The CRISPR Enzyme Cpf1 as a Tool for Gene Regulation

A Major Qualifying Project submitted to the faculty of Worcester Polytechnic Institute in partial fulfillment of the requirements for the Degree of Bachelor of Science

Submitted by:
Kyle Morrison

Project Advisors:
Destin Heilman (WPI)
Scot Wolfe (UMMS)

April 26th, 2017

Abstract

Clustered regularly-interspaced short palindromic repeats (CRISPR), a bacterial immune system fully characterized less than a decade ago, has become indispensable in the fields of biology over the past several years as it has been engineered into a tool for the targeted modification of DNA. In particular, the RNA-guided, single-enzyme CRISPR/Cas9 system has been used extensively to introduce mutations and create fusion proteins with targetable functions. However, the new CRISPR enzyme Cpf1 provides a much more robust tool for researchers, with the novel ability to process pre-CRISPR RNAs (pre-crRNAs) independently and to generate staggered cuts in double-stranded DNA. These capabilities, among other nuances of the system, open the door to multiplexed gene targeting and streamlined experimental design. Already, this new system has been used to introduce mutations to various ends, but it has only just begun to be used for gene regulation. To develop Cpf1 into a scalable tool for such studies, the DNA and RNA catalytic domains in three Cpf1 species (*Acidaminococcus sp.* (As), *Francisella novicida* (Fn), *Lachnospiraceae bacterium* (Lb)) were mutated. The results of an analysis of the catalytic function of these mutants verified previous studies that identified the necessary catalytic residues for DNase function (As: E993; Fn: E1006; Lb: E925) and RNase function (As: H800; Fn: H843; Lb: H759). Furthermore, I show that the deoxyribonuclease-dead "dCpf1" mutants can be fused to activator and repressor domains, and introduced into mammalian HEK293T cells with only slightly reduced levels of expression and with nuclear localization maintained. Activation of CXCR4 expression in the HEK293T cell line with designed crRNA guides was not observed, underscoring the complexity of regulatory studies in mammalian systems. These results advance Cpf1 as a tool for precisely controlling the regulation of endogenous genes, identify potential pitfalls for future resolution and will serve to guide researchers in the quest to expand the capabilities of Cpf1.

Table of Contents

Acknowledgements

I would like to recognize several people for their contributions to this work and my development as a scientist over the course of this work:

Destin Heilman, for his guidance, both personal and professional, in writing this paper, managing the stresses of a research project, providing critical feedback, and keeping the goal in sight. I certainly wouldn't have the "skin of a rhinoceros," nor would I have an understanding of technical writing, without his help.

Scot Wolfe for his scientific guidance and mentorship. From helping me design experiments and interpreting results to getting me to think about the entire system and the context in which my experiments were taking place, his assistance was vital to the direction of the project. He provided a supportive and collaborative environment that I will be searching to replicate in all my future endeavors.

Pengpeng Liu and **Kevin Luk** for mentoring me. Their willingness to help me learn, understand results, develop and run experiments, and avoid pitfalls speaks volumes of their character. This paper would not exist without their help. **Pengpeng** supplied the wild type AsCpf1, FnCpf1, and LbCpf1 pCSDEST constructs as well as the corresponding sequencing primers, the DNMT1/EMX1 Cpf1 crRNA/pre-crRNA plasmids and the DNMT1/EMX1 sequencing primers, wild type SpCas9 and associated guide RNAs. **Kevin** supplied DNMT1/EMX1 Cpf1 crRNA plasmids and DNMT1/EMX1 sequencing primers, and he performed the Cpf1 immunofluorescence staining seen in Figure 2.

Samantha Kwok, for helping me get experiments moving in spite of my restricted schedule. I wouldn't have been able to move through things as quickly if not for her.

Ben Roscoe, **Sukanya Iyer**, and **Fatih Bolukbasi** for their career advice, experimental design advice, help interpreting results, technique suggestions and assistance, and patience. **Ben** supplied the pLV hUbC-dCas9-VPR-T2A-GFP construct, as well as the pCSDEST_dCas9_KRAB plasmid that were used to amplify VPR and KRAB, respectively. He also supplied the CXCR4 FW primer. **Sukanya** supplied HEK293T cells for me, which enabled me to push my experiments forward on time.

Introduction

The clustered, regularly interspaced short palindromic repeats (CRISPR) system has revolutionized the fields of molecular biology, genetics, and epigenetics since its characterization as an RNA-guided DNA endonuclease in 2012 (1). Just two years later, almost 600 papers had been published mentioning CRISPR (2). Today, a search on PubMed for "CRISPR" returns over 5,000 results, 2,107 of which are from 2016 alone (3). It is already being investigated as a method for the mass engineering of mosquito genomes to wipe out diseases (4). To quote Dr. Schimenti of Cornell University, "I've seen two huge developments since I've been in science: CRISPR and PCR" (2).

In nature, CRISPR systems are a kind of adaptive immune system found in bacteria (1). When a bacterium that harbors a CRISPR system is infected by a bacteriophage, a CRISPR-associated protein (Cas1) binds to invasive foreign DNA proximal to a recognition sequence, called a protospacer-adjacent motif (PAM) (5). This PAM is specific to the species of bacteriophage that infect the CRISPR-harboring bacteria, and allows the bacterium to distinguish self and foreign DNA molecules with high (99.6%) fidelity (5, 6). Together with another protein, Cas2, the Cas1/Cas2 complex acts as a nuclease, degrading the foreign DNA. This generates small DNA fragments, known as spacers, which are then integrated by this same complex into a CRISPR locus found in the bacterial genome (5, 7). This locus contains spacers acquired from foreign-DNA encounters interspersed with direct repeats, sequences that are repeated in close proximity to one another (7). Spacers are inserted towards the 5' end of the locus, immediately downstream of the "leader" sequence, a sequence that the Cas1/Cas2 complex uses to orient spacer insertion (8) (See Figure 1). This insertion is accompanied by the duplication of the first repeat to generate a completely direct-repeat flanked region (5).

When expressed, this constitutively transcribed CRISPR locus produces a long mRNA transcript (9), containing all of the integrated spacers and their associated direct repeats. Each spacer and its flanking direct repeats are collectively referred to as pre-CRISPR RNAs (pre-crRNAs). The direct repeats are recognized by another Cas protein (Cas5d) and cleaved, producing mature crRNAs that are complementary to the original DNA sequence, or "protospacer" (1, 10). These crRNAs then bind with several other Cas proteins, assembling into an enzymatic complex containing the multi-purpose Cas3 enzyme (11). This critical Cas protein acts as an ATP-dependent helicase and DNA nuclease that unwinds the target foreign DNA double helix (11). This allows the crRNA to Watson-Crick base pair to the complementary foreign DNA strand next

to the PAM, forming an R-loop structure (11). This allows the complex to cleave and inactivate the target DNA (1, 12) (See Figure 1). Importantly, this enzyme complex has much greater catalytic activity than the Cas1/Cas2 complex, making it very effective at recognizing and cleaving target DNA. In short, CRISPR is a form of "molecular memory" for bacteria that allows them to recognize and respond to repeat infections, as well as acquire resistance to new threats based on DNA sequences and pass this resistance on to daughter cells during replication.

Although fascinating as a tour-de-force of the evolutionary push for novel cellular machinery, the components of the CRISPR system are what make these adaptive immune proteins so valuable. The crRNA/Cas protein enzymatic complex is an RNA-guided endonuclease that allows for the targeted cleavage of any DNA sequence that contains a PAM followed by sequence complementary to that encoded by the crRNA. This means that crRNAs can be *designed* to target a specific region in a genome or plasmid.

However, the CRISPR system described above, a Type I system requiring multiple proteins for the final foreign DNA-interfering nuclease function and crRNA interaction, is cumbersome at best. Type II systems, on the other hand, provide a much simpler system for engineering. In particular, the Type II CRISPR/Cas9 system is one such system of significant interest in the field. Found in many species of bacteria, particularly *Streptococcus pyogenes* (Sp), this Type II system uses a single Cas protein, Cas9, to perform the helicase and nuclease functions performed by the entire multi-protein enzymatic interfering complex in Type I systems (1).

Cas9 also makes use of crRNAs to form R-loops with target DNA in a PAM-dependent manner, but the mechanism for pre-crRNA processing is somewhat different. In the Cas9 system, pre-crRNAs interact with another RNA molecule, known as a trans-activating CRISPR RNA (tracrRNA). This tracrRNA provides necessary secondary structure for enzymatic interaction with Cas9 as well as endogenous RNase III, which recognizes and cleaves pre-crRNAs at the spacer-flanking direct repeats (1, 12). The Cas9 protein unwinds the target DNA double helix, allowing the crRNA to Watson-Crick base pair with the target sequence immediately proximal to the PAM recognized by Cas9 (1). Once recognition has occurred, nuclease domains that resemble those of two proteins, the RuvC protein involved with Holliday junction resolution (13) and the HNH family of nucleases (14), will nick the sense and antisense strands respectively, resulting in a blunt-end double-stranded break (DSB) (1).

Compared to Type I systems that require multiple proteins forming an effector complex, Type II systems are much more straightforward and are thus more accessible to the scientific

community. They provide a platform for single enzyme-based RNA-guided DNA cleavage that can be tuned solely based on the crRNA sequence. This has led to the development of powerful Cas9-based tools for introducing deletions into target genomes. Additionally, cellular DNA repair mechanisms allow targeted sequence insertion into the genome. This is accomplished by co-delivering Cas9 with the desired oligonucleotide containing regions homologous to the target site, which the target cell will use as a template for homologous recombination-based repair upon the introduction of a double-strand break (15). These techniques have also led to the generation of mice carrying desired mutations (16).

Furthermore, deoxyribonuclease-dead versions of Cas9 (17) have been generated, turning it into an extremely malleable system. This catalytically inactive enzyme becomes an easily-directed DNA binding protein with the ability to localize any fused protein to the target DNA region. This opens the door for a plethora of applications, one of which is gene regulation with activator/repressor fusion constructs. In the Cas9 system, this has already been used to probe gene regulatory networks (18, 19) and up- and down-regulate particular genes (20). With guide RNAs and Cas9, it seems as though we have accomplished our "long sought-after goal" of DNA-localized function (21). How could it get much better?

Enter CRISPR from Prevotella and Francisella 1 (Cpf1): a new Type II CRISPR system that has recently been described (22) that utilizes a single large CRISPR enzyme that has been identified in several bacterial species, including *Acidaminococcus sp.* (As), *Francisella novicida* (Fn), and *Lachnospiraceae bacterium* (Lb). The effector of a CRISPR system like Cas9, Cpf1 is a ~1300 amino acid residue protein with a nuclease domain and crRNA-recognition motifs (23). However, several aspects of this system are remarkably different from the CRISPR/Cas9 system (Figure 2). The Cpf1 crRNA resembles a streamlined version of the Cas9 crRNAs. Cas9 crRNAs interact with tracrRNAs at their 3' ends, or downstream of the spacer, to provide the necessary secondary structure for recognition by Cas9 (23). Cpf1 crRNAs, on the other hand, contain a simple hairpin loop immediately 5' to the spacer that is sufficient for recognition, although the sequence is slightly different between species.

Cpf1 also contains a RuvC deoxyribonuclease domain like Cas9 (23), but this acts as the sole catalytic domain. Cpf1 produces staggered cuts several nucleotides distal to the PAM by cleaving one strand, undergoing a conformational change, and cleaving the other strand (22, 23, 24). Cas9, on the other hand, generates blunt-end double-strand breaks proximal to the PAM through independent RuvC- and HNH-domain-mediated cleavage of the plus and minus strands,

respectively (1) (Figure 2). This staggered cutting simplifies engineered gene insertion greatly, as sticky end-based non-homologous end joining is much more likely to incorporate the desired insert properly and in the desired orientation than blunt-end-based insertion (22). Additionally, PAMs recognized by Cpf1 are T-rich (23) compared to the 5'-NGG-3' PAM of the *Sp.* Cas9 system (15, 19), which expands the number and types of sequences that can be targeted.

