The CRISPR Journal
Volume 2, Number 5, 2019
Mary Ann Liebert, Inc.
DOI: 10.1089/crispr.2019.0026

The CRISPR Journal

ORIGINAL ARTICLE

Expanding the CRISPR Toolbox with ErCas12a in Zebrafish and Human Cells

Wesley A. Wierson,[1,†] Brandon W. Simone,[2,†] Zachary WareJoncas,[2] Carla Mann,[1] Jordan M. Welker,[1] Bibekananda Kar,[2] Michael J. Emch,[2] Iddo Friedberg,[3] William A.C. Gendron,[2] Michael A. Barry,[2] Karl J. Clark,[2] Drena L. Dobbs,[1] Maura A. McGrail,[1] Stephen C. Ekker,[2,*] and Jeffrey J. Essner[1,*]

Abstract

CRISPR and CRISPR-Cas effector proteins enable the targeting of DNA double-strand breaks to defined loci based on a variable length RNA guide specific to each effector. The guide RNAs are generally similar in size and form, consisting of a ~20 nucleotide sequence complementary to the DNA target and an RNA secondary structure recognized by the effector. However, the effector proteins vary in protospacer adjacent motif requirements, nuclease activities, and DNA binding kinetics. Recently, ErCas12a, a new member of the Cas12a family, was identified in *Eubacterium rectale*. Here, we report the first characterization of ErCas12a activity in zebrafish and expand on previously reported activity in human cells. Using a fluorescent reporter system, we show that CRISPR-ErCas12a elicits strand annealing mediated DNA repair more efficiently than CRISPR-Cas9. Further, using our previously reported gene targeting method that utilizes short homology, GeneWeld, we demonstrate the use of CRISPR-ErCas12a to integrate reporter alleles into the genomes of both zebrafish and human cells. Together, this work provides methods for deploying an additional CRISPR-Cas system, thus increasing the flexibility researchers have in applying genome engineering technologies.

Introduction

CRISPR systems have been widely adopted in zebrafish research due to their efficacy and ease of reprogramming DNA binding activity, which is mediated by a single chimeric single guide RNA (sgRNA) molecule.[1–3] The CRISPR toolbox continues to expand, with the identification of systems that display varying protospacer adjacent motif (PAM) requirements and produce a different double-strand break (DSB) architecture.[4] While Cas9 proteins often hydrolyze DNA, leaving a blunt-ended cut three base pairs (bp) upstream from the 5′ end of the PAM sequence, Cas12a proteins from *Acidaminococcus* and *Lachnospiraceae* spp. hydrolyze DNA in a staggered fashion, cutting on the 3′ side of the PAM and leaving 5′ single-stranded overhangs of four nucleotides.[5] The architecture of the DSB and end resection products is a critical determinant of DNA repair pathway activation.[6] For example, inducing DNA overhangs with staggered CRISPR-Cas9 nickases targeted to opposite strands can stimulate precision genome engineering using oligonucleotides.[7] Thus, there is demand for CRISPR variants that generate different DSB architectures and elicit more predictable repair outcomes for precision genome engineering.

CRISPR-Cas12a activity has been reported in zebrafish by injection of sgRNA/Cas12a protein ribonucleoprotein (RNP) complexes.[8,9] In these studies, it was shown that Cas12a-mediated DNA cleavage could be further enhanced by a 34°C heat-shock treatment or by co-targeting of nuclease dead Cas9 (dCas9) to the Cas12a target site, indicating that DNA melting is a potential rate limiting step for Cas12a. Cas12a resulted in increased efficiency of oligonucleotide incorporation compared to Cas9 into genomic cut sites by homology-directed repair, suggesting the two enzymes may employ distinct mechanisms and result in different genomic DNA end resection products.[8] A CRISPR-Cas12a system was identified in *Eubacterium rectale* and found to be ancestrally related to the type V

Departments of [1]Genetics, Development and Cell Biology and [3]Veterinary Microbiology and Preventive Medicine, Iowa State University, Ames, Iowa; and [2]Department of Biochemistry and Molecular Biology, Mayo Clinic, Rochester, Minnesota.
[†]These authors contributed equally to this work.

*Address correspondence to: *Stephen C. Ekker, PhD, Department of Biochemistry and Molecular Biology, Mayo Clinic, 200 1st St. SW, 1342C Guggenheim, Rochester, MN 55905,* E-mail: ekker.stephen@mayo.edu; *or Jeffrey J. Essner, PhD, Department of Genetics, Development and Cell Biology, Iowa State University, 2213 Pammel Drive, Ames, IA 50011-2140,* E-mail: jessner@iastate.edu

© Wesley A. Wierson *et al*. 2019; Published by Mary Ann Liebert, Inc. This Open Access article is distributed under the terms of the Creative Commons License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

class II family of CRISPR proteins, sharing the greatest similarity with Cas12a from *Acidaminococcus* sp. (AsCas12a; Fig. 1A). *E. rectale* Cas12a (ErCas12a) is a 1262 amino acid protein that is reported to recognize a 5′-YTTN PAM and uses a 42 or 56 base guide RNA to recognize and catalyze site-specific DNA cleavage.[10] However, in this study we find that ErCas12a elicits robust nuclease activity with a TTTT PAM suggesting that the PAM requirement may actually be YTTN. CRISPR-ErCas12a is active in human cells,[10] but its application for gene targeting *in vivo* and *in vitro* has yet to be described.

Microhomology mediated end joining (MMEJ) is a DNA repair pathway that uses short regions of homology between resected DNA ends to drive repair, leading to predictable outcomes after nuclease targeting.[11–13] MMEJ-based gene editing and targeted integration using MMEJ has been described in zebrafish and mammalian cells, although the length of the homology arms appears to be more similar to those employed in single-strand annealing (SSA).[14–17] To use MMEJ/SSA for gene targeting, nuclease targeting of a donor plasmid *in vivo* liberates a donor cassette, exposing homology arms that are complementary to the nuclease cut site in the genome, and thereby directs integration into genomic DNA. A more general term for the use of this strategy for gene targeting is ''homology-mediated end joining'' (HMEJ).[18] We recently reported high-efficiency gene targeting by HMEJ with short homology arms called GeneWeld.[19] GeneWeld involves the simultaneous cleavage of a donor plasmid and a genomic target with designer nucleases to reveal 24 or 48 bp homology arms that can be used by the cellular MMEJ or SSA machinery for homology-directed repair, leading to donor integration. GeneWeld works with genomic DSB generated by either transcription activator-like effector nucleases or CRISPR-Cas9, but other nucleases have not yet been tested.

Here, we report the successful application of ErCas12a for gene targeting in zebrafish, widely used to model development and disease, and in human cells. Consistent with its similarity to other Cas12a proteins, high ErCas12a activity requires a 34°C heat-shock treatment in zebrafish. However, injection of mRNA encoding for ErCas12a is sufficient to induce DSB activity, and pre-crRNA can serve as an effective RNA guide for ErCas12a, in contrast to reports using previously described Cas12a and related systems.[8,20] We developed a universal pre-crRNA (U-pre-crRNA) for ErCas12a, which can be used in both zebrafish and human cells for donor DNA cleavage, and we show that CRISPR-ErCas12a potently induces strand-annealing mediated repair (SAMR) in a genomic reporter locus at efficiencies greater than CRISPR-Cas9. Additionally, CRISPR-ErCas12a promotes GeneWeld activity at rates similar to CRISPR-Cas9. Finally, we apply GeneWeld with ErCas12a in human cells and demonstrate ErCas12a-mediated targeted integration activity at several human loci, including the safe harbor locus *AAVS1*.

Methods

Zebrafish husbandry

Zebrafish were maintained in Aquatic Habitats (Pentair) housing on a 14 h/10 h light/dark cycle. Wild-type WIK were obtained from the Zebrafish International Resource Center. All experiments were carried out under approved protocols from Iowa State University IACUC.

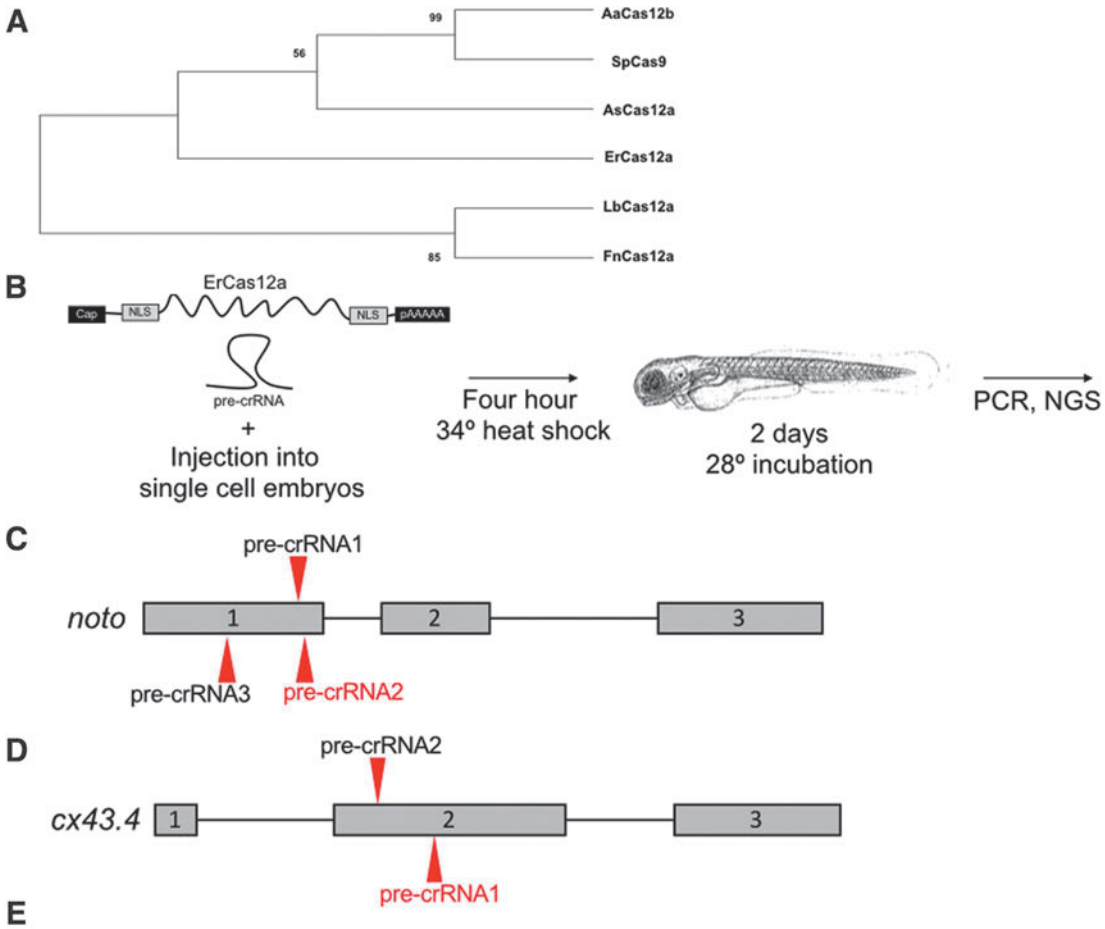
nErCas12an cDNA cloning

gBlocks were ordered from IDT, with zebrafish codon optimized ErCas12A cDNA sequences based on Inscripta public disclosure and the addition of dual NLS sequences at the 5′ and 3′ end of the cDNA (Supplementary Table S1). Three gBlock dsDNA templates were amplified with KOD Hot-Start DNA polymerase (EMD Millipore) using primers ErCas12af1/ErCas12aAr1, ErCas12af2/ErCas12ar2, and ErCas12af3/ErCas12ar3 cut with respective restriction enzymes, and four-part restriction cloning into NcoI/SacII cut pT3TS vector backbone was performed (Plasmid #46757; Addgene). Sequence of ERCAS12A using primer walking and a full annotation was confirmed.

