocycling program: initial denaturation 98°C for 3 minutes, 31 cycles of 98°C for 20 s, 62°C for 20 s, 72°C for 30 s, final extension at 72°C for 5 min, and hold at 12°C. DNA amplification 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 adapters 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 DRI-crRNA targeting the AAVS1 locus (Supplementary Table S8) was in vitro transcribed as described above. Different amounts (5 pmol, 10 pmol, 20 pmol, 40 pmol and 80 pmol) 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 × 10⁵ 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 pmol) complexed with the Albino crRNA at 2.5-fold excess, performed microinjection of Cas12a protein and crRNAs into one-cell-stage zebrafish embryos as previously 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 EMXIS1; 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 EMXIS1 target sites (Supplementary Figures S2A–D and S3A–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-
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), truncated Direct-Repeat crRNA with a 3′ truncated Direct-Repeat (DFR-crRNA), and full-length Direct-Repeat crRNA with a 3′ full-length Direct-Repeat (DFR-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 DR1-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 DR1-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.

Genome editing in zebrafish embryos

Only a single previous study has successfully utilized Lb-Cas12a 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 process 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 ±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).
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 LbCas12a: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 heat shock (B) or with heat shock (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 heat shock (E) or with heat shock (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 ± 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 (Nucleo-plasmin 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 (AAVS1) 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 cr-RNAs). 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 DRf-crRNA and the DRt-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 isoteric 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 it is possible that 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 heat shock 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 is observed for LbCas12a suggests that this system may

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

1. Zetsche,B., Gootenberg,J.S., Abudayyeh,O.O., Slaymaker,I.M.,
Makarova,K.S., Esvelt,K.M., Joung,J.W., van der Oost,J.,
Regan,A. et al. (2015) Cas9 is a single RNA-guided endonuclease of a
class 2 CRISPR–Cas system. Cell, 163, 759–771.

2. Shmakov,S., Abudayyeh,O.O., Makarova,K.S., Wolf,Y.I.,
Gootenberg,J.S., Semenova,E., Minakhin,L., Joung,J.W.,
characterization of diverse class 2 CRISPR–Cas systems. Mol. Cell, 60,
385–397.

3. Shmakov,S., Smargon,A., Scott,D., Cox,D., Pyzochna,N., Yan,W.,
Abudayyeh,O.O., Gootenberg,J.S., Makarova,K.S., Wolf,Y.I. et al.

classification and evolution of CRISPR–Cas systems. Curr. Opin.
Microbiol., 37, 67–78.

5. East-Seletsky,A., O’Connell,M.R., Burstein,D., Knott,G.J. and
Doudna,J.A. (2017) RNA targeting by functionally orthogonal type

Dy,A.J., Joung,J., Verdin,E., Donghia,N., Daringer,N.M.,
Freije,C.A. et al. (2017) Nucleic acid detection with
CRISPR–Cas3a/C2c2. Science, 356, 438–442.

7. Zhang,Y., Long,C., Li,H., McAnally,J.R., Baskin,K.K.,
Steinfeld,I., Lunstad,B.D., Kaiser,R.J., Wilkens,A.B.
RNA-guided genome editing in human primary cells. Nat.

8. Rahdar,M., McMahon,M.A., Prakash,T.P., Swayne,E.E.,

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.

Chemically modified guide RNAs enhance CRISPR–Cas genome

Variant-aware saturating mutagenesis using multiple Cas9 nucleases
identifies regulatory elements at trait-associated loci. Nat. Genet.,
49, 634–643.

12. Canver,M.C., Smith,E.C., Sher,F., Pinello,L., Sanjana,N.E.,
(2015) BCL11A enhancer dissection by Cas9-mediated in situ

Variant-aware saturating mutagenesis using multiple Cas9 nucleases
identifies regulatory elements at trait-associated loci. Nat. Genet.,
49, 634–643.

(2015) Functional footprinting of regulatory DNA. Nat. Methods,
12, 927–930.

15. Kleinstiver,B.P., Tsai,S.Q., Prew,M.S., Nguyen,N.T., Welch,M.M.,
Characterization of diverse class 2 CRISPR–Cas systems. Cell
759–771.

16. Gootenberg,J.S., Abudayyeh,O.O., Makarova,K.S., Wolf,Y.I.,
Gootenberg,J.S., Semenova,E., Minakhin,L., Joung,J.W.,
characterization of diverse class 2 CRISPR–Cas systems. Mol. Cell, 60,
385–397.

17. Yan,W.X., Mirzazadeh,R., Garnerone,S., Scott,D., Schneider,M.W.,
BLISS is a versatile and quantitative method for genome-wide

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) [R01HL03766 to S.A.W., N.D.L., R01AI117839 to S.A.W., J.L., R35HL140017 to N.D.L., U01HG07910 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].