Most importantly, Cpf1 also possesses a ribonuclease domain, which enables the enzyme to generate fully mature crRNAs from available pre-crRNAs completely independent of tracrRNAs (22). The Cpf1 pre-crRNA, with a simple hairpin loop immediately upstream of the spacer, is further preceded upstream by a directly adjacent repeat recognition sequence (RRS), which serves as the site of structure- and sequence-dependent cleavage by the Cpf1 RNase domain to release a mature crRNA that can be used for subsequent targeting (26).

Herein lies the greatest advantage that Cpf1 has over Cas9: to edit multiple sites with Cas9, several separate guide RNAs have to be delivered, either as separate plasmids, transcripts, or entirely separate ribonucleoprotein (RNP) complexes. Additionally, tracrRNA must be co-delivered to catalyze pre-crRNA processing and generate functional Cas9 guides. In this case, the activity of endogenous RNase enzymes must be relied upon (16). In Cpf1, however, a single transcript containing all of the desired crRNAs in RRS-flanked pre-crRNA form can be used with Cpf1, which will process the transcript into each independent crRNA and subsequently target the desired loci (27).

Like Cas9, deoxyribonuclease dead versions of Cpf1 have also been generated, and their capabilities as the same class of tool as dCas9-fusions have begun to be explored. Although in its infancy, dCpf1 has already been used as a targetable effector for gene repression in *Arabidopsis* (28) and CRISPRi (steric-hindrance-based CRISPR interference) in *Escherichia coli* (29).

While there are many appealing targets for regulation with a programmable system like Cas9 or Cpf1, one of particular interest is the surface protein CXCR4 (CD184). CXCR4 is a chemokine receptor that is best known for its role in T-tropic HIV strain entry into human CD4+ T-cells (30), as well as its prevalence in various types of cancers (31). Regulation of endogenous proteins is a crucial step forward in developing a clinically-relevant tool, and a molecule like CXCR4 that is common to mammalian cells yet data suggests is nonessential for cell viability (32) provides a safe target. Repression of CXCR4 has been previously described in HeLa cells with dCas9 fused to a Krüppel associated box (KRAB) repressor domain (33), and activation of various genes has been

achieved in HEK293T cells with dCas9 fused to a VP64-p65-Rta (VPR) tripartite activator (34), but no mammalian activation or repression has been shown with Cpf1.

As powerful as CRISPR systems are, they are unfortunately not perfect. The dependence of the systems on complementary base-pairing and PAM recognition to induce enzymatic activity means that any thermodynamically feasible near-perfect matches between the guide RNA and DNA are possible. This can lead to cleavage or localization of fused functional domains at unintended loci, known generally as "off-target" effects. Off-target effects have been reported for the CRISPR/Cas9 system over the years (35) and pose a very real problem if CRISPR systems are to be clinically relevant (36).

Furthermore, in initial experiments the Wolfe lab has observed some toxicity upon delivery of Cpf1 plasmid in both bacterial and mammalian *in vitro* systems. It is hypothesized by our lab that this might relate to the RNase domain of Cpf1, but investigation of this has only just begun at the time of this writing. *In vivo*, others have reported Cpf1 toxicity as well (31). This is again very topical if these systems are to be used in a clinical setting; unwanted cell death purely as a result of treatment is a problem worth resolving.

To take advantage of the rise of Cpf1 and to lay the groundwork for its future use as a powerful tool, I show here that I have generated ribonuclease-dead and deoxyribonuclease-dead versions of As, Fn, and Lb Cpf1 that exhibit none of their respective catalytic activities *in vitro*. I further demonstrate that this dCpf1 can be fused to VPR and KRAB domains and introduced into mammalian cells with minimal additional toxicity and no effect on nuclear localization. This work furthers the development of dCpf1 as a tool for mammalian gene regulation.

Materials and Methods

Generation and Verification of Nuclease-Dead Cpf1 Constructs

In a 50 µL reaction, 100 ng of template pCSDEST_pmd262 plasmid containing Cpf1 was combined with 5 µL each of 5 µM forward and reverse primers, along with 5 U of Phusion polymerase (NEB) in 1x HF Buffer or 5 U of Q5 polymerase (NEB) in 1x Q5 Reaction Buffer and the remaining relevant reaction components per the manufacturer's protocol. QuikChange primers were designed to have 15-20 complementary nucleotides on either side of the inserted base mismatches.

Initial denaturation was performed at 98°C for 30s, followed by 35 cycles of denaturation at 98°C for 10s and a combined annealing-extension step at 72°C for 4 minutes. After cycling, a final extension time of 5 minutes at 72°C was followed by ramping to 4°C and then storage at -20°C. 1 µL was used to transform XL-10 Gold Cells (NEB). Plasmid DNA was isolated using the GenElute Plasmid Miniprep Kit (Sigma) according to the manufacturer's protocol.

Mutants were verified through Sanger sequencing (Genewiz) using GenomeCompiler (Twist Bioscience) with custom primers and plasmid integrity was confirmed with PvuI (NEB) (AsCpf1), XmnI (NEB) (FnCpf1), or AhdI and BamHI-HF (NEB) (LbCpf1) restriction digestion and gel electrophoresis (1% agarose in TAE) with ethidium bromide (Thermofisher) according to the manufacturer's protocols.

Alternative Generation and Verification of Remaining dCpf1 Constructs

For templates that proved difficult to amplify with the QuikChange method, oligonucleotides were designed to substitute in for the target region through Gibson Assembly. Primers were designed to amplify the plasmid using Phusion (NEB) with complementary overhangs to the annealed oligonucleotides, which were generated by combining 1 µL of each 100 µM oligonucleotide suspension with 5 U of T4 Polynucleotide Kinase (NEB) in 1x T4 PNK Buffer (NEB) in a 25 µL reaction. Samples were incubated at 37°C for 40 minutes, followed by inactivation at 95°C and ramping to 12°C. Plasmid backbone was amplified using the same PCR program as the QuikChange method.

The products of these reactions were then combined in a Gibson cloning reaction using the Gibson Assembly Cloning Kit (NEB), with oligo inserts diluted 1:3 and plasmid vector diluted 1:2. Samples were incubated at 50°C for 1 hour, then 2 µL were used to transform XL-10 Gold Cells

(NEB). Plasmid DNA was isolated using the GenElute Plasmid Miniprep Kit (Sigma) according to the manufacturer's protocol. Mutants were verified using the same method as described above.

Generation and Verification of dCpf1 Effector Constructs

Primers were designed to amplify VPR and KRAB genes as well as the entire Cpf1 vector. The 5' and 3' ends were designed to contain homology domains between the fragments, and to incorporate protein linker regions and restriction sites. Fragments were amplified by Phusion (NEB) PCR using 10 ng of template and 5 µL each of 5 µM of each primer in a 50 µL reaction.

Initial denaturation was performed at 98°C for 30s, followed by 30 cycles of denaturation at 98°C for 10s, annealing at 72°C for 30s, and by extension for 3 minutes (Cpf1 vector) or 45s (VPR/KRAB) at 72 °C. After cycling, a final extension time of 5 minutes (Cpf1) or 2 minutes (VPR/KRAB) at 72°C was followed by ramping to 4 °C and then storage at -20 °C. Fragment amplification was confirmed by gel electrophoresis (1% agarose in TAE) with ethidium bromide (Thermofisher).

Amplified fragments were then assembled using the Gibson Assembly Cloning Kit (NEB), with KRAB/VPR fragments diluted 1:3 and Cpf1 amplicons diluted 1:2 (approximately 5:1 mass (by picomoles) ratio). Samples were incubated at 50°C for 1 hour, then 2 µL were used to transform XL-10 Gold Cells (NEB). Plasmid DNA was isolated using the GenElute Plasmid Miniprep Kit (Sigma) according to the manufacturer's protocol. Mutants were confirmed through Sanger sequencing (Genewiz) using GenomeCompiler (Twist Bioscience) with custom primers.

Cell Culture, Transfection, and Harvest

HEK293T cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) with 10% FBS and 1% penicillin-streptomycin. 100 ng of wild type/mutant Cpf1 or wild type Cas9, 100 ng of sgRNA/crRNA, and 50 ng each of mCherry and pmd5 using Polyfect Transfection Reagent (Qiagen) by adding 4 µL of Polyfect to the DNA mixture followed by 30 µL of DMEM without serum or antibiotics (for flow cytometry experiments, mCherry was excluded and 100 ng of pmd5 were used to maintain total plasmid load and transfection efficiency). Samples were incubated for 15 minutes and then added dropwise to 1.8×10^5 cells in 24-well format. CRISPR enzymes and sgRNAs/crRNAs were designed to be constitutively expressed in culture.

Nuclease-Dead Cpf1 Functional Assay

72 hours post-transfection, cells were harvested with trypsin (Gibco), washed twice with 1x phosphate-buffered saline (PBS) at 700-1000 xg for 5 minutes, and frozen overnight at -20°C. Genomic DNA was isolated using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma) according to the manufacturer's protocol.

A Phusion (NEB) PCR using 50-100 ng of genomic DNA was run in a 25 µL reaction with 1.5 µL each of 5 µM forward and reverse primers targeting DNMT1 and EMX1, 3% DMSO, and 0.5 U of Phusion polymerase in 1x HF Buffer. Initial denaturation was performed at 98°C for 10s, followed by 30-35 cycles of denaturation at 98°C for 10s, annealing at 65°C for 20s, and extension for 20s at 72°C. After cycling, a final extension time of 5 minutes at 72°C was followed by ramping to 4°C and then storage at -20°C.

For genomic DNA that proved difficult to amplify specifically, a nested PCR approach was used. A set of outer primers was used in the same reaction as above with annealing for 30s and extension for 30s to generate an initial amplicon, which was then used as template for the original PCR outlined above, with a higher annealing temperature of 67°C.

PCR products were then denatured and reannealed by incubation at 95°C for 5 minutes followed by ramping down to 25°C by -5°C/minute and holding at 4°C. 500 ng of reannealed products were digested with 5 U of T7 Endonuclease I (NEB) in a 20 µL reaction for 45 minutes at 37°C. Digest results were visualized using gel electrophoresis (1-1.2% agarose TAE) with ethidium bromide (ThermoFisher).

Immunostaining

48 hours post-transfection, growth media was removed from the cells, which were then fixed with 4% formaldehyde in 1x PBS at room temperature for 15 minutes. Fixing solution was then removed, and cells were washed three times with 1x PBS for 5 minutes each at room temperature. Following fixing and washing, samples were blocked for 1 hour with 2% Bovine Serum Albumin (BSA) and 0.3 % Triton X-100 (Bio-Rad) in 1x PBS. After blocking, the blocking media was removed and the samples were incubated overnight at 4°C in the dark in 1x blocking buffer with a 1:500 dilution of anti-HA Tag mouse primary antibody (Invitrogen). Samples were then washed three times with 1x PBS for 5 minutes each, followed by incubation with a 1:2000 dilution of Alexa Fluor 488 Donkey anti-mouse secondary antibody (Invitrogen) in 1x blocking buffer for 1 hour at 4°C in the dark. A final 3x PBS wash for 5 minutes each was then performed. Slides were

then covered with Prolong Gold Antifade Reagent with DAPI (ThermoFisher) and immunofluorescence was then assayed on a Zeiss light microscope.

