

Electronic Thesis and Dissertation Repository

1-31-2017 12:00 AM

Dual-active genome-editing reagents

Jason M. Wolfs
The University of Western Ontario

Supervisor
David Edgell
The University of Western Ontario

Graduate Program in Biochemistry
A thesis submitted in partial fulfillment of the requirements for the degree in Doctor of
Philosophy
© Jason M. Wolfs 2017

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Biochemistry Commons](#)

Recommended Citation

Wolfs, Jason M., "Dual-active genome-editing reagents" (2017). *Electronic Thesis and Dissertation Repository*. 4546.
<https://ir.lib.uwo.ca/etd/4546>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

ABSTRACT

Manipulation of complex genomes has many beneficial downstream applications in agriculture and human gene therapy. Precise genome-editing requires the introduction of a specific DNA double-strand break at a locus of interest, in turn inducing host DNA repair pathways to cause gene knockout through non-homologous end-joining or gene repair using homologous recombination and donor template. No matter the application, the field has depended on a few reagents to introduce precise double-strand breaks in host genomes. LAGLIDADG homing endonucleases or meganucleases harness the natural properties of these rare-cutting enzymes to target precise sequences in a complex genome. Other successful reagents are derived from a type IIS restriction endonuclease, FokI, fused to various DNA-binding architectures such as zinc finger domains and transcription activator-like effector domains. However, the discovery of clustered regularly interspaced short palindromic repeat-associated protein, CRISPR-Cas9, has dominated the field with its ease of design requiring a single RNA molecule to target the sequence of interest. Even with a handful of reagents to choose from, no one reagent is suitable for every application as every reagent has its own set of limitations and advantages.

Here we present another potential genome-editing reagent derived from a GIY-YIG homing endonuclease, I-TevI, fused to all four DNA-targeting proteins described above. First, we demonstrate that I-TevI is a portable nuclease domain that can be targeted using Zinc-Fingers and LAGLIDADG proteins. Using these new reagents, we were able to further characterize I-TevI specificity using high throughput *in vitro* and *in vivo* screens to highlight important sequence requirements for targeting. Using this knowledge, we systematically engineered new I-TevI variants with altered specificity to broaden the number of targets available for I-TevI-derived reagents. We incorporated these new I-TevI variants into a more versatile dual-active nuclease, TevCas9, capable of introducing two double-strand breaks at a single target site. This dual cleavage event is capable of excising a short DNA fragment from the target site and is unique to I-TevI derived fusions. We envisioned that the monomeric, sequence-specific I-TevI catalytic domain would improve current tools by providing additional specificity and the ability to introduce dual-cleavage event would present unique applications for genome engineering.

KEYWORDS

GIY-YIG homing endonuclease, LAGLIDADG homing endonuclease, Meganucleases, I-TevI, I-OnuI, double-strand break, genome-editing, clustered regularly interspaced short palindromic repeats (CRISPR), dual-active endonuclease

CO-AUTHORSHIP

For the peer-reviewed publications presented in Chapters 2, 3, 4 and 5, Jason Wolfs performed the research with exceptions noted below. Jason Wolfs and David Edgell conceived and designed the experiments, analyzed the data and David Edgell wrote a majority of the manuscripts.

Chapter 2 – Ben Kleinstiver designed and tested I-TevI/Zinc-finger chimeras. Alanna Roberts, Sherry Hu and Tomasz Kolaczyk helped perform the *in vivo* bacterial activity assays. Ben Kleinstiver performed *in vitro* Tev-ZFE assays.

Chapter 3 – Matthew DaSilva and Sarah Meister helped profile the 64-nucleotide combinations in the I-TevI cleavage motif in bacteria and yeast, respectively. Xu Wang was instrumental in helping set up HEK 293 assays to test dual-active MegaTev constructs in mammalian cells.

Chapter 4- Marcon Laforet helped profile MegaTev cleavage specificity for altered G8 I-TevI anchor position. Jenny Zhang helped characterize I-TevI variants in bacteria.

Chapter 5- Thomas Hamilton helped purify TevCas9 and perform *in vitro* cleavage assays. Jeremy Lant help perform T7E1 assays for TevCas9 NQO2_54 off-target sites. Jenny Zhang helped characterize I-TevI variants in HEK 293 cells. Lousia Salemi was instrumental in HEK 293 mammalian assays for TevCas9. Gregory Gloor and David Edgell analyzed next-generation reads for TevCas9 and *in vitro* MegaTev spacer region profile.

DEDICATION

This thesis is dedicated to all my lab-mates, mentors, and family who
have provided endless guidance to help achieve
my dreams and aspirations

Occam's Razor:

“If you have two equally likely solutions to a problem, choose the simplest.”

ACKNOWLEDGMENTS

First of all, I have to honour my supervisor, Dr. David Edgell, for his superior guidance and mentorship as I would not be where I am today without him. The sarcasm can be intimidating at first but one learns quickly that Dave is very kind-hearted and always available to help. I am lucky to have Dr. David Edgell as a supervisor to provide great insight into my work over the years.

Thank you to my advisory committee, Dr. David Haniford and Dr. Gregory Gloor, for helpful discussions about my graduate research. To past and present members of the Edgell lab that have helped me stay motivated over the years. Particularly, to my mentor Dr. Benjamin Kleinstiver who took me under his wing and projected me to where I am today. Starting in the Edgell lab was extremely overwhelming, trying to keep up with Ben, but I am extremely grateful for everything he taught me and for the many publications I was able to be a part of. Dr. Benjamin Kleinstiver was an amazing role model that drove me to work harder each and every day, I could not have been as successful without his guidance and mentorship. Thank you to Tom McMurrugh for being my lab-mate for life, for your helpful discussion and of course your endless comic relief throughout my graduate career. I wish you all the best to the future CEO Dr. McMurrugh, paving the direction of a big biotech or investment firm. To all other Edgell lab members, thank you for helping me manoeuvre down this winding road they call a PhD.

Finally, I would like to acknowledge my friends and family for supporting me every step of the way. Your endless support and love helped me through this rollercoaster ride called life. Thank you Kavita for being there through this whole journey and convincing me to get two loving dogs that brought a smile to my face even on the toughest days. Thank you everyone in helping me complete this achievement and I look forward to many more years to come.

TABLE OF CONTENTS

CERTIFICATE OF EXAMINATION.....	ii
ABSTRACT.....	iii
KEYWORDS.....	iv
CO-AUTHORSHIP.....	v
DEDICATION.....	vi
ACKNOWLEDGMENTS.....	vii
TABLE OF CONTENTS.....	viii
LIST OF TABLES.....	xiii
LIST OF FIGURES.....	xiv
LIST OF APPENDICES.....	xvi
LIST OF ABBREVIATIONS AND SYMBOLS.....	xvii
CHAPTER 1: INTRODUCTION.....	1
1.1 Homing endonucleases.....	2
1.1.1 Mobility and evolution.....	4
1.1.2 GIY-YIG homing endonucleases.....	5
1.1.3 LAGLIDADG homing endonucleases.....	7
1.2 DNA Sequence specificity for site-specific proteins.....	9
1.2.1 Restriction endonuclease sequence specificity.....	9
1.2.2 Homing endonuclease sequence specificity.....	10
1.2.3 ZF and TALE DNA-binding sequence specificity.....	11
1.2.4 CRISPR-Cas9 sequence specificity.....	12
1.3 Genome editing.....	13
1.3.1 Engineered LAGLIDADG homing endonucleases.....	15

1.3.2	Zinc-finger and TALE nucleases.....	17
1.3.3	CRISPR-Cas9 nucleases.....	18
1.4	Scope of the thesis: dual-active nucleases utilizing the GIY-YIG I-TevI nuclease domain for genome engineering.....	20
1.5	References.....	22

CHAPTER 2: MONOMERIC SITE-SPECIFIC NUCLEASE FOR GENOME

EDITING.....	32	
2.1	Introduction.....	32
2.2	Material and methods.....	33
2.3	Results.....	34
2.3.1	Construction and validation of GIY-zinc finger endonucleases.....	34
2.3.2	Tev-ZFEs function as monomers to cleave at a specific sequence.....	36
2.3.3	The Tev-ZFEs function in a yeast-based recombination assay.....	39
2.3.4	The I-TevI nuclease domain is portable to the LAGLIDADG architecture.....	39
2.3.5	A 5'-CNNNG-3' cleavage motif is not limiting for targeting.....	41
2.4	Discussion.....	43
2.5	References.....	46

CHAPTER 3: MEGATEVS: SINGLE-CHAIN DUAL NUCLEASES FOR EFFICIENT GENE

DISRUPTION	50	
3.1	Introduction.....	50
3.2	Material and Methods.....	51
3.2.1	Bacterial strains and plasmid construction.....	52
3.2.2	<i>In vitro</i> randomized substrate selection.....	52
3.2.3	Cleavage assays on radiolabeled substrate.....	53
3.2.4	Modified two-plasmid target site screen.....	53
3.2.5	Yeast β -galactosidase repair assay.....	54

3.2.6	DdeI-resistance assays with plasmid substrates.....	54
3.2.7	Surveyor assays with integrated targets.....	54
3.3	Results.....	55
3.3.1	MegaTevs: chimeric fusions of GIY-YIG and meganuclease components.....	55
3.3.2	Dual active site MegaTevs for highly efficient targeted deletions.....	57
3.3.3	Requirements of the I-TevI CNNNG cleavage motif <i>in vivo</i>	61
3.3.4	MegaTevs selects for the appropriately spaced cleavage motif from a random substrate.....	63
3.3.5	Nucleotide preference within the CNNNG motif and at flanking positions.....	65
3.3.6	Testing for off-target cleavage.....	67
3.4	Discussion.....	69
3.5	References.....	70

CHAPTER 4: BIASING GENOME-EDITING EVENTS TOWARDS PRECISE LENGTH DELETIONS

	WITH AN RNA-GUIDED TEVCAS9 DUAL NUCLEASE (PART1: I-TEVI VARIANTS).....	76
4.1	Introduction.....	76
4.2	Material and Methods.....	77
4.2.1	Bacterial strains and plasmid construction.....	77
4.2.2	<i>In vitro</i> enrichment assay.....	79
4.2.3	Bacterial two-plasmid selection.....	79
4.3	Results.....	80
4.3.1	Identifying and validating I-TevI nucleotide preference in the DNA spacer.....	80
4.3.2	Position G8 of the DNA spacer acts as an anchor for the linker.....	82
4.3.3	Isolation of I-TevI linker variants that can cleave substrates with substitutions in critical positions.....	84
4.3.4	I-TevI linker variants cleave target sites derived from human genes.....	86
4.4	Discussion.....	88
4.5	References.....	92

CHAPTER 5: BIASING GENOME-EDITING EVENTS TOWARDS PRECISE LENGTH DELETIONS WITH AN RNA-GUIDED TEVCAS9 DUAL NUCLEASE (PART2: TEVCAS9 FUSIONS)	96
5.1 Introduction	96
5.2 Material and Methods	97
5.2.1 Bacterial strains and plasmid construction	97
5.2.2 Protein purification	97
5.2.3 gRNA visualization	98
5.2.4 TevCas9 endonuclease assays	98
5.2.5 HEK293 transfections and genomic extractions	98
5.2.6 T7 endonuclease I mismatch repair assays and restriction endonuclease assays	99
5.2.7 Binding models for MegaTev and TevCas9	99
5.2.8 Next generation DNA sequencing	99
5.3 Results	100
5.3.1 Fusion of I-TevI to Cas9 creates an RNA-guided dual nuclease	100
5.3.2 TevCas9 functions robustly in HEK 293 cells	102
5.3.3 TevCas9 generates defined length deletions in HEK 293 cells	104
5.3.4 Deep sequencing reveals minimal processing of TevCas9 deletions	104
5.3.5 Positioning of the I-TevI cleavage motif and gRNA biases in-frame to out- of-frame deletions	106
5.4 Discussion	108
5.5 References	111
CHAPTER 6: DISCUSSION	116
6.1 Considering off-target cleavage of genome editing reagents	116
6.2 Implication of I-TevI biology for genome engineering	120
6.3 Potential for dual-active GIY-YIG endonucleases	123
6.4 References	125
APPENDICES	129
CURRICULUM VITAE	182

LIST OF TABLES

Table 1.1: Properties of genome-editing reagents.....19

LIST OF FIGURES

Figure 1.1: Mobility of homing endonucleases.....	3
Figure 1.2: Structures for GIY-YIG and LAGLIDADG homing endonucleases.....	6
Figure 1.3: Crystal structures of various DNA binding domains and nucleases.....	8
Figure 1.4: Schematic of two outcomes of genome-editing.....	14
Figure 1.5: Schematic of various genome-editing reagents.....	16
Figure 2.1: Design and functionality of Tev-ZFEs.....	35
Figure 2.2: TevN201-ZFE is a monomer with a preferred cleavage site.	37
Figure 2.3: Tev-ZFEs can induce recombination in a eukaryotic system.....	40
Figure 2.4. Design and functionality of Tev-LHEs.....	42
Figure 2.5: Cleavage requirements do not limit GIY-ZFE and GIY-LHE applicability.....	44
Figure 3.1: MegaTev activity in yeast and HEK 293T cells.....	56
Figure 3.2: Dual active site MegaTev activity <i>in vitro</i>	58
Figure 3.3: Dual active site MegaTev activity in HEK 293 cells.....	60
Figure 3.4: MegaTev nucleotide preference within the CNNNG cleavage motif.....	62
Figure 3.5: MegaTev selects for an appropriately spaced cleavage motif.....	64
Figure 3.6: Nucleotide preference within the N8 randomized library.....	66
Figure 3.7: Surveyor assay for off-target cleavage.....	68
Figure 4.1: Schematic of modular genome-editing nucleases based on the I-TevI nuclease and linker domains.	78
Figure 4.2: Determination of I-TevI nucleotide preference in the DNA spacer.....	81
Figure 4.3: Determination of I-TevI distance preference from G8 position in the DNA spacer...82	
Figure 4.4: Identification and characterization of I-TevI linker variants with altered DNA preferences.	85
Figure 4.5: I-TevI linker variants cleavage activity for various human gene sequences.	87

Figure 4.6: I-TevI linker variants function in the context of Tev-mTALENs.	89
Figure 4.7: Overview of TevCas9 targeting model.....	91
Figure 5.1: Purification and characterization of a TevCas9 dual nuclease.....	101
Figure 5.2: Activity and specificity of TevCas9 in HEK 293 cells.....	103
Figure 5.3: TevCas9 generates deletions of precise lengths in HEK 293 cells.....	105
Figure 5.4: TevCas9 can bias the proportion of in-frame to out-of-frame indels.....	107
Figure 5.5: Model of how TevCas9 biases DNA repair outcomes.....	109
Figure 6.1: Schematic of off-target screening techniques using next-generation sequencing...	118
Figure 6.2: Considerations when targeting I-TevI derived nucleases.....	122
Figure 6.3: Potential genome engineering outcomes using TevCas9 dual-active nucleases.....	124

LIST OF APPENDICES

Appendix S1: Copyright permissions for Chapters 2-5	129
S1.2 Permission for Chapter 2.....	129
S1.3 Permission for Chapter 3.....	130
S1.4 Permission for Chapter 4 and 5.....	131
Appendix S2: Supplemental information for Chapter 2	132
S2 Supplemental material and methods.....	132
S2 Supplemental figures.....	136
S2 Supplemental tables.....	140
S2 Supplemental references.....	156
Appendix S3: Supplemental information for Chapter 3	157
S3 Supplemental figures.....	157
S3: Supplemental Tables.....	161
Appendix S4: Supplemental information for Chapter 4 and 5	167
S4 Supplemental Figures.....	167
S4 Supplemental Tables.....	173

LIST OF ABBREVIATIONS AND SYMBOLS

aa	amino acid
Bmo-ZFE	I-Bmol/zinc-finger fusion endonuclease
bp	base pair
Cas	CRISPR-associated protein
CRISPR	clustered regularly interspaced short palindromic repeat-associated protein
CS	cleavage site
DNA	deoxyribonucleic acid
DSB	double-strand break
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
GIY-HE	GIY-YIG homing endonuclease
GIY-LHE	GIY-YIG/LAGLIDADG homing endonuclease fusion
GIY-ZFE	GIY-YIG zinc-finger endonuclease
HDR	Homology-directed repair
HE	homing endonuclease
HEK293	human embryonic kidney
HR	homologous recombination
IPTG	isopropyl β -D-1-thiogalactopyranoside
KD	dissociation constant
kDa	kilodalton
LB	Luria-Bertani
LHE	LAGLIDADG homing endonuclease
MegaTAL	LAGLIDADG homing endonuclease/ transcription activator-like effector fusion
MegaTev	Tev/LAGLIDADG homing endonuclease (meganuclease) fusion
MgCl ₂	magnesium chloride
min	minute
N	nucleotide
NaCl	sodium chloride
NHEJ	nonhomologous end joining
NMR	nuclear magnetic resonance
nt	nucleotide
ONPG	ortho-nitrophenyl- β -galactoside
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB ID	protein data bank identity
pKa	acid dissociation constant
PSSM	position-specific scoring matrix
PWM	position weight matrix

R	purine
RE	restriction endonuclease
RNA	ribonucleic acid
RVD	repeat-variable diresidue
s	second
SDS	sodium dodecyl sulfate
T7E1	T7 endonuclease 1
TAE	tris base, acetic acid, and EDTA buffer
TALE	transcription activator-like effector
TALEN	TAL effector nuclease
TC-R27A	R27A TevCas9
TC-V	V117F TevCas9
TC-VK	V117F/K135N TevCas9
TC-VKN	V117F/K135R/N140S TevCas9
td	thymidylate synthase allele of bacteriophage T4
Tev-LHE	Tev LAGLIDADG homing endonuclease
Tev-ZFE	Tev zinc-finger endonuclease
T _m	melting temperature
tris-HCl	tris(hydroxymethyl)aminomethane hydrochloride
WT	wild-type
YPD	yeast extract peptone dextrose
ZF	zinc finger
ZFN	zinc-finger nuclease
ω	omega factor

Chapter 1

1 Introduction

Macromolecules have evolved to recognize DNA and perform a multitude of functions from DNA organization, DNA repair, gene expression and DNA hydrolysis (1–7). These processes require macromolecules, such as RNA and proteins, to recognize structural and biochemical features of DNA and bind either non-specifically through phosphate-backbone contacts or specifically through hydrogen bonding interactions (8–12). The requirement for non-specific or sequence-specific contacts is highly role dependent. Histones, a conserved chromosomal packing protein in eukaryotic cells, are examples of proteins that have evolved to bind DNA non-specifically, recognizing primarily the phosphate backbone to help package the large DNA molecule into the nucleus via nucleosomes (9). Other proteins, such as a majority of transcription factors and endonucleases, have evolved to locate specific sequences for gene regulation and DNA hydrolysis, respectively. Whatever the function, there has always been the intriguing question how these proteins locate their specific site in complex genomes (13). Some groups have tried to decipher the amino acid/nucleic acid binding code is but with a large variety of interacting domain structures, determining a simplistic code appears to be a difficult task (5, 8). Better understanding of protein specificity could benefit the genome engineering field greatly, by aiding in engineering novel reagents targeted to new sequences (14–16). Understanding specificity of the LAGLIDADG homing endonuclease (LHE) family (alternatively known as Meganucleases) has been pivotal in engineering new endonucleases with alternative target sites (17–19). The focus of this thesis is to characterize and decipher the DNA recognition of a GIY-YIG homing endonuclease (GIY-HE), I-TevI, and adapt it to novel DNA targets for genome engineering applications.