Injection protocol

Linear, purified pT3TS-nCas9n or pT3TS-nErCas12an was used as template for *in vitro* transcription of capped, polyadenylated mRNA with the Ambion T3TS mMessage mMachine Kit. mRNA was purified using the Qiagen miRNeasy Kit. The Cas9 universal sgRNAs were generated

FIG. 1. Characterization and activity of CRISPR-ErCas12a in zebrafish. **(A)** Phylogenetic relationship between known CRISPR associated proteins and ErCas12a. Evaluation of branching was ensured by bootstrap statistical analysis (1,000 replications). **(B)** Workflow showing dual NLS ErCas12a mRNA and pre-crRNA injection into single-cell animals. Animals are heat shocked for 4 h and then allowed to develop normally until DNA is isolated and analyzed. **(C)** Schematic and sequences of pre-crRNA used to target *noto*. **(D)** Schematic and sequences of pre-crRNA used to target *cx43.4*. **(E)** Results displaying the wild-type and top five mutated alleles after AmpliconEZ analysis of two *noto* pre-crRNAs and one *cx43.4* pre-crRNA.



E

Target	Sequence	Indel Length	Microhomology Signature	% of Total Read Count	Total mutant
	noto target 1				
noto target 1 WT:	TTTACTCGCAGATGCCACACTTCGCGTACAGCCAAAGCATCATGCAAACTCAG	N/A	N/A	38.85	61.15%
Allele 1:	TTTACTCGCAGATGCCACA----------GCCAAAGCATCATGCAAACTCAG	-11	ACA	9.85	
Allele 2:	TTTACTCGCAGATGCCACAC---------AGCCAAAGCATCATGCAAACTCAG	-9	AC	7.84	
Allele 3:	TTTACTCGCAGATGCCA----------AAGCATCATGCAAACTCAG	-17	GCCA	5.23	
Allele 4:	TTTACTCGCAGATGCCACA----------GCATCATGCAAACTCAG	-17	GCCACA	1.75	
Allele 5:	TTTACTCGCAGATGCCACAC----------CAAAGCATCATGCAAACTCAG	-12	N/A	1.35	
	noto target 3				
noto target 3 WT:	CCAGGCGGAGATGAGAGAACGAACAAACGCGTACCGGAGCATAACCAACCAAA	N/A	N/A	10.47	89.53%
Allele 1:	CCAGGCGGAGATGAGAGAACGAACAA----------GCATAACCAACCAAA	-12	N/A	12.13	
Allele 2:	CCAGGCGGAGATGAGAGAACGAACA----------TAACCAACCAAA	-16	CA	7.67	
Allele 3:	CCAGGCGGAGATGAGAGAACGA----------GCATAACCAACCAAA	-16	GA	6.12	
Allele 4:	CCAGGCGGAGATGAGAGAACGAACAA----------TAGCATAACCAACCAAA	-10	N/A	3.38	
Allele 5:	CCAGGCGGAGATGAGAGAACGAAC----------GAGCATAACCAACCAAA	-12	N/A	2.06	
	cx43.4 target 2				
cx43.4 target 2 WT:	TTTGACTGTTGTGGGGGGAGAATCGATATACTACGATGAACAGAGCAAATTTG	N/A	N/A	73.68	24.32%
Allele 1:	TTTGACTGTTGTGGGGGGA---------TACTACGATGAACAGAGCAAATTTG	-9	N/A	1.01	
Allele 2:	TTTGACTGTTGTGGGGGGAGA----------TGAACAGAGCAAATTTG	-15	GA	0.65	
Allele 3:	TTTGACTGTTGTGGGGGGAG--------TACTACGATGAACAGAGCAAATTTG	-8	N/A	0.48	
Allele 4:	TTTGACTGTTGTGGGGGGAGA-------TACTACGATGAACAGAGCAAATTTG	-7	N/A	0.42	
Allele 5:	TTTGACTGTTGTGGGGGGAG-TTC---ATACTACGATGAACAGAGCAAATTTG	-4	N/A	0.34	

using cloning free sgRNA synthesis and purified using the Qiagen miRNeasy Kit. All ErCas12a pre-crRNA and crRNA was ordered as custom RNA oligos from Synthego, with sequences described in the Supplementary Tables. In all experiments, 2 nL of 75 pg/pl of ErCas12a and Cas9 were injected. Unless otherwise noted, 2 nL of 12.5 pg/pl pre-crRNA and/or U-pre-crRNA were injected.

Heat shock protocol

Immediately after injection, embryos were placed in a 34°C incubator for 4 h. At 4 h, embryos were sorted for fertilization, and fertilized embryos were moved to a 28°C incubator as normal.

DNA isolation and PCR genotyping

Genomic DNA for PCR was extracted by digestion of single embryos in 50 mM NaOH at 95°C for 30 min and neutralized by addition of 1/10th volume 1 M Tris-HCl, pH 8.0. GoTaq Green was used as the DNA polymerase master mix with the primers listed in Supplementary Table S2. AmpliconEZ from GeneWiz was used for NGS sequencing (see below) using primers ErCas12anoto1fEZ and ErCas12anoto1rEZ for *noto* target 1, ErCas12anoto3fEZ and ErCas12anoto3rEZ for *noto* target 3, and cxm7gRNA2fEZ, cxm7gRNA2rEZ for *cx43.4* listed in Supplementary Table S2. GFP+ embryo 5′ junction fragments for *noto* target 1 and target 3 were PCR-amplified with primer notojxnf and gfp5′r listed in Supplementary Table S2. GFP+ embryo 3′ junction fragments for *noto* target 3 were PCR-amplified with primer GFP3′F and notojxnr listed in Supplementary Table S2. All junction fragment products were cloned into pCR4-TOPO vector and sequenced (Invitrogen).

Donor vector preparation

Donor vectors were prepared and purified, as described previously.[19] Homology arms are built as follows. One arm begins 13 bp 3′ of the PAM, while the other arm begins immediately outside of the 3′ end of the crRNA target site. See Figures 3 and 4 for homology arm design, and Supplementary Table S3 for GeneWeld homology arm oligos used for Golden Gate cloning. Gene targeting oligos and donor vector sequences are listed in Supplementary Tables S2 and S3.

Generating noto:RFP-DR48

Zebrafish RFP assay generation, injection, and line isolation. pPRISM-V3 was PCR amplified with v3f and v3r to remove the ocean-POUT terminator and add SgrAI and SpeI cloning sites (Supplementary Table S2). Bactin 3′ UTR was PCR amplified using KOD polymerase with primers bactinf and bactinr to add SgrAI and SpeI enzyme sites for sticky end cloning. pPRISM-V3 (pout negative) amplicon and bactin 3′ UTR were cut with SgrAI and SpeI. After ligation with Fisher Optizyme T4 Ligase and sequence verification to create pPRISM-V3-bactin, pPRISM-V3-bactin-SSA-DR48 was created by simultaneously adding the NBM and creating a direct repeat with phosphorylated primers DRf and DRr using pPRISM-V3-bactin with KOD polymerase followed by Fisher Optizyme T4 Ligation and sequence verification. To target these constructs to *noto*, homology domains up- and downstream of a genomic CRISPR-Cas9 target site were chosen, as described in Wierson *et al.*[19] Oligos flhv35aflh, flhv35bflh, flhv33aflh, and flhv33bflh, containing the gene targeting information, were added to pPRISM-V3-bactin RFP-DR48 using Golden Gate cloning, as described in Wierson *et al.*[19] The RFP cassette was liberated from the donor using the same *noto* sgRNA used to cut the genome. Gamma-crystalin:eGFP+ embryos were sorted and raised to adulthood, outcrossed to generate the F1 generation, and outcrossed again to generate lines of F2s.

Southern blot analysis

Genomic Southern blot and copy number analysis was conducted, as previously described.[21] PCR primers used for genomic- and donor-specific probes are listed in Supplementary Table S2.

Cell culture

HEK293T cells were obtained from ATCC (CRL-3216). Cells were maintained in Dulbecco's modified Eagle's medium (Gibco #11995-040) supplemented with 10% fetal bovine serum (Gibco #26140079) and 1% penicillin/streptomycin (Gibco #15140-122). Media was changed every 2–3 days and re-plated at a final dilution of 1:10 maintained at about 750,000 cells/mL.

DNA isolation and sequence analysis

DNA from whole-cell populations was purified using the Qiagen DNeasy Blood & Tissue Kit (Qiagen 69504). PCR amplification was performed with MyTaq DNA Polymerase (Bioline BIO-21108) and purified with the Qiagen QIAquick PCR Purification Kit (Qiagen 28104). Samples used for ICE analysis were submitted to GeneWiz Sanger Sequencing service.

Cloning in vitro *ErCas12a construct targeting AAVS1 (pErCas12aErCas12a-AAVS1)*

Due to redundant restriction sites in the guide scaffold and ErCas12a protein, the pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA plasmid (Addgene

#61591) was digested with BsaI and NotI first to insert the ErCas12a secondary structure and sgRNA targeting AAVS1 with ErCas12a sgRNA AAVS1 top and ErCas12a sgRNA AAVS1 bottom (termed AAV: ErCas12a AAVS1 sgRNA; Table 3). Following this, ErCas12a as well as the Xenopus globin 5′ UTR and both N and C termini SV40 NLS signals were amplified from ''T3TS nErCas12an'' using PCR primers ErCas12a AgeI forward and ErCas12a BamHI reverse (Supplementary Table S1). The resulting 3.9 kbp PCR fragment was cloned into an Agilent pSC Strataclone PCR cloning vector (termed ErCas12a Strataclone) to amplify the fragment with the desired restriction sites. ErCas12a Strataclone was digested with AgeI and BamHI to isolate ErCas12a with ends compatible with the AAV:ErCas12a AAVS1 sgRNA construct. The px601 plasmid was likewise digested with AgeI and BamHI to remove SaCas9 and replace it with ErCas12a. Plasmids were screened for insertion of both ErCas12a as well as the AAVS1 targeting sgRNA (pErCas12a-AAVS1) and amplified with the Qiagen Endotoxin Free Maxiprep Kit (Qiagen 12362).

pErCas12a-AAVS1-U

pErCas12a-AAVS1-U was generated by inserting ErCas12a pre-U-crRNA into AAV 601 by digesting AAV601 with BsaI+ NotI and inserting annealed oligos ErCas12a UgRNA + Scaffold Top and ErCas12a UgRNA + Scaffold Bottom. The U6 promoter, the crRNA scaffold, and the U-pre-crRNA were PCR amplified from AAV 601 using ErCas12a U6 + ugRNA Primer PfoI Fw and ErCas12a U6 + ugRNA Primer PfoI Rev. This PCR amplicon as well as pErCas12a-AAVS1 were digested with PfoI and ligated with T4 DNA ligase to generate pErCas12a-AAVS1-pre-crRNA1-U-pre-crRNA.