CXCR4 crRNA Functional Assay

72 hours post-transfection, cells were harvested with trypsin (Gibco), washed twice with 1x phosphate-buffered saline (PBS) at 700-1000 xg for 5 minutes, and frozen overnight at -20°C. Genomic DNA was isolated using the GenElute Mammalian Genomic DNA Miniprep Kit (Sigma) according to the manufacturer's protocol.

Several PCR programs were attempted using Phusion polymerase (NEB) and Q5 polymerase (NEB). Annealing temperatures ranging from 65-69°C, annealing times from 15s-30s, and initial denaturation times of 10s, 15s, and 3 minutes were used.

The program that specifically amplified negative control DNA was initial denaturation at 98°C for 15s, followed by 35 cycles of denaturation at 98°C for 10s, annealing at 66°C for 20s, and extension at 72°C for 20s. A final extension step of 5 minutes at 72°C was then followed by ramping to 4°C and subsequent storage at -20°C. This was performed in a 25 µL reaction using 100-200 ng of genomic DNA, 5 U of Q5 polymerase, 1x Q5 Enhancer, 1.5 µL each of 5 µM forward and reverse primers in 1x Q5 Reaction Buffer.

Due to the difficulty in producing clean amplicons, T7EI assays were not run on these samples, as a relatively homogeneously-sized population of amplicons is essential to interpretation of assay results.

Quantification of Gene Activation

72 hours post-transfection, cells were harvested with trypsin (Gibco), washed twice with 1x PBS at 700-1000 xg for 5 minutes, and frozen overnight at -20°C. Cells were then resuspended to a concentration of $1x10^6$ cells/mL in staining buffer (1x PBS with 0.1% BSA). CXCR4 antibody (PE Rat anti-human CD184, CLONE, BD Pharmingen) was added to a final concentration of ~2.5 ug/mL and samples were incubated covered, on ice for 45 minutes. Cells were then washed twice with 1x PBS at 700-1000 xg for 5 minutes, and finally resuspended in 250 µL of staining buffer. When used, 4 µL of 7AAD (BD Pharmingen) were added to samples and then incubated covered, on ice for 5 minutes. Samples were analyzed on a BD LSRII flow cytometer (BD Biosciences).

Results

Cas9 has been engineered into a powerful tool for DNA manipulation, as well as gene regulation through catalytic inactivation and subsequent fusion to activation or repression domains, but Cpf1 has only just begun to be utilized in this fashion. Control of gene expression has not yet been demonstrated in mammalian cells.

To develop a Cpf1-based tool of this class, the deoxyribonuclease (glutamic acid) and ribonuclease (histidine) catalytic residues for AsCpf1 (glutamic acid, position 993 - E993; histidine, position 800 - H800), FnCpf1 (E1006, H843), and LbCpf1 (E925, H759) were substituted for alanine, which has been previously shown to be sufficient for loss of catalytic function (20, 21) (Figure 3A). To introduce the necessary point mutations, QuikChange primers were designed to have the desired base-pair substitutions flanked on either side by regions of complementarity to the Cpf1 gene to facilitate thermodynamically stable Watson-Crick base-pairing during mutagenic PCR (Figure 3B). For mutants that proved difficult to generate through this QuikChange method, Cpf1 amplification primers were designed to PCR amplify the plasmid for use in Gibson Assembly with custom oligonucleotides built to insert the desired base pair substitutions by homology. Mutants were Sanger sequenced to confirm successful nucleotide substitutions (Figure 3C, Supplemental Figure 1) and plasmids were further verified by restriction digest to confirm the absence of cloning scars (Figure 3D).

To confirm that these amino acid substitutions were sufficient for functional knockout of deoxyribonuclease and ribonuclease activity, HEK293T cells were co-transfected with crRNAs targeting the DNMT1 gene or the EMX1 gene (20), along with the mutant Cpf1 plasmids. When building a tool for the regulation of gene expression, it is crucial to determine that mutants also localize to the nucleus. This was confirmed by immunofluorescent staining and subsequent light microscopy of the built-in HA tags on the Cpf1 constructs, which showed that the nuclease mutations had no effect on nuclear localization and expression when compared to wild type SpCas9 and wild type AsCpf1 (Figure 4). However, transfection rates appeared to be slightly lower (Figure 4).

Furthermore, to avoid unintended mutagenesis when localizing future regulatory constructs, it is imperative that the As-E993A-Cpf1 (AsE), Fn-E1006A-Cpf1 (FnE), and Lb-E925A-Cpf1 (LbE) DNase mutants are catalytically inactive. To verify this, HEK293T cells were again transfected with DNMT1 or EMX1 LbCpf1 crRNAs (Figure 5B) and mutant Cpf1 plasmids. The genomic DNA was harvested and analyzed for insertion or deletion (indel) mutations by PCR

amplification of the gene. The PCR products, which are potentially slightly heterogeneous in length (~6 bp) due to indels, are then denatured to allow the strands to separate. This is followed by cooling, which allows slightly mismatched strands to anneal to one another. The reannealed products are then treated with T7 endonuclease I, which will cleave only mismatched products. By this T7EI mismatch assay, the mutants failed to produce cleaved DNA products on an agarose gel when compared to wild type Cpf1 and negative controls (Figures 5E-5F), confirming previous research that suggested one mutation is sufficient for loss of deoxyribonuclease function. Interestingly, these data show that AsCpf1 and FnCpf1 can produce indels at target sites using crRNAs designed for LbCpf1 (Figure 5E), although the activity of FnCpf1 in this experiment appeared lower than AsCpf1. This suggests that the structure of the mature crRNA might be sufficient for the induction of DNA cleavage.

The As-H800A-Cpf1 (AsH) RNase mutant also did not produce indels relative to wild type AsCpf1 and to negative controls when co-transfected with DNMT1 or EMX1 pre-crRNAs (Figure 5A) requiring RNase-domain-mediated RRS processing (Figures 5C-5D). However, the AsH mutant was able to produce indels when co-transfected with mature crRNAs, by the same mismatch assay, when compared to wild type AsCpf1 (Figures 5C-5D). FnCpf1 and LbCpf1 were observed to be capable of generating indels when supplied with mature AsCpf1 crRNA, but not with AsCpf1 pre-crRNA when compared to the negative control (Figures 5C-5D). These data also correlate with previous research, confirming the indispensability of the H800 residue to catalysis in AsCpf1. Additionally, these data suggest that although crRNA structure may be sufficient for the generation of indels in all three species of Cpf1, pre-crRNA structure does not appear to be sufficient for the induction of RNase-domain-mediated processing.

The deoxyribonuclease-dead constructs for each species of Cpf1 were further engineered using primers designed for use in Gibson Assembly such that KRAB or VPR domains could be amplified from source plasmids and attached to the C-terminus of the dCpf1 genes by homologous recombination (Figures 6A-6B). These constructs were designed to retain the 3xHA tag/2xNLS at the C-terminus of the expressed protein, which are critical for nuclear localization (Figure 6A). Sanger sequencing was used to verify that constructs were assembled properly and in the correct orientation (Figure 6C, Supplemental Figure 1).

To effectively induce the activation of a gene, it is important to avoid interfering with the cellular machinery involved in transcription. For effective repression, however, steric interference is often synergistic with the function of a repression domain. For these reasons, crRNAs targeting

5’-TTTN-3’ PAM regions upstream, downstream, and proximal to the transcription start site for the human (HG19) CXCR4 gene were designed based off of previous studies (31) (Figures 7A and 7B).

The LbE-VPR construct, along with CXCR4 crRNAs targeting upstream of the transcript start site, were then delivered to HEK293T cells by transfection. 72 hours later, cells were stained with fluorescent antibody and analyzed with flow cytometry to determine the CXCR4 protein levels. Upon analysis, little to no difference in CXCR4 staining was observed when compared to the non-transfected control and the no-DNA transfection control (Figure 7C) 72 hours post-transfection. All polyfect-treated samples saw reduced overall CXCR4 positive staining compared to the no polyfect-treated, PE/CXCR4-stained sample. A slight increase in cell death was observed when the LbE-VPR construct was delivered without crRNAs and with crRNAs 7 and 10 when compared to the no polyfect-treated, 7AAD-stained sample (Figure 7D). This cell death was also higher than that observed for SpCas9 with EMX1 guide RNA-treated cells, as well as wild type LbCpf1.

To confirm that this negative result was not due to a lack of protein expression and/or localization, HEK293Ts were transfected with LbE-VPR with CXCR4 crRNA 8 and stained for the presence of the HA epitope tag 48 hours later. This showed that LbE-VPR is expressed and present in the nucleus of transfected cells, although rates of expression were significantly lower than what was observed for SpCas9 transfected with EMX1 guide RNA (Figure 8). Additionally, the LbE deoxyribonuclease-dead mutant appeared to be expressed in slightly lower levels than previously seen (Figure 4), suggesting that transfection efficiency is somewhat variable.

I then attempted to investigate the binding affinity of LbCpf1 for the CXCR4 locus with the designed crRNAs by PCR and subsequent T7EI assay. Unexpectedly, the region proved exceptionally difficult to amplify for treated samples, but not untreated samples, under various experimental conditions (Figure 9). This occurred with three biologically replicative transfections and subsequent genomic DNA purifications.

Discussion

Cpf1 has emerged as a novel tool for gene insertion, knockout generation, and has begun to be investigated as a tool for gene regulation. However, no studies have shown the potential for Cpf1 as such a regulatory tool in mammalian systems; all of the work in this area has so far been exclusive to the various species of Cas9. Here, I verify previous studies confirming the residues essential to DNA and RNA catalysis in AsCpf1, FnCpf1, and LbCpf1 and advance Cpf1 towards being used as a tool for the activation/repression of genes in mammalian cells.

Compared to dCas9-based gene regulation techniques, deoxyribonuclease-dead Cpf1 has several potential advantages. Not only is the enzyme smaller and relies on a simpler crRNA for its targeting, but the additional functionality in the ribonuclease domain opens the door for multiplexed gene regulation, wherein multiple genes could be targeted for activation or repression at once. This could be particularly useful in controlling the cell fate decisions in stem cells, where activating multiple regulatory genes simultaneously could simplify direct reprogramming.

Aside from being a direct competitor to Cas9-based approaches, Cpf1's unique T-rich PAM requirement broadens the range of target sites that can be accessed. With comparable functionality between the two systems, this means that substituting Cpf1 for Cas9 and vice-versa is feasible if PAM restrictions become a concern in experimental design.

It is interesting that all three species of Cpf1 can not only produce indels using their own mature crRNAs, but also with those built for AsCpf1 and LbCpf1. The guides were designed to target 5'-TTTN-3' PAMs, which are required for AsCpf1 and LbCpf1 target recognition, whereas only 5'-TTN-3' PAMs are required for FnCpf1 function. This suggests that Cpf1 recognition of crRNAs based upon structure might be sufficient for enzymatic function. Furthermore, this seems to indicate that AsCpf1 and LbCpf1 may be more structurally and functionally similar to each other than to FnCpf1. A quantitative analysis of the behavior observed here might simplify future experiments that involve multiple species of Cpf1 if the activity with one of the species of guides proves to be satisfactory for producing indels.

Due to time constraints, it was not possible to test the repressive capabilities of the LbE-KRAB fusion protein, compare the activities of the three species of dCpf1-effector constructs, or resolve the lack of CXCR4 upregulation by the LbE-VPR construct

To test the LbE-KRAB construct's repressive capabilities, the Wolfe lab has been investigating the Jurkat immortalized human T-cell line, which expresses CXCR4 very highly, as a

potential model. For the reasons outlined at the beginning of this report, this would be a safe target for downregulation. A GFP stably-expressing HEK293T cell line would also be useful in a proof-of-concept study, although the targeting of an endogenous human gene might prove more foundational for future research studies.