Using a well-studied GIY-HE, I-TevI, I sought to fuse the nuclease and linker domain to various DNA-binding platforms and demonstrate their genome-editing potential. To first demonstrate that I-TevI nuclease and linker domain were portable to various DNA-binding platforms, we created I-TevI fusion proteins to zinc-finger (ZF) DNA-binding domains and LHEs to create GIY-YIG zinc-finger endonucleases (GIY-ZFE) and GIY-YIG/LAGLIDADG homing endonucleases (GIY-LHE, or alternatively, MegaTev). As described in Chapter 2, we investigate I-TevI as a portable nuclease domain and optimize I-TevI fusions using *in vitro* biochemical assays and *in vivo* bacterial and yeast selection assays. In Chapter 3, we examine a potential unique application for MegaTev, a

fusion of two nuclease domains capable of creating two double-strand breaks (DSBs) at a single target site. Using biochemical and mammalian cell assays, we investigate if MegaTev could help escape a persistent cleavage and re-ligation cycling event resulting from unproductive modifications at the target site. This cyclic cleavage event has been described previously for single cutting genome-editing reagents (20). In addition, we probed the cleavage site specificity of the I-TevI nuclease domain in an attempt to define the activity for all 64-nucleotide combinations in its 5 base pair (bp) 5'-CNNNG-3' cleavage motif. Our next goal was to develop a more versatile I-TevI genome-editing reagent, as described in Chapter 4 and 5. We deciphered targeting guidelines for I-TevI nuclease and linker domains and applied these rules to identify novel I-TevI target sites. To aid in increasing applicability of I-TevI derived reagents, we created fusions to a fourth DNA-binding platform, the clustered regularly interspaced short palindromic repeat-associated protein CRISPR-Cas9 protein. By understanding the targeting requirements of the GIY-HE, I-TevI, we hope to transform this HE into a malleable genome-editing reagent that can be targeted to various human gene targets.

1.1 Homing endonucleases

HEs are considered selfish genetic elements that have evolved for the sole purpose of propagating their own DNA. HEs are found in all three domains of life. Although viewed as selfish elements, there is some evidence of providing beneficial properties to the host (21, 22). One LHE in particular, HO, is an example of a beneficial HE as it is responsible for mating-type switch in *Saccharomyces cerevisiae* (23, 24). Another HE first described in the 1980s was originally known as the genetic element ' ω '. It was present in some *Saccharomyces cerevisiae* strains and was observed to transfer when a ω^- strain was crossed with a ω^+ strain (25, 26). Not too long after, it was discovered that a LHE known as I-SceI was responsible for this observation. The transfer of the genetic information by a HE was termed "homing" process (27, 28). The homing process begins with the introduction of a double-strand break (DSB) by an HE, subsequently inducing homologous recombination (HR) to repair the break utilizing the HE open reading frame (ORF) (Figure. 1.1) (27–29). Therefore, homing endonuclease specificity is an important aspect to ensure successful gene transfer and to prevent toxicity to the host organism. To ensure the correct sequence is cleaved, homing endonucleases have evolved to recognize long

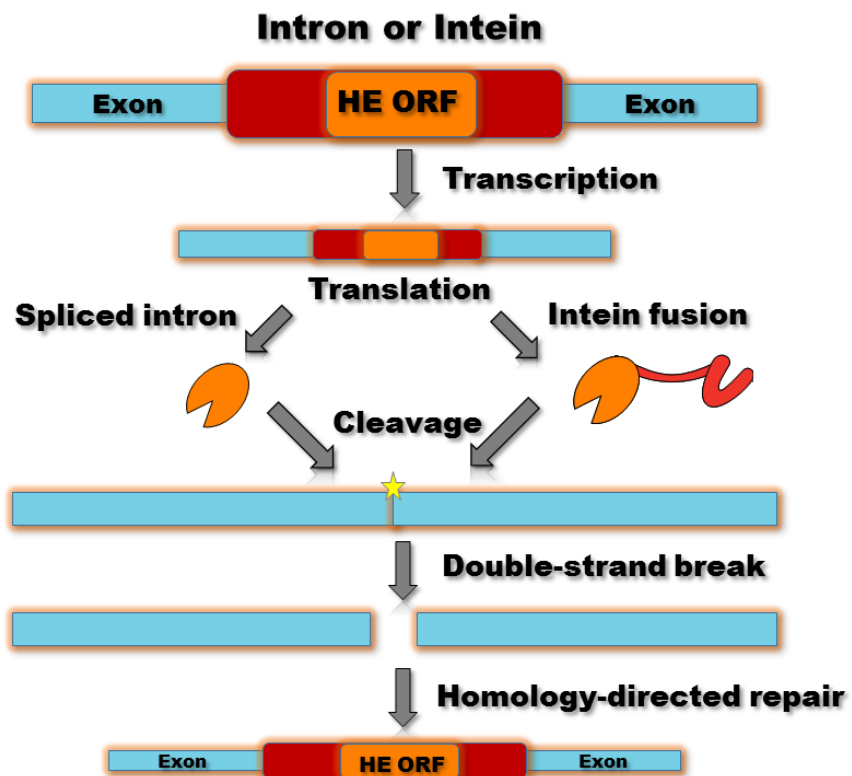


Figure 1.1: Mobility of homing endonucleases

Homing endonuclease expression from an intron or intein results in double-strand break formation at an intronless sequence resulting in lateral gene transfer through homologous recombination. The HE ORF DNA is used as the repair template to propagate the HE coding sequence.

sequences of DNA, from 12-40 bp (30–34). There are five distinct families of homing endonucleases characterized by their conserved nuclease active-site core motifs known as LAGLIDADG, GIY-YIG, HNH, HIS-Cys box and PD-(D/E)XK. For the purpose of this thesis, our focus is on the largest family, the LAGLIDADG or Meganuclease homing endonucleases, and the GIY-HEs.

1.1.1 Mobility and evolution

HEs ORFs occur in many different forms including introns, inteins and freestanding. Once translated HEs scan for their homing site to introduce a DSB and promote their own proliferation. As with group I introns and inteins, the transfer of biological information is initiated solely by the endonuclease, whether as a free or fusion protein from an intron- or intein-encoded sequence, respectively (Figure 1.1) (7, 27, 29, 35). The cleavage of the intron-less homing site is then converted through host repair pathways to an intron or intein containing allele with the HE ORF (Figure 1.1). HEs and their hosts possess a very dynamic relationship assures that they rapidly spread through a population. However, once this occurs there is no selective pressure to keep a functional HE, so the intron/ORF is lost. This results in an increase of intron-less sites available for cleavage and reactivates the HE for site re-invasion (36).

Although HEs target long sequences of DNA in their host genome, they must maintain a certain level of sequence tolerance or malleability to adapt to mutations at their homing sites. The extra flexibility in sequence recognition may also aid in targeting new sites in a gene duplication event. As proposed for the LHE family, single-motif enzymes form dimers bringing two “LAGLIDADG motifs” together to form an active protein against pseudopalindromic targets (37). Interestingly, it is proposed that double-motif monomers evolved from a duplication event, allowing further relaxation of sequence requirements for symmetry, expanding the substrate repertoire (19, 38, 39). Evolution of GIY- HEs is thought to be distinct from LHEs, because the GIY-HE DNA-binding domain is distant from the catalytic domain, therefore the catalytic domain is more promiscuous and can utilize a large variety of DNA-binding proteins (40, 41). As part of this thesis, we expand on this hypothesis for the GIY-HE I-TevI, by demonstrating its ability to use various DNA-binding platforms to target novel DNA sequences.

1.1.2 GIY-YIG homing endonuclease

The GIY- HE family is a well characterized HE family defined by their conserved GIY-X(10-11)-YIG motif. The conserved GIY-YIG motif is involved in folding and catalytic activity of the N-terminal catalytic domain. One of the best-studied GIY-YIG HEs, I-TevI, was found in the thymidylate synthase (*td*) gene of bacteriophage T4 and has been used to studies GIY-YIG HE structure and function. Limited proteolysis and footprinting analysis first demonstrated I-TevI was a bipartite enzyme, where the N-terminal catalytic domain is connected by a flexible linker region to the C-terminal DNA-binding domain (Figure 1.2) (40). NMR and X-ray crystallography studies were used to solve the structure of 1-96 amino acids (aa) of the catalytic domain and 148-245 aa of the C-terminal DNA-binding domain in complex with DNA provided a closer look at I-TevI structure (42–45). The catalytic domain consisted of a globular structure with the conserved GIY-YIG motif located in the first two β -strands of the core $\beta\beta\alpha\alpha\beta\alpha$ fold (46, 47). The catalytic domain was crystalized in the absence of DNA substrate, presumably due to the weak binding affinity of the catalytic domain for DNA (46). Surprisingly, the C-terminal DNA-binding domain contained a non-canonical zinc finger domain from residues 148-169 (42). Previous studies have demonstrated that the zinc finger region acts as a molecular ruler in determining the distance for I-TevI cleavage and does not make any specific DNA contacts (42, 46). Interestingly, the zinc finger region was required for our shortest, most active I-TevI fusions containing aa 1-169 suggesting it must be important for DSB formation (48).

I-TevI recognizes a relatively long target site of about ~40 bp, making mostly minor groove and phosphate backbone contacts along the full length of its homing site (49, 50). Mutagenesis and chemical interference studies have demonstrated that I-TevI can tolerate high abundance of nucleotide substitutions at its homing site (49, 50). Mapping the cleavage site for I-TevI identified a dinucleotide requirement, 5'-CNNNG-3', within the 5 bp cleavage motif (51). Not surprisingly, I-TevI has evolved to recognize conserved codons found within the thymidylate synthase gene active site such as the C and G nucleotide in the cleavage motif of the conserved glutamine (CAR) and arginine (CGN or AGR) codons (51). Probing the specificity of the I-TevI linker region resulted in similar findings, as described in Chapter 4. Previous studies have shown that I-TevI induces a ~40° bend toward the major groove and a similar homing site distortion has been reported for the LHEs and restriction

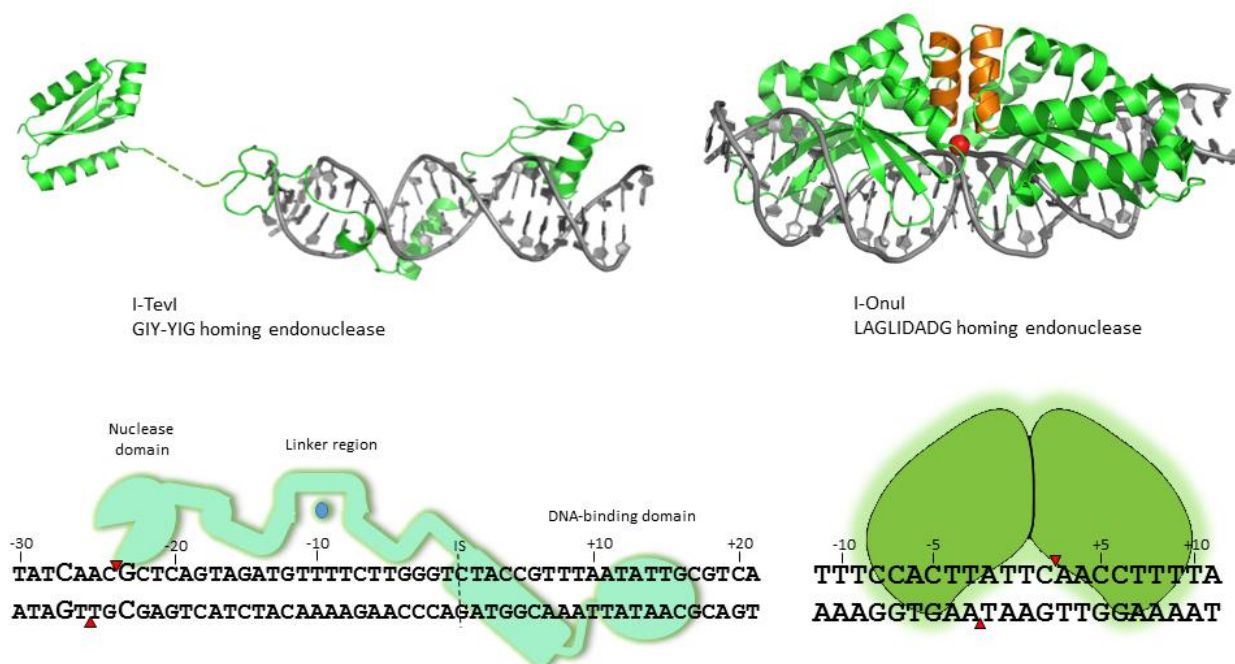


Figure 1.2: Structures for GIY-YIG and LAGLIDADG homing endonucleases

Crystal structures and schematics for the GIY-YIG homing endonuclease I-TevI and LAGLIDADG homing endonuclease I-OnuI. LAGLIDADG alpha-helices critical for the two domain interactions and forming the active sites are highlighted in orange. Schematic representations with red triangles indicating cleavage sites.

endonucleases (REs) (30, 38, 49). The DNA distortion by homing endonucleases is thought to help expose the phosphate backbone to catalytic residues, helping I-TevI sequentially nick the bottom and top strands (30). Biochemical assays have demonstrated that I-TevI binds DNA as a monomer with a single active site to create a DSB (30). To better study the mechanism of GIY-YIG HEs, an alternative protein, I-BmoI, with ~750-fold reduced specific activity than I-TevI was utilized (52). I-BmoI targets a homologous thymidylate synthase gene from *Bacillus mojavensis* containing a similar bipartite structure as I-TevI (53). Both enzymes sequentially nick their target sites leaving a 2 bp 3' overhang, but the reduced specific activity of I-BmoI permitted study of the mechanism of GIY-HEs in more detail (52, 54). Using biochemical assays such as OP-Cu footprinting and gel filtration, I-BmoI was demonstrated to bind DNA as a monomer and sequentially nick DNA with only a single active site (55). In addition, a critical guanine (G) contacted by the I-BmoI linker region is responsible for anchoring the linker region on DNA allowing the catalytic domain to undergo a conformational change to cleave both strands of DNA (55). This critical G anchor position by the I-BmoI linker region parallels an important G requirement for I-TevI cleavage described in Chapter 4.

1.1.3 LAGLIDADG homing endonuclease

LHEs, often referred to as Meganucleases, are the best-studied group with the largest number of characterized proteins. The members are segregated into two main groups depending on whether they have a single or double copy of the conserved LAGLIDADG motif. I-CreI is a well-studied single copy LAGLIDADG protein that requires dimerization to form an active enzyme (56). Other members of the LHE family are monomeric enzymes containing two copies of the LAGLIDADG motif expressed as a single polypeptide (19, 57). Surprisingly, with a sequence similarity of only ~30%, both homodimeric and monomeric LHEs form very similar 3D structures (34, 19, 58, 59). The two LAGLIDADG motifs are located in α -helices and have a critical role bringing the two halves of the enzyme together to form the catalytic core at the base of the enzyme. Many crystal structures have been solved showing an α/β topology, where the β -sheets form a saddle containing a large number of basic and polar residues for DNA contacts (Figure 1.2) (34, 37, 19, 58). The saddle sits on the ~20 bp homing site and projects amino acid residues into the major groove to locate the homing site. Similar to the GIY-HEs, LHEs place a great deal of strain on the DNA substrate to help facilitate DNA binding and scissile phosphate positioning in the active site

(Figure 1.3) (38). This strain helps LAGLIDADG cleave across the minor groove leaving a 4 bp 3' overhang. Their compact size containing both the DNA recognition and catalytic activity packed into ~300 amino acids is just one characteristic that has highlighted LHEs as potential genome-editing reagents.

1.2 DNA sequence specificity for site-specific proteins

There has been a large effort to try and decipher a protein-DNA binding code as it would be helpful in many downstream applications to help determine transcription binding sites or potential off-target binding sites. However, proteins contain diverse tertiary and quaternary structures that makes deciphering a simple code for sequence-specific contacts a very tough task (8). Surprisingly, about two-thirds of all protein-DNA interactions are from van der Waals contacts, while about one-sixth are from hydrogen and water-mediated bonds (8). Diverse protein-DNA interactions present a problem when mutating proteins with the goal of changing overall specificity. Unfortunately, altering the specificity of DNA-binding proteins is not as simple as mutating a couple aa-DNA contacts as this may disrupt protein intramolecular interactions. As described in previous studies, mutating HEs or zinc-finger (ZF) domains and changing only DNA-contacting residues often reduces binding affinity or cleavage activity (19, 60). Another consideration is the general flexibility of the DNA sequence. It is well known that HEs and REs distort their DNA target sites and a general protein flexibility must be maintained for effective binding and cleavage (11, 30). This presents an indirect readout mechanism that is very difficult to characterize and generalize across all sequence-specific DNA-binding proteins. A third problem was observed when creating de-novo DNA-contacting transcription activator-like effector (TALE) and ZF domains. In both cases, intramolecular interactions within the protein structure affected the intermolecular DNA-binding contacts (10, 61–63). Although large barriers exist when attempting to change specificity, high throughput protein engineering methods have aided in overcoming some problems with creating novel DNA-binding proteins.

1.2.1 Restriction endonuclease sequence specificity

Restriction endonucleases (REs) are some of the most specific DNA-binding endonucleases as single bp substitutions can abolish activity. Therefore, REs have transformed into a massive industry allowing scientists to readily manipulate DNA sequences and clone their gene of

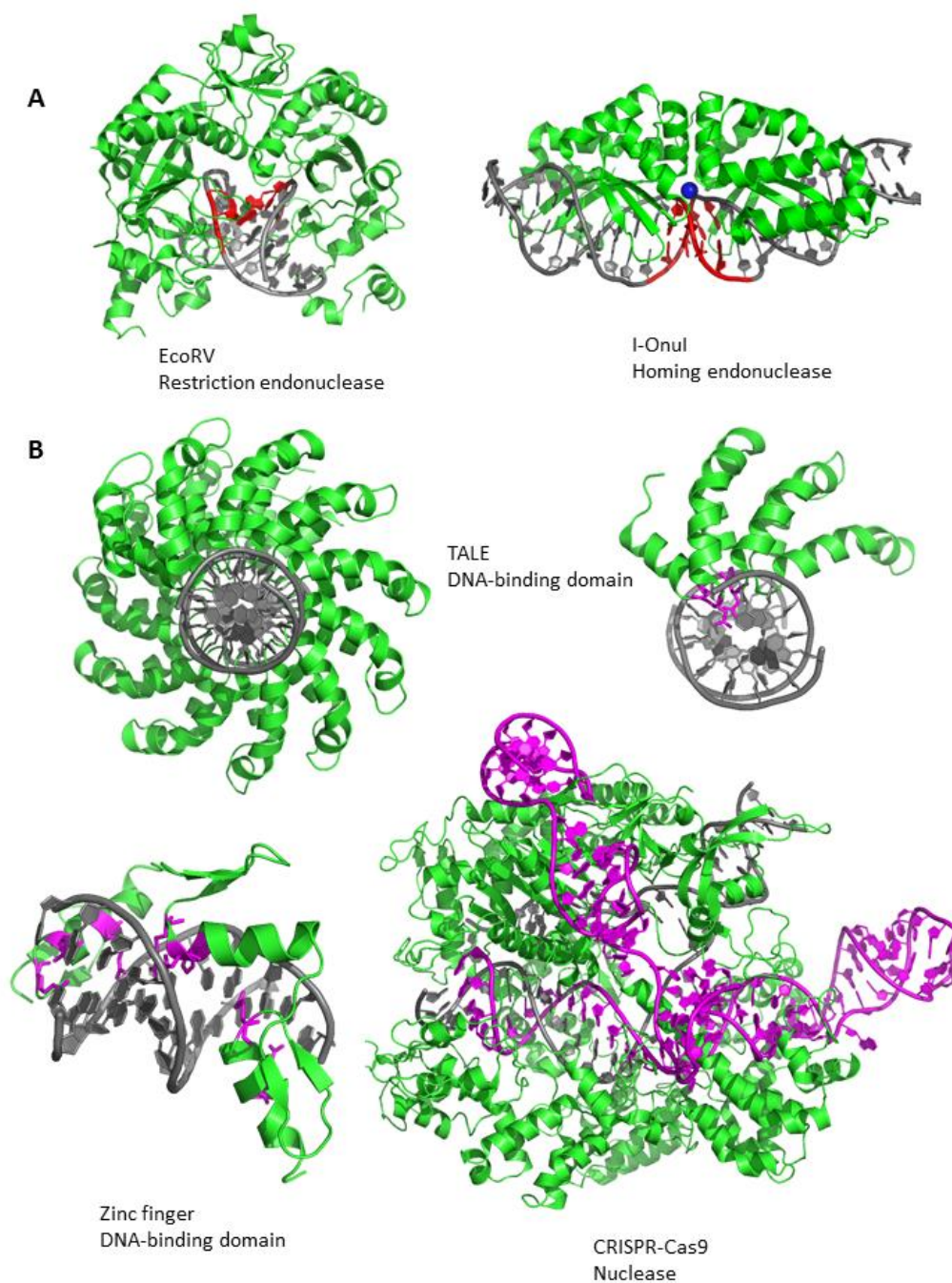


Figure 1.3 Crystal structures of various DNA binding domains and nucleases

Crystal structure of DNA bound proteins highlighting the diversity in DNA recognition. (A) Restriction endonuclease I-EcoRV and homing endonuclease I-OnuI bound to DNA with indirect readout portion of DNA highlighted in red. (B) Crystal structure of transcription activator-like effector (TALE), zinc-fingers and Cas9 bound to DNA with contacting residues or RNA molecule highlighted in purple.

interest. Although REs may be known mainly for cloning purposes, they were originally discovered in the mid 1950's as a bacterial defense mechanism (64–66). Their natural purpose was to identify foreign DNA and cleave it to protect the host from invading DNA molecules (67, 68). Therefore, REs have evolved to recognize more frequent short DNA sequences from 4–8 bps. In addition, there was evolutionary pressure to become ultra-specific, as introducing a DSB in the host genome can be detrimental. Mutagenesis studies making single substitutions in the recognition sequence of EcoRI resulted in a reduction in k_{cat}/K_M cleavage from 10^5 to 10^9 (69, 70). It was observed that single nucleotide substitutions affected DNA cleavage rate larger than DNA binding of the enzyme. This stringent discrimination at the EcoRI recognition site is a result of both direct and indirect readout of the DNA target site (69). A recurring theme is the requirement for DNA distortion, observed in both HEs and REs, to properly bind and access the scissile phosphates for cleavage (Figure 1.3). In the case of EcoRI, the DNA distortion allows access for a bundle of four α -helices to contact the major groove (71). EcoRI saturates 16 out of the 18 possible hydrogen-bonding positions in its 6 bp recognition sequence and makes additional van der Waals' contacts to all thymidine methyl groups (71). Almost complete saturation of target site contacts is the main reason why REs are so intolerant to substitutions. However, there are examples of REs that do not completely saturate their recognition site, such as EcoRV. EcoRV only contacts 4 positions in its target site, as there are no observable sequence-specific interactions with the central TA in the recognition site (11). The central two bps could be an indirect readout mechanism as severe DNA distortion is observed upon EcoRV binding (11, 72). Interestingly, LAGLIDADG homing endonucleases are thought to have an indirect readout mechanism for the central 4 bps in their target site, as no contacts have been observed in crystal structures (58, 73–75).