Generating knock-in cassette for AAVS1

The 24 and 48 bp homology arm CMV/GFP/Zeocin resistance knock-in cassettes were generated by designing PCR primers complementary to the psiRNA-SV40 Early PolyA GFPzeo plasmid (Invivogen) flanked by the 48 bp of homology and the UgRNA sequence separated by a 3 bp spacer. The left 48HA forward primer AAVSI T1 L48HA and the right 48HA reverse primer AAVS1 T1 R48HA[22] were used to amplify the CMV/GFP/Zeocin resistance cassette containing the homology arms and the UgRNA target sequence (Supplementary Tables S2 and S3). This 2.6 kbp PCR fragment was subsequently cloned into a Strataclone PCR cloning vector and screened for the insert. Sequence confirmed knock-in constructs were amplified with the Qiagen Maxiprep Kit and termed ''pGFP::Zeo-48.''

Generating ErCas12a EZ clone

nErCas12an was PCR amplified from pT3TSnErCas12an with Platinum Taq DNA Polymerase HiFi (Thermo Fisher Scientific #11304011) with ErCas12a AgeI Fw and ErCas12a BamHI Rev and PCR cloned into the Agilent Strataclone vector (Agilent #240205) to generate ErCas12a Strataclone. ErCas12a Strataclone was subjected to site-directed mutagenesis with ErCas12a SDM remove BsaI Top and ErCas12a SDM remove BsaI bottom with the Q5 SDM kit (New England Biolabs #E0554S) to generate ErCas12a Strataclone no BsaI. The nErCas12an fragment lacking the BsaI site was removed from the ErCas12a Strataclone no BsaI backbone by digesting with AgeI and BamHI. Likewise, the pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA; U6::BsaI-sgRNA plasmid (Addgene #61591) was digested with AgeI and BamHI and had nErCas12an inserted into it to generate p601nErCas12an no BsaI. The crRNA scaffold was inserted into the p601nErCas12an no BsaI by digesting with BsaI and NotI and inserted the annealed oligos ErCas12a EZ clone scaffold top and ErCas12a EZ clone scaffold bottom to generate ErCas12a EZ Clone.

Cloning in vitro *ErCas12a construct targeting CCR5 and TRAC*

Cr-RNA oligos corresponding to the genomic target site (e.g., *CCR5* sgRNA 1 Top + *CCR5* sgRNA 1 Bottom) were annealed in a thermocycler according to the Zhang lab protocol. Briefly, oligos were annealed by incubating at 37°C for 30 min followed by 95°C for 5 min and then ramped down to 25°C at 5°C/min. The annealed oligo duplex was cloned into ErCas12a EZ clone Digested with BsaI and ligated by T4 DNA ligase.

pMiniCAAGs:RFP-DR48 RFP assay generation, transfection, and analysis

pPRISM-V3-bactin SSA-DR48 was PCR amplified with Platinum Taq DNA polymerase HiFi (Thermo Fisher Scientific #11304011) with Broken RFP Transfer to Tol2 XhoI Fw and Broken RFP Transfer to Tol2 BglII Rev to add XhoI and BglII restriction sites to the RFP cassette. pkTol2C-EGFP as well as the RFP amplicon were digested with XhoI and BglII to create compatible cohesive ends and ligated with T4 DNA ligase to generate pkTol2CBrokenRFP. In order to the construct to express episomally in cell culture systems, a Kozak sequence was added by digesting pkTol2CBrokenRFP with EcoRI and XhoI and restriction cloning the annealed oligos Kozak Seq Top and Kozak Seq Bottom to generate pMiniCAAGs:RFP-DR48.

ErCas12a constructs targeting the UgRNA were generated by digesting ErCas12a EZ clone with BsaI and cloning in the annealed oligos ErCas12a UgRNA BsaI Top and ErCas12a UgRNA BsaI bottom to generate pErCas12a-U-pre-crRNA. Cas9 constructs targeting the UgRNA were generated by digesting lentiCRISPR v2 (Addgene #52961) with BsmBI and cloning in the annealed oligos Cas9 UgRNA BsmBI Top and Cas9 UgRNA BsmBI Bottom to generate pCas9-UgRNA. To control for promoter expression between ErCas12a and Cas9, the CMV promoter was added to pCas9-UgRNA by digesting ErCas12a EZ clone with XbaI and AgeI to remove the CMV promoter and pCas9-UgRNA with NheI and AgeI to replace the native EF1 alpha promoter.

HEK 293T cells were transfected with 5 μg pMiniCAAGs:RFP-DR48 and 5 μg of either pErCas12a-U-pre-crRNA or pCas9-UgRNA with the Etta H1 electroporator as described above. Cells were assessed for RFP expression 48 h later by flow cytometry with 584 nm emission and 607 nm detection.

Transfection

HEK293T cells used for indel acquisition assays were transfected with Lipofectamine 3000 (Thermo Fisher Scientific #L3000008) according to manufacturer's protocol with 5 μg of ErCas12a plasmid targeting each site. Cells used for targeted integration assays were transfected with Etta H1 electroporator with the following parameters: 200 V, 784 ms interval, six pulses, 1,000 μs pulse duration, at a concentration of 20E6 cells/mL at the volume of 100 μL in Etta EB electroporation buffer. Cells are recovered post electroporation by incubating at 37°C for 5–10 min before being plated on a six-well tissue culture plate at a density of about 1.5E6 cells/mL.

GeneWiz AmpliconEZ: DNA library preparation and Illumina sequencing

DNA library preparations, sequencing reactions, and initial bioinformatics analysis were conducted at GENEWIZ, Inc. DNA library preparation, clustering, and sequencing reagents were used throughout the process using the NEBNext Ultra DNA Library Prep Kit, following the manufacturer's recommendations (Illumina). End-repaired adapters were ligated after adenylation of the 3′ ends followed by enrichment by limited cycle PCR. DNA libraries were validated on the Agilent TapeStation, quantified using the Qubit 2.0 Fluorometer (Invitrogen), and multiplexed in equal molar mass. The pooled DNA libraries were loaded on the Illumina instrument according to the manufacturer's instructions. The samples were sequenced using a 2×250 paired-end (PE) configuration. Image analysis and base calling were conducted by the Illumina Control Software on the Illumina instrument.

Data analysis

The raw Illumina reads were checked for adapters and quality via FastQC. The raw Illumina sequence reads were trimmed of their adapters and nucleotides with poor quality using Trimmomatic v0.36. Paired sequence reads were then merged to form a single sequence if the forward and reverse reads were able to overlap. The merged reads were aligned to the reference sequence, and variant detection was performed using the GENEWIZ proprietary Amplicon-EZ program.

Phylogeny and homology analysis

The amino acid sequence (ErCas12a) was used for homology search by the Blastp tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) against the non-redundant (nr) database. The homologous sequences were then retrieved from the NCBI database. Multiple sequence alignment was done using the CLUSTAL Omega webserver (www.ebi.ac.uk/Tools/msa/clustalo/) The phylogenetic tree was constructed using the maximum likelihood method based on JTT matrix-based model in MEGA v7.0.18.[23] Evaluation of branching was ensured by bootstrap statistical analysis (1,000 replications).

Phylogenetic analysis

The phylogenetic tree relationship of ErCas12a from *E. rectale* and other related Cas proteins available in the NCBI database was constructed using the maximum likelihood method with MEGA v7. The numbers above and below the branch points specify the confidence levels meant for the relationship of the paired sequences as estimated by the bootstrap analysis. Branch lengths are measured as the number of substitutions per site. Cas protein sequences used for phylogenetic analysis are CRISPR-associated endonuclease Cas12b from *Alicyclobacillus acidoterrestris*, CRISPR-associated Endonuclease Cas9 from *Streptococcus pyogens*, CRISPR-associated Endonuclease Cpf1 from *Acidaminococcus* sp. BV3L6, type V CRISPR-associated protein Cpf1 from *Lachnospiraceae* bacterium ND2006, and type V CRISPR-associated protein Cpf1 from *Francisella tularensis.*

Cas variant homology alignments

Multiple sequence alignment of ErCa12a with homologous Cas proteins such as CRISPR-associated endonuclease Cas12b from *A. acidoterrestris*, CRISPR-associated Endonuclease Cas9 from *S. pyogens*, CRISPR-associated

Endonuclease Cpf1 from *Acidaminococcus* sp. BV3L6, type V CRISPR-associated protein Cpf1 from *Lachnospiraceae* bacterium ND2006, and type V CRISPR-associated protein Cpf1 from *F. tularensis*. Six representative sequences were used for alignments with Clustal Omega webserver.

Determining nuclease genome coverage

Sites were discovered using the *fuzznuc* nucleotide pattern sequence analyzer from EMBOSS v6.5.7.[24] Human sites were determined from whole-genome assembly GRch38 downloaded from Ensembl release 97 (ftp://ftp.ensembl.org/pub/release-97/fasta/homo_sapiens/dna/Homo_sapiens.GRCh38.dna.primary_assembly.fa.gz). Zebrafish sites were determined from whole-genome assembly GRCz11 downloaded from Ensembl release 97 (ftp://ftp.ensembl.org/pub/release-97/fasta/danio_rerio/dna/Danio_rerio.GRCz11.dna.primary_assembly.fa.gz). Matching hits were tabulated using *egrep* for each pattern set.

Results

ErCas12a induces indels in zebrafish embryos

We used informatic analyses to categorize a new potential gene editor (GenBank Accession no. MH347339.1) that aligns as a novel Cas12a protein: ErCas12a (Fig. 1A and Supplementary Fig. S1). To determine the utility of ErCas12a targeting in the human and zebrafish genomes, we enumerated the amount of ErCas12a's YTTN PAM sites compared to Cas9's NGG PAM sites. We found that across the zebrafish genome, ErCas12a has 168,074,842 potential YTTN PAM targets, and Cas9 has 95,376,529 NGG PAMs. In humans, ErCas12a has 345,066,663 YTTN PAMs as opposed to Cas9's 304,420,059 NGG PAMs. This expanded coverage for ErCas12a increases the genomic accessibility in AT-rich genomes, such as zebrafish, otherwise inaccessible to Cas9.

To assess the activity of ErCas12a, we first synthesized a codon-optimized version for expression in eukaryotes and added dual SV40 nuclear localization signals (NLS) at the N- and C-termini, as dual NLS was shown to increase efficacy of Cas9 in zebrafish.[3] To test ErCas12a gene-editing activity *in vivo*, ErCas12a mRNA was co-injected with a pre-crRNA into single-cell zebrafish embryos followed immediately by heat shock at 34°C for 4 h (Fig. 1B). Three pre-crRNAs and one crRNA were designed to target exon 1 of the zebrafish *noto* gene (Fig. 1C). Co-injection of ErCas12a mRNA with *noto*-pre-crRNA1 and incubation at 28°C did not result in gene-editing activity, similar to previous reports using Cas12a (Moreno-Mateos *et al.*[8] and data not shown). However, injection of ErCas12a mRNA with *noto*-pre-crRNA1 or *noto*-pre-crRNA3 followed by a 4 h 34°C heat shock resulted in phenotypes characteristic of biallelic loss of *noto*, including loss of the notochord and a shortened tail,[25] at efficiencies ranging from 7% to 61% (Supplementary Fig. S2A). We noted no difference in nuclease activity between two concentrations of pre-crRNA and chose the lower dose moving forward to decrease the potential for RNA toxicity. Polymerase chain reaction (PCR) across these individual target sites and subsequent heteroduplex mobility shift assays revealed the presence of indels characteristic of repair by non-homologous end joining (NHEJ) activity (Supplementary Fig. S2B), which we then confirmed and quantified by next-generation sequencing (NGS), as described below. *noto*-pre-crRNA2 and *noto*-crRNA1 did not elicit a phenotype, and heteroduplex mobility shift assays across those targets indicate they are inactive sgRNAs (data not shown). Efficiencies of biallelic *noto* inactivation at the same target varied considerably from injection to injection (Supplementary Fig. S2), consistent with previous reports in human cells and zebrafish using CRISPR-Cas12a[8,20,26,27] To confirm that our results were not specific to a single locus, we injected ErCas12a mRNA and pre-crRNA to target two sites in the *cx43.4* gene (Fig. 1D). PCR across the individual target sites and subsequent restriction fragment length polymorphism (RFLP) analysis demonstrated that *cx43.4*-pre-crRNA2 was active (Supplementary Fig. S2D), while *cx43.4*-pre-crRNA1 was inactive (data not shown).