Comparing the relative capacities of the different species of the dCpf1-effector constructs is imperative for building the most potent tool for gene regulation possible. The Wolfe lab has observed more robust catalytic and binding activity with AsCpf1 relative to both LbCpf1 and FnCpf1, and other groups have reported on the high specificity of AsCpf1 (36). This comparative analysis could be carried out with relative simplicity upon functional validation of crRNAs at the target site for each species.

However, this type of functional validation did not prove as simple as it initially appeared. Exhaustive attempts were made to assess the activity of LbCpf1 by T7EI assay at the CXCR4 locus when targeted with the designed crRNAs, but the genomic region proved highly difficult to amplify from LbCpf1/crRNA-treated samples with both Phusion and Q5 polymerases under a myriad of conditions. Several primer sets were designed, all producing clean PCR products on genomic DNA purified from untreated cells under optimized conditions. Upon use of LbCpf1/crRNA treated genomic DNA as PCR template, however, the same primers under the same conditions produced amplicons of varying sizes and abundances. It is possible, due to the high GC content and long stretches of adenines and thymines in the CXCR4 gene, as well as the presence of an upstream SINE repeat element, that induced DSBs are compromising the integrity of the genomic DNA. Large-scale deletions could conceivably remove primer sites or affect PCR amplicon size, which could generate the highly convoluted results.

Taken together with the nuclear localization of the fusion proteins, there are a few potential reasons for the lack of increased CXCR4 expression. It is unknown whether or not LbCpf1 effectively binds the CXCR4 target sites with the provided crRNAs. Successfully determining the activity of LbCpf1 at the target sites using a T7EI assay would answer this question, as would a ChIP-qPCR analysis to address whether or not LbCpf1 can bind to the targeted region with the supplied crRNAs.

Another possibility is the fusion protein design, which contained six amino acid residues for the linker region between Cpf1 and the effector domains (VPR or KRAB). This may prove to be too inflexible for the effector domains to function properly, and inserting additional residues to lengthen the linker region and reduce the rigidity of the construct might lead to effective function.

Although the LbE-VPR fusion construct localizes to the nucleus in transfected cells, the low expression levels might make it difficult to quantify any effect that the constructs may be having on CXCR4 expression. Optimizing transfection conditions to maximize expression might resolve this, or putting the gene under the control of a stronger promoter. Another option that could result in more robust activation would be to generate Cpf1-effector stably-expressing HEK293T cell lines that are under antibiotic selection, thus ensuring the presence and expression of the LbE-VPR protein. These lines could then be transiently exposed to the desired crRNAs followed by experimental readout 48-72 hours later.

The designed crRNAs may also be a potential source of complication. The regions the crRNAs target may not be far enough upstream from the transcription start site and related promoter elements to effectively upregulate CXCR4 transcription. In the referenced study that these crRNAs were designed to mimic (31), the SpdCas9 guides were used for CRISPRi and SpdCas9-KRAB-mediated repression. As a consequence, these were likely optimized for interference behavior and by generating similar constructs here, there might be steric hindrance of the transcription machinery taking place. Designing additional guide constructs that target further upstream of the current region targeted could solve this.

Lastly, it is also possible that cellular machinery is interfering with the function of the construct. The presence of histones near the CXCR4 gene would make access to it extremely difficult. Considering HEK293T cells were chosen as a target for this activation study due to their low surface levels of CXCR4, this is entirely plausible and would mean that another cell type, a different gene target, or targeted histone modifications might be necessary to observe an increase in CXCR4 expression.

However, it is difficult to determine from the results presented here what is occurring, as all the transfected samples (including the polyfect-only control) stained less positive for CXCR4 than the PE antibody control. Again, ChIP-qPCR analysis could shed some light on what is happening, as it could be used to ascertain whether or not the LbE-VPR and can be bound to the region surrounding the CXCR4 transcription start site/promoter region by immunoprecipitation and subsequent PCR amplification. Given the precedent for the use of dCas9 as a tool for activation, designing guide RNAs for use with a pre-defined dCas9-VPR construct would serve as a good positive control.

CRISPR systems have made genetic engineering, gene regulation, and many other kinds of studies much more affordable and accessible to the scientific community. Cpf1 shows promise as the

next step in the evolution of these tools, and may ultimately eclipse Cas9 in its usefulness. Although this study underscores how complex working with mammalian cells can be, it is crucial that these initial steps be taken and the groundwork laid in order to push Cpf1 forward as the next powerful tool for genome editing and modulation.

Figures

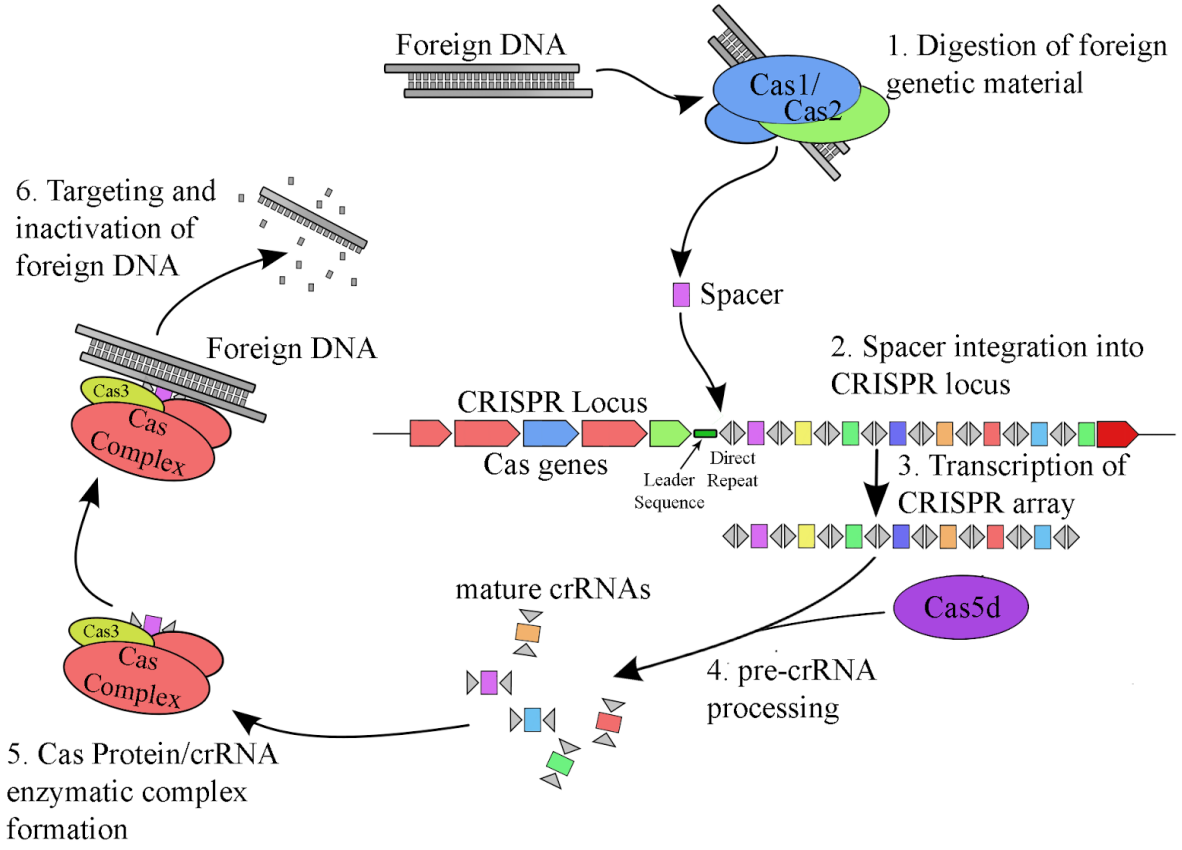


Figure 1. Diagram of CRISPR system function in nature. Figure depicting the basics of Type I CRISPR bacterial adaptive immune systems. Adapted from Horvath et al (37).

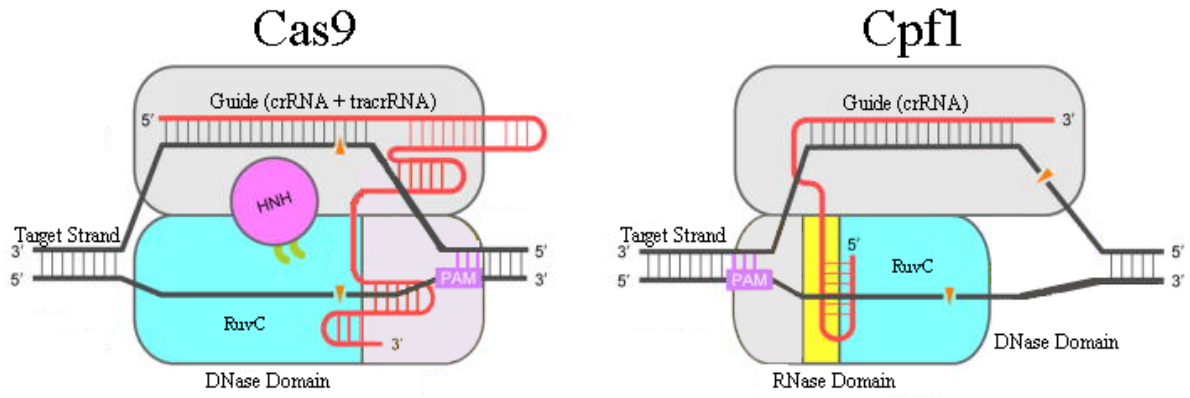
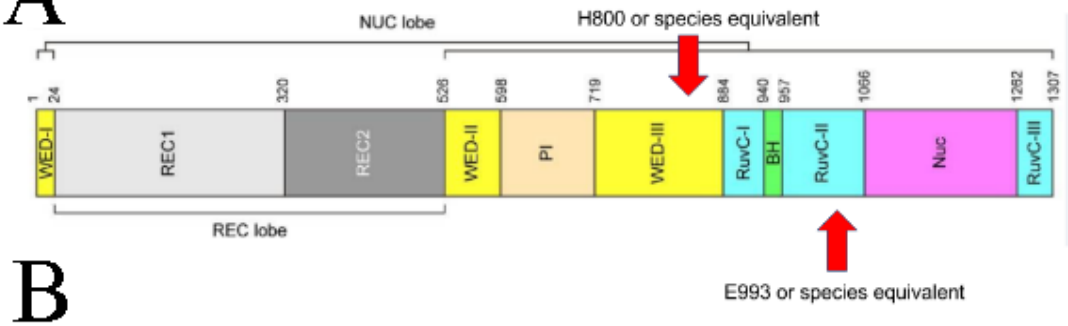


Figure 2. Comparison of Cas9 and Cpf1. Figure depicting the basic structure of each CRISPR enzyme in complex with double-stranded DNA after R-loop formation. Adapted from Yamano et al (21).