1.2.2 Homing endonuclease sequence specificity

Altering HE sequence specificity can be a very complex challenge even with a large number of crystal structures readily available. With at least 15 crystal structures for LHEs published to date, the base-specific residues and hydrogen binding contributions that identify the ~20 bp homing site are known. Unlike REs, HEs only contact 65–75% of the possible hydrogen donors and acceptors in the major groove and one third of phosphate backbone contacts (59). This makes sense evolutionarily, as HEs need to remain flexible to individual polymorphisms that occur at their

homing site to prevent loss of binding or cleavage. Therefore, LHEs may target long sequences ~20 bp, but sacrifice some specificity by under-saturating DNA-protein contacts to maintain the ability to adapt to polymorphisms (58, 76, 77). In addition, LAGLIDADG contain a central 4-bp region that has no direct contacts but is restricted by an indirect readout mechanism. Substitutions at these positions have minimal effects on enzyme binding affinity but cause significant reduction in DNA cleavage, presumably because DNA flexibility disrupts metal coordination by catalytic residues (75). Some groups have utilized computer modeling to study LHE binding and help predict residue combinations to help alter specificity (78–80). One example was using the homodimer I-MsoI, the Thr83 residue was predicted to change specificity at +6 and -6 and an additional Lys28 mutation would reduce activity on the cognate homing site (Ashworth 2006). These two mutations enhanced activity on the mutant target site and reduced the activity on the cognate site by 4,000 fold (79).

GIY-HEs, such as I-TevI and I-BmoI, make mostly minor groove contacts and can tolerate multiple nucleotide substitutions at their target site (30, 49). No full-length crystal structure in complex with DNA has been solved, leaving little information on direct or indirect contacts. Biochemical assays using I-BmoI identified a critical guanine base required to anchor the linker region and thought to promote rearrangement of the catalytic domain (55). In addition, the catalytic domains of both I-BmoI and I-TevI both have inherent sequence specificity for 5'-NNNNG-3' and 5'-CNNNG-3', respectively (51, 52). However, little is known about how these catalytic domains discriminate the identified sequence-specific requirements.

1.2.3 ZF and TALE DNA-binding sequence specificity

Zinc fingers are a common DNA-binding platform found in more than a thousand characterized transcription factors (81). Each zinc finger is very small consisting of a simple $\beta\beta\alpha$ fold and recognizing a 3 bp target site (82). Therefore, ZF transcription factors often have multiple fingers that wrap around DNA in a spiral manner. Each finger recognizes the 3 bp by inserting the α -helix into the major groove to contact DNA bases (Figure 1.3) (10, 83). There are three main positions in the alpha helix that recognize each base, the -1 residue contacts 3' base, position 3 contacts the central base and position 6 contacts the 5' base. A simple rule-based system for zinc-finger specificity to govern targeting new DNA sequences was not accomplished as no single approach was successful at predicting all possible DNA sequences for new ZFs (84). Swapping the order of

the fingers resulted in significant context-dependent effects that were difficult to predict (84). Although disappointing, complex protein interactions were found in crystal structures revealing a cross-strand interactions through position 2 (10). In addition, the insertion of linker regions between adjacent zinc-fingers along with mutational analysis has demonstrated to destabilize neighboring zinc-finger protein-DNA complexes (85–87). Since no simple code has been described, robust screening is required to isolate highly specific ZF domains and these are typically biased to G/C rich target sites, limiting programmability (87).

Transcription activator-like effector (TALE) was the most recent protein-DNA binding architecture introduced to the genome-editing field. Naturally, TALEs are transcription factors utilized by the bacterial plant pathogen, *Xanthomonas*, who injects the protein into the plants to hijack the plants resources. The first indication of a TALE DNA-binding code was found in the *avrBs³* gene as it contained a repetitive region that encoded for 34 aa repeats (88). Comparing the AvrBs3 and AvrXa10 TALE repeats highlighted a hypervariable di-residue (RVD) region at the 12th and 13th position in the repeats suggesting that the repeats must determine TALE specificity (Figure 1.3) (89, 90). The true breakthrough in deciphering the TALE code was from two independent studies, describing how the hypervariable residues 12 and 13 correlated with specific bases in the target sequence (62, 63). For example, NI (Asn-Ile) aa coded for adenine and HD (His-Asp) aa coded for cytosine binding. However, assembling novel TALEs to target any sequence of interest was not quite as easy as initially thought, as the putative repeat at the beginning of the TALE required a thymine base. Also similar to ZF domains, TALEs had some unexplainable context dependent interactions. A simple strategy of increasing the number of repeats to increase specificity was not plausible either as the longer TALEs could tolerate more mismatches because moderate affinity was maintained through matching repeat-DNA combinations. Therefore, a set of rules were defined to help construct reliable TALEs including: (i) minimal number of repeats are required; (ii) non-matching repeat-DNA combinations contribute more strongly to overall interaction; (iii) different repeat types exhibit different DNA-interaction binding strength; (iv) and effect of individual non-matching repeat-DNA combinations are position and context-dependent (91).

1.2.4 CRISPR-Cas9 sequence specificity

Clustered regularly interspaced short palindromic repeat-associated protein (CRISPR) evolved as a prokaryotic adaptive immunity to recognize and cleave invading DNA (92). CRISPR-associated (cas) genes are expressed and function in all steps of host immunity including processing RNA molecules used for recognizing invading DNA and endonucleases used to cleave exogenous DNA. The Type II CRISPR system is of particular interest as the targeting endonuclease, Cas9, uses a single polypeptide loaded with a targeting RNA molecule (crRNA) and a stabilizing RNA molecule known as the tracrRNA (Figure 1.3) (93). Since Cas9 uses a RNA molecule to target DNA, it depends on Watson-Crick rules for targeting which is different than all previously described DNA binding platforms used for genome-editing. One restriction for Cas9 targeting is the requirement for a protospacer associated motif (PAM) which evolved to limit off-target cleavage by Cas9 (94). The most extensively studied CRISPR-Cas9 is from *Streptococcus pyogenes* and requires a NGG PAM motif (95). The PAM motif varies depending on the organism, as different Cas9 orthologs contain different PAM motifs. Cas9 binds a crRNA to target a 20 bp DNA sequence and introduce a blunt DSB 3 bp upstream from the PAM motif. Although the crRNA is 20 bp in length, the specific recognition is limited to 7-12 nucleotides adjacent to the PAM motif as multiple mutations are tolerated in the crRNA (96–100). Early studies identified that Cas9 could tolerate many insertions or deletions (indels) in its Watson-Crick binding site. This was problematic, as a short recognition sequence in combination with flexible binding of the RNA-DNA combination results in a large number of unwanted cleavage sites. Efforts to mitigate off-target cleavage by Cas9 have been performed and are discussed in Chapter 1.3.4.

1.3 Genome editing

The ability to modify complex genomes has many downstream applications in agriculture, research and human disease treatment. The first therapeutic trial involving a human gene was approved by the FDA in the early 1990s to treat adenosine deaminase deficiency (ADA-SCID) (101). This monogenetic disease leads to severe immunodeficiency and was thought to be treatable using an *ex vivo* retroviral treatment, where the adenosine deaminase wild-type gene is integrated back into the patients (101). Initial treatments were not as effective as researchers hoped, and detrimental results were observed in some cases where patients either reacted poorly to the retrovirus or random insertions caused leukemia (101, 102). Even with unsettling first results, there have been over 1000 approved gene therapies to treat a wide range of human diseases (103). The

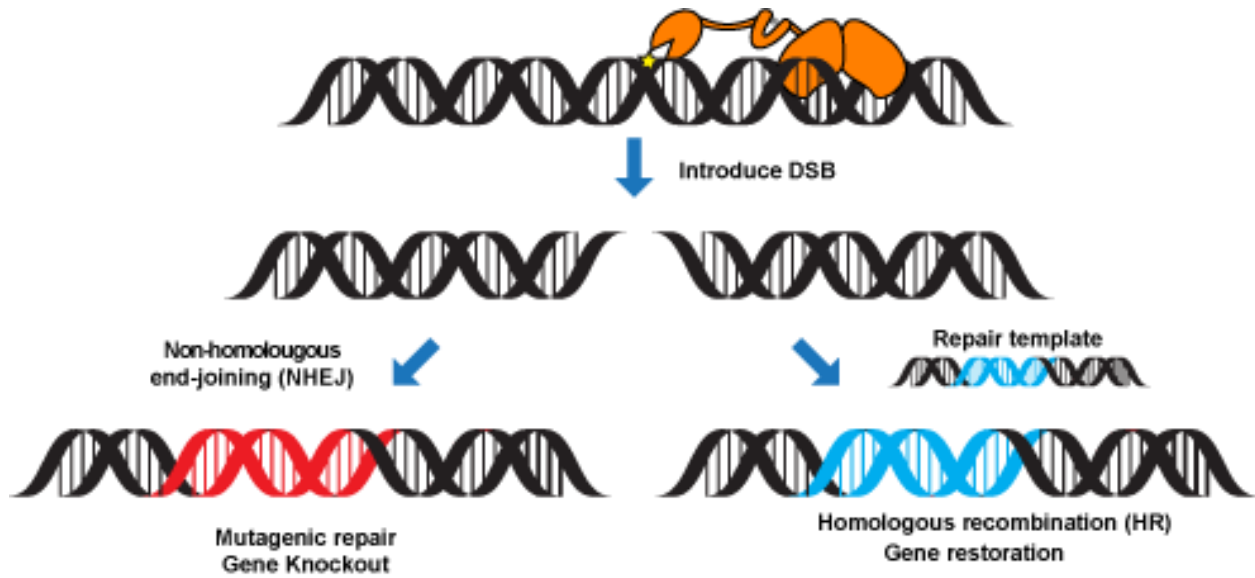


Figure 1.4: Schematic of two outcomes of genome-editing. After the introduction of a double-strand break, the DNA can be repaired through NHEJ or HR. If no repair template is present, Ku proteins will recognize the free ends and they will be ligated back together through NHEJ. If a template DNA is provided then the host can use homologous recombination for gene correction.

largest fear annotated with integrating vector gene therapy continues to be insertional mutagenesis into important host genes, as DNA integration is essentially random (104, 105). In the last 15 years, genome engineering has exploded into an exciting field composed using a number of site-specific reagents (Table 1.1). Genome-editing reagents have replaced retroviral integrases in non-specific ‘gene augmentation’, as they are the most efficient method to achieve targeted integration through HR (106). Genome-editing reagents take advantage of host repair pathways, in order to correct or knockout genes of interest through the HR or non-homologous end joining pathway (NHEJ), respectively (Figure 1.4). First generation genome-editing reagents used the FokI Type IIs RE, fusing the nuclease domain to various DNA-binding platforms such as ZF and TALE domains to target various sequences. The FokI nuclease domain is a non-specific nuclease domain that requires dimerization for efficient DNA cleavage (107, 108). One of the first proof of concept experiments using FokI/zinc-finger fusions, known as ZFNs, successfully cleaved the ‘common gamma chain’ (IL2ry) gene in cultured human cells (109). Initial studies in human embryonic kidney (HEK293) cells demonstrated that ZFNs could induce an 1% gene correction efficiency at their target site (110–112). Around the same time, engineered LHEs were developed to recognize novel targets and were demonstrated to induce modifications at intended target sites (19, 41, 113, 114). In 2010, the genome engineering field transitioned to TALE/FokI fusions, called TALENs, as the repeat:DNA 1:1 code was deciphered and offered a more programmable DNA-binding domain. However, TALENs had their own disadvantages and were short lived with the emergence of the CRISPR-Cas9 system in 2013. This RNA-guided genome-editing platform now dominates the field of genome engineering because of its ease of use. However, it was not long before specificity issues emerged in the new CRISPR-Cas9 system.

1.3.1 Engineered LAGLIDADG homing endonucleases

The first nuclease to successfully modify mammalian cells was the I-SceI LHE, presenting an exciting future for site-specific genome-editing. When I-SceI was co-transfected with a donor template in mammalian cells, a 100-fold increase in gene conversion (gene was repaired using template DNA) was observed (115). The compact size containing both catalytic and DNA-binding capabilities made it an attractive genome-editing reagent. Successful gene modification in mammalian cells was only the beginning, since researchers next needed to demonstrate that LHEs

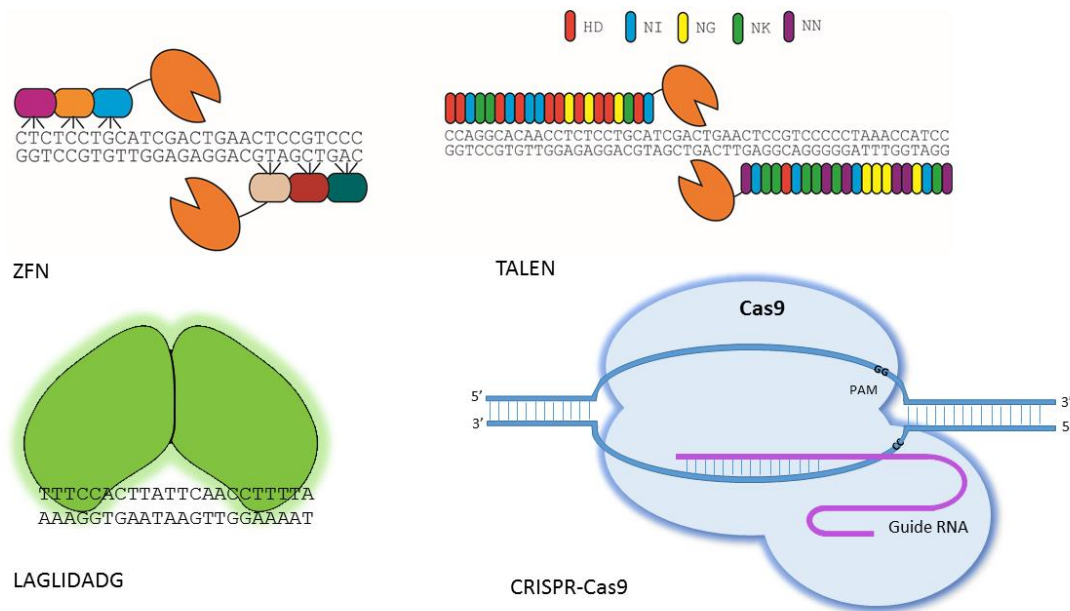


Figure 1.5: Schematic of various genome-editing reagents.

Dimeric genome-editing reagents, zinc-finger nucleases (ZFN) and transcription activator-like effector (TALEN) reagents are derived from the non-specific Type II restriction enzyme FokI fusions. Monomeric genome-editing reagents depicted below with engineered LAGLIDADG homing endonuclease in green and Cas9 RNA-guided nuclease in blue.

could be reprogrammed to target different sequences. The compact structure of LHEs—two DNA-binding saddles flanking the central catalytic core—has made engineering these enzymes a daunting task. The first two engineered LHE described were chimeras, H-DreI and DmoCre, created by fusing the N- and C- terminus of two different LHEs (77, 116). Fusing LHEs halves together to create novel engineered LHE was further explored by Baxter et al. 2012, they found roughly 50% of chimeras created were active enzymes. With a family of over 300 members characterized to date, being able to create chimeras could exponentially increase the number of available targets.

Other studies have focused on engineering specific members in the LHE family, including the homodimeric enzyme I-CreI and monomeric I-OnuI. This approach used rational re-design creating libraries of locally altered specificity modules, designed to change specificity for 3 bp at a time. Directed evolution of LHE has been successful in creating novel LAGLIDADG targets in various human genes including MAO-B, RAG-1 and XPC (17, 15, 19, 117). Since catalytic activity is intrinsically related to DNA-binding, reprogramming LHEs often reduces or abolishes LHE activity. Other implications to consider when engineering LHE is that binding can involve partially folded protein requiring the two LAGLIDADG halves to communicate with each other to locate the correct target. Also, a conformational change in the DNA target site occurs upon LHE binding requiring intermolecular interactions between the two LAGLIDADG halves. Some pivotal high throughput techniques including *in vitro* compartmentalization and yeast surface display has made protein engineering of LAGLIDADG more feasible (118, 119). In addition, to enhance LAGLIDADG activity and specificity, recent groups have fused engineered LAGLIDADG enzymes to TALE domains to create “MegaTALs”. However, the technical skills involved in engineering LHE have limited it to a few specialized labs and in turn has limited their application as a genome-editing reagents.

1.3.2 Zinc-finger and TALE nucleases

Some of the most successful genome-editing reagents utilized the Type II restriction enzyme FokI fused to various DNA-binding platforms described in Chapter 1.2.3. Some important characteristics of the FokI nuclease include a bi-partite structure, requirement for dimerization for cleavage, and a non-specific nuclease domain. The FokI nuclease domain was fused to DNA-binding platforms ZFs and TALEs to create zinc-finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), respectively. One caveat was FokI required

dimerization for efficient cleavage making the design more complex, as two ZFNs or TALENs were required to bind in a “head-to head” fashion for efficient cleavage. The first ZFNs presented many problems with specificity due to both the catalytic domain and DNA-binding domain. As described in Chapter 1.2.2, the ZF DNA specificity code was tough to predict and ZFNs required lots of optimization for efficient cleavage. Similar to engineered LAGLIDADG, the development of high-throughput assembly and screening brought some success to creating active ZFNs (120–123). Unfortunately, undesired cleavage was observed as the ZFNs dimerization occurred at weakly bound sites (122, 124). Although ZFNs were not very specific, this first generation reagent helped pioneer the genome-editing field, developing a lot of the current techniques that are still used today.