To gain a better understanding of the frequency of ErCas12a insertion/deletion (indel) events, we used Illumina NGS to analyze the efficiencies of mutation at the active *noto* and *cx43.4* target sites. DNA amplicons from five *noto* mutant embryos injected with ErCas12a mRNA and *noto*-pre-crRNA1 or *noto*-pre-crRNA3 and five *cx43.4* pre-crRNA1 injected embryos were selected randomly for NGS. As expected for biallelic inactivation at *noto*, ~61% of alleles at *noto*-pre-crRNA1 showed indels, characteristic of NHEJ and/or MMEJ after nuclease targeting (Fig. 1E). The more efficient *noto*-pre-crRNA3 showed ~90% of alleles containing indels (Fig. 1E). However, in agreement with the RFLP analysis, the majority (~74%) of alleles sequenced at *cx43.4* had the wild-type sequence, indicating that not all targeting events with ErCas12a are efficient enough to produce biallelic mutation of the target locus (Fig. 1E). Taken together, these data indicate CRISPR-ErCas12a introduced as an RNA system can be active in zebrafish with a 34°C heat shock, and that pre-crRNA can serve as an active RNA guide for directing ErCas12a activity to genomic target sites.

ErCas12a elicits SAMR more efficiently than SpCas9 in a genomic reporter system in zebrafish

We next employed a stably integrated red fluorescent protein (RFP) reporter system to compare visually the efficiencies with which CRISPR-Cas9 and CRISPR-ErCas12a differentially elicit DNA repair using SAMR. We used GeneWeld to create a transgenic line of zebrafish that contained a single copy of an SAMR reporter, *noto:*RFP-DR48 (Fig. 2A and Supplementary Fig. S3). The RFP coding sequence is disrupted by a target cassette engineered with Cas9 and ErCas12a sites that is flanked by 48 bp of direct repeat sequence. A DSB and end resection in the target liberates the 48 nucleotide direct repeats, which can restore the RFP reading frame after SAMR, leading to a semi-quantitative read-out of repair efficiencies. Injection of single-cell *noto:*RFP-DR48 transgenic embryos with Cas9 mRNA and universal guide RNA (UgRNA) results in RFP expression in the notochord, indicative of SAMR after CRISPR-Cas9 targeting (Fig. 2B and Supplementary Table S4).

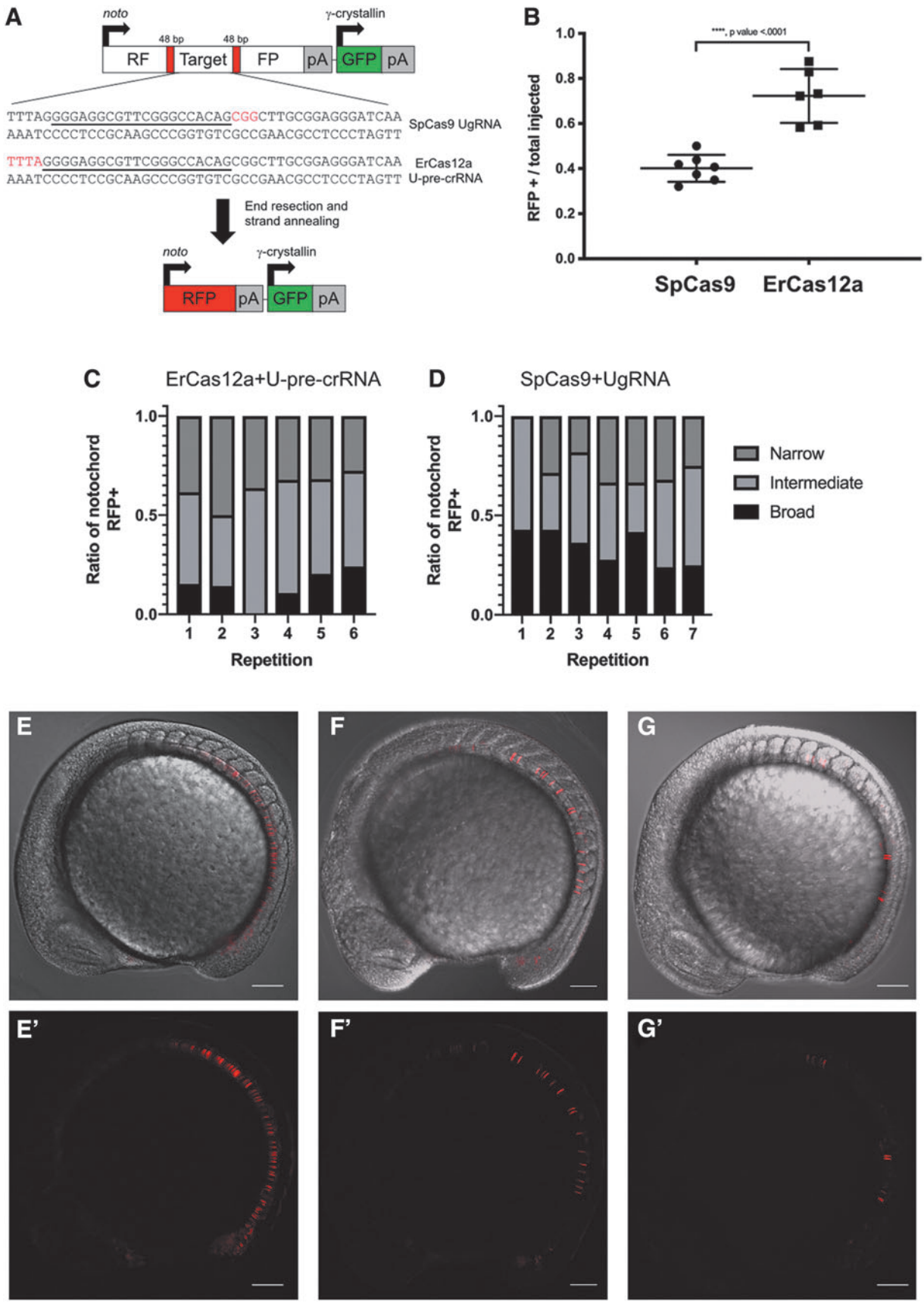
We then tested whether CRISPR-ErCas12a promotes SAMR in the *noto:*RFP-DR48 assay. We designed a U-pre-crRNA, with no predicted off-target sites in zebrafish or human cells, to direct ErCas12a activity to *noto:*RFP-DR48. Injection of ErCas12a mRNA and U-pre-crRNA resulted in RFP expression in the notochord (Fig. 2B and C). Based on the percentage of injected animals with RFP+ notochords, ErCas12a elicited SAMR at a higher frequency than SpCas9 (Fig. 2B–D). Injected embryos were highly mosaic and qualitatively sorted into three classes of notochord RFP expression pattern: broad, intermediate, and narrow (Fig. 2E–G). While ~70% of ErCas12a-injected animals showed RFP+ cells in the notochord as opposed to ~40% in SpCas9-injected embryos, ErCas12a repair events are equally mosaic (Figure 2B–D). These results indicate that *noto:*RFP-DR48 is a viable assay for screening the propensity of designer nucleases to elicit SAMR and that in this context ErCas12a preferentially activates SAMR at rates higher than Cas9.

Using ErCas12a for precise integrations in zebrafish

We leveraged the activity of the U-pre-crRNA to determine whether CRISPR-ErCas12a can catalyze targeted integration of fluorescent reporters using GeneWeld.[19] First, we injected nErCas12an, *noto*-pre-crRNA1, U-pre-crRNA, and a green fluorescent protein (GFP) reporter plasmid programmed with the U-pre-crRNA target site and 24 bp of homology 5′ to the noto-crRNA1 genomic site to promote gene targeting at *noto* (Supplementary Fig. S4A and Supplementary Table S5). Fluorescent positive notochord cells were observed at the 18-somite stage, indicating in-frame integration of the GFP cassette at *noto* (Supplementary Fig. S4B and B′). On average, 24% of embryos injected display targeted *noto* integrations (Supplementary Fig. S4C). However, notochords were highly mosaic, indicating a low efficiency of somatic integration activity in tissue types that form the notochord (Supplementary Fig. S4B and B′). GFP was often observed outside of the notochord yet in the mesoderm, as expected with biallelic disruption of *noto* and transfating of notochord cells.[21] Junction fragment PCR was conducted on GFP+ embryos to confirm that integration was directed by the programmed homology. The expected PCR band was recovered in GFP+ embryos, while no PCR band was detected in the control experiment performed without the genomic pre-crRNA (Supplementary Fig. S4D and E). DNA sequencing confirmed precise junctions at the 5′ end of the integration (Supplementary Fig. S4F).

We next designed a GeneWeld donor with both 5′ and 3′ homology domains flanking the GFP cassette to promote precise repair at both sides of the integration site (Fig. 3A).[19] Targeting *noto* using pre-crRNA3 with GeneWeld resulted in an average of 31% of embryos with GFP+ notochords (Fig. 3B, B′, and C and Supplementary Table S5). Most notochords were highly mosaic and displayed GFP expression indicative of lost cell fate, as in *noto* target 1 targeting (Supplementary Fig. S4B), although some events were recovered in which >90% of the notochord expressed GFP (Fig. 3B). Predicted 5′

FIG. 2. Using *noto*:RFP-DR48 to assay the propensity of ErCas12a and SpCas9 to elicit strand annealing in zebrafish. **(A)** Schematic of *noto*:RFP-DR48 showing the location of the 48 bp direct repeats flanking both the SpCas9 and ErCas12a universal RNA cursor sites (underline) and protospacer adjacent motif (PAM) sites (red text). **(B)** Data plot showing the ratio of injected animals displaying red fluorescent protein (RFP) in the notochord out of the total carrying the transgene. Data plot represents the mean ± *SD*. *p*-Values calculated with one-tailed Student's *t*-test. **(C)** Qualitatively scored ratios of notochord converted to RFP+ after SpCas9 induced strand-annealing mediated repair (SAMR). **(D)** Qualitatively scored ratios of notochord converted to RFP+ after ErCas12a induced SAMR. **(E–G** and **E′-G′)** Representative embryos for broad, intermediate, and narrow conversion of RFP in the notochord.