A



B

Name	Sequence
As Cpf1 (E993)	5'-[...]ATGATCCACTACCAGGCCGTGGTGGTGCTGGAGAACCTGAATTTCGGCTT TAAGAGCAAGAGGACCGGC[...]-3'
Fn Cpf1 (E1006)	5'-[...]ATCGAATACAATGCCATTGTGGTGTTCGAGGATCTGAACTTCGGCTTTAA GAGGGGGCGCTTTAAG[...]-3'
Lb Cpf1 (E925)	5'-[...]GTGGAGAAGTACGATGCCGTGATCGCCCTGGAGGACCTGAACTCTGGCT TTAAGAATAGCCGCGTGAAG[...]-3'
As Cpf1 (H800)	5'-[...]CGCCCTAAGTCCAGGATGAAGAGGATGGCACACCGGCTGGGAGAGAAG ATGCTGAACAAGAAG[...]-3'
Fn Cpf1 (H843)	5'-[...]CGGAAGCAGTCTATTCCTAAGAAAATCACTCACCCAGCTAAGGAGGCCA TCGCTAACAAGAAC[...]-3'
Lb Cpf1 (H759)	5'-[..]GCCTCCCTGAAGAAGGAGGAGCTGGTGGTGCACCCAGCCAACTCCCCTA TCGCCAACAAGAAT[...]-3'
As_E993A_FW	5'-CCGTGGTGGTGCTGGCGAACCTGAATTTCGG-3'
As_E993A_RV	5'-CCGAAATTCAGGTTCGCCAGCACCACCACGG-3'
As_H800A_FW	5'-TCCAGGATGAAGAGGATGGCAGCCCGGCTGGGAGAGAAGATG-3'
As_H800A_RV	5'-CATCTTCTCTCCCAGCCGGGCTGCCATCCTCTTCATCCTGGA-3'
As_H_Open_FW	5'-CCGGCTGGGAGAGAAGATG-3'
As_H_Open_Rv	5'-TGCCATCCTCTTCATCCTGGA-3'
As_Nuclease_Seq	5'-CCACCACGGCAAGCCTAATCTG-3'
Fn_E1006A_FW	5'-GCCATTGTGGTGTTCGCGGATCTGAACTTCGGC-3'
Fn_E1006A_RV	5'-GCCGAAGTTCAGATCCGCGAACACCACAATGGC-3'
Fn_H843A_FW	5'-GCAGTCTATTCCTAAGAAAATCACTGCCCCAGCTAAGGAG-3'
Fn_H843A_RV	5'-CTCCTTAGCTGGGGCAGTGATTTTCTTAGGAATAGACTGC-3'
Fn_Nuclease_Seq	5'-CGGAGAGGCCGAACTGTTTTAC-3'
Lb_E925A_FW	5'-CGTGATCGCCCTGGCGGACCTGAACTCTG-3'

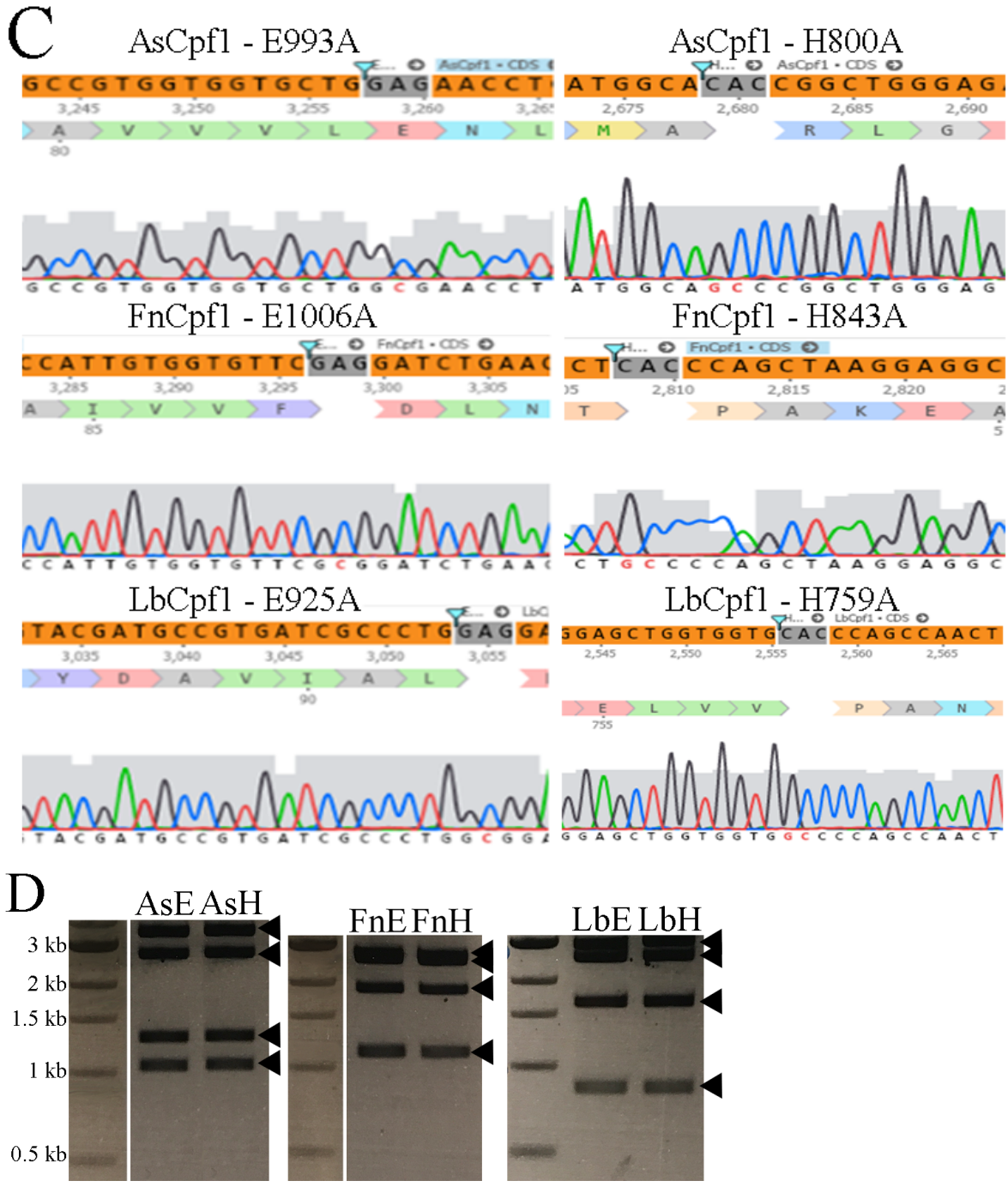


Figure 3. Generation of dCpf1 constructs. A) Schematic of the Cpf1 protein domains (residue numbers for AsCpf1, adapted from Yamano et. al) (21), with mutation locations labeled by red arrows. B) Sequences for As, Fn, and Lb Cpf1 catalytic regions and relevant mutagenic primers/oligonucleotides/sequencing primers. Target codons are underlined, base pair substitutions are bolded. Mutants were generated by QuikChange PCR or by Gibson Assembly and subsequent transformation of and plasmid purification from NEB XL-10 Gold cells. C) Chromatograms/sequences aligned to *in silico* constructs to verify DNase and RNase mutants. D) Mutant plasmid integrity was confirmed by restriction digest. AsE and AsH mutants by PvuI digest, expected band sizes: 3279 bp, 2559 bp, 1280 bp, 1045 bp. FnE and FnH mutants by XmnI digest, expected band sizes: 2604 bp, 2220 bp, 1983 bp, 1335 bp. LbE and LbH mutants by AhdI and BamHI-HF digest, expected band sizes: 2900 bp, 2510 bp, 1603 bp, 853 bp.

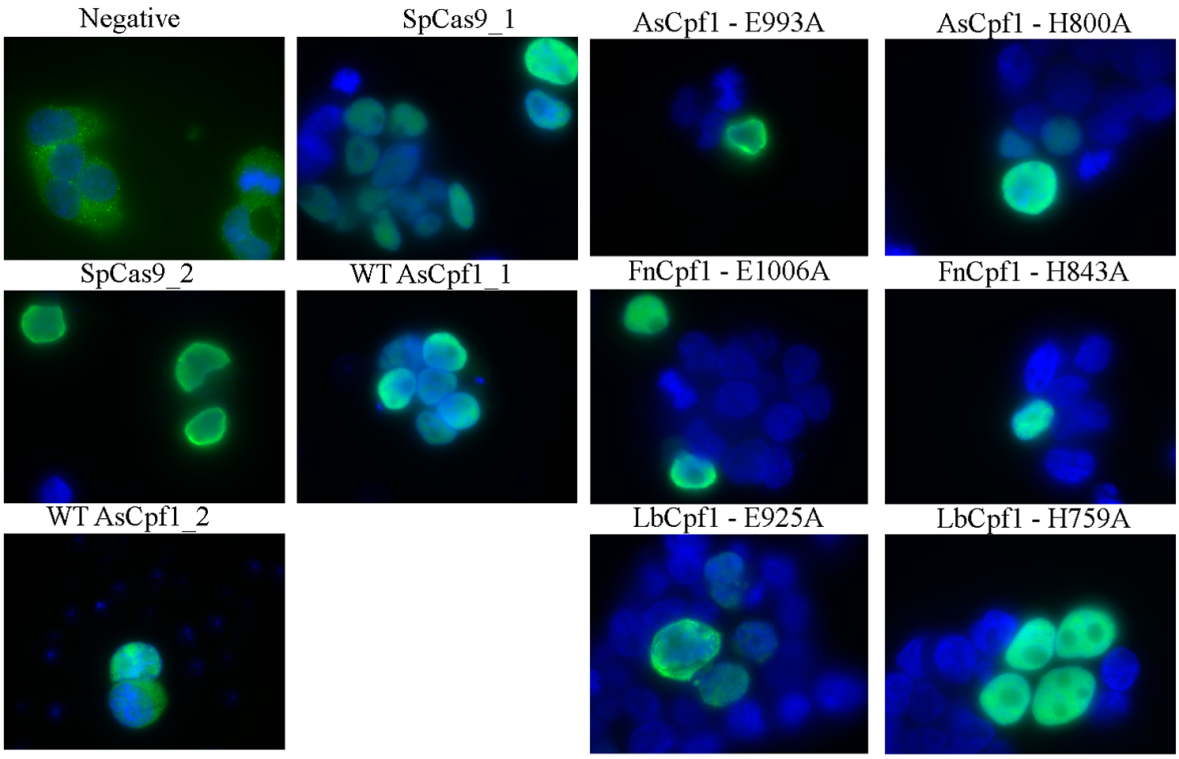
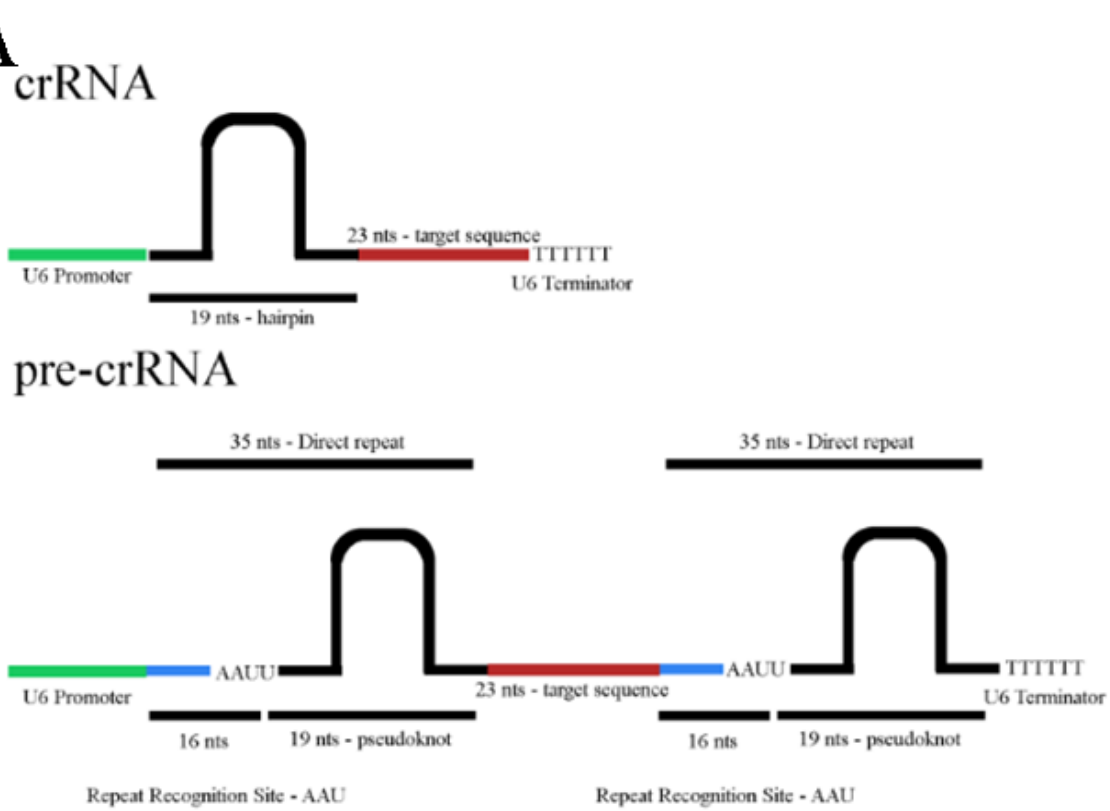