The second generation FokI fusions were more customizable utilizing the TALE DNA-binding domain to create TALENs. As described in Chapter 1.2.3, TALEs advantage over ZF domains was their 1:1 protein-DNA code, where each repeat-variable diresidue (RVD) dictated a single bp contact. The discovery of the TALE protein-DNA code in 2010 changed the genome engineering field, as TALEs were quickly fused to the FokI nuclease domain and demonstrated to efficiently cleave desired targets in a wide range of model organisms (125–128). Although TALENs were more programmable, they presented new challenges including the large size ~130 kDa and repeat sequences made it difficult to assemble/clone and deliver using lentiviral systems (129, 130). Many protocols were developed to overcome the assembly problem including Golden Gate (GG) assembly and ligation independent cloning (126, 131). This allowed for easier assembly of a large number of TALENs at a single time. Packaging problems were also addressed using an alternative scaffold with arrays that are non-repetitive sequences the BurrH domain called BuDs (132). In addition, BuDs circumvented another restriction of classic TALENs requiring a 5' thymine (T) due to a non-canonical repeat at the start of classic TALENs (133). TALEs have been used in an array of applications apart from nuclease reagents, playing a role in gene regulation by fusing TALEs to various transcription factors (125, 134–136). The TALEN platform did have some success in genome-editing as they were more specific than ZFNs but were quickly replaced by CRISPR-Cas9 technologies because of its ease of use and programmability.

Table 1.1: Properties of genome-editing reagents

Genome-editing reagent	Nuclease programmability	Specificity	Binding	Cleavage requirements	Overhang
Engineered LAGLIDADG	Difficult to re-target, requires extensive protein engineering	Good specificity, especially when linked to TALE proteins (<u>MegaTALs</u>)	Uses β -sheets to bind the major groove	Functions as a monomer with a single active site flanked by DNA-binding regions	4-bp 3' overhang
ZFN	Limited by ZF targeting, requires some protein engineering to optimize	Not very specific, longer ZFs have failed to mitigate off-target cleavage	Places an α -helices perpendicular to the major groove to project residues to make base-specific contacts	Non-specific <u>FokI</u> nuclease, requires dimerization for efficient cleavage	4-bp 5' overhang
TALEN	More modular programming. Requires 5' thymine for efficient binding	Good specificity, fewer off-target events than ZFNs	Uses α -helices to protrude a di-residue loop region into the major groove	Non-specific <u>FokI</u> nuclease, requires dimerization for efficient cleavage	4-bp 5' overhang
CRISPR-Cas9	Expressing different gRNA is simple and cheap, restricted by the PAM motif	Poor specificity at first, improvements made with Cpf1 and 2 nd generation high-fidelity Cas9	Uses classic Watson-Crick base-pairing using an RNA molecule to bind DNA	Functions as a monomer with active sites in two domains; the HNH and <u>RuvC</u> domain	Blunt
TevCas9	Guided by the Cas9 gRNA, inherent specificity of <u>I-TevI</u> can be restrictive	Unknown, highly dependent on Cas9 specificity	Binding is largely contributed to Cas9 through the DNA-RNA complex	Functions as a monomer, Cas9 cleavage independent from <u>I-TevI</u> . <u>I-TevI</u> has inherent specificity that must be considered	2-bp 3' overhang and blunt

1.3.4 CRISPR-Cas9 nucleases

The introduction of the CRISPR-Cas9 system into the genome-editing field, it has made an astonishing impact on the scientific community. The main reason this platform has been so successful is its “ease-of-use”, governed by classic Watson-Crick base pairing to target DNA sequences. Therefore, changing the target sequence did not require complex protein engineering but a simple change in a RNA sequence. One pivotal study to transition Cas9 into an easy-to-use genome-editing reagent, was fusing the crRNA to the stabilizing tracrRNA to create one RNA molecule for gene targeting known as the guide RNA (gRNA) (96). Although Cas9 could be readily adapted to various sequences, it still had its own limitations that needed to be overcome. First, all Cas9 known to-date require a short PAM sequence adjacent to the intended target site. For *Streptococcus pyogenes*, spCas9, the most widely used CRISPR system in genome-editing requires a NGG PAM motif at its target site (95). To expand the NGG requirement, a protein engineering approach was used to evolve new Cas9 with novel PAM sequences such as NAG (14). Another limitation was that the short 20 bp Watson-Crick target site presented off-target cleavage problems. The true specificity was limited to 7-12 nucleotides adjacent to the PAM motif as the more distal base-pairing was mismatch tolerant (96–100). Therefore, 12 bps plus the 3 bps for the PAM motif was not specific enough for the human genome ($4^{15} = 1.07 \times 10^9$). Many approaches were devised to overcome this problem, one approach used Cas9 nickase mutants as single-strand nicks are less detrimental to the genome but off-target cleavage was still observed (137). Another study created dimeric FokI-Cas9 fusions using inactive Cas9 proteins to reduce off-target cleavage but this also reduced the number of available targets (138). Utilizing truncated gRNAs with only 17-bp or 18-bp of crRNA sequence aided in reducing the overall binding affinity allowing fewer mismatches but did not completely abolish off-target cleavage (139). Finally, through mutating the non-specific contacts, a high-fidelity Cas9 with significantly decreased off-target cleavage was developed (140, 141). Other applications of Cas9 included using a multiplex gRNA expression system to target and knock-out multiple genes simultaneously (99). This approach could be used to study protein pathways or redundant proteins. The success of CRISPR technology has scientific articles using CRISPR being published almost daily, with over 15 model organisms modified by CRISPR-Cas9 (142).

1.4 Scope of the thesis: dual-active nucleases utilizing the GIY-YIG I-TevI nuclease domain for genome engineering

This thesis focuses on the GIY-HE, I-TevI, with the goal to further characterize its catalytic and linker specificity and apply this knowledge to I-TevI derived genome-editing reagents. There have been a number of biochemical and structural studies characterizing the catalytic and linker domain specificity and function, but due to the inherent toxicity of native I-TevI, all previous studies were performed *in vitro* (30, 46, 49, 51). Biochemical studies highlighted regions in the native I-TevI homing site that were hypo- or hyper-sensitive to chemical modifications but sequence specificity was still not well characterized (49). Cleavage specificity of I-TevI was determined almost 20 years ago, having a requirement for a 5'-CNNNG-3' cleavage motif but the lack of sensitivity of the *in vitro* assay could not determine discrimination in the central 3 bp at the cleavage motif (51). By fusing the native I-TevI nuclease and linker domain to different DNA-binding platforms, we de-toxified the catalytic domain allowing us to study the specificity of I-TevI *in vivo*. We exploit one of the first I-TevI fusions to the LHE I-OnuI, referred to as MegaTev, to study the inherent specificity of the catalytic and linker domain of I-TevI in more depth. We hoped that knowledge gained from this platform could be used for other Tev-derived genome-editing reagents.

Our first goal was to demonstrate that I-TevI could be fused to different DNA-binding platforms similar to FokI derived ZFNs and TALENs, to create a monomeric genome-editing reagent. The unique aspects of I-TevI fusions compared to FokI fusions included its inherent cleavage motif specificity 5'-CNNNG-3', its ability to function as a monomer resulting in a more simplistic design and an opportunity of creating a dual-active endonuclease that can create two DSB at a single target site. A dual-active endonuclease could excise a short DNA fragment from the genome, resulting in the free DNA ends being ligated together *in vivo*. We envisioned this would increase the efficiency of gene knockouts as persistent cleavage and re-ligation from single cutting endonucleases would not occur. Previous studies utilized the DNA 5'-exonuclease Trex2 to increase gene modification for LAGLIDADG and ZFNs (20, 143). All other genome-editing reagents, MegaTALs, ZFNs, TALENs, and Cas9 create only a single DSB at their target site. Although they all vary in efficiency ranging from 15-80% depending on the target site and reagent, the indels created are random and unpredictable (108, 144–146). We proposed that our dual-active

endonucleases are able to bias these events to further favour gene-knockout as the excised fragment creates a predictable repair event between the two free ends.

Here, we develop I-TevI fusions to programmable DNA-binding platforms and investigate the specificity of the I-TevI nuclease and linker domain *in vivo*. With this fundamental knowledge we were able to engineer novel I-TevI mutants with enhanced activity and altered specificity. Using the dual-active nuclease platforms, MegaTev and TevCas9 fusions, could prove advantageous over traditional reagents because their ability to create more predictable modifications at their target sites. The genome-editing field is always in search of the most specific, easy-to use, and efficient reagents. It is difficult to engineer a reagent that checks all three boxes but our most recent TevCas9 reagent is at least 2 out of 3, being relatively easy-to-use and more efficient than Cas9 alone.

1.5 References

1. Chapman, J.R., Taylor, M.R.G. and Boulton, S.J. (2012) Playing the End Game: DNA Double-Strand Break Repair Pathway Choice. *Mol. Cell*, **47**, 497–510.
2. Krejci, L., Altmannova, V., Spirek, M. and Zhao, X. (2012) Homologous recombination and its regulation. *Nucleic Acids Res.*, **40**, 5795–5818.
3. Jasin, M. and Rothstein, R. (2013) Repair of strand breaks by homologous recombination. *Cold Spring Harb. Perspect. Biol.*, **5**:a012740.
4. Boch, J. and Bonas, U. (2010) Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu. Rev. Phytopathol.*, **48**, 419–36.
5. Gupta, A., Christensen, R.G., Bell, H. a, Goodwin, M., Patel, R.Y., Pandey, M., Enameh, M.S., Rayla, A.L., Zhu, C., Thibodeau-Beganny, S., *et al.* (2014) An improved predictive recognition model for Cys2-His2 zinc finger proteins. *Nucleic Acids Res.*, **42**, 4800–4812.
6. Stoddard, B.L. (2011) Homing endonucleases: From microbial genetic invaders to reagents for targeted DNA modification. *Structure*, **19**, 7–15.
7. Hafez, M., Hausner, G. and Bonen, L. (2012) Homing endonucleases: DNA scissors on a mission. *Genome*, **55**, 553–569.
8. Luscombe, N.M., Laskowski, R. a and Thornton, J.M. (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucleic Acids Res.*, **29**, 2860–2874.
9. Luscombe, N.M., Austin, S.E., Berman, H.M. and Thornton, J.M. (2000) An overview of the structures of protein-DNA complexes. *Genome Biol.*, **1**, 1-37.
10. Pavletich, N.P. and P.C.O. Zinc Finger-DNA Recognition: Crystal Structure of a Zif268-DNA Complex at 2.1 Å. *Science* **252**, 809-817.
11. Winkler, F.K., Banner, D.W., Oefner, C., Tsernoglou, D., Brown, R.S., Heathman, S.P., Bryan, R.K.,

- Martin,P.D., Petratos,K. and Wilson,K.S. (1993) The crystal structure of EcoRV endonuclease and of its complexes with cognate and non-cognate DNA fragments. *EMBO J.*, **12**, 1781–1795.
12. Moure,C.M., Gimble,F.S. and Quiococho,F.A. (2008) Crystal structures of I-SceI complexed to nicked DNA substrates: Snapshots of intermediates along the DNA cleavage reaction pathway. *Nucleic Acids Res.*, **36**, 3287–3296.
 13. Halford,S.E. and Marko,J.F. (2004) How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.*, **32**, 3040–3052.
 14. Kleinstiver,B.P., Prew,M.S., Tsai,S.Q., Topkar,V. V, Nguyen,N.T., Zheng,Z., Gonzales,A.P., Li,Z., Peterson,R.T., Yeh,J.R., *et al.* (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*, **523**, 481–485.
 15. Arnould,S., Perez,C., Cabaniols,J.P., Smith,J., Gouble,A., Grizot,S., Epinat,J.C., Duclert,A., Duchateau,P. and Paques,F. (2007) Engineered I-Crel Derivatives Cleaving Sequences from the Human XPC Gene can Induce Highly Efficient Gene Correction in Mammalian Cells. *J. Mol. Biol.*, **371**, 49–65.
 16. Li,H., Ulge,U.Y., Hovde,B.T., Doyle,L.A. and Monnat,R.J. (2012) Comprehensive homing endonuclease target site specificity profiling reveals evolutionary constraints and enables genome engineering applications. *Nucleic Acids Res.*, **40**, 2587–2598.
 17. Grizot,S., Smith,J., Daboussi,F., Prieto,J., Redondo,P., Merino,N., Villate,M., Thomas,S., Lemaire,L., Montoya,G., *et al.* (2009) Efficient targeting of a SCID gene by an engineered single-chain homing endonuclease. *Nucleic Acids Res.*, **37**, 5405–5419.
 18. Arnould,S., Chames,P., Perez,C., Lacroix,E., Duclert,A., Epinat,J.C., Stricher,F., Petit,A.S., Patin,A., Guillier,S., *et al.* (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J. Mol. Biol.*, **355**, 443–458.
 19. Takeuchi,R., Lambert,A.R., Mak,A.N.-S., Jacoby,K., Dickson,R.J., Gloor,G.B., Scharenberg,A.M., Edgell,D.R. and Stoddard,B.L. (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 13077–82.
 20. Certo,M.T., Ryu,B.Y., Annis,J.E., Garibov,M., Jarjour,J., Rawlings,D.J. and Scharenberg,A.M. (2011) Tracking genome engineering outcome at individual DNA breakpoints. *Nat. Methods*, **8**, 671–6.
 21. Lambowitz,A.M. and Belfort,M. (1993) Introns as mobile genetic elements. *Annu. Rev. Biochem*, **62**, 587-622.
 22. Roberts,R.J. and Halford,S.E. (1993) Type II Restriction Endonucleases. *Nucleases*, **6**, 35-88.
 23. Syvanen,M. (1984) The evolutionary implications of obile elements. *Ann. Rev. Genet.* **18**, 271-93.
 24. Jin,Y., Binkowski,G., Simon,L.D. and Norris,D. (1997) Ho Endonuclease Cleaves MAT DNA in Vitro by an Inefficient Stoichiometric Reaction Mechanism *. **272**, 7352–7359.
 25. Dujon,B. (1980) Sequence of the intron and flanking exons of the mitochondrial 21S rRNA gene of yeast strains having different alleles at the ω and rib-1 loci. *Cell*, **20**, 185–197.
 26. Jacquier,A. and Dujon,B. (1985) An intron-encoded protein is active in a gene conversion process that spreads an intron into a mitochondrial gene. *Cell*, **41**, 383–394.

27. Dujon,B. (1989) Group I introns as mobile genetic elements: Facts and mechanistic speculations - a review. *Gene*, **82**, 91–114.
28. Belfort,M. and Perlman,P.S. (1995) Mechanisms of intron mobility. *J. Biol. Chem.*, **270**, 30237–30240.
29. Colleaux,L., d’Auriol,L., Betermier,M., Cottarel,G., Jacquier,A., Galibert,F. and Dujon,B. (1986) Universal code equivalent of a yeast mitochondrial intron reading frame is expressed into *E. coli* as a specific double strand endonuclease. *Cell*, **44**, 521–533.
30. Mueller,J.E., Smith,D., Bryk,M. and Belfort,M. (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the td homing site. *EMBO J.*, **14**, 5724–35.
31. Belfort,M. and Roberts,R.J. (1997) Homing endonucleases: Keeping the house in order. *Nucleic Acids Res.*, **25**, 3379–3388.
32. Plessis,A., Perrin,A., Haber,J.E. and Dujon,B. (1992) Site-specific recombination determined by I-SceI, a mitochondrial group I intron-encoded endonuclease expressed in the yeast nucleus. *Genetics*, **130**, 451–460.
33. Perrin, a, Buckle,M. and Dujon,B. (1993) Asymmetrical recognition and activity of the I-SceI endonuclease on its site and on intron-exon junctions. *EMBO J.*, **12**, 2939–2947.
34. Jurica,M.S., Monnat,R.J. and Stoddard,B.L. (1998) DNA recognition and cleavage by the LAGLIDADG homing endonuclease I-CreI. *Mol. Cell*, **2**, 469–476.
35. Sussman,D., Chadsey,M., Fauce,S., Engel,A., Bruett,A., Monnat,R., Stoddard,B.L. and Seligman,L.M. (2004) Isolation and characterization of new homing endonuclease specificities at individual target site positions. *J. Mol. Biol.*, **342**, 31–41.
36. Posey,K.L., Koufopanou,V., Burt,A. and Gimble,F.S. (2004) Evolution of divergent DNA recognition specificities in VDE homing endonucleases from two yeast species. *Nucleic Acids Res.*, **32**, 3947–3956.
37. Heath,P.J., Stephens,K.M., Monnat,R.J. and Stoddard,B.L. (1997) The structure of I-CreI, a group I intron-encoded homing endonuclease. *Nat. Struct. Biol.*, **4**, 468–476.
38. Wende,W., Grindl,W., Christ,F., Pingoud,A. and Pingoud,V. (1996) Binding, bending and cleavage of DNA substrates by the homing endonuclease PI-SceI. *Nucleic Acids Res.*, **24**, 4123–4132.
39. Lykke-Andersen,J., Garrett,R.A. and Kjems,J. (1996) Protein footprinting approach to mapping DNA binding sites of two archaeal homing enzymes: Evidence for a two-domain protein structure. *Nucleic Acids Res.*, **24**, 3982–3989.
40. Derbyshire,V., Kowalski,J.C., Dansereau,J.T., Hauer,C.R. and Belfort,M. (1997) Two-domain structure of the td intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *J. Mol. Biol.*, **265**, 494–506.
41. Doyon,J.B., Pattanayak,V., Meyer,C.B. and Liu,D.R. (2006) Directed evolution and substrate specificity profile of homing endonuclease I-SceI. *J. Am. Chem. Soc.*, **128**, 2477–2484.
42. Van Roey,P., Waddling,C.A., Fox,K.M., Belfort,M. and Derbyshire,V. (2001) Interwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J.*, **20**, 3631–3637.