A
noto
γ-crystallin
48 bp
48 bp
RF
Target
FP
pA
GFP
pA
TTTAGGGGAGGCGTTCGGGCCACAGCGGCTTGCGGAGGGATCAA
AAATCCCCTCCGCAAGCCCGGTGTCGCCGAACGCCTCCCTAGTT
SpCas9 UgRNA
TTTAGGGGAGGCGTTCGGGCCACAGCGGCTTGCGGAGGGATCAA
AAATCCCCTCCGCAAGCCCGGTGTCGCCGAACGCCTCCCTAGTT
ErCas12a
U-pre-crRNA
End resection and
strand annealing
noto
γ-crystallin
RFP
pA
GFP
pA
B
****, p value <.0001
RFP + / total injected
1.0
0.8
0.6
0.4
0.2
0.0
SpCas9
ErCas12a
C
ErCas12a+U-pre-crRNA
Ratio of notochord RFP+
1.0
0.5
0.0
1 2 3 4 5 6
Repetition
D
SpCas9+UgRNA
Ratio of notochord RFP+
1.0
0.5
0.0
1 2 3 4 5 6 7
Repetition
Narrow
Intermediate
Broad
E
F
G
E'
F'
G'

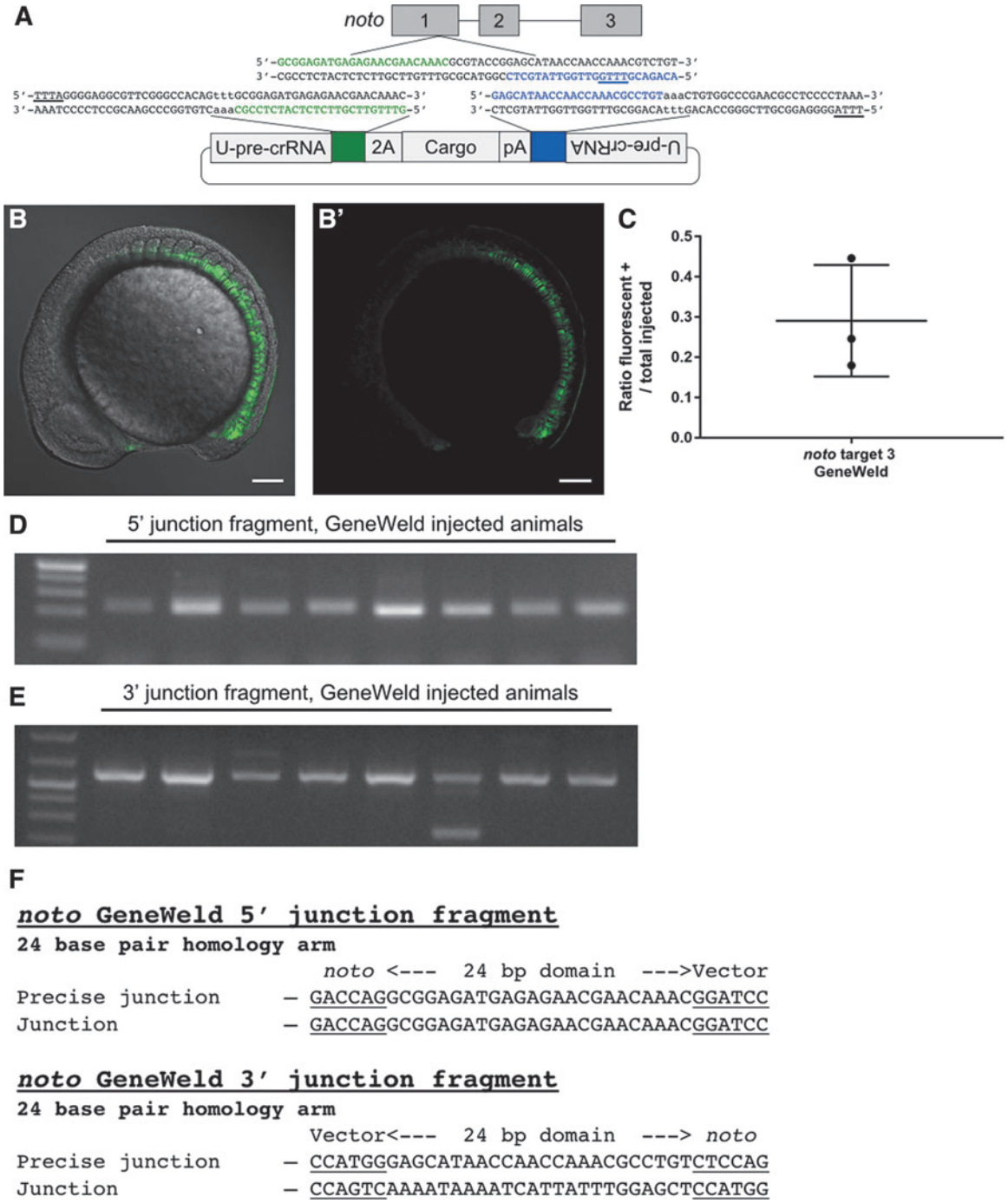


FIG. 3. Targeting *noto* with GeneWeld. **(A)** Schematic of *noto* showing designed homology for precise integration using ErCas12a and the U-pre-crRNA. Green is designed 5′ homology. Blue is designed 3′ homology. The PAM for ErCas12a targeting in the genome and donor is underlined. **(B** and **B′)** Confocal Z-stack image showing broad green fluorescent protein (GFP) expression in the embryo. Scale bar is 100 μm. **(C)** Data plot showing the ratio of embryos with GFP expression in the notochord out of total injected embryos. Data plot represents the mean $\pm$ *SD*. **(D)** Gel showing 5′ junction fragment expected after precise integration using GeneWeld. **(E)** Gel showing 3′ junction fragment expected after precise integration using GeneWeld. **(F)** DNA sequencing of lane 5 in **(D)** and **(E)** showing a precise integration using the programmed homology.

and 3′ junction fragments were recovered by PCR in GFP+ embryos (Fig. 3D and E), and junction fragment sequencing from a single embryo demonstrated precise integration at both ends of the cassette (Fig. 3F). These data indicate that ErCas12a can be an effective nuclease for catalyzing integration in zebrafish using GeneWeld. However, we note both qualitative as well as quantitative differences. For example, the requirement for the heat-shock step imposes a potential delay on ErCas12a activity compared to SpCas9, and this kinetic difference can explain the higher mosaicism for Ercas12a compared to SpCas9 in this integration method.

ErCas12a induces DSB in human cells

We developed an all-in-one vector for expression of the dual nuclear localization signal ErCas12a and a pre-crRNA *in vitro* (Fig. 4A). We targeted two therapeutically relevant safe harbor loci: *AAVS1* and *CCR5*[28] (Fig. 4B and D). We additionally targeted the *TRAC* locus due to its reported significance in generating chimeric antigen receptor T cells (Fig. 4E).[29] *In vitro* cleavage activity of ErCas12a with our expression system was assessed using a T7 endonuclease (T7EI) assay to determine whether a DSB and subsequent indels were induced at the *AAVS1* locus after transfection into HEK293T cells. The expected banding pattern was apparent in DNA amplified from pErCas12a-AAVS1-pre-crRNA1 transfected cells but not control cell DNA, indicating that ErCas12a was active at the AAVS1 target site (Fig. 4C). No indel activity was detected at the *AAVS1* pre-crRNA2 site by T7E1 assay (data not shown). These data suggest that the all-in-one expression system is functional *in vitro.*

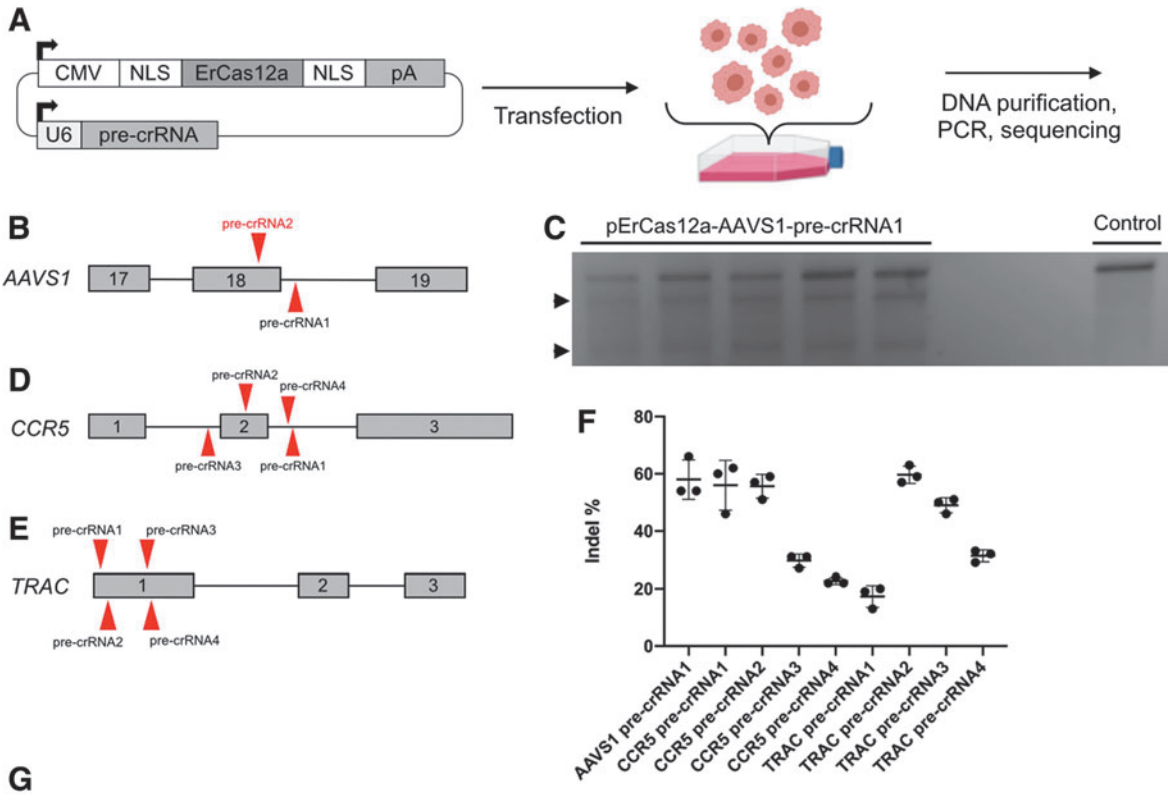
To gain a quantitative understanding of the cleavage activity of ErCas12a at the *AAVS1*, *CCR5*, and *TRAC* target sites, PCR amplicons were submitted for Sanger sequencing and analyzed using ICE software, which infers CRISPR activity from sequencing trace reads.[30] ICE analysis showed targeting *AAVS1* with ErCas12a, and pre-crRNA1 created indels in 58% of sequenced amplicons (Fig. 4F and Supplementary Table S6). Targeting *CCR5* with ErCas12a at four sites resulted in a range of mutagenesis efficiency from 20% to 60% (Fig. 4F and Supplementary Table S6). Similarly, we tested four pre-crRNAs targeting *TRAC* and found the most efficient resulted in 60% of sequenced amplicons contained indels (Fig. 4F and Supplementary Table S6). Because ICE is based on Sanger sequencing data, which considers fewer reads than NGS, it is likely to underestimate the true percentage of indels. We therefore measured indel activity at the *AAVS1*-pre-crRNA1 target site using Illumina sequencing, which showed that ~74% of recovered alleles were edited (Fig. 4G). Taken together, these data demonstrate that ErCas12a is active in human cells at clinically relevant genomic loci.