Figure 4. Nuclease-dead Cpf1 mutants localize to the nucleus. As-E993A-Cpf1, As-H800A-Cpf1, Fn-E1006A-Cpf1, Fn-H843A-Cpf1, Lb-E925A-Cpf1, and Lb-H759A-Cpf1 localize to the nucleus in HEK293T cells. Cells were transfected using the Polyfect reagent (Qiagen) to deliver 100 ng of plasmid encoding for each CRISPR enzyme. Images were taken 48h post transfection. The blue channel was used for DAPI staining and the green channel for the secondary, Alexa Fluor 488, anti-anti-HA stain. Untransfected cells were used as a background staining control, and cells treated with SpCas9 were used as a positive control.

A



B

Name	Sequence
DNMT1_target_3	5'-**TTTC**CTGATGGTCCATGTCTGTTACTC-3'
DNMT1_Outer_FW	5'-TCTGACCTCAGCCAGAAGTCCCG-3'
DNMT1_Outer_RV	5'-ACCGTTTTGGGCTCTGGGACTC-3'
DNMT1_Inner_FW	5'-CAGAAACAGGGGTGACGGGAG-3'
DNMT1_Inner_RV	5'-CTGTGAGGATTGAGTGAGTTGCACG-3'
EMX1_target_1	5'-**TTTC**TCATCTGTGCCCCTCCCTCCCTG-3'
EMX1_Outer_FW	5'-GGCTCTCTCATTTACTACTCACATCCAC-3'
EMX1_Outer_RV	5'-AACAAAAGGGAGATTGGAGACACG-3'
EMX1_Inner_FW	5'-TGTCCTCTTCCTGCCCTGCCATC-3'
EMX1_Inner_RV	5'-CAGCCAGCCCATTGCTTGTCCC-3'

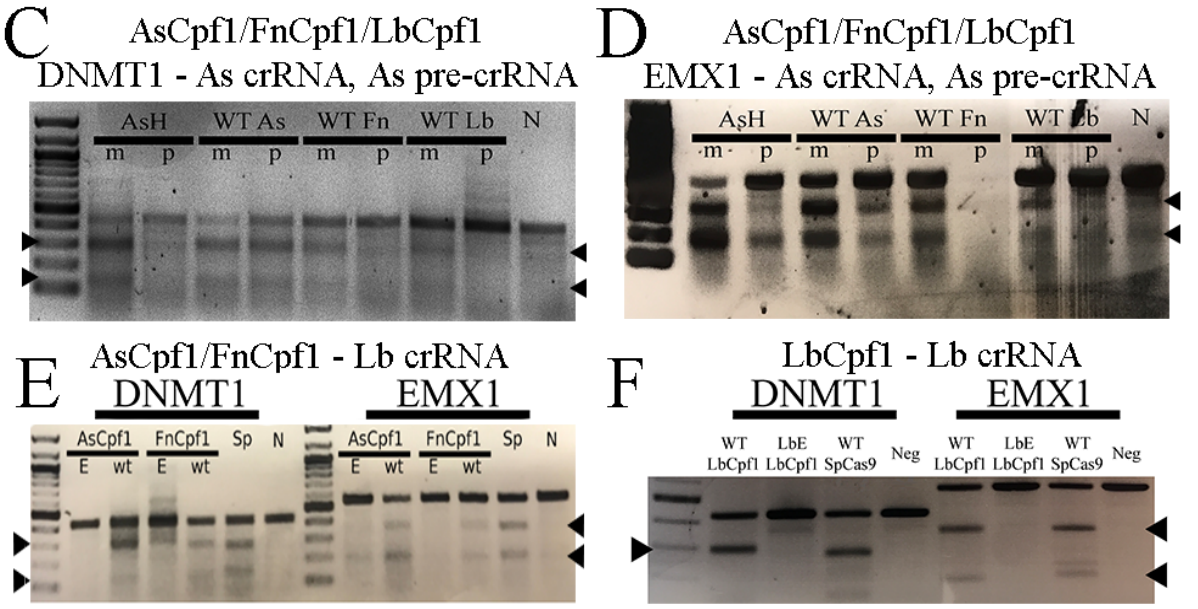
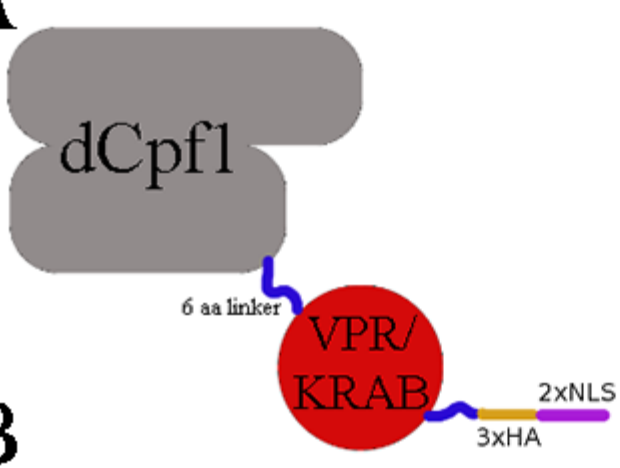


Figure 5. dCpf1 mutants are catalytically inactive. A) Schematic of the crRNAs/pre-crRNAs used in this study. B) DNMT1 and EMX1 primers and target sequences used in this study. TTTN PAMs are underlined and bolded. C) The AsH mutant fails to produce indels by T7 endonuclease 1 assay when delivered with pre-crRNAs, whereas wild type Cpf1 enzymes do, when targeted to DNMT1 and D) EMX1. FnCpf1 and LbCpf1 produce indels with the supplied crRNAs, but not the pre-crRNAs, when compared to controls. Sp is SpCas9, which was used as a positive indel-producing control, with its own sgRNA. N is negative (no enzyme or guide delivered), H is the RNase mutant, wt is the wild type enzyme, p is for samples that received pre-crRNA plasmid, and m is for samples that received the mature crRNA plasmid. Experiment was performed using guides designed for AsCpf1. WT FnCpf1/pre-crRNA-treated sample did not produce sufficient amplicon for T7EI assay. E) The AsE and FnE mutants and the F) LbE mutant are deoxyribonuclease-dead by DNMT1 and EMX1 T7EI assay when compared to controls. E is the DNase mutant. Experiment was performed using guides designed for LbCpf1.

A



B

Name	Sequence
All_Cpf1_Amplif_FW	5'-AAAAGGCCGGCGGCCAC-3'
All_Cpf1_KRAB_Amplif_RV	5'-TTCGTGGCCGCCGGCCTTTTTCCACCGGCGGATACCAGCCAAGGTTCTTC-3'
All_Cpf1_VPR_Amplif_RV	5'-TTCGTGGCCGCCGGCCTTTTTCCACCGGCGGAAAACAGAGATGTGTCGAAGATGGACAGTCC-3'
As_Amplif_RV	5'-GTTGCGCAGCTCCTGGATGTAGG-3'
As_KRAB_Amplif_FW	5'-ACATCCAGGAGCTGCGCAACGGAAGCGCTAGCGGTTCAATGGACGCGAAATCACTTAC-3'
As_VPR_Amplif_FW	5'-ACATCCAGGAGCTGCGCAACGGAAGCGCTAGCGGTTCAAGCGGTTCCGGACGGGCT-3'
Fn_Amplif_RV	5'-GTTATTTCTATTCTGGACAAACTCGAAGTATTCCTC-3'
Fn_KRAB_Amplif_FW	5'-GTTTGTCCAGAATAGAAATAACGGAAGCGCTAGCGGTTCAATGGACGCGAAATCACTTAC-3'
Fn_VPR_Amplif_FW	5'-GTTTGTCCAGAATAGAAATAACGGAAGCGCTAGCGGTTCAAGCGGTTCCGGACGGGCT-3'
Lb_Amplif_RV	5'-GTGCTTCACGCTGGTCTGGGC-3'
Lb_KRAB_Amplif_FW	5'-CCCAGACCAGCGTGAAGCACGGAAGCGCTAGCGGTTCAATGGACGCGAAATCACTTAC-3'
Lb_VPR_Amplif_FW	5'-CCCAGACCAGCGTGAAGCACGGAAGCGCTAGCGGTTCAAGCGGTTCCGGACGGGCT-3'
Lb_End_Seq	5'-ATCGCCATCTCTAACAAGGAG-3'
KRAB_End_Seq	5'-ATGTGATCCTTAGGCTCG-3'
VPR_Beg_Seq	5'-GACCTACGAGACATTCAAG-3'
VPR_Mid_Seq	5'-CTCCAGGAAGTCCATGGGC-3'
VPR_End_Seq	5'-CGAGATTCTGGATACCTTCC-3'