43. Van Roey,P., Meehan,L., Kowalski,J.C., Belfort,M. and Derbyshire,V. (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat. Struct. Biol.*, **9**, 806–11.
44. Liu,Q., Dansereau,J.T., Puttamadappa,S.S., Shekhtman,A., Derbyshire,V. and Belfort,M. (2008) Role of the Interdomain Linker in Distance Determination for Remote Cleavage by Homing Endonuclease I-TevI. *J. Mol. Biol.*, **379**, 1094–1106.
45. Kowalski,J.C., Belfort,M., Stapleton,M.A., Holpert,M., Dansereau,J.T., Pietrokovski,S., Baxter,S.M. and Derbyshire,V. (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: Coincidence of computational and molecular findings. *Nucleic Acids Res.*, **27**, 2115–2125.
46. Dean,A.B., Stanger,M.J., Dansereau,J.T., Van Roey,P., Derbyshire,V. and Belfort,M. (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 8554–61.
47. Bujnicki,J.M., Rotkiewicz,P., Kolinski, a and Rychlewski,L. (2001) Three-dimensional modeling of the I-TevI homing endonuclease catalytic domain, a GIY-YIG superfamily member, using NMR restraints and Monte Carlo dynamics. *Protein Eng.*, **14**, 717–21.
48. Kleinstiver,B.P., Wolfs,J.M., Kolaczyk,T., Roberts,A.K., Hu,S.X. and Edgell,D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 8061–6.
49. Bryk,M., Quirk,S.M., Mueller,J.E., Loizosi,N., Lawrence,C. and Belfort,M. (1993) The td intron endonuclease I-TevI makes extensive sequence tolerant contacts across the minor groove of its DNA target. *EMBO J.*, **12**, 2141–2149.
50. Bryk,M., Belisle,M., Mueller,J.E. and Belfort,M. (1995) Selection of a remote cleavage site by I-tevl, the td intron-encoded endonuclease. *J. Mol. Biol.*, **247**, 197–210.
51. Edgell,D.R., Stanger,M.J. and Belfort,M. (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J. Mol. Biol.*, **343**, 1231–1241.
52. Edgell,D.R., Stanger,M.J. and Belfort,M. (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-Bmol. *Curr. Biol.*, **13**, 973–978.
53. Edgell,D.R. and Shub,D.A. (2001) Related homing endonucleases I-Bmol and I-TevI use different strategies to cleave homologous recognition sites. *Proc. Natl. Acad. Sci. U. S. A.*, **98**, 7898–903.
54. Bell-Pedersen,D., Quirk,S.M., Bryk,M. and Belfort,M. (1991) I-TevI, the endonuclease encoded by the mobile td intron, recognizes binding and cleavage domains on its DNA target. *Proc. Natl. Acad. Sci. U. S. A.*, **88**, 7719–23.
55. Kleinstiver,B.P., Wolfs,J.M. and Edgell,D.R. (2013) The monomeric GIY-YIG homing endonuclease I-Bmol uses a molecular anchor and a flexible tether to sequentially nick DNA. *Nucleic Acids Res.*, **41**, 5413–5427.
56. Thompson,A.J., Yuan,X., Kudlicki,W. and Herrin,D.L. (1992) Cleavage and recognition pattern of a double-strand-specific endonuclease (I-Crel) encoded by the chloroplast 23S rRNA intron of *Chlamydomonas reinhardtii*. *Gene*, **119**, 247–251.
57. McConnell Smith,A., Takeuchi,R., Pellenz,S., Davis,L., Maizels,N., Monnat,R.J. and Stoddard,B.L. (2009) Generation of a nicking enzyme that stimulates site-specific gene conversion from the I-Anil

- LAGLIDADG homing endonuclease. *Proc. Natl. Acad. Sci. U. S. A.*, **106**, 5099–104.
58. Moure, C.M., Gimble, F.S. and Quijoch, F.A. (2003) The crystal structure of the gene targeting homing endonuclease I-SceI reveals the origins of its target site specificity. *J. Mol. Biol.*, **334**, 685–695.
 59. Chevalier, B.S., Monnat, R.J. and Stoddard, B.L. (2001) The homing endonuclease I-CreI uses three metals, one of which is shared between the two active sites. *Nat. Struct. Biol.*, **8**, 312–316.
 60. Gupta, A., Meng, X., Zhu, L.J., Lawson, N.D. and Wolfe, S.A. (2011) Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Res.*, **39**, 381–392.
 61. Scholze, H. and Boch, J. (2010) TAL effector-DNA specificity. *Virulence*, **1**, 428–432.
 62. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
 63. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
 64. Vasu, K. and Nagaraja, V. (2013) Diverse functions of restriction-modification systems in addition to cellular defense. *Microbiol. Mol. Biol. Rev.*, **77**, 53–72.
 65. LURIA, S.E. and HUMAN, M.L. (1952) A nonhereditary, host-induced variation of bacterial viruses. *J. Bacteriol.*, **64**, 557–569.
 66. BERTANI, G. and WEIGLE, J.J. (1953) Host controlled variation in bacterial viruses. *J. Bacteriol.*, **65**, 113–121.
 67. Pingoud, A., Fuxreiter, M., Pingoud, V. and Wende, W. (2005) Type II restriction endonucleases: Structure and mechanism. *Cell. Mol. Life Sci.*, **62**, 685–707.
 68. Yuan, R. (1981) Structure and mechanism of multifunctional restriction endonucleases. *Ann. Rev. Biochem.* **50**: 285–315.
 69. Lesser, D.R., Kurpiewski, M.R. and Jen-Jacobson, L. (1990) The energetic basis of specificity in the EcoRI endonuclease–DNA interaction. *Science*, **250**, 776–86.
 70. Alves, J., Selent, U. and Wolfes, H. (1995) Accuracy of the EcoRV restriction endonuclease: Binding and cleavage studies with oligodeoxynucleotide substrates containing degenerate recognition sequences. *Biochemistry*, **34**, 11191–11197.
 71. Kim, Y., Grable, J., Love, R., Greene, P. and Rosenberg, J. (1990) Refinement of EcoRI endonuclease crystal structure: a revised protein chain tracing. *Science*, **249**, 1307–1309.
 72. Martin, a M., Sam, M.D., Reich, N.O. and Perona, J.J. (1999) Structural and energetic origins of indirect readout in site-specific DNA cleavage by a restriction endonuclease. *Nat. Struct. Biol.*, **6**, 269–277.
 73. Baxter, S., Lambert, A.R., Kuhar, R., Jarjour, J., Kulshina, N., Parmeggiani, F., Danaher, P., Gano, J., Baker, D., Stoddard, B.L., *et al.* (2012) Engineering domain fusion chimeras from I-OnuI family LAGLIDADG homing endonucleases. *Nucleic Acids Res.*, **40**, 7985–8000.
 74. Molina, R., Redondo, P., Stella, S., Marenchino, M., D’Abramo, M., Gervasio, F.L., Charles Epinat, J., Valton, J., Grizot, S., Duchateau, P., *et al.* (2012) Non-specific protein-DNA interactions control I-CreI

- target binding and cleavage. *Nucleic Acids Res.*, **40**, 6936–6945.
75. Lambert,A.R., Hallinan,J.P., Shen,B.W., Chik,J.K., Bolduc,J.M., Kulshina,N., Robins,L.I., Kaiser,B.K., Jarjour,J., Havens,K., *et al.* (2016) Indirect DNA Sequence Recognition and Its Impact on Nuclease Cleavage Activity. *Structure*, **24**, 862–873.
 76. Moure,C.M., Gimble,F.S. and Quiocho,F. a (2002) Crystal structure of the intein homing endonuclease PI-SceI bound to its recognition sequence. *Nat. Struct. Biol.*, **9**, 764–770.
 77. Chevalier,B.S., Kortemme,T., Chadsey,M.S., Baker,D., Monnat,R.J. and Stoddard,B.L. (2002) Design, activity, and structure of a highly specific artificial endonuclease. *Mol Cell*, **10**, 895–905.
 78. Ulge,U.Y., Baker,D.A. and Monnat,R.J. (2011) Comprehensive computational design of mCrel homing endonuclease cleavage specificity for genome engineering. *Nucleic Acids Res.*, **39**, 4330–4339.
 79. Ashworth,J., Havranek,J.J., Duarte,C.M., Sussman,D., Monnat,R.J., Stoddard,B.L. and Baker,D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. *Nature*, **441**, 656–9.
 80. Ashworth,J., Taylor,G.K., Havranek,J.J., Quadri,S.A., Stoddard,B.L. and Baker,D. (2010) Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. *Nucleic Acids Res.*, **38**, 5601–5608.
 81. Ding,G., Lorenz,P., Kreutzer,M., Li,Y. and Thiesen,H.J. (2009) SysZNF: The C2H2 zinc finger gene database. *Nucleic Acids Res.*, **37**, 267–273.
 82. Lee,M., S, Garry,P., Kizhake,V., David,A. and Peter,E. (1989) Three-Dimensional Solution Structure of a Single Zinc Finger DNA-Binding Domain.
 83. Suzuki,M., Gerstein,M. and Yagi,N. (1994) Stereochemical basis of DNA recognition by Zn fingers. *Nucleic Acids Res.*, **22**, 3397–405.
 84. Desjarlais,J.R. and Berg,J.M. (1993) Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc. Natl. Acad. Sci. U. S. A.*, **90**, 2256–2260.
 85. Choo,Y. and Klug,A. (1993) A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. *Nucleic Acids Res.*, **21**, 3341–3346.
 86. Clemens, K.R, Zhang, P., Liao, X., McBryant, S.J., Wright, P.E. and Gottesfeld, J.M. (1994). Relative Contributions of the Zinc Fingers Transcription Factor IIIA to the Energetics of DNA binding. *J. Mol. Biol.* **244**, 23-35.
 87. Laity,J.H., Chung,J., Dyson,H.J. and Wright,P.E. (2000) Alternative splicing of Wilms' tumor suppressor protein modulates DNA binding activity through isoform-specific DNA-induced conformational changes. *Biochemistry*, **39**, 5341–5348.
 88. Bonas,U., Stall,R.E. and Staskawicz,B. (1989) Genetic and structural characterization of the avirulence gene *avrBs3* from *Xanthomonas campestris* pv. *vesicatoria*. *Mol. Gen. Genet.*, **218**, 127–136.
 89. Herbers, K., Conrads-Strauch, J and Bonas,U. (1992) Race-Specificity of Plant Resistance to Bacterial Spot Disease ...
 90. Hopkins,C.M., White,F.F., Choi,S.H., Guo,A., Leach,J.E. and C.M. Hopkins, F. F. White, S.-H. Choi, A. Guo,J.E.L. (1992) Identification of a family of avirulence genes from *xanthomonas oryzae* pv.

- oryzae. *Mol. Plant Microbe Interact.*, **5**, 451–459.
91. Scholze,H. and Boch,J. (2010) TAL effector-DNA specificity. *Virulence*, **1**, 428–432.
 92. Mojica,F.J.M., Díez-Villaseñor,C., García-Martínez,J. and Soria,E. (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J. Mol. Evol.*, **60**, 174–182.
 93. Gasiunas,G., Barrangou,R., Horvath,P. and Siksnys,V. (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, E2579–86.
 94. Cencic,R., Miura,H., Malina,A., Robert,F., Ethier,S., Schmeing,T.M., Dostie,J. and Pelletier,J. (2014) Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLoS One*, **9**, 1–13.
 95. Mojica,F.J.M., Díez-Villaseñor,C., García-Martínez,J. and Almendros,C. (2009) Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, **155**, 733–740.
 96. Jinek,M., Chylinski,K., Fonfara,I., Hauer,M., Doudna,J.A. and Charpentier,E. (2012) A Programmable Dual-RNA-Guided. **337**, 816–822.
 97. Saprunauskas,R., Gasiunas,G., Fremaux,C., Barrangou,R., Horvath,P. and Siksnys,V. (2011) The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic Acids Res.*, **39**, 9275–9282.
 98. Semenova,E., Jore,M.M., Datsenko,K. a, Semenova,A., Westra,E.R., Wanner,B., van der Oost,J., Brouns,S.J.J. and Severinov,K. (2011) Interference by clustered regularly interspaced short palindromic repeat (CRISPR) RNA is governed by a seed sequence. *Proc. Natl. Acad. Sci. U. S. A.*, **108**, 10098–10103.
 99. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A., *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–23.
 100. Jiang,W., Bikard,D., Cox,D., Zhang,F. and Marraffini,L. A. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*, **31**, 233–239.
 101. Blaese,R.M., Culver,K.W., Miller,A.D., Carter,C.S., Fleisher,T., Clerici,M., Shearer,G., Chang,L., Chiang,Y., Tolstoshev,P., *et al.* (1995) T lymphocyte-directed gene therapy for ADA- SCID: initial trial results after 4 years. *Science*, **270**, 475–480.
 102. Cavazzana-calvo,M. and Hacein-bey,S. (2000) Gene Therapy of Human Severe Combined Immunodeficiency (SCID)– X1 Disease. *Science (80-)*, **288**, 669–672.
 103. Wirth,T., Parker,N. and Yla-Herttua,S. (2013) History of gene therapy. *Gene*, **525**, 162–169.
 104. Donsante,A., Vogler,C., Muzyczka,N., Crawford,J., Barker,J., Flotte,T., Campbell-Thompson,M., Daly,T. and Sands,M. (2001) Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. *Gene Ther.*, **8**, 1343–1346.
 105. Li,H., Malani,N., Hamilton,S.R., Schlachterman,A., Bussadori,G., Edmonson,S.E., Shah,R., Arruda,V.R., Mingozzi,F., Wright,J.F., *et al.* (2011) Assessing the potential for AAV vector genotoxicity in a murine model. *Blood*, **117**, 3311–3319.

106. Urnov,F.D., Rebar,E.J., Holmes,M.C., Zhang,H.S. and Gregory,P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, **11**, 636–646.
107. Durai,S., Mani,M., Kandavelou,K., Wu,J., Porteus,M.H. and Chandrasegaran,S. (2005) Zinc finger nucleases: Custom-designed molecular scissors for genome engineering of plant and mammalian cells. *Nucleic Acids Res.*, **33**, 5978–5990.
108. Porteus,M.H. and Carroll,D. (2005) Gene targeting using zinc finger nucleases. *Nat. Biotechnol.*, **23**, 967–973.
109. Urnov,F.D., Miller,J.C., Lee,Y.L., Beausejour,C.M., Rock,J.M., Augustus,S., Jamieson, a C., Porteus,M.H., Gregory,P.D. and Holmes,M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646–651.
110. Porteus,M.H. and Baltimore,D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.
111. Alwin,S., Gere,M.B., Guhl,E., Effertz,K., Barbas,C.F., Segal,D.J., Weitzman,M.D. and Cathomen,T. (2005) Custom zinc-finger nucleases for use in human cells. *Mol. Ther.*, **12**, 610–617.
112. Porteus,M.H. (2006) Mammalian gene targeting with designed zinc finger nucleases. *Mol. Ther.*, **13**, 438–446.
113. Seligman,L.M., Chisholm,K.M., Chevalier,B.S., Chadsey,M.S., Edwards,S.T., Savage,J.H. and Veillet,A.L. (2002) Mutations altering the cleavage specificity of a homing endonuclease. *Nucleic Acids Res.*, **30**, 3870–3879.
114. Pâques,F. and Duchateau,P. (2007) Meganucleases and DNA double-strand break-induced recombination: perspectives for gene therapy. *Curr. Gene Ther.*, **7**, 49–66.
115. Rouet,P., Smih,F. and Jasin,M. (1994) Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. *Proc. Natl. Acad. Sci. U. S. A.*, **91**, 6064–8.
116. Epinat,J.C., Amould,S., Chames,P., Rochaix,P., Desfontaines,D., Puzin,C., Patin,A., Zanghellini,A., Pâques,F. and Lacroix,E. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res.*, **31**, 2952–2962.
117. Smith,J., Grizot,S., Arnould,S., Duclert,A., Epinat,J.C., Chames,P., Prieto,J., Redondo,P., Blanco,F.J., Bravo,J., *et al.* (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res.*, **34**, 1–12.
118. Takeuchi,R., Choi,M. and Stoddard,B.L. (2014) Redesign of extensive protein-DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization. *Proc. Natl. Acad. Sci. U. S. A.*, **111**, 4061–6.
119. Jarjour,J., West-Foyle,H., Certo,M.T., Hubert,C.G., Doyle,L., Getz,M.M., Stoddard,B.L. and Scharenberg,A.M. (2009) High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucleic Acids Res.*, **37**, 6871–6880.
120. Zykovich,A., Korf,I. and Segal,D.J. (2009) Bind-n-Seq: high-throughput analysis of in vitro protein-DNA interactions using massively parallel sequencing. *Nucleic Acids Res.*, **37**, e151.
121. Miller,J.C., Holmes,M.C., Wang,J., Guschin,D.Y., Lee,Y.-L., Rupniewski,I., Beausejour,C.M., Waite,A.J., Wang,N.S., Kim,K. a, *et al.* (2007) An improved zinc-finger nuclease architecture for

- highly specific genome editing. *Nat. Biotechnol.*, **25**, 778–85.
122. Cathomen,T. and Keith Joung,J. (2008) Zinc-finger Nucleases: The Next Generation Emerges. *Mol. Ther.*, **16**, 1200–1207.
 123. Maeder,M.L., Thibodeau-beganny,S., Osiak,A., Wright,D.A., Anthony,R.M., Eichtinger,M., Jiang,T., Foley,J.E., Winfrey,R.J., Townsend,J.A., *et al.* (2008) Rapid ‘open-source’ engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Ther.*, **31**, 294–301.
 124. Halford,S.E., Catto,L.E., Pernstich,C., Rusling,D. a and Sanders,K.L. (2011) The reaction mechanism of FokI excludes the possibility of targeting zinc finger nucleases to unique DNA sites. *Biochem. Soc. Trans.*, **39**, 584–588.
 125. Christian,M., Cermak,T., Doyle,E.L., Schmidt,C., Zhang,F., Hummel,A., Bogdanove,A.J. and Voytas,D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 756–761.
 126. Cermak,T., Doyle,E.L., Christian,M., Wang,L., Zhang,Y., Schmidt,C., Baller,J.A., Somia,N. V., Bogdanove,A.J. and Voytas,D.F. (2011) Erratum: Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting (Nucleic Acids Research (2011) 39 (e82) DOI: 10.1093/nar/gkr218). *Nucleic Acids Res.*, **39**, 7879.
 127. Carlson,D.F., Tan,W., Lillico,S.G., Stverakova,D., Proudfoot,C. and Christian,M. (2012) Ef fi cient TALEN-mediated gene knockout in livestock. 10.1073/pnas.1211446109/- /DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1211446109.
 128. Zhang,Y., Zhang,F., Li,X., Baller,J.A., Qi,Y., Starker,C.G., Bogdanove,A.J. and Voytas,D.F. (2013) Transcription Activator-Like Effector Nucleases Enable Efficient Plant Genome Engineering. *Plant Physiol.*, **161**, 20–27.
 129. Yang,L., Guell,M., Byrne,S., Yang,J.L., De Los Angeles,A., Mali,P., Aach,J., Kim-Kiselak,C., Briggs,A.W., Rios,X., *et al.* (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.*, **41**, 9049–9061.
 130. Holkers,M., Maggio,I., Liu,J., Janssen,J.M., Miselli,F., Mussolino,C., Recchia,A., Cathomen,T. and Gonçalves,M.A.F. V (2013) Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res.*, **41**.
 131. Schmid-Burgk,J.L., Schmidt,T., Kaiser,V., Höning,K. and Hornung,V. (2013) A ligation-independent cloning technique for high-throughput assembly of transcription activator–like effector genes. *Nat. Biotechnol.*, **31**, 76–81.
 132. Juillerat,A., Bertonati,C., Dubois,G., Guyot,V., Thomas,S., Valton,J., Beurdeley,M., Silva,G.H., Daboussi,F. and Duchateau,P. (2014) BurrH: a new modular DNA binding protein for genome engineering. *Sci. Rep.*, **4**, 3831.
 133. Stella,S., Molina,R., Lopez-Mendez,B., Juillerat,A., Bertonati,C., Daboussi,F., Campos-Olivas,R., Duchateau,P. and Montoya,G. (2014) BuD, a helix-loop-helix DNA-binding domain for genome modification. *Acta Crystallogr. Sect. D Biol. Crystallogr.*, **70**, 2042–2052.
 134. Kleinstiver,B.P., Wang,L., Wolfs,J.M., Kolaczyk,T., McDowell,B., Wang,X., Schild-Poulter,C., Bogdanove,A.J. and Edgell,D.R. (2014) The I-TevI nuclease and linker domains contribute to the specificity of monomeric TALENs. *G3 (Bethesda)*, **4**, 1155–65.

135. Lin,J., Chen,H., Luo,L., Lai,Y., Xie,W. and Kee,K. (2015) Creating a monomeric endonuclease TALE-I-SceI with high specificity and low genotoxicity in human cells. *Nucleic Acids Res.*, **43**, 1112–1122.
136. Li,T., Huang,S., Jiang,W.Z., Wright,D., Spalding,M.H., Weeks,D.P. and Yang,B. (2011) TAL nucleases (TALNs): Hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res.*, **39**, 359–372.
137. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
138. Tsai,S.Q. and Joung,J.K. (2014) What’s changed with genome editing? *Cell Stem Cell*, **15**, 3–4.
139. Fu,Y., Sander,J.D., Reyon,D., Cascio,V.M. and Joung,J.K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.*, **32**, 279–84.
140. Kleinstiver,B.P., Pattanayak,V., Prew,M.S., Tsai,S.Q., Nguyen,N.T., Zheng,Z. and Keith Joung,J. (2016) High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490–495.
141. Slaymaker,I.M., Gao,L., Zetsche,B., Scott,D.A., Yan,W.X. and Zhang,F. (2015) Rationally engineered Cas9 nucleases with improved specificity. *Science*, **351**, 84–88.
142. Carroll,D. (2014) Genome engineering with targetable nucleases. *Annu. Rev. Biochem.*, **83**, 409–439.
143. Delacôte,F., Perez,C., Guyot,V., Duhamel,M., Rochon,C., Ollivier,N., Macmaster,R., Silva,G.H., Pâques,F., Daboussi,F., *et al.* (2013) High Frequency Targeted Mutagenesis Using Engineered Endonucleases and DNA-End Processing Enzymes. *PLoS One*, **8**.
144. Fu,Y., Foden,J. a, Khayter,C., Maeder,M.L., Reyon,D., Joung,J.K. and Sander,J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–6.
145. Tsai,S.Q., Zheng,Z., Nguyen,N.T., Liebers,M., Topkar,V. V, Thapar,V., Wyvekens,N., Khayter,C., lafrate,A.J., Le,L.P., *et al.* (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*, **33**, 187–197.
146. Beurdeley,M., Bietz,F., Li,J., Thomas,S., Stoddard,T., Juillerat,A., Zhang,F., Voytas,D.F., Duchateau,P. and Silva,G.H. (2013) Compact designer TALENs for efficient genome engineering. *Nat. Commun.*, **4**, 1762.