ErCas12a elicits SAMR more efficiently than SpCas9 in an episomal reporter system in human cells

As ErCas12a was shown to induce SAMR in a stable transgenic in zebrafish at least as efficiently as Cas9, we next tested whether the same was also true in human cells with an episomal analog of the zebrafish experiment. Episomal reporters have been used routinely in cell culture systems to assay DNA repair pathways and identify the proteins involved.[31–34] We modified the *noto*:RFP-DR48 cassette for use as a reporter in mammalian cell culture, creating pMiniCAAGs:RFP-DR48 (Fig. 5A). As observed in zebrafish, transfection of pMiniCAAGs:RFP-DR48 with a nuclease and requisite sgRNA into human cells is expected to promote a DSB at the nuclease target site, resulting in a fluorescent readout of nuclease activity if SAMR is used to repair the DSB. Transfection of HEK293T cells with pMiniCAAGs:RFP-DR48 and pErCas12a-U-pre-crRNA or pCas9-UgRNA resulted in RFP+ cells, demonstrating efficient SAMR with both ErCas12a and Cas9 nucleases (Fig. 5A). Quantification of RFP+ cells by flow cytometry indicated SAMR induction was significantly greater ($p < 0.0044$) with ErCas12a compared to SpCas9 (Fig. 5B and Supplementary Fig. S5). Together, these results show ErCas12a nuclease has similar cleavage activity in human cells and zebrafish, and is more efficient than Cas9 in mediating SAMR *in vivo* and *in vitro*.

ErCas12a promotes precise integration in human cells

Because of *AAVS1*'s utility in therapeutic expression of transgenes, we next wanted to determine whether ErCas12a could mediate precise transgene insertion at the pre-crRNA1 site in *AAVS1*.[35–37] To test this, we used the GeneWeld strategy to mediate integration of a pCMV:GFP::Zeo reporter containing 48 bp 5′ and 3′ arms homologous to the *AAVS1* pre-crRNA1 site (Fig. 6A and B). HEK293T cells were transfected with the pCMV:GFP:Zeo donor plasmid and the all-in-one pErCas12a-AAVS1-U plasmid that expresses ErCas12a, *AAVS1* pre-crRNA1, and U-pre-crRNA (pErCas12a-AAVS1-U). Transfections were also done with the pCMV:GFP:Zeo-48 donor plasmid and pErCas12a-AAVS1 pre-crRNA1 without the U-pre-crRNA, or pCMV:GFP:Zeo-48 donor plasmid alone. GFP+ cells were isolated by fluorescence-activated cell sorting



G

Target	Sequence	Indel Length	Microhomology Signature	% of Total Read Count	Total mutant
AAVS1 target 1					
AAVS1 target 1 WT	TTTAGGACGGTGCGATCCCACCCCTCCGTCGGGGCTTTGCCTGCTGGGTGCCC	N/A	N/A	25.28	74.72%
Allele 1:	TTTAGGACGGTGCGAT-----CCCTCCGTCGGGGCTTTGCCTGCTGGGTGCCC	-5	CCC	13.26	
Allele 2:	TTTAGGACGGTGCGAT----CCCCTCCGTCGGGGCTTTGCCTGCTGGGTGCCC	-4	N/A	8.05	
Allele 3:	TTTAGGA----------CGGGGCTTTGCCTGCTGGGTGCCC	-22	CGG	1.99	
Allele 4:	TTTAGGACGGTGCGATC--------CCGTCGGGGCTTTGCCTGCTGGGTGCCC	-8	CC	1.83	
Allele 5:	TTTAGGACGGTGCGAT------CCTCCGTCGGGGCTTTGCCTGCTGGGTGCCC	-6	CC	1.51	

FIG. 4. ErCas12a activity in human HEK293T cells. **(A)** Schematic showing the workflow for ErCas12a and pre-crRNA expression in human cells from a *cis* expression plasmid. After transfection, cells are recovered and then analyzed. **(B)** Schematic of two pre-crRNAs designed to target *AAVS1*. **(C)** T7E1 assay showing *AAVS1*-pre-crRNA1 is active. Arrowheads are cleaved bands, indicating nuclease activity. **(D)** Schematic of four pre-crRNAs designed to target *CCR5*. **(E)** Schematic of four pre-crRNAs designed to target *TRAC*. **(F)** Graph showing percentage of indels after using pre-crRNAs to target *AAVS1*, *CCR5*, and *TRAC* in HEK293T cells as determined by ICE analysis. Data plot represents the mean ± *SD*. **(G)** Results displaying the wild-type and top five mutated alleles after AmpliconEZ analysis of *AAVS1* targeted DNA.

(FACS) 1 day post electroporation to control for plasmid delivery, and then screened for stable integration by flow cytometry 2 weeks post transfection to measure dropout of GFP+ cells (Fig. 6C and Supplementary Fig. S6). The data revealed a significant 2.5-fold increase ($p < 0.0003$) in the number of cells that remained GFP+ when transfected with the donor molecule, ErCas12a and the pre-crRNA, compared to the donor alone control (Fig. 6D and Supplementary Fig. S6B–C″). GFP+ cell populations were screened for targeted integration at the *AAVS1* pre-crRNA1 site by junction fragment PCR and sequencing. The 5′ junction analysis showed precise repair at the integration site in the cell population transfected with pErCas12a-AAVS1-U and the donor (Fig. 6E and F) but was absent in the donor alone population. These data suggest ErCas12a is a tractable

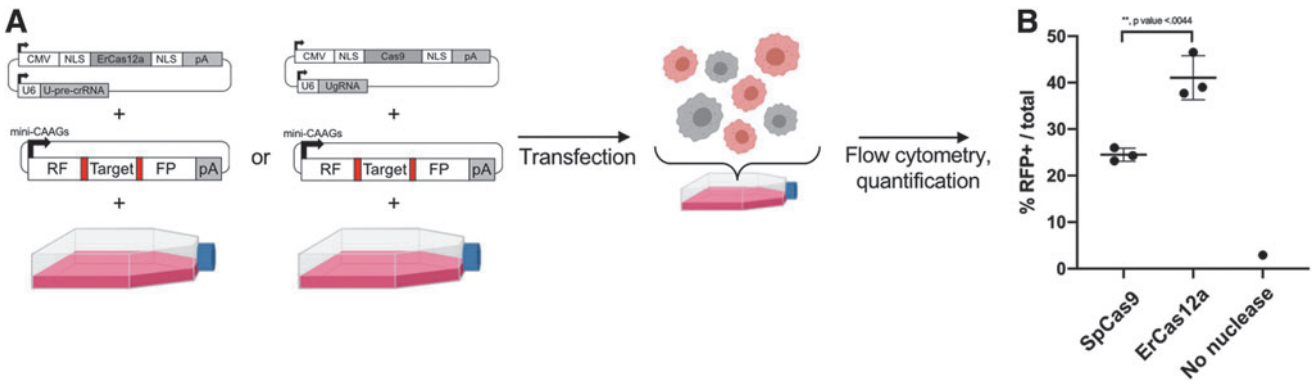


FIG. 5. Using pMini-CAAGs::RFP-DR48 to assay the propensity of ErCas12a and SpCas9 to elicit strand annealing in HEK293T cells. **(A)** Schematic showing the workflow for determining SAMR in human cells. An all-in-one expression plasmid for ErCas12a or SpCas9 is transfected along with pMini-CAAGs::RFP-DR48. Cells are allowed to recover and are then isolated with fluorescence-activated cell sorting for RFP+ cells. **(B)** Quantification of RFP+ cells per total cells analyzed by flow cytometry. Data plot represents the mean ± *SD*. *p*-Values calculated with one-tailed Student's *t*-test.

nuclease for GeneWeld-mediated precision integration at the *AAVS1* safe harbor locus in human cells.

Discussion

In this report, we establish that CRISPR-ErCas12a is an active nuclease in zebrafish and explore its use in human cells, demonstrating its ability to serve as an efficient alternative to CRISPR-Cas9 for generating HMEJ-mediated gene targeting events *in vitro* and *in vivo*. Targeting in zebrafish embryos, we created frame shift indel alleles in the *noto* and *cx43.4* at up to ~90% and ~24%, respectively. ErCas12a activity is temperature-dependent in zebrafish and requires a 34°C heat shock for activity. In a genomic reporter assay, ErCas12a elicits SAMR in a reporter assay at levels nearly twofold greater than Cas9 in both zebrafish and human cells. ErCas12a is also a viable nuclease for mediating gene targeting using the GeneWeld strategy in zebrafish and human cells. We detected robust DSB induction at three therapeutically relevant human loci and demonstrated that ErCas12a is effective for directing GeneWeld integrations at the *AAVS1* locus in human cells. These results suggest that ErCas12a could be employed as a tractable nuclease for gene editing across multiple species for both basic research and clinical applications.

In prior reports of CRISPR-Cas12a activity in zebrafish, appreciable nuclease activity with a sgRNA was only observed after RNP delivery and heat shock.[8,20] While still requiring a heat shock but in contrast to previously described Cas12a systems, ErCas12a functions with a sgRNA when delivered as mRNA. Cas12a activity was shown to be dependent on the stability of the bound pre-crRNA or crRNA, and although pre-crRNA was ineffective for Cas12a nuclease activity under standard conditions, longer heat shocks rescued activity, indicating a likely kinetics issue with the use of pre-crRNA. In our experiments, pre-crRNA injection with ErCas12a mRNA permitted nuclease activity, highlighting a difference between other Cas12a related proteins and ErCas12a, either in their ability to bind the RNA guide stably and access DNA or in the stability of their respective pre-crRNA structures. Because the use of multiple crRNAs per target gene in tandem dramatically increases the possibility of off-target effects, it is notable that in this study, ErCas12a displays efficient nuclease activity with only a single pre-crRNA per target gene.

While ErCas12a showed enhanced activation of SAMR in comparison with Cas9 in our *noto*:RFP-DR48 assay, targeted integration via GeneWeld using ErCas12a in zebrafish displayed no substantive difference compared to our previous report using Cas9.[19] The increased SAMR frequency detected using the reporter assay may be because repair of DNA *in cis* is more efficient than *in trans*, and thus the outcome of strand annealing in a genomic reporter will not translate to targeted integrations.