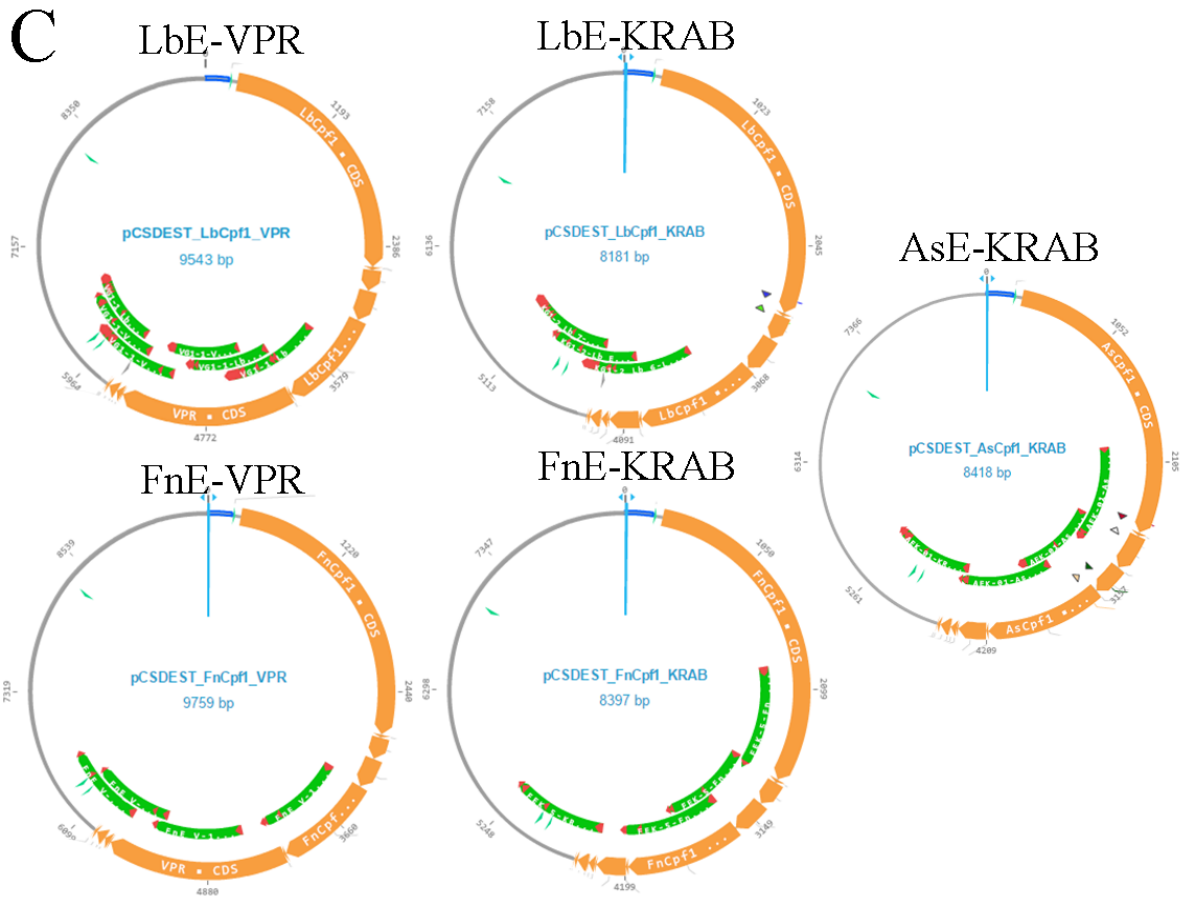
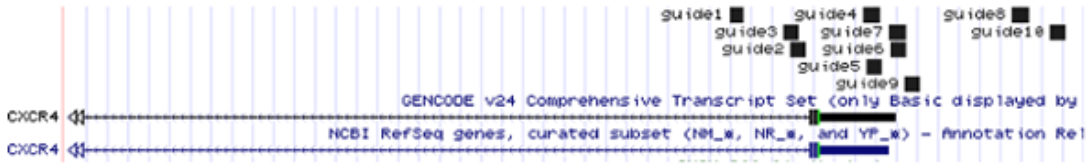


Figure 6. Generation of dCpf1-VPR and dCpf1-KRAB fusion proteins. A) Schematic of fusion protein constructs. Cpf1 is representative of As, Fn, and Lb Cpf1, and the purple region is the linker protein. B) Primers used for Gibson Assembly and sequencing. Regions that overlap with As/Fn/Lb/All Cpf1 are highlighted in green/orange/yellow/purple, regions that overlap with KRAB are highlighted in red, and regions that overlap with VPR are highlighted in blue. Additional sequencing primers can be found in the Supplemental Information section. C) Sequencing results aligned to *in silico* designs to verify fusion construct identity. Green signifies regions of the sequencing results that match the *in silico* constructs. Red signifies regions that do not match. All such red regions were inspected (by chromatograms) and determined to be non-specific reads as the sequencing run began or ended, the induced point mutations for the dCpf1, or were determined to be correct upon inspection of the fluorescence peaks.

A

Name	Sequence
CXCR4_crRNA_1	5'-GAGTACGGGTACCTCCAATG-3'
CXCR4_crRNA_2	5'-TGACAAAGCAGGTTGAAACTGGA-3'
CXCR4_crRNA_3	5'-AACCTGCTTTGTCATAAATGTAC-3'
CXCR4_crRNA_4	5'-TTGGCTGCGGCAGCAGGTAGCAA-3'
CXCR4_crRNA_5	5'-CTACCTGCTGCCGCAGCCAACAA-3'
CXCR4_crRNA_6	5'-TGGCCGCGGCCGGACTTTTATAA-3'
CXCR4_crRNA_7	5'-TATAAAAGTCCGGCCGCGGCCAG-3'
CXCR4_crRNA_8	5'-CGGGTGGTCGGTAGTGAGTCCGG-3'
CXCR4_crRNA_9	5'-ATAAAAACACGCTCCGAGCGCGG-3'
CXCR4_crRNA_10	5'-CTAAGTTTGAGGGAAGCGGGAT-3'
CXCR4_FW	5'-GGCTGCGCTCTAAGTTCAAACG-3'
CXCR4_RV	5'-GAAGACAGGTGGGAAGCGC-3'

B



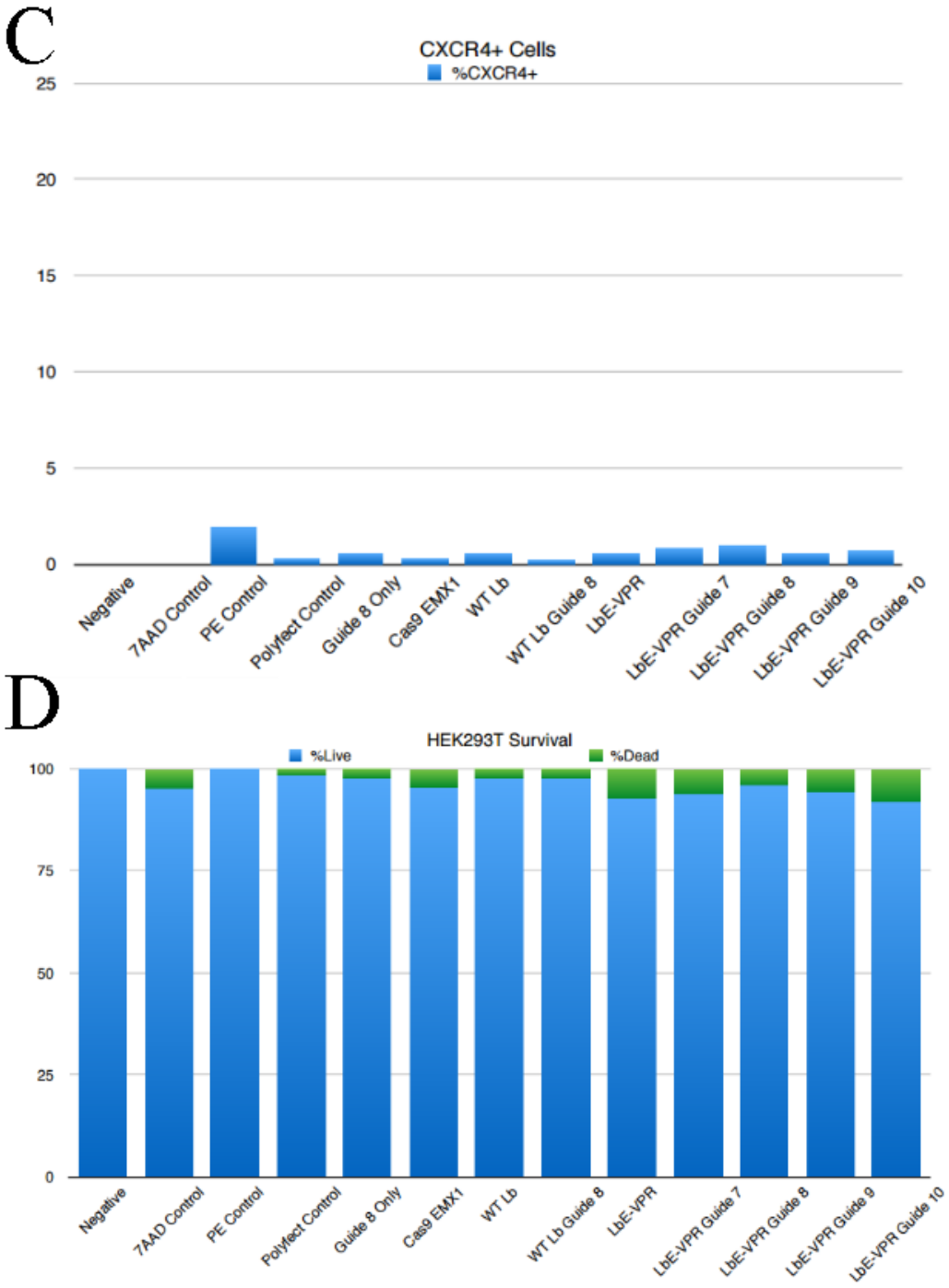


Figure 7. Effect of LbCpf1 constructs on CXCR4 expression. A) CXCR4 guide sequences and PCR primer sequences. Some guides were designed to target the CXCR4 region in the antisense direction. B) CXCR4 guide sequences aligned to the human genome using BLAT. C) CXCR4 positive staining in HEK293Ts treated with nothing (Negative, 7AAD control, PE control), just Polyfect, just Polyfect and CXCR4 crRNA 8, Cas9 and an EMX1 guide RNA, just wild type (WT) LbCpf1, WT LbCpf1 with CXCR4 crRNA 8, LbE-VPR without any crRNAs, and then LbE-VPR with crRNAs 7-10. No significant increase in CXCR4 expression was observed in the treated samples. D) HEK293T survival upon delivery of the same constructs as in panel C. LbCpf1 constructs exhibit some toxicity (~10%) upon delivery.

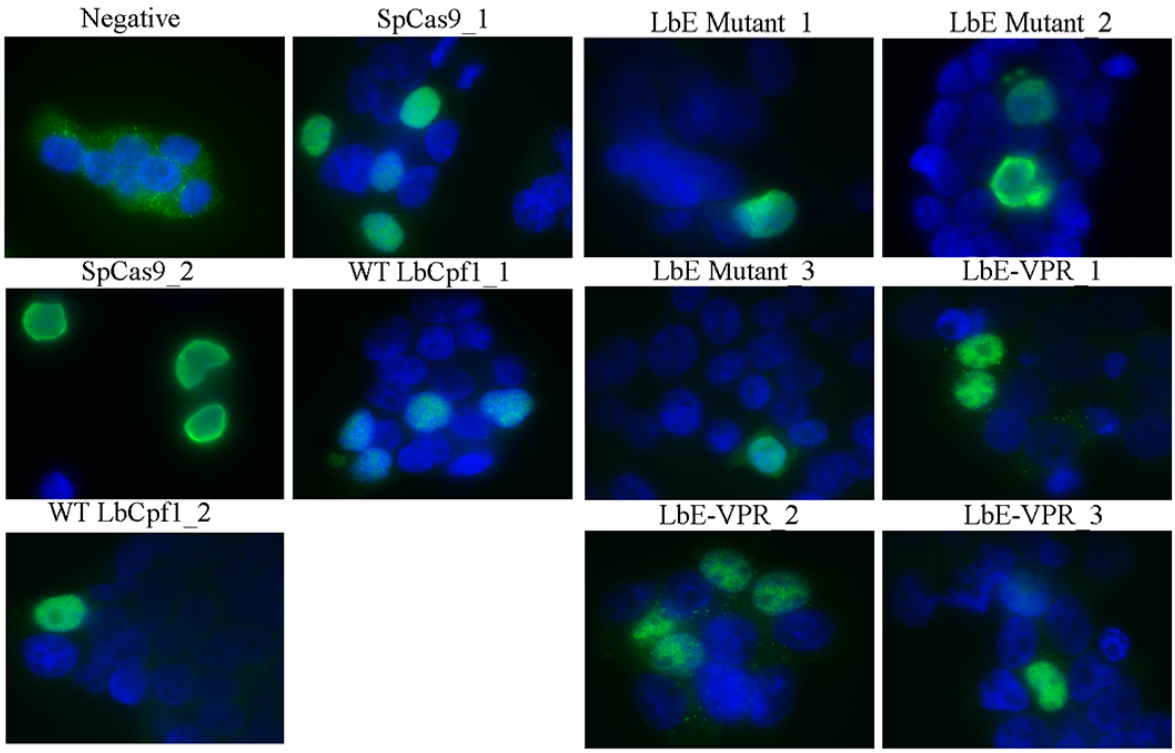


Figure 8. LbE mutant and LbE-VPR fusion construct localize to the nucleus. LbE nuclease-dead mutant and LbE-VPR fusion protein localize to the nucleus when delivered to HEK293T cells with CXCR4 crRNA 8 by transfection. Cells were transfected using the Polyfect reagent (Qiagen) to deliver 100 ng of plasmid encoding for each CRISPR enzyme, as well as 100 ng of crRNA/guide RNA where applicable. Images were taken 48h post transfection. The blue channel was used for DAPI staining and the green channel for the secondary, Alexa Fluor 488, anti-anti-HA stain. Untransfected cells were used as a background staining control, and cells treated with SpCas9 were used as a positive control. LbE mutants appeared less expressed compared to previous experiments, suggesting efficiency of construct delivery is variable.