Chapter 2

2 Monomeric site-specific nuclease for genome editing

The work presented in this chapter is reproduced (with permission, Appendix S1) from:

Kleinstiver, B.P., Wolfs, J.M., Kolaczyk, T., Roberts, A.K., Hu, S.X., Edgell, D.R.
(2012) Monomeric site-specific nucleases for genome editing. *Proceedings of the National Academy of Sciences USA* 109(21):8061-6

2.1 Introduction

Precise genome editing often requires the introduction of a double-strand break (DSB) at defined positions (1-3), and two distinct site-specific DNA endonuclease architectures have been developed towards this goal. One of these architectures relies on reprogramming the DNA-binding specificity of naturally occurring LAGLIDADG homing endonucleases (LHEs) to target desired sequences (4, 5). The other architecture utilizes the reprogrammable DNA-binding specificity of zinc-finger proteins or TAL-effector domains that are fused to the non-specific nuclease domain of the type IIS restriction enzyme FokI to create chimeric zinc-finger nucleases (ZFNs) or TAL effector nucleases (TALENs) (6-8). Regardless of the architecture, the underlying biology of the component proteins imposes design challenges and the relative merits of the LHE and the ZFN/TALEN architectures are the subject of much debate in the literature (6, 9). One notable constraint imposed by the FokI nuclease domain is the requirement to function as a dimer to efficiently cleave DNA (10, 11). For any given DNA target, this necessitates the design of two distinct ZFNs (or two TALENs), such that each pair of zinc finger or TAL effector domains is oriented for FokI dimerization and DNA cleavage (12).

Expanding the repertoire of DNA nuclease domains with distinctive properties is necessary to facilitate the development of new genome editing reagents. Indeed, a number of recent studies have explored the potential of the PvuII restriction enzyme as an alternative site-specific nuclease domain for genome editing applications (13, 14). The PvuII chimeras, however, share similar design constraints as ZFNs and TALENs, requiring two nuclease fusions for precise targeting. In considering alternative nuclease domains for genome editing, we were intrigued by the properties of the GIY-YIG nuclease domain that is associated with a variety of proteins of diverse cellular functions (15). The small (~100 aa) globular GIY-YIG domain is characterized by a structurally conserved central three-stranded antiparallel β sheet, with catalytic residues positioned to utilize a

single metal ion to promote DNA hydrolysis (16-18). Intriguingly, the GIY-YIG homing endonucleases, typified by the isoschizomers I-TevI and I-BmoI (19), bind DNA as monomers (20), and generate a DSB with 2-nt, 3' overhangs. It is unknown, however, if GIY-YIG homing endonucleases function as monomers in all steps of the reaction, as the oligomeric status during cleavage has yet to be studied. Notably, GIY-YIG homing endonucleases prefer a specific DNA sequence to generate a DSB (21, 22). For I-TevI, the bottom (↑) and top (↓) strand nicking sites lie within a 5'-CN↑NN↓G-3' motif (CNNNG), with the critical G optimally positioned ~28 bp from where the H-T-H module of the I-TevI DNA-binding domain interacts with substrate (21, 22). From an engineering perspective, the modularity and sequence specificity of the GIY-YIG nuclease domain makes it an appealing candidate to create new chimeric endonucleases. Indeed, swapping of the I-BmoI and I-TevI catalytic and DNA-binding domains suggested that the GIY-YIG nuclease domain could be fused to unrelated DNA-targeting platforms (23).

To highlight the genome engineering potential of the GIY-YIG nuclease domain, we fused the domain to 3-member zinc fingers to construct GIY-YIG zinc finger endonucleases (GIY-ZFEs). The GIY-ZFEs are active in bacterial and yeast cells, and *in vitro* data show that they function catalytically as monomers and retain the cleavage specificity associated with the parental GIY-YIG nuclease domain. The GIY-YIG nuclease domain is also portable to the LHE platform, as we constructed monomeric GIY-LHEs that are active *in vivo* and possess ~18-bp binding specificity. We selected LHEs as a DNA targeting domain because of the greater sequence specificity compared to 3-member zinc fingers, the ability to reprogram LHE DNA-binding specificity (24-26), and recent success in generating PuvII-LHE fusions (13). Collectively, our data highlight the unique biochemical properties of the GIY-YIG nuclease domain as an alternative to the FokI nuclease domain for genome editing applications.

2.2 Material and methods

See detailed *Supplemental materials and methods*. Briefly, Tev-ZFE and Tev-LHE fusions and hybrid target sites were modeled in PyMOL using the I-TevI 130C (PDB 1I3J), Zif268 (PDB 1AAY), and I-OnuI (PDB 3QQY) cocrystal structures (25, 27, 28). The *in vivo* activity of fusions was determined using a two-plasmid bacterial selection (31) or yeast-based reporter assay (that was used to calibrate activity of Tev-ZFEs and Tev-TALs) against a characterized ZFN (35). TevN201-ryA was purified using nickel affinity chromatography to determine the *in vitro*

biochemical properties of Tev-ZFEs. Cleavage assays were performed as described (43). A custom Perl script was created to determine CNNNG occurrences relative to 8,829 predicted ZFN sites on zebrafish chromosome 1 (40).

2.3 Results

2.3.1 Construction and validation of GIY-zinc finger endonucleases

To create novel chimeric enzymes, we modeled GIY-zinc finger endonucleases (GIY-ZFEs) using existing crystal structures of the I-TevI 130C DNA binding domain and the Zif268 zinc finger (27, 28). One notable feature of our constructs is the polarity, as the I-TevI nuclease domain is fused to the N-terminal end of the ryA protein to mimic its native orientation, unlike FokI constructs that are fused to the C-terminal end of zinc-finger proteins. We modeled the Zif268 zinc finger in place of the H-T-H at the C-terminus of I-TevI, providing the rationale to subsequently fuse various lengths of the I-TevI N-terminal region to the ryA zinc finger that targets a sequence in the *Drosophila rosy* gene to create Tev-ryA zinc finger endonucleases (Tev-ZFEs, Fig. 2.1A) (29). The Tev-zinc finger DNA substrates (TZ) consisted of 30 to 38 bps of the I-TevI *td* homing site joined to the 9-bp ryA target site. The TZ substrates differ in the distance of the CNNNG cleavage motif relative to the ryA-binding site (Fig. 2.1B). Each TZ substrate possesses a single zinc finger targeting sequence, rather than two head-to-head zinc finger sites necessary for efficient ZFN cleavage. A similar set of I-BmoI-ryA fusions (Bmo-ZFEs) and substrates (BZ) were constructed (Fig. S1).

We tested the activity of the GIY-ZFEs using a well-described two-plasmid bacterial selection system, where survival is dependent on the endonuclease cleaving a target plasmid (30, 31). Eight Tev-ZFEs were tested on seven TZ substrates cloned into the reporter plasmid (Fig. 2.1B and C, Table S2.1). In general, the survival of all Tev-ZFEs was highest against TZ substrates where the preferred CNNNG motif was positioned between 33 and 35-bp from the ryA binding site. Low survival (~4-6%) was observed for all Tev-ZFEs against the TZ1.32 substrate, while none survived on the TZ1.30 substrate. Likewise, there was no survival against the longer substrates, with the exception that the longest fusion (TevS206-ryA) exhibited ~22% survival against the TZ1.36 substrate. No survival was observed when the Tev-ZFEs were tested against the target plasmid without a target site (p11lacYwtx1). Mutation of the catalytic arginine 27 of the I-TevI nuclease domain to alanine to create TevR27A-ryAs showed that survival is dependent on GIY-YIG

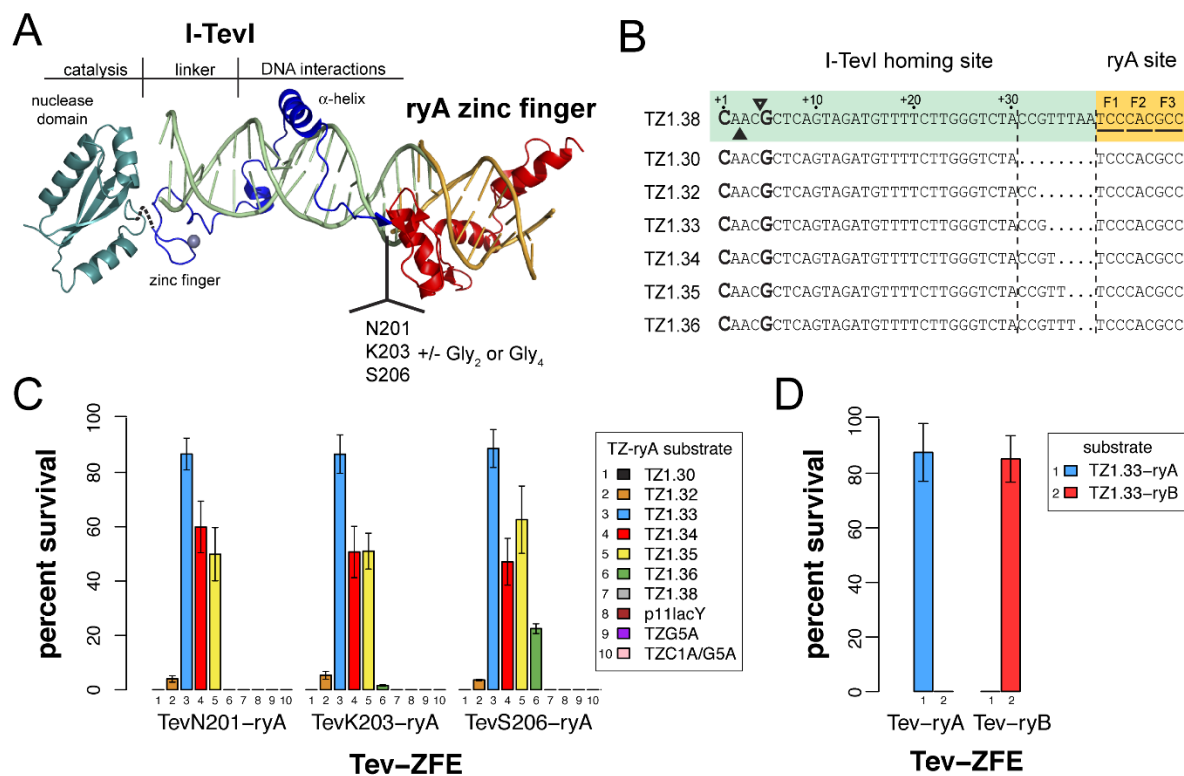


Figure 2.1: Design and functionality of Tev-ZFEs

(A) Modeling of a Tev-zinc finger fusion with DNA substrate (light green) using structures of the I-TevI catalytic domain in green (PDB 1MK0), the I-TevI DNA-binding domain co-crystal in blue (PDB 1I3J), and the Zif268 co-crystal in red (PDB 1AAY) (B) The TZ-ryA substrate is colored according to the structural model. Shown is the top strand of the I-TevI *td* homing site substrate fused to the 5' end of the ryA-binding site for all wild-type substrates tested. The substrate is numbered from the first base of the *td* homing site sequence (the numbering scheme is reverse of that used for the native *td* homing site). The substrates tested differ by insertion or deletion of *td* sequence at the junction of the *td*/ryA sites. (C) Percent survival of three representative Tev-ryA ZFEs in the bacterial two-plasmid selection. All Tev-ryA ZFEs were tested against plasmids containing various length substrates (TZ1.30-1.38), plasmids lacking a target site (p11lacY), and TZ1.33 plasmids with single or double mutations in the CNNNG motif (G5A and C1A/G5A) (Table S1). (D) Percent survival of TevN201-ryA and TevN201-ryB ZFEs on their cognate and reciprocal target sites. Data are plotted with standard deviation for $n \geq 3$.

nuclease activity as none of the Tev-R27A constructs survived (Table S2.1). We also constructed and tested a fusion of the TevN201 domain to a different 3-member zinc finger, the ryB zinc finger, creating TevN201-ryB. The TevN201-ryB showed survival in the bacterial selection assay against a corresponding TZ-ryB target, indicating that the I-TevI nuclease domain can function in the context of two different 3-member zinc fingers, but did not survive when tested against the TZ-ryA substrate (Fig. 2.1D). Likewise, the TevN201-ryA fusion did not survive against the TZ-ryB substrate, indicating that the zinc finger alone directs DNA-binding. We also tested the Bmo-ZFEs in the genetic selection, but did not observe significant survival for any of the fusions, consistent with the ~750-fold reduced activity of wild-type I-BmoI relative to I-TevI (32). However, as described below, enzymatic activity was detected *in vitro* using purified Bmo-ZFEs. Collectively, these data show that two different GIY-YIG nuclease domains could be fused to zinc finger DNA binding domains to create active site-specific chimeric nucleases.

2.3.2 Tev-ZFEs function as monomers to cleave at a specific sequence

To study the GIY-ZFE biochemical characteristics in more detail, we purified TevN201-ryA for cleavage assays and *in vitro* mapping. We first performed cleavage assays to determine the relationship between TevN201-ryA enzyme concentration and initial reaction velocity using a plasmid substrate with a single TZ-ryA target site. The reaction progress curves indicated an initial burst of cleavage followed by a slower rate of product accumulation (Fig. 2.2A), consistent with product release being the rate-limiting step. The initial burst phase was used to estimate initial velocity, and plotting against protein concentration yielded a linear relationship (Fig. 2.2A), suggesting that DNA hydrolysis catalyzed by TevN201-ryA is first order with respect to protein concentration.

The model TZ-ryA substrates were designed as a single ryA zinc finger site fused to the I-TevI target sequence. To determine if cleavage by TevN201-ryA was influenced by additional Tev-ryA target sites, we constructed two-site plasmids that differed in whether the target sites were in the same or opposite orientations relative to each other. The single- or two-site plasmids were used in time-course cleavage assays under single-turnover conditions (~10-fold molar excess of protein to substrate) to determine reaction rates. As shown in Fig. 2.2B, cleavage of the one-site plasmid yielded $k_{\text{obs}(1\text{-site})} = 0.099 \pm 0.001 \text{ s}^{-1}$, and cleavage of the two-site plasmids with target sites in the opposite or same (Fig. S2.2B) orientations generated very similar rate constants, $k_{\text{obs}(2\text{-site})} = 0.088$

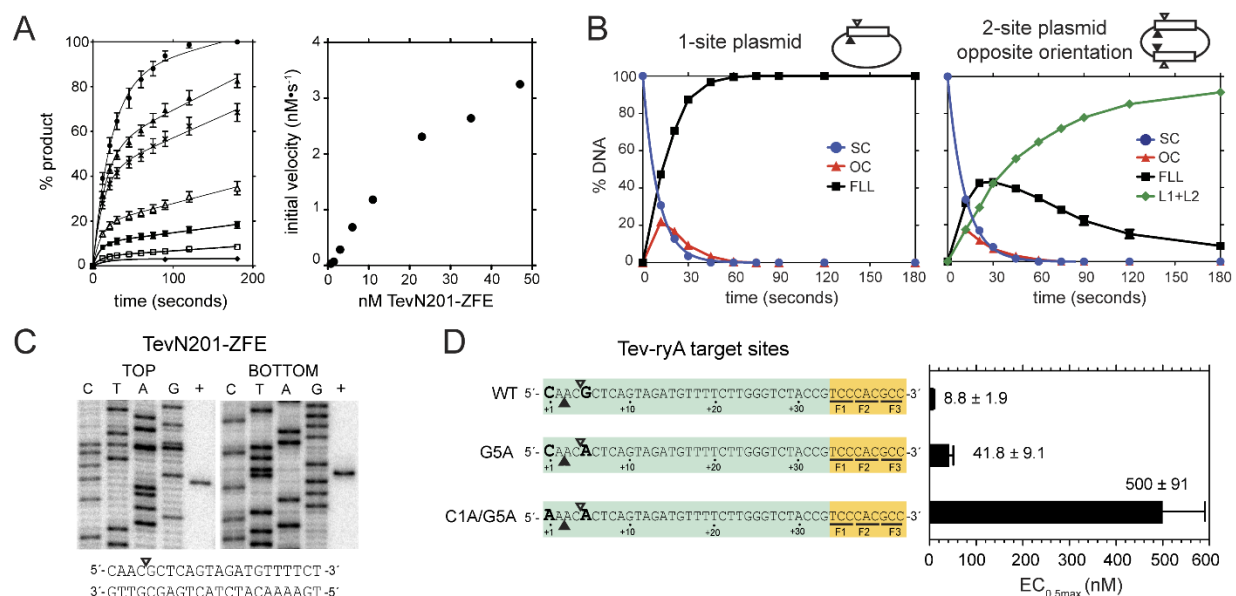


Figure 2.2: TevN201-ZFE is a monomer with a preferred cleavage site.

(A) *Left panel*: plot of initial reaction progress for seven TevN201-ZFE concentrations expressed as percent linear product. Protein concentrations from highest to lowest are 47 nM, 32.5 nM, 23 nM, 11 nM, 6 nM, 3 nM, and 0.7 nM. *Right panel*: graph of initial reaction velocity (nM s⁻¹) versus TevN201-ZFE concentration (nM). (B) Graphical representation of cleavage assays with 90 nM TevN201-ZFE and 10 nM one- or two-site TZ1.33 plasmids (*left and right panels*, respectively). The two-site plasmid had the TZ-ryA sites in the opposite (shown) or same (Fig. S2B) orientation. SC, supercoiled; OC, open-circle (nicked); FLL, full-length linear; L1+L2, linear products. (C) Mapping of TevN201-ZFE cleavage sites on the TZ1.33 substrate, with top and bottom cleavage sites indicated below on the TZ-ryA substrate by open and closed triangles, respectively. (D) Activity of TevN201-ZFE on the wild-type TZ1.33, or the TZ1.33 G5A and TZ1.33 C1A/G5A mutant substrates. A graph of EC_{0.5max} determinations for each substrate is shown to the right, with EC_{0.5max} values in nM. Data are plotted as averages of three independent replicates with standard deviations

$\pm 0.001 \text{ s}^{-1}$ and $0.089 \pm 0.001 \text{ s}^{-1}$, respectively, to the one-site plasmid. In contrast, similar experiments with FokI showed a significant rate enhancement for two-site plasmids relative to one-site plasmids, consistent with FokI functioning as a dimer (33). We conclude that cleavage by TevN201-ryA is non-cooperative and that efficient DNA hydrolysis does not require two sites, consistent with TevN201-ryA functioning catalytically as a monomer.

The I-TevI nuclease domain preferentially cleaves DNA within a 5'-CN \uparrow NN \downarrow G-3' motif, with \uparrow and \downarrow representing the bottom- and top-strand nicking sites, respectively (22). Wild-type I-TevI defaults to cleave at the correct distance on substrates *in vitro* when this motif is moved closer to, or distant from, the primary binding site, whereas mutants in the I-TevI specific zinc finger cleave at the correct sequence rather than the correct distance on mutant substrates (34). To determine the cleavage preference of the TevN201-ryA construct, we mapped the bottom- and top-strand nicking sites using strand-specific end-labeled substrates to the CNNNG motif (Fig. 2.2C). Combined with data from the genetic assays showing no survival on substrates that displace the CNNNG motif from an optimal position, our data suggests that in the context of a ryA fusion, the TevN201 domain acts as a molecular ruler with a distance preference.