Reported levels of mosaicism in gene targeting using HMEJ in zebrafish vary greatly in the literature.[15,19] Although ErCas12a is an effective catalyst for GeneWeld integrations, mosaicism of expression is qualitatively higher than when using Cas9 as the GeneWeld nuclease.[19] Consistent with the observation that SAMR in *noto*:RFP-DR48 is mosaic, it was reported recently that translation of Cas9 mRNA and subsequent gene editing

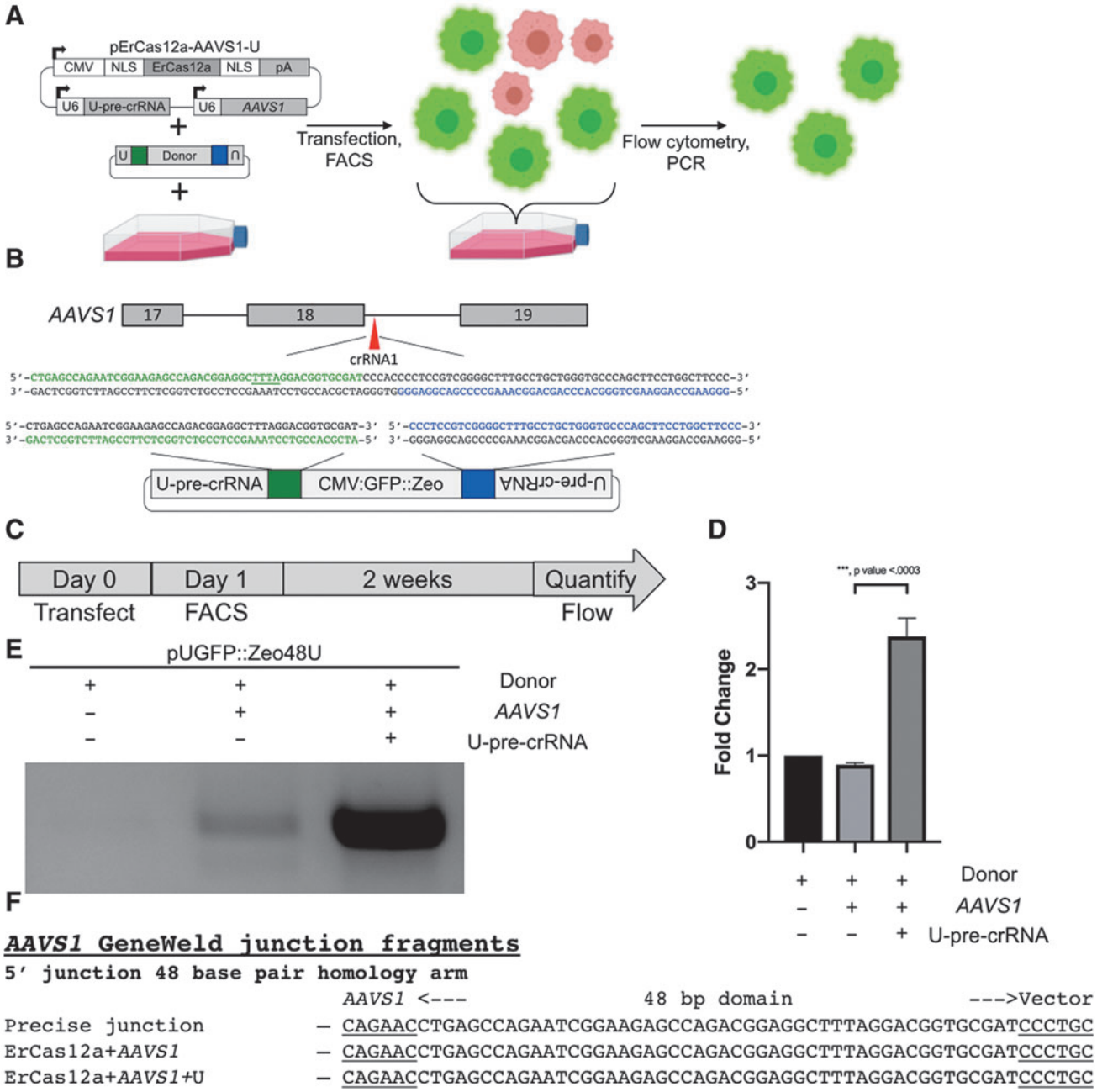


FIG. 6. Using ErCas12a for targeted integration in human cells. **(A)** Schematic showing the workflow for targeted integration at *AAVS1* in HEK293T cells. **(B)** Schematic of *AAVS1* showing designed homology for precise integration using ErCas12a and the U-pre-crRNA to liberate the cassette. Green block, 5′ homology; blue block, 3′ homology. The PAM for ErCas12a targeting in the genome and donor is underlined. **(C)** Schematic for flow analysis after transfection of GeneWeld reagents for targeting *AAVS1*. **(D)** Bar graph showing the ratio of cells with GFP expression as determined by flow cytometry. Graph represents the mean ± *SD*. *p*-Values calculated with one-tailed Student's *t*-test. **(E)** Junction fragment gel showing the expected 5′ integration amplicon. No amplicon is seen when transfecting donor alone. **(F)** DNA sequencing of lanes in **(E)** showing a precise integration using the programmed homology.

after one-cell stage injection is not complete until the 16- or 32-cell stage, while RNP injections result in appreciable nuclease activity by the two- to four-cell stage.[38] Thus, *noto*:RFP-DR48 is an assay system where injection conditions can be optimized to enhance somatic gene targeting and decrease mosaicism in cell types that arise from the mesoderm. Also, *noto*:RFP-DR48 represents a novel reporter in zebrafish for determining strategies for altering DSB repair pathway choice, such as using small molecules or dominant-negative proteins.

Previously reported Cas12a orthologs have displayed relatively poor editing efficiency in mammalian systems.[26,27,39,40] AsCas12a and LbCas12a are among the best characterized Cas12a systems but have very modest activity in human cells.[26,40] To address this issue, additional Cas12a systems with enhanced activity and differential targeting ability such as FnCas12a were engineered and characterized. Although FnCas12a uses a more common TTN PAM sequence, its cleavage activity is still modest.[39] Editing efficiency of the Cas12a systems has since been enhanced by engineering the crRNA to increase activity without a loss of specificity.[41] Additionally, recent work has identified particular point mutations in Cas12a variants that enhance the targeting range, activity, and fidelity of the nuclease.[42] This suggests the possibility of a future ErCas12a variant in which cleavage kinetics could be leveraged with broadened PAM targeting to target regions typically restricted to SpCas9. Going forward, it will be interesting to evaluate the effects of various modifications of crRNAs and the protein itself to optimize mutagenesis and gene targeting.

The five-base thymidine stretch in the secondary structure of the ErCas12a crRNA could cause cryptic termination of guide expression[43] and lead to decreased expression of the guide RNA, potentially decreasing nuclease activity. Testing crRNA designs that remove this thymidine stretch may increase crRNA expression and nuclease efficiency. It has been shown previously that a single U6 promoter is sufficient to express multiple crRNAs with Cas12a for multiplex gene editing due to the intrinsic self-processing activity of the nuclease.[2] It is of interest for future investigations to determine if ErCas12a has similar properties in order to leverage it for multiplex gene editing.

Interestingly, the RuvC domain in Cas12a family members is sufficient to elicit indiscriminate ssDNA and ssRNA cleavage.[44] Though ErCas12a shares only 31% sequence homology with Cas12a orthologue AsCas12a (Supplementary Fig. S1) and the RuvC domain is conserved among Cas12a orthologues and ErCas12a, it is possible that long-range interactions at the catalytic pocket differ in ErCas12a due to the relatively low amino acid similarity. Therefore, further studies are needed to elucidate if ErCas12a shares similar nonspecific nuclease activity.

It is notable that in this study, several of the guide RNAs selected had no activity. When designing guide RNAs for CRISPR systems, the researcher must consider two primary factors: on-target (efficiency) and off-target (specificity) activity of the guide. In this study, we visually interrogated the target loci for the 5′-YTTN PAM sequence due to the lack of precedent for ErCas12a guide RNA design. Previous studies with SpCas9 have shown that nuclease activity has variable activity across different target sites, with certain genomic target sites having inherently more activity,[2,3,45–50] and it is likely that similar limitations exist with ErCas12a. This and other work has established approximate guidelines for designing SpCas9 guide RNAs, including avoiding poly-T sequences, limiting GC content, and avoiding a G immediately upstream of the NGG PAM.[51–53] Based on these studies, several guide RNA design tools such as GuideScan[54] and sgRNA Scorer[55] were developed to assist in designing efficient and specific guide RNAs. With further experimentation, it is likely that design platforms will likewise be established for ErCas12a guide RNAs. Tools such as DeepCRISPR[56] are utilizing deep learning algorithms to determine ideal guide RNAs for Cas9 and could prove to be powerful assets in determining optimal ErCas12a crRNA design.

Though Cas12a systems have been shown to be amenable to HDR-mediated integration, these transgene integration events have been shown to be highly dependent on using RNP complexes in concert with modified crRNAs to show appreciable activity.[27] Here, we demonstrate ErCas12a's utility in generating robust editing efficiency without chemically modified crRNAs, showing its utility in streamlined and accessible gene editing for most research applications. Further, we show ErCas12a mediates precise integration without engineering of the nuclease or the corresponding guide RNA with mRNA and plasmid DNA.

While we show that ErCas12a is compatible with GeneWeld, we also demonstrate that in the absence of the UgRNA liberating the donor, there is still integration of the transgene at *AAVS1*. When GeneWeld is used with Cas9, the donor homology arms can include a portion of the genomic target but exclude the 3′ PAM sequence and prevent nuclease targeting at the homology arms. However, because ErCas12a utilizes a 5′ PAM and induces a distal DSB, it is likely that the homology arm is being targeted by ErCas12a and linearizing the donor at the 5′ end. In our design, the homology arms contain up to 17 bp of the target region that may be

sufficient for a Cas12a-like seed region for the crRNA and facilitate a DSB.[44,57] These data and previous studies of other Cas12a proteins suggest that ErCas12a may elicit measurable off-target activity. As such, future studies of ErCas12a should include in-depth analysis for off-target activities, as has been conducted for other CRISPR systems.[46,58] In addition to further optimization of both the crRNA and donor constructs, ErCas12a can be readily be adapted for precise integration in mammalian systems.

Conclusion

Alternative nuclease systems beyond canonical CRISPR-Cas9 are of interest in the precision therapeutics field as they expand the available toolbox for precision gene editing. Here, we demonstrate effective genome editing in both zebrafish and human cells using the newly described CRISPR-ErCas12a system, which increases the number of accessible regions in the genome due to its AT-rich PAM sequence. In zebrafish, CRISPR-ErCas12a promotes efficient somatic mutation and HMEJ-mediated precise integration of donor cassettes. In human cells, CRISPR-ErCas12a efficiently targeted therapeutically relevant loci, including the safe harbor loci *AAVS1* and *CCR5*, and facilitated transgene integration. Our data suggest that, with further optimization, ErCas12a may be an invaluable tool for gene editing in both basic research and future clinical gene therapy applications.

Acknowledgments

The authors would like to acknowledge Inscripta for providing their *E. rectale* ErCas12a (Mad7) sequence as an open source nuclease for the gene-editing community. pX601-AAV-CMV::NLS-SaCas9-NLS-3xHA-bGHpA;U6::BsaI-sgRNA was a gift from Dr. Feng Zhang. We thank Dr. Hirotaka Ata for experimental design discussions, Dr. Karthik Murugan for providing insight into *in vitro* knock in experimental results, and Gabriel Martinez Galvez in assisting in analyzing NGS data. We would further like to thank the Mayo Clinic Microscopy and Cell Analysis Core for conducting flow cytometry and FACS. This work was supported by NIH grants OD020166 and GM63904 and the Mayo Foundation.

Author Disclosure Statement

W.A.W., Z.W.J., B.W.S., S.C.E., and J.J.E. have a financial interest in LEAH Laboratories, a licensee from Mayo Clinic/Iowa State University for the filed GeneWeld patents. W.A.W., Z.W.J., S.C.E., and J.J.E. have a financial interest in LifEngine Biotechnologies, a licensee from Mayo Clinic/Iowa State University for the filed GeneWeld patents. J.J.E. and S.C.E. have financial interests in Recombinetics, Inc.

Funding Information

Funding for this article was provided by the Mayo Clinic, Ekker/Clark Lab (NIH grants GM63904 and OD020166) and by Iowa State University, Essner/McGrail Lab (NIH Grant OD020166).