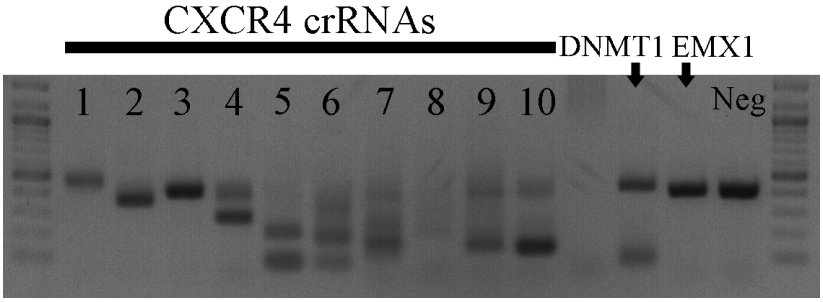


Figure 9. CXCR4 gene amplification by PCR. Representative gel of attempts to amplify the ~480 bp region surrounding the transcription start site of the CXCR4 gene from genomic DNA purified from LbCpf1 and crRNA-treated HEK293T cells. Negative (Neg, non-LbCpf1-treated) genomic DNA produced clean PCR products, whereas LbCpf1 and crRNA-treated samples produced a myriad of amplicons, unsuitable for T7EI analysis. Several PCR programs and primer sets were tried, as well as both Phusion polymerase (NEB) and Q5 polymerase (NEB) with various reaction components. The image shown is from a program that used the primers in Figure 7A, 5 U of Q5 polymerase with 1x Q5 Enhancer in 1x Q5 Reaction Buffer according to the manufacturer's protocol. Reactions were denatured at 98°C for 15s, followed by 35 cycles of 98°C for 10s, 66°C for 20s, and 72°C for 20s. A final extension at 72°C for 5 minutes was followed by ramping to 4°C before storage at -20°C and subsequent analysis by gel electrophoresis. Products of various lengths are present, except for EMX1 and non-treated CXCR4 amplicons.

Supplemental Information

Name	Sequence
As_Seq_1R	5’-CAGGGCGTTCCTTGTCTCC-3’
As_Seq_2	5’-GCCGACCAGTGCCTGCAGCTGG-3’
As_Seq_3	5’-CCCTGCCACACAGATTCATCC-3’
As_Seq_4	5’-GAATAAGGAGAAGAACAATGGC-3’
As_Seq_5	5’-GCTGAACAAGAAGCTGAAGG-3’
As_Seq_6	5’-CCAAGATGGGCACCCAGTCTGG-3’
As_Seq_7	5’-CTACATCCAGGAGCTGCGCAAC-3’
Fn_Seq_1R	5’-GGTTGTCATCGTCGCTCTTC-3’
Fn_Seq_2	5’-CGAGGAAATTCTGAGCTCCG-3’
Fn_Seq_3	5’-GAGAATACCAAGCGCAAGGG-3’
Fn_Seq_4	5’-GGATCTGCTGGACCAGACC-3’
Fn_Seq_5	5’-TGAAAGTTACATCGACAGCG-3’
Fn_Seq_6	5’-CCATGAGATTGCAAAGCTGG-3’
Fn_Seq_7	5’-ACGGAAACTTCTTCGACAGC-3’
Lb_Seq_1R	5’-CTCGGTTCTGGTTTTCTTCC-3’
Lb_Seq_2	5’-CAACGACGTGCTGCACAGC-3’
Lb_Seq_3	5’-TGAACAAGAACAGCGAGATC-3’
Lb_Seq_4	5’-GAAGTACGCCAAGTGCCTGC-3’
Lb_Seq_5	5’-CAAGATCAATACAGAGGTGC-3’
Lb_Seq_6	5’-CCAAGTATACCAGCATCGCCG-3’
Lb_Seq_7	5’-TACCCATACGATGTTCCAG-3’

Supplemental Figure 1. Additional primers used for sequencing of AsCpf1, FnCpf1, and LbCpf1 genes in the pCSDEST vector.

References

1. Gasiunas, G., Barrangou, R., Horvath, P., Siksnys, V. 2012. Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci. 109 (39): E2579-E2586.
2. Ledford, H. 2015. CRISPR, the disruptor. Nature News Feature. 522 (7554). http://www.nature.com/news/crispr-the-disruptor-1.17673
3. 2017. CRISPR - PubMed - NCBI. Web. Retrieved from: https://www.ncbi.nlm.nih.gov/pubmed/?term=CRISPR
4. Hammond, A. et al. 2016. A CRISPR-Cas9 gene drive system targeting female reproduction in the malaria mosquito vector *Anopheles gambiae*. Nat Biotech. 34: 78-83.
5. Nunez, J. et al. 2014. Cas1–Cas2 complex formation mediates spacer acquisition during CRISPR–Cas adaptive immunity. Nat Struc & Mol Bio. 21: 528-534.
6. Stern, A. et al. 2010. Self-targeting by CRISPR: gene regulation or autoimmunity? Trends in Genetics. 26 (8): 335-340.
7. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D., Horvath, P. 2007. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. Science. 315 (5819): 1709-1712.
8. Yosef, I., et al. 2012. Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. Nucleic Acids Res. 40: 5569-5576.
9. Marraffini, L. and Sontheimer, E. 2010. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. Nat Rev Genet. 11 (3): 181-190.
10. Garside, E. et al. 2012. Cas5d processes pre-crRNA and is a member of a larger family of CRISPR RNA endonucleases. RNA. 18: 2020-2028.
11. Sinuknas, T. et al. 2011. Cas3 is a single-stranded DNA nuclease and ATP-dependent helicase in the CRISPR/Cas immune system. EMBO Journal. 30: 1335-1342.
12. Deltcheva, E., Chylinski, K., Sharma, C., Gonzales, K., Chao, Y., Pirzada, Z., Eckert, M., Vogel, J., Charpentier, E. 2011. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature. 471: 602-607.
13. Kowalczykowski, S. 2000. Initiation of genetic recombination and recombination-dependent replication. Trends Biochem Sci. 25 (4): 156-165.
14. Mehta, P. et al. 2004. HNH family subclassification leads to identification of commonality in the His-Me endonuclease superfamily. Protein Sci. 13 (1): 295-300.
15. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., Zhang, F. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science. 339 (6121): 819-823.
16. Wang H., Yang H., Shivalila C., Dawlaty M., Cheng A., Zhang F., Jaenisch R. 2013. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. 153 (4): 910-918.
17. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J., Charpentier, E. 2012. A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 337 (6096): 816-821.

18. Parnas O., Jovanovic M., Eisenhaure T., Herbst R., Dixit A., Ye C., Przybylski D., Platt R., Tirosh I., Sanjana N., Shalem O., Satija R., Raychowdhury R., Mertins P., Carr S., Zhang F., Hacohen N., Regev A. 2015. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. Cell. 162 (3): 675-686.
19. Kearns, N., Genga, R., Enuameh, M., Garber, M., Wolfe, S., Maehr, R. 2014. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development. 141 (1): 219-223.
20. Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., Marraffini, L. 2013. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nuc Acid Res. 41 (15): 7429-7437.
21. Zhang, F., Cong, L. et al. 2011. Programmable Sequence-Specific Transcriptional Regulation of Mammalian Genome Using Designer TAL Effectors. Nat Biotech. 29 (2): 149-153.
22. Zetsche, B. et al. Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. Cell. 163 (3): 759-771.
23. Yamano, T., Nishimasu, H., Zetsche, B., Hirano, H., Slaymaker, I., Li, Y., Fedorova, I., Nakane, T., Makarova, K., Koonin, E., Ishitani, R., Zhang, F., Nureki, O. 2016. Crystal Structure of Cpf1 in Complex with Guide RNA and Target DNA. Cell. 165 (4): 949-962.
24. White, M. 2016. Cpf1 shape-shifts for streamlined CRISPR cleavage. Nat Struc & Mol Bio. 23: 365-366.
25. Fonfara, I., Le Rhun, A., Chylinski, K., Makarova, K., Lecrivain, A.L., Bzdrenga, J., Koonin, E., Charpentier, E. 2014. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. Nuc Acid Res. 42 (4): 2577-2590.
26. Fonfara, I. et al. 2015. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. Nature. 532: 517-521.
27. Zetsche, B. et al. 2017. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. Nat Biotech. 35 (1): 31-34.
28. Tang, X., Lowder, L. et al. 2017. A CRISPR-Cpf1 system for efficient genome editing and transcriptional repression in plants. Nature Plants 3 (Article 17018).
29. Kim, S. et al. 2017. Efficient Transcriptional Gene Repression by Type V-A CRISPR-Cpf1 from *Eubacterium eligens*. ACS Synth. Bio. Just Accepted Manuscript.
30. Moriuchi, M., Moriuchi, H., Turner, W., Fauci, A. 1997. Cloning and analysis of the promoter region of CXCR4, a coreceptor for HIV-1 entry. Journ Imm. 159 (9): 4322–4329.
31. Sun, X., Cheng, G., Hao, M., Zheng, J., Zhou, X., Zhang, J., Taichman, R., Pienta, K., Wang, J. 2010. CXCL12/CXCR4/CXCR7 Chemokine Axis and Cancer Progression. Cancer Metas Rev. 29 (4): 709-722.
32. Balabanian, K., Lagane, B. et al. 2005. The Chemokine SDF-1/CXCL12 Binds to and Signals through the Orphan Receptor RDC1 in T Lymphocytes. Journ. Biol. Chem. 280: 35760-35766.

33. Gilbert, L., Larson, M., Morsut, L., Liu, Z., Brar, G., Torres, S., Stern-Ginossar, N., Brandman, O., Whitehead, E., Doudna, J., Lim, W., Weissman, J., Qi, L. 2013. CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell. 154 (2): 442-451.
34. Chavez, A., Scheiman, J., Vora, S. et al. 2015. Highly efficient Cas9-mediated transcriptional programming. Nature 12 (4): 326-328.
35. Pattanayak, V., Lin, S., Guilinger, J., Ma, E., Doudna, J., Liu, D. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat Biotech. 31 (9): 839-843.
36. Liang, P. et al. 2015. CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. Prot. & Cell. 6 (5): 363-372.
37. Kim, Y. et al. 2016. Generation of knockout mice by Cpf1-mediated gene targeting. Nat Biotech. 34: 808-810.
38. Kleinstiver, B. et al. 2016. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. Nat Biotech. 34: 869-874.
39. Horvath, P. et al. 2010. CRISPR/Cas, the Immune System of Bacteria and Archaea. Science. 327 (5962): 167-170.