To further demonstrate TevN201-ryA cleavage preference, we introduced mutations in the CNNNG motif that were previously shown to drastically reduce I-TevI cleavage efficiency (Fig. 2.2D) (21, 22). Significantly, we observed no survival in the two-plasmid selection assay on plasmids carrying either the single G5A (CNNNA) or double C1A/G5A (ANNNA) substitutions (Fig. 2.1C), equivalent to positions C-27 and G-23 of the I-TevI *td* substrate, respectively. We also performed *in vitro* cleavage assays on wild type and mutant substrates with increasing concentrations of TevN201-ZFE to determine the amount of protein required for half-maximal cleavage ($EC_{0.5\max}$). As shown in Fig. 2.2D, ~60 fold and ~4.7 fold more protein were required to achieve half-maximal cleavage of the double- and single-mutant substrates relative to the wild-type substrate. The greater substrate discrimination observed in the genetic assay likely reflects lower *in vivo* protein concentrations than those used for *in vitro* cleavage assays. These results show that the TevN201-ryA fusion retains the cleavage specificity of the parental I-TevI enzyme and that double nucleotide substitutions significantly reduce cleavage efficiency. To determine if Bmo-ZFEs also retained substrate specificity, the bottom- and top-strand nicking sites of the BmoN221-ryA fusion were mapped to a 5'-NN \uparrow NN \downarrow G-3' motif, consistent with the cleavage site

preference of I-BmoI (Fig. S2.1D) (19).

2.3.3 The Tev-ZFEs function in a yeast-based recombination assay

To extend the *in vivo* relevance of the Tev-ZFE fusions, we utilized a well-described yeast-based recombination assay to test Tev-ZFE function in a eukaryotic system (35). This assay provides a quantitative β -galactosidase readout if the nuclease cleaves its target site that is positioned between a partially duplicated *lacZ* gene. Furthermore, the assay allowed us to calibrate TevN201-ryA activity relative to a homodimeric FokI-Zif268 control with previously measured *in vivo* activity sufficient to induce recombination events for genome engineering applications (35). As shown in Fig. 2.3, the level of β -galactosidase activity for the TevN201-ryA fusion on its cognate TZ-ryA substrate was ~1.4-fold higher than the Zif268 ZFN control. The TevN201-ryA or Zif268 ZFN constructs displayed no activity on each other's substrates, and activity was dependent on a functional I-TevI nuclease domain, as the TevN201R27A catalytic mutant was unable to induce recombination. Furthermore, TevN201-ryA activity was not observed on mutant substrates where one or both of the critical residues of the CNNNG motif were mutated in the TZ1.33 substrate. Collectively, these assays show that the I-TevI nuclease domain functions in a eukaryotic system with activity on par to a characterized ZFN.

2.3.4 The I-TevI nuclease domain is portable to the LAGLIDADG architecture

To demonstrate that the I-TevI nuclease domain functions in the context of DNA-targeting platforms other than 3-member zinc fingers, we constructed fusions of the domain to a catalytically inactive monomeric single-chain LAGLIDADG homing endonuclease (Tev-LHE). As with the Tev-ZFE constructs, we modeled a Tev-LHE chimera using the co-crystal of I-OnuI with its DNA substrate such that the I-TevI nuclease and linker domains were fused to the N-terminus of I-OnuI, which is partially disordered in the structure (Fig. 2.4A) (25). Based on this model, we fused TevN201G₄ and TevK203 fragments to a catalytically dead I-OnuI E1 E22Q mutant. A series of model DNA substrates were constructed by fusing the *td* target site to the I-OnuI E1 binding site in the human MAO-B gene, differing in the position of the CNNNG cleavage motif relative to the I-OnuI E1 site (TO1.12 to TO1.30) (Fig. 2.4B).

In the bacterial two-plasmid selection, we found that the TevN201G₄-Onu and TevK203-Onu fusions were active against a range of DNA substrates. Notably, the fusions displayed maximal

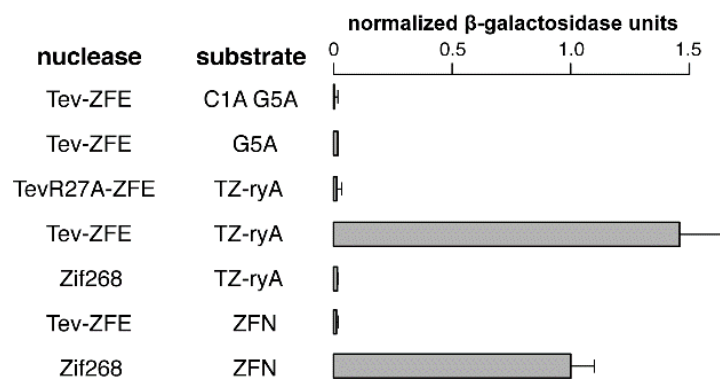


Figure 2.3: Tev-ZFEs can induce recombination in a eukaryotic system

Shown are normalized β -galactosidase units from a yeast-based recombination assay for the indicated nuclease/substrate combinations. Activity was normalized to a homodimeric FokI-Zif268 ZFN positive control. Data are plotted with standard deviation for $n = 4$.

survival on longer targets (TO1.26, TO1.28, and TO1.30), and lower survival against shorter targets (TO1.18 and TO1.20 targets). The two groups of substrate differ by approximately one helical turn of DNA, meaning that the preferred CNNNG motif would be presented on the same face of substrate even though the motif is closer to the I-OnuI E1 binding site on the shorter targets. Similar periodic cleavage patterns have been observed *in vitro* with I-TevI on substrates with a displaced CNNNG motif (36). This result also implies that the N-terminus of I-OnuI possesses inherent flexibility to allow the I-TevI nuclease domain to search out the CNNNG motif, in contrast to the ruler-like behaviour of the Tev-ZFE constructs, likely because the zinc finger N-terminus is inflexible. Importantly, the Tev-Onu fusions were not active against the TZ-ryA zinc finger substrates (Table S2.2), showing that the LHE, and not the I-TevI linker, directs DNA targeting. Survival was also dependent on an active I-TevI nuclease domain, as TevR27A fusions in the context of the I-OnuI E22Q mutant did not survive (Fig. 2.4D). Conversely, the targeting and activity of wild-type I-OnuI E1 was not affected by fusion of the I-TevI domain, as the TevR27A-OnuWT fusions survived against TO substrates (Fig. 2.4D).

The apparent flexibility of the N-terminus and the greater specificity of I-OnuI prompted us to test fusions containing shorter fragments of the I-TevI nuclease domain (Fig. 2.4A). Based on structural and genetic data, we constructed TevS114-Onu, TevD127-Onu, TevN140-Onu, TevN169-Onu, and TevD184G₂-Onu fusions, progressively removing amino acid residues of I-TevI that make specific base-pair contacts to the *td* substrate (28)(Fig. 2.4A). Notably, the TevS114, TevD127, TevN140 and TevN169 removed the α -helix that binds in the minor groove, as well as residues shown by structural data to make base-specific contacts (28). The TevS114 fusion point lies at the boundary of the deletion tolerant region of the I-TevI linker, and represents a functionally minimal GIY-YIG nuclease domain (36, 37). We found that the shorter fusions were not active against the longer TO1.28 and TO1.30 substrates, yet displayed the same periodic activity on the shorter substrates (Fig. 2.4C and Table S2.2). A single exception was the TevD184G₂ fusion that showed low survival against the TO1.22 substrates, against which no other fusion survived. No survival was observed on mutant substrates that contained single (CNNNA) or double (ANNNA) mutations in the CNNNG motif, recapitulating the necessity for an appropriately positioned CNNNG as seen with the Tev-ZFE fusions.

2.3.5 A 5'-CNNNG-3' cleavage motif is not limiting for targeting

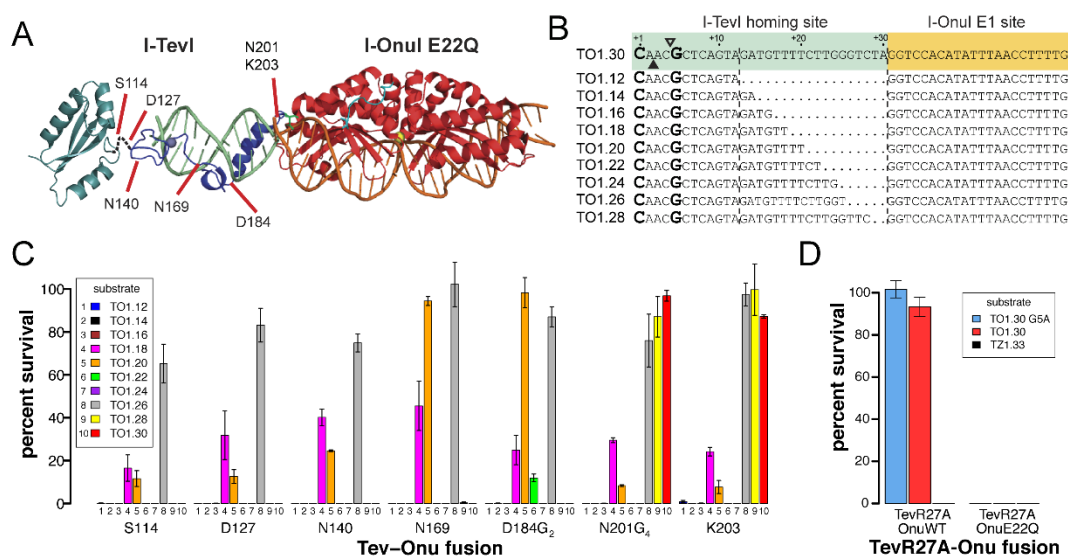


Figure 2.4. Design and functionality of Tev-LHEs

(A) Modeling of a Tev-Onu E1 fusion with DNA substrate (light green) using structures of the I-TevI catalytic domain in green (PDB 1MK0), the I-TevI DNA-binding domain co-crystal in blue (PDB 113J), and the I-OnuI co-crystal in red (PDB 3QQY). Shown are fusion points at which the I-TevI fragment has been shortened. (B) The Tev-Onu E1 (TO) substrate is colored according to the structural model. Shown is the top strand of the I-TevI *td* homing site substrate fused to the 5' end of the Onu E1-binding site. The substrates are numbered from the first base of the *td* homing site sequence and differ by the deletion of *td* nucleotides at the junction of the *td*/Onu E1 sites. (C) Percent survival of Tev-LHEs in the bacterial two-plasmid selection with various length target sites (TO1.12-1.30). All Tev-LHEs tested were in the I-OnuI E1 E22Q background. (D) Percent survival of TevR27A(N201G₄)-OnuE1 and TevR27A(N201G₄)-OnuE1(E22Q) on TO1.30, TO1.30G5A, and TZ1.33. Data are plotted with standard deviation for $n = 3$.

An important consideration in the design of GIY-ZFEs or GIY-LHEs for genome-editing applications are the targeting requirements, notably the need for the CNNNG di-nucleotide cleavage motif (Fig. 2.5A). In a complex genome of $\sim 3 \times 10^9$ bp, the statistically predicted occurrence of the CNNNG motif is once every 15 bp assuming a 50% GC content. To determine if the frequency of the CNNNG motif would be limiting for targeting applications, we examined 35 bp flanking 8,829 computationally predicted ZFN sites on zebrafish chromosome 1 for the occurrence of the CNNNG motif (38). As shown in Fig. 2.5B, the motif is highly represented at all positions within a 35-bp window relative to the ZFN sites. Of the 8,829 sites examined, 88% (7,845) of ZFN sites possessed at least one motif within 35 bp of the predicted binding site (Fig. 2.5C). These requirements contrast sharply with those of the recently described PvuII-LHEs and PvuII-ZFNs that require the 6-bp 5'-CAGCTG-3' PvuII site in addition to the LHE or ZF binding site (13, 14). Of the 8,829 ZFN sites, 97% lacked a PvuII site within the 35-bp window (Fig. S2.3). Thus, the requirement for a di-nucleotide cleavage motif in the context of a GIY-ZFE or GIY-LHE will not severely limit potential targeting sites.

2.4 Discussion

Here, we provide evidence that the GIY-YIG nuclease domain is a potential alternative to the currently used FokI nuclease domain for genome editing applications. We show that the I-TevI GIY-YIG nuclease domain is portable to two reprogrammable DNA-binding scaffolds, the 3-member zinc fingers and LAGLIDADG homing endonucleases. The Tev-ZFE and Tev-LHE fusions are active *in vitro* and *in vivo*, with the activity of the Tev-ZFE in a yeast-based recombination assay on par with that of a characterized ZFN. We foresee the monomeric nature of the Tev-LHEs and Tev-ZFEs as a key advantage over existing ZFNs and TALENs, as a single fusion protein need be designed to target a given sequence, rather than two ZFNs or TALENs required to promote dimerization of the FokI nuclease domain (12). Moreover, the fact that the I-TevI nuclease domain possesses a preferred cleavage motif adds another layer of specificity to targeting requirements, potentially limiting DSBs at off-target sites that do not possess the cleavage motif.

One targeting consideration for chimeric GIY-YIG endonucleases is the DNA sequence requirement of the I-TevI linker. The I-TevI linker is a complex structure, consisting of defined structural elements with distinct roles in I-TevI function (28, 34, 36). The primary role of the linker

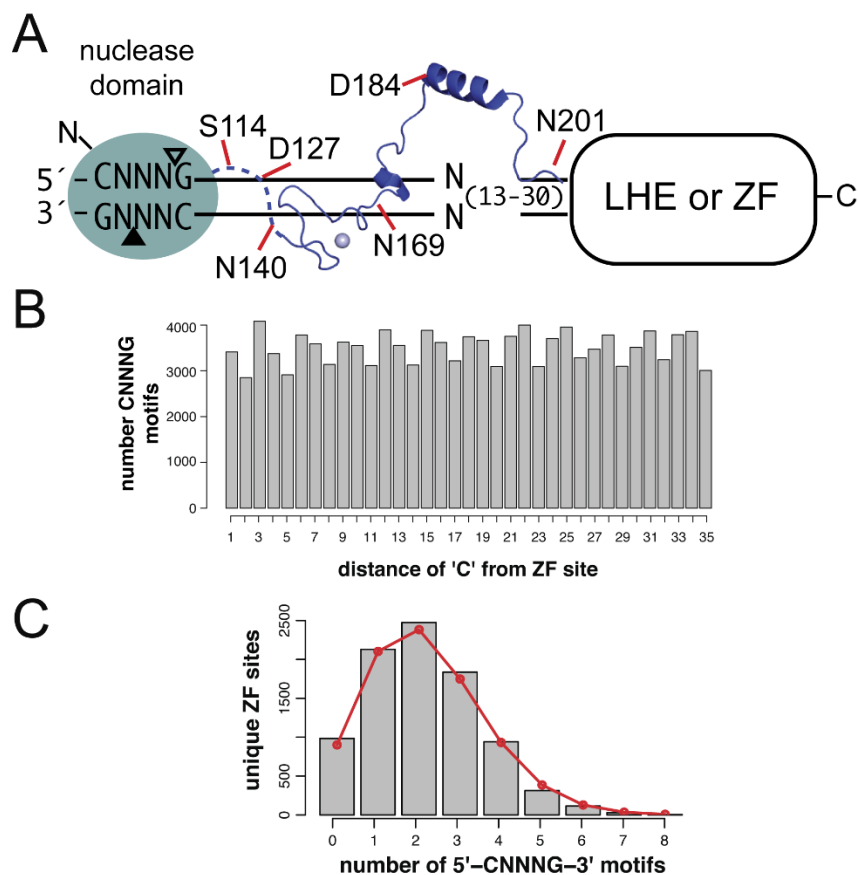


Figure 2.5: Cleavage requirements do not limit GIY-ZFE and GIY-LHE applicability

(A) A diverse set of monomeric and sequence specific reagents can be generated by fusing distinct GIY-YIG domain linker lengths to engineered DNA-binding platforms, including zinc-finger arrays and inactive LAGLIDADGs. (B) Shown is the distribution of the CNNNG motif in a 35-bp window flanking 8,829 predicted ZFN sites on zebrafish chromosome 1. The number of occurrences of the 'C' of the motif at each distance is indicated. (C) Unique ZFN sites were grouped according to the number of occurrences of the CNNNG motif in the 35-bp window. The red line is the expected number of ZFN sites for each group based on a binomial distribution.

is to position the nuclease domain on substrate for cleavage at the CNNNG motif, which is found at a defined distance from the binding site on naturally occurring I-TevI substrates. However, the linker can direct the nuclease domain *in vitro* to search out displaced CNNNG motifs on both native and non-native substrates with insertions or deletions, albeit with reduced cleavage efficiency. Our Tev-LHE fusions recapitulate this distance versus sequence behaviour *in vivo*, as the fusions can cleave displaced CNNNG motifs with a periodicity that parallels the helical nature of DNA. We partially attribute this ability of the Tev-Onu fusions to the flexible N-terminus of I-OnuI. The substrate flexibility of different length Tev-Onu fusions is an important consideration for targeting, as CNNNG motifs at various positions relative to the LHE binding site would be accessible by the choice of the appropriate Tev-LHE fusion. In contrast, the apparently inflexible N-terminus of the 3-member zinc fingers constrains cleavage to a distance of 33-36 bp from the ryA-binding site, mimicking the spacing of the CNNNG motif on native *td* substrate. Our longest Tev-ZFE and Tev-LHE fusions encompass all the known elements of the I-TevI linker that make multiple base-specific and non-specific contacts to DNA (28). However, biochemical studies revealed that I-TevI retains significant cleavage activity on substrates with multiple substitutions in the central region of its cognate DNA substrate that is contacted by the linker, equivalent to positions 6-33 of our longest chimeric substrates (39). The shortest Tev-LHE fusions do not contain any linker elements that are known to make base-specific DNA contacts, and cleave only at the preferred CNNNG motif. This observation implies that the I-TevI linker may contact substrate nucleotides adjacent to the CNNNG motif. Potential contacts may play a role in the positioning of the nuclease domain, rather than being necessary for cleavage, and any preference may be related to regulating the position of the nuclease domain on substrate or the maintenance of DNA-structure (36).

Future work on Tev-ZFEs and Tev-LHEs will require a detailed dissection of binding affinity and specificity, and characterization of cellular toxicity that results from cleavage at off-target sites. In their current form, the targeting specificity of the Tev-ZFEs is a function of the 3-zinc finger domain, which could be further enhanced by addition of zinc fingers to generate a 4-, 5-, or 6-zinc finger fusion, as has been done with a variety of ZFNs (40). In contrast, the ~18-bp specificity of LHEs is sufficient to direct targeting and cleavage at endogenous loci in human cells. LHEs, however, are tolerant of nucleotide substitutions within their recognition sequence, and I-OnuI E1 cleaves off-target sites that differ by one or two nucleotide substitutions (25). In the context of

Tev-LHEs, decoupling of DNA-cleavage and DNA-binding activity by using a catalytically dead LHE scaffold, combined with the requirement for a preferred I-TevI CNNNG cleavage motif, would significantly reduce cleavage at off-target sites (Fig. S2.4). Another advantage of the decoupled activities of Tev-LHEs is that they would not require re-optimization of catalytic activity that is often necessary in LHEs that have been reprogrammed to bind non-native target sites (25, 41). Similar to the exploration of alternative DNA-binding platforms (2), it is imperative to incorporate nuclease domains with distinct biochemical properties into the genome engineering pipeline to create highly precise tools. With further optimization, the I-TevI nuclease domain may become an alternative to the FokI-derived ZFNs and TALENs.