Supplementary Material

Supplementary Figure S1
Supplementary Figure S2
Supplementary Figure S3
Supplementary Figure S4
Supplementary Figure S5
Supplementary Figure S6
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Table S4
Supplementary Table S5
Supplementary Table S6

References

1. Jinek M, Chylinski K, Fonfara I, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012;337:816–821. DOI: 10.1126/science.1225829.
2. Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* 2013;339:819–823. DOI: 10.1126/science.1231143.
3. Jao LE, Wente SR, Chen W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci U S A* 2013;110:13904–13909. DOI: 10.1073/pnas.1308335110.
4. Adli M. The CRISPR tool kit for genome editing and beyond. *Nat Commun* 2018;9:1911. DOI: 10.1038/s41467-018-04252-2.
5. Zetsche B, Gootenberg JS, Abudayyeh OO, et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 2015;163:759–771. DOI: 10.1016/j.cell.2015.09.038.
6. Truong LN, Li Y, Shi LZ, et al. Microhomology-mediated end joining and homologous recombination share the initial end resection step to repair DNA double-strand breaks in mammalian cells. *Proc Natl Acad Sci U S A* 2013;110:7720–7725. DOI: 10.1073/pnas.1213431110.
7. Ran FA, Hsu PD, Lin CY, et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell 2013;154:1380–1389. DOI: 10.1016/j.cell.2013.08.021.
8. Moreno-Mateos MA, Fernandez JP, Rouet R, et al. CRISPR-Cpf1 mediates efficient homology-directed repair and temperature-controlled genome editing. *Nat Commun* 2017;8:2024. DOI: 10.1038/s41467-017-01836-2.
9. Fernandez JP, Vejnar CE, Giraldez AJ, et al. Optimized CRISPR-Cpf1 system for genome editing in zebrafish. *Methods* 2018;150:11–18. DOI: 10.1016/j.ymeth.2018.06.014.
10. Maksimova E, Schiel JA, Zhao G, et al. *In Vitro* and *In Vivo* CRISPR Gene Editing with MAD7. Cambridge, United Kingdom: Horizon Discovery, 2018.
11. Thyme SB, Schier AF. Polq-mediated end joining is essential for surviving DNA double-strand breaks during early zebrafish development. *Cell Rep* 2016;15:707–714. DOI: 10.1016/j.celrep.2016.03.072.
12. He MD, Zhang FH, Wang HL, et al. Efficient ligase 3-dependent microhomology-mediated end joining repair of DNA double-strand breaks in zebrafish embryos. *Mutat Res* 2015;780:86–96. DOI: 10.1016/j.mrfmmm.2015.08.004.
13. Ata H, Ekstrom TL, Martinez-Galvez G, et al. Robust activation of microhomology-mediated end joining for precision gene editing applications. *PLoS Genet* 2018;14:e1007652. DOI: 10.1371/journal.pgen.1007652.
14. Aida T, Nakade S, Sakuma T, et al. Gene cassette knock-in in mammalian cells and zygotes by enhanced MMEJ. *BMC Genom* 2016;17:979. DOI: 10.1186/s12864-016-3331-9.

15. Hisano Y, Sakuma T, Nakade S, et al. Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Sci Rep* 2015;5:8841–8841. DOI: 10.1038/srep08841.
16. Nakade S, Tsubota T, Sakane Y, et al. Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat Commun* 2014;5:5560. DOI: 10.1038/ncomms6560.
17. Bhargava R, Onyango DO, Stark JM. Regulation of single-strand annealing and its role in genome maintenance. *Trends Genet* 2016;32:566–575. DOI: 10.1016/j.tig.2016.06.007.
18. Yao X, Wang X, Hu X, et al. Homology-mediated end joining-based targeted integration using CRISPR/Cas9. *Cell Res* 2017;27:801–814. DOI: 10.1038/cr.2017.76.
19. Wierson WA, Welker JM, Almeida MP, et al. GeneWeld: a method for efficient targeted integration directed by short homology. *bioRxiv* 2018:431627. DOI: 10.1101/431627.
20. Liu P, Luk K, Shin M, et al. Enhanced Cas12a editing in mammalian cells and zebrafish. *Nucleic Acids Res* 2019;47:4169–4180. DOI: 10.1093/nar/gkz184.
21. McGrail M, Hatler JM, Kuang X, et al. Somatic mutagenesis with a Sleeping Beauty transposon system leads to solid tumor formation in zebrafish. *PLoS One* 2011;6:e18826. DOI: 10.1371/journal.pone.0018826.
22. Field AC, Vink C, Gabriel R, et al. Comparison of lentiviral and sleeping beauty mediated $\alpha\beta$ T cell receptor gene transfer. *PLoS One* 2013;8:e68201. DOI: 10.1371/journal.pone.0068201.
23. Kumar S, Stecher G, Tamura K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular Biology and Evolution* 2016;33:1870–1874. DOI: 10.1093/molbev/msw054.
24. Rice P, Longden I, Bleasby A. EMBOSS: The European Molecular Biology Open Software Suite. *Trends Genet* 2000;16:276–277. DOI: 10.1016/S0168-9525(00)02024-2.
25. Talbot WS, Trevarrow B, Halpern ME, et al. A homeobox gene essential for zebrafish notochord development. *Nature* 1995;378:150–157. DOI: 10.1038/378150a0.
26. Kim D, Kim J, Hur JK, et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* 2016;34:863–868. DOI: 10.1038/nbt.3609.
27. Kim HK, Song M, Lee J, et al. *In vivo* high-throughput profiling of CRISPR-Cpf1 activity. Nat Methods 2017;14:153–159. DOI: 10.1038/nmeth.4104.
28. Papapetrou EP, Schambach A. Gene insertion into genomic safe harbors for human gene therapy. *Mol Ther* 2016;24:678–684. DOI: 10.1038/mt.2016.38.
29. Eyquem J, Mansilla-Soto J, Giavridis T, et al. Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* 2017;543:113–117. DOI: 10.1038/nature21405.
30. Hsiau T, Conant D, Maures T, et al. Inference of CRISPR edits from Sanger trace data. *bioRxiv* 2019:251082. DOI: 10.1101/251082.
31. Pierce AJ, Johnson RD, Thompson LH, et al. XRCC3 promotes homology-directed repair of DNA damage in mammalian cells. *Genes Dev* 1999;13:2633–2638. DOI: 10.1101/gad.13.20.2633.
32. Certo MT, Ryu BY, Annis JE, et al. Tracking genome engineering outcome at individual DNA breakpoints. *Nat Methods* 2011;8:671–676. DOI: 10.1038/nmeth.1648.
33. Gunn A, Stark JM. I-SceI-based assays to examine distinct repair outcomes of mammalian chromosomal double strand breaks. *Methods Mol Biol* 2012;920:379–391. DOI: 10.1007/978-1-61779-998-3_27.
34. Bennardo N, Stark JM. ATM limits incorrect end utilization during non-homologous end joining of multiple chromosome breaks. *PLoS Genet* 2010;6:e1001194. DOI: 10.1371/journal.pgen.1001194.
35. Smith JR, Maguire S, Davis LA, et al. Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration. *Stem Cells* 2008;26:496–504. DOI: 10.1634/stemcells.2007-0039.
36. Yang L, Soonpaa MH, Adler ED, et al. Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cell-derived population. *Nature* 2008;453:524–528. DOI: 10.1038/nature06894.
37. Zou J, Mali P, Huang X, et al. Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood* 2011;118:4599–4608. DOI: 10.1182/blood-2011-02-335554.
38. Zhang Y, Zhang Z, Ge W. An efficient platform for generating somatic point mutations with germline transmission in the zebrafish by CRISPR/Cas9-mediated gene editing. *J Biol Chem* 2018;293:6611–6622. DOI: 10.1074/jbc.RA117.001080.
39. Tu M, Lin L, Cheng Y, et al. A "new lease of life": FnCpf1 possesses DNA cleavage activity for genome editing in human cells. *Nucleic Acids Res* 2017;45:11295–11304. DOI: 10.1093/nar/gkx783.
40. Kleinstiver BP, Tsai SQ, Prew MS, et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol* 2016;34:869–874. DOI: 10.1038/nbt.3620.
41. Bin Moon S, Lee JM, Kang JG, et al. Highly efficient genome editing by CRISPR-Cpf1 using CRISPR RNA with a uridinylate-rich 3′-overhang. *Nat Commun* 2018;9:3651. DOI: 10.1038/s41467-018-06129-w.
42. Kleinstiver BP, Sousa AA, Walton RT, et al. Engineered CRISPR–Cas12a variants with increased activities and improved targeting ranges for gene, epigenetic and base editing. *Nat Biotechnol* 2019;37:276–282. DOI: 10.1038/s41587-018-0011-0.
43. Gao Z, Herrera-Carrillo E, Berkhout B. Delineation of the exact transcription termination signal for type 3 polymerase III. *Mol Ther Nucleic Acids* 2018;10:36–44. DOI: 10.1016/j.omtn.2017.11.006.
44. Chen JS, Ma E, Harrington LB, et al. CRISPR-Cas12a target binding unleashes indiscriminate single-stranded DNase activity. *Science* 2018;360:436–439. DOI: 10.1126/science.aar6245.
45. Wilson LOW, O'Brien AR, Bauer DC. The current state and future of CRISPR-Cas9 gRNA design tools. *Front Pharmacol* 2018;9:749. DOI: 10.3389/fphar.2018.00749.
46. Fu Y, Foden JA, Khayter C, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* 2013;31:822–826. DOI: 10.1038/nbt.2623.
47. Fu Y, Sander JD, Reyon D, et al. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* 2014;32:279–284. DOI: 10.1038/nbt.2808.
48. Mali P, Yang L, Esvelt KM, et al. RNA-guided human genome engineering via Cas9. *Science* 2013;339:823–826. DOI: 10.1126/science.1232033.
49. Doench JG, Hartenian E, Graham DB, et al. Rational design of highly active sgRNAs for CRISPR-Cas9–mediated gene inactivation. *Nat Biotechnol* 2014;32:1262–1267. DOI: 10.1038/nbt.3026.
50. Koike-Yusa H, Li Y, Tan EP, et al. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* 2013;32:267–273. DOI: 10.1038/nbt.2800.
51. Ren X, Yang Z, Xu J, et al. Enhanced specificity and efficiency of the CRISPR/Cas9 system with optimized sgRNA parameters in Drosophila. *Cell Rep* 2014;9:1151–1162. DOI: 10.1016/j.celrep.2014.09.044.
52. Shalem O, Sanjana NE, Hartenian E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 2014;343:84–87. DOI: 10.1126/science.1247005.
53. Wong N, Liu W, Wang X. WU-CRISPR: characteristics of functional guide RNAs for the CRISPR/Cas9 system. *Genome Biol* 2015;16:218. DOI: 10.1186/s13059-015-0784-0.
54. Perez AR, Pritykin Y, Vidigal JA, et al. GuideScan software for improved single and paired CRISPR guide RNA design. *Nat Biotechnol* 2017;35:347–349. DOI: 10.1038/nbt.3804.
55. Chari R, Yeo NC, Chavez A, et al. sgRNA Scorer 2.0: a species-independent model to predict CRISPR/Cas9 activity. *ACS Synth Biol* 2017;6:902–904. DOI: 10.1021/acssynbio.6b00343.
56. Chuai G, Ma H, Yan J, et al. DeepCRISPR: optimized CRISPR guide RNA design by deep learning. *Genome Biol* 2018;19:80. DOI: 10.1186/s13059-018-1459-4.
57. Swarts DC, Van Der Oost J, Jinek M. Structural basis for guide RNA processing and seed-dependent DNA targeting by CRISPR-Cas12a. *Mol Cell* 2017;66:221–233.e224. DOI: 10.1016/j.molcel.2017.03.016.
58. Cho SW, Kim S, Kim Y, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* 2014;24:132–141. DOI: 10.1101/gr.162339.113.