2.5 References

1. Bibikova M, Beumer K, Trautman JK, & Carroll D (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* 300(5620):764.
2. Bogdanove AJ & Voytas DF (2011) TAL effectors: customizable proteins for DNA targeting. *Science* 333(6051):1843-1846.
3. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, & Gregory PD (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* 11(9):636-646.
4. Marcaida MJ, Munoz IG, Blanco FJ, Prieto J, & Montoya G (2010) Homing endonucleases: from basics to therapeutic applications. *Cell Mol Life Sci* 67(5):727-748.
5. Stoddard BL (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* 19(1):7-15.
6. Cathomen T & Joung JK (2008) Zinc-finger nucleases: the next generation emerges. *Mol Ther* 16(7):1200-1207.
7. Christian M, *et al.* (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* 186(2):757-761.
8. Kim YG, Cha J, & Chandrasegaran S (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* 93(3):1156-1160.
9. Halford SE, Catto LE, Pernstich C, Rusling DA, & Sanders KL (2011) The reaction mechanism of FokI excludes the possibility of targeting zinc finger nucleases to unique DNA sites. *Biochem Soc Trans* 39(2):584-588.
10. Bitinaite J, Wah DA, Aggarwal AK, & Schildkraut I (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A* 95(18):10570-10575.

11. Vanamee ES, Santagata S, & Aggarwal AK (2001) FokI requires two specific DNA sites for cleavage. *J Mol Biol* 309(1):69-78.
12. Smith J, *et al.* (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res* 28(17):3361-3369.
13. Fonfara I, Curth U, Pingoud A, & Wende W (2011) Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res* 40(2):847-860.
14. Schierling B, *et al.* (2011) A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res*:doi.10.1093/nar/gkr1112.
15. Dunin-Horkawicz S, Feder M, & Bujnicki JM (2006) Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* 7:98.
16. Mak AN, Lambert AR, & Stoddard BL (2010) Folding, DNA recognition, and function of GIY-YIG endonucleases: crystal structures of R.Eco29kI. *Structure* 18(10):1321-1331.
17. Sokolowska M, Czapińska H, & Bochtler M (2011) Hpy188I-DNA pre- and post-cleavage complexes--snapshots of the GIY-YIG nuclease mediated catalysis. *Nucleic Acids Res* 39(4):1554-1564.
18. Van Roey P, Meehan L, Kowalski JC, Belfort M, & Derbyshire V (2002) Catalytic domain structure and hypothesis for function of GIY-YIG intron endonuclease I-TevI. *Nat Struct Biol* 9(11):806-811.
19. Edgell DR & Shub DA (2001) Related homing endonucleases I-BmoI and I-TevI use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci U S A* 98(14):7898-7903.
20. Mueller JE, Smith D, Bryk M, & Belfort M (1995) Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *EMBO J* 14(22):5724-5735.
21. Bryk M, Belisle M, Mueller JE, & Belfort M (1995) Selection of a remote cleavage site by I-TevI, the *td* intron-encoded endonuclease. *J Mol Biol* 247(2):197-210.
22. Edgell DR, Stanger MJ, & Belfort M (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J Mol Biol* 343:1231-1241.
23. Liu Q, Derbyshire V, Belfort M, & Edgell DR (2006) Distance determination by GIY-YIG intron endonucleases: discrimination between repression and cleavage functions. *Nucleic Acids Res* 34(6):1755-1764.
24. Rosen LE, *et al.* (2006) Homing endonuclease I-CreI derivatives with novel DNA target specificities. *Nucleic Acids Res* 34(17):4791-4800.

25. Takeuchi R, *et al.* (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc Natl Acad Sci U S A* 108(32):13077-13082.
26. Jacoby K, *et al.* (2012) Expanding LAGLIDADG endonuclease scaffold diversity by rapidly surveying evolutionary sequence space. *Nucleic Acids Res*:doi.10.1093/nar/gkr1303.
27. Elrod-Erickson M, Rould MA, Nekludova L, & Pabo CO (1996) Zif268 protein-DNA complex refined at 1.6 Å: a model system for understanding zinc finger-DNA interactions. *Structure* 4(10):1171-1180.
28. Van Roey P, Waddling CA, Fox KM, Belfort M, & Derbyshire V (2001) Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *EMBO J* 20(14):3631-3637.
29. Bibikova M, Golic M, Golic KG, & Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161(3):1169-1175.
30. Chen Z & Zhao H (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res* 33(18):e154.
31. Kleinstiver BP, Fernandes A, Gloor GB, & Edgell DR (2010) A unified genetic, computational, and experimental framework identifies non-conserved residues as critical for function of the homing endonuclease I-BmoI. *Nucleic Acids Res* 38:2411-2427.
32. Edgell DR, Stanger MJ, & Belfort M (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr Biol* 13(11):973-978.
33. Catto LE, Ganguly S, Milsom SE, Welsh AJ, & Halford SE (2006) Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res* 34(6):1711-1720.
34. Dean AB, *et al.* (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc Natl Acad Sci U S A* 99(13):8554-8561.
35. Cermak T, *et al.* (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39(12):e82.
36. Liu Q, *et al.* (2008) Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *J Mol Biol* 379(5):1094-1106.
37. Kowalski JC, *et al.* (1999) Configuration of the catalytic GIY-YIG domain of intron endonuclease I-TevI: coincidence of computational and molecular findings. *Nucleic Acids Res* 27(10):2115-2125.
38. Foley JE, *et al.* (2009) Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS One* 4(2):e4348.
39. Bryk M, *et al.* (1993) The *td* intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J* 12(5):2141-2149.

40. Shimizu Y, *et al.* (2011) Adding fingers to an engineered zinc finger nuclease can reduce activity. *Biochemistry* 50(22):5033-5041.
41. Takeuchi R, Certo M, Caprara MG, Scharenberg AM, & Stoddard BL (2009) Optimization of *in vivo* activity of a bifunctional homing endonuclease and maturase reverses evolutionary degradation. *Nucleic Acids Res* 37(3):877-890.

Chapter 3

3 MegaTevs: Single chain dual-active site nucleases for efficient gene disruption

The work presented in this chapter is reproduced (with permission, Appendix S1) from:

Wolfs, J.M., Dasilva, M., Meister, S.E., Wang, X., Schild-Poulter, C. and Edgell, D.R. (2014) MegaTevs: Single-chain dual nucleases for efficient gene disruption. *Nucleic Acids Res.*, **42**, 8816–8829.

3.1 Introduction

The rapid pace of development in the genome-editing field has led to a number of competing technologies, each with their benefits and limitations (1,2). The technologies can be broadly characterized based on the nuclease domain used to introduce a double-strand break (DSB) or nick at a target site. Two common reagents are zinc-finger nucleases (ZFNs) and TAL effector nucleases (TALENs) that utilize the dimeric and non-specific FokI nuclease domain (3-6). Two head-to-head ZFN or TALENs pairs must be designed to target a single site, and positioned such that the FokI domains can dimerize to introduce a DSB (7-9), typically with 4-nt 5' overhangs. The non-specific cleavage activity of the FokI nuclease domain facilitates targeting of a wide range of sequences, but can lead to off-target cleavage (10-12). The recently developed CRISPR/Cas9 system has received significant attention due to the ease of programming targeting (13,14). In this system, a RNA guide molecule (the crRNA) targets the Cas9 nuclease to a DNA target through a RNA/DNA heteroduplex (15-18). A blunt-ended DSB results from two independent nicking reactions, one by a HNH nuclease domain, and the other by a RuvC-like domain. An alternative nuclease architecture utilizes the naturally occurring meganucleases, or LAGLIDADG family homing endonucleases, which are typically encoded within self-splicing group I introns and inteins, and are characterized by an extensive protein-DNA interface (19,20). The nuclease active site is formed at the interface of two parallel α -helices, with cleavage generating a DSB with 4-nt 3' overhangs. Recently, a number of recombinase (21,22) and sequence-specific nuclease domains have been developed as alternatives to the non-specific FokI nuclease domain (23-26). In particular, we and others showed that the monomeric and sequence tolerant GIY-YIG nuclease domain from the homing endonuclease I-TevI could be fused to zinc fingers, meganucleases, and

TAL effectors to create novel monomeric enzymes (27,28). The I-TevI-based reagents are active on substrates that contain a preferred CNNNG cleavage motif, generating 2-nt 3' overhangs.

Regardless of the technology, one common application of genome-editing nucleases is the generation of gene disruptions whereby mutagenic repair at the targeted DSB introduces frame-shift mutations into a coding region. The mutagenic DNA repair events occur in the absence of an exogenously provided DNA repair template, and result mainly from the non-homologous end-joining (NHEJ) pathway (29-31). However, many DSBs are repaired without mutation, as the compatible cohesive ends generated by the nucleases are re-ligated through the canonical NHEJ pathway, leading to a cycle of persistent cleavage and precise repair events that are non-productive for genome engineering. One strategy to bias repair events towards gene disruption is to co-express a DNA end-processing enzyme with the genome-editing nuclease (32,33). For instance, Trex2, a 3'-5' exonuclease, dramatically increases gene disruption when co-expressed with ZFNs, TALENs, and meganucleases by processing of DSBs before DNA repair. One potential limitation of this strategy is the requirement to transfect the Trex2 coding region with the ZFN, meganuclease or dimeric TALEN constructs, which may be problematic in size-constrained vectors. Overexpression of Trex2 could also enhance mutagenic repair at unwanted off-target sites, although no increases in cellular toxicity or off-target cleavages were observed with Trex2 overexpressing cell lines (32,33). Gene disruption could be enhanced by targeting two reagents to the same locus, positioning two DSBs to effectively excise the intervening sequence and introduce a deletion. Such multiplexing of genome-editing reagents is constrained by the dimeric architecture of the ZFNs and TALENs (34), and in the case of the CRISPR/Cas9 system, requires the use of nicking variants and dual-guide RNAs (35,36).

Here, we propose an alternative strategy for gene disruption by coupling two different nuclease active sites into a single polypeptide. The MegaTev architecture is the fusion a meganuclease (Mega) with the nuclease domain derived from the GIY-YIG homing endonuclease I-TevI (Tev). The two active sites are positioned ~30-bp apart on DNA substrate, and generate two DSBs with non-compatible cohesive ends. The dual active MegaTev shows high gene disruption activity in HEK 293 cells without overexpression of DNA-end processing enzymes.

3.2 Material and methods

3.2.1 Bacterial strains and plasmid construction

Escherichia coli DH5 α (New England Biolabs) was used for plasmid amplification, ER2566 (New England Biolabs) for protein expression and BW25141 (λ DE3) for bacterial two-plasmid selections (Supplemental Table S1) (37). Tev-Onu and Tev-Ltr fusions were cloned into pACYCDuet-1 using 5' NcoI and 3' XhoI sites as previously described (27). For the yeast DNA repair assay (3,38), the Tev-Onu and Tev-Ltr genes were amplified using Phusion DNA polymerase (New England Biolabs) with a 3' primer that introduced a C-terminal SV40 nuclear localization sequence (NLS) (primers are listed in Supplemental Table S2). PCR products were cloned into the NcoI/SalI sites of pGPD. The backbone target site plasmid for the yeast assay was created by amplifying a 300-bp fragment of the pTox plasmid and cloning it BglII/SpeI into pCP5.1 to create pCPTox. All target sites were subsequently cloned into pCPTox using *in vivo* homology directed repair. For mammalian assays, human codon optimized Tev-Onu fusions (synthesized by IDT-DNA) were PCR amplified and cloned PstI and RsrII into pExodus. Tev-Onu fusions were cloned in-frame with a mCherry gene linked by a T2A peptide sequence from *Thosea asigna* virus to separate the translated proteins. The TO15 target site was subcloned into the pMSCVpuro retroviral vector (Clontech) using BglII and XhoI sites to integrate into genomic DNA in HEK 293 cells. To generate target sites for episomal plasmid assays, substrates were cloned the SacI/XhoI sites of pcDNA3(+) vector. Constructs were confirmed by sequencing.

3.2.2 *In vitro* randomized substrate selection

A list of randomized target site oligonucleotides is found in Supplemental Table S2. The target site plasmid library for the randomized cleavage motif plus 3 bps of the spacer (N8) were constructed in the pSP72 backbone as described (39). The library complexity was estimated to be $\sim 6.4 \times 10^4$ for the N8 library based on the number of independent transformants, and from analyses of next-generation sequencing data.

Cleavage assays were performed with 23 nM of Tev169-Onu E22Q and 10 nM N8 plasmid in NEBuffer 3 (50 mM Tris-HCl pH 7.9, 100 mM NaCl, 10 mM MgCl₂ and 1 mM DTT) at 37°C for 5 mins. The Tev169-Onu fusions were purified as described (27) (Supplemental Fig. S1). Samples were prepared for Ion Torrent sequencing at the London Regional Genomics Centre by PCR amplification of the target site region from the input plasmid library, and from the plasmids isolated

after three rounds of selection using PWO (Roche) with barcoded primers. The sequencing data were parsed with custom Perl scripts that checked for anchor sequences either side of the randomized region, confirmed that the sequence between the anchors corresponding to the randomized region was 8 nts in length, and then extracted the randomized region for further analyses. For each round of selection, counts for each nucleotide j per position i were determined, and then converted to proportions using the centred log-ratio transformation

$$C_{i,j} = \log_2(p_{i,j}) - \text{mean } \log_2(p_j)$$

Nucleotide selection was then determined by taking the difference in proportions for each nucleotide per position between the final round of selection and the input library. A positive value indicates selection or enrichment for a particular nucleotide relative to the input library, and a negative value indicates selection against a particular nucleotide relative to the input. The enrichment values were plotted in heatmap format using R and ggplot2 (40,41), and enrichment values were considered significant if they were > 2 standard deviations from the mean enrichment value for each dataset.

3.2.3 Cleavage assays on radiolabelled substrate

In vitro cleavage assays were performed on internally radiolabelled substrates that were PCR amplified with [α - 32 P] dCTP. PCR products were loaded onto an 8% (w/v) polyacrylamide gel (29:1 acrylamide/bisacrylamide), run at 40 mA for ~1.5 hrs, gel purified, eluted overnight at 42°C in 5 mL of TE pH 8.0, and concentrated into 50- μ l volume. Cleavage reactions were performed in 20 μ L reaction volumes with NEBuffer 3 (50 mM Tris-HCl pH 7.9, 100 mM NaCl, 10 mM MgCl₂ and 1 mM DTT), 0.1 pmol of substrate and 2 pmols of Tev169-Onu fusion protein. Cleavage reactions were incubated at 37°C for 1, 5, 10 and 25 mins before stopping the reaction with 6 μ L of 100 mM EDTA and 5 μ L of loading dye containing 0.5% SDS. Mutant Tev169-Onu fusions were incubated for 1 hr at 37°C before stopping the reaction with 6 μ L of 100 mM EDTA and 5 μ L of loading dye containing 0.5% SDS. The entire reaction was loaded on a 15% (w/v) polyacrylamide gel (29:1 acrylamide/bisacrylamide) and electrophoresed at 40 mA for ~1.5 hrs. The gel was removed from the apparatus and soaked in 10% glycerol plus 8% acetic acid before drying on Whatman paper and visualized using a phosphorimager (GE Healthcare).

3.2.4 Modified two-plasmid target site screen

The 64-triplet variants for the CNNNG cleavage motif were screened using a modified two-plasmid selection (42). Transformants were gridded onto selective plates (LB plus 25 µg/mL chloramphenicol and 10 mM L-(+)-arabinose) and non-selective plates (LB plus 25 µg/mL chloramphenicol, 50 µL/mL kanamycin and 0.2% glucose) and incubated overnight at 37°C. Plasmids isolated from survivors and non-survivors were sequenced to identify the NNN variant of the CNNNG motif.

3.2.5 Yeast β -galactosidase repair assay

This assay was performed as described (27). Briefly, YPH499(a) containing target site constructs were mated in triplicate with YPH500(α) harbouring the MegaTev constructs. After an overnight selection for diploids, cells were assayed for β -galactosidase activity using orthonitrophenol (ONPG). Activity was normalized to either a validated homodimeric zinc-finger nuclease (Zif268) or the wild-type TP15 substrate depending on the assay.

3.2.6 DdeI-resistance assays with plasmid substrates

HEK 293T cells were cultured in Dulbecco's modified eagle medium (DMEM) supplemented with 10% fetal bovine serum (FBS). Approximately 2.5×10^6 million cells were seeded 24 hrs before transfection on 6 cm plates. Cells were co-transfected with 3 µg of pExodus Tev169-Onu E22Q and 3 µg of pcDNA3(+) TO15 using calcium phosphate and incubated at 37°C with 5% CO₂ for 16 hrs before changing media. After 48 hrs, plasmid was isolated from HEK 293T cells using the BioBasic miniprep kit. Target sites were PCR amplified, separated on a 1% agarose gel, and gel purified. After gel purification, 250 ng of PCR product was incubated with 2 U of DdeI (New England Biolabs) in NEBuffer2 for 1 hr at 37°C. Digests were electrophoresed on a 1.5% agarose gel, stained with ethidium bromide, and analyzed on an AlphaImagerTM3400 (Alpha Innotech).

3.2.7 Surveyor assays with integrated targets

Target site integration into HEK 293 cells was performed using the Phoenix Amphi retroviral packaging cell line. To accomplish this, 8 µg of pMSCV TO15 was transfected into Phoenix cells using calcium phosphate and incubated at 37°C with 5% CO₂ for 48 hrs. Media was removed and filtered through a 0.45 µm filter into a falcon tube containing 6 µL of 4 mg/mL polybrene

(hexadimethrine bromide), and 6 mls of virus solution was used to infect HEK 293 cells to create the integrated cell line (HEK 293-TO15). Approximately 24 hrs before transfections, $\sim 2.5 \times 10^6$ HEK 293-TO15 cells were seeded on 6 cm plates, and subsequently transfected with 6 μg of pExodus Tev169-Onu or pExodus I-SceI. After 48 hrs, the HEK 293-TO15 cells were harvested and total genomic DNA isolated. Two rounds of nested PCR were performed, and gel purified PCR products were boiled at 95°C for 10min, then cooled slowly to 50°C before flash freezing at -20°C for 2 mins. To assay for indels, 200 ng of PCR product was incubated with 2 U of T7 endonuclease I (New England Biolabs) in NEBuffer2 for 1 hr at 37°C, separated on a 1.5% agarose gel, and analyzed using an AlphaImagerTM3400 (Alpha Innotech).

3.3 Results

3.3.1 MegaTevs: Chimeric fusions of GIY-YIG and meganuclease components

To determine if the I-TevI nuclease domain could function in the context of different meganucleases, we fused residues 1-169 of I-TevI (Tev169) to the native N-terminus of the catalytically inactive I-LtrI variant, I-LtrI E29Q, to create Tev169-Ltr E29Q (Tev-xLtr) (Fig. 3.1A). Along with the previously constructed Tev169-Onu E22Q (Tev169-xOnu) (27), we assayed the activity of both MegaTevs using a yeast recombination assay where a target site is positioned between a partially duplicated *lacZ* gene. Cleavage of the target site induces the single-strand annealing pathway to reconstitute a functional *lacZ* gene resulting in β -galactosidase activity. We tested activity on hybrid target sites consisting of the native I-TevI CNNNG cleavage motif (5'-CAACG-3') and DNA spacer derived from the phage T4 thymidylate synthase (*td*) gene fused to either the I-OnuI or I-LtrI binding site (TO or TL, respectively) (Fig. 3.1A). The substrates differed in the length of the DNA spacer (from 11 to 21-bp) separating the I-TevI CNNNG cleavage motif from the I-OnuI or I-LtrI binding site. As shown in Figure 3.1B, Tev169-xOnu and Tev169-xLtr activity was highest with a DNA spacer length of 15 bp, agreeing with results from profiling DNA spacer length requirements of the Tev-xOnu construct in a bacterial two-plasmid survival assay (27). Furthermore, mutating the critical cleavage CNNNG motif to ANNNA abolished activity for both Tev169-xOnu and Tev169-xLtr on the 15-bp spacer substrate [TO15CS(-) and TL15CS(-)], demonstrating the I-TevI nuclease domain maintains cleavage specificity in the context of a meganuclease fusion. To demonstrate that the MegaTevs are directed to their target sites by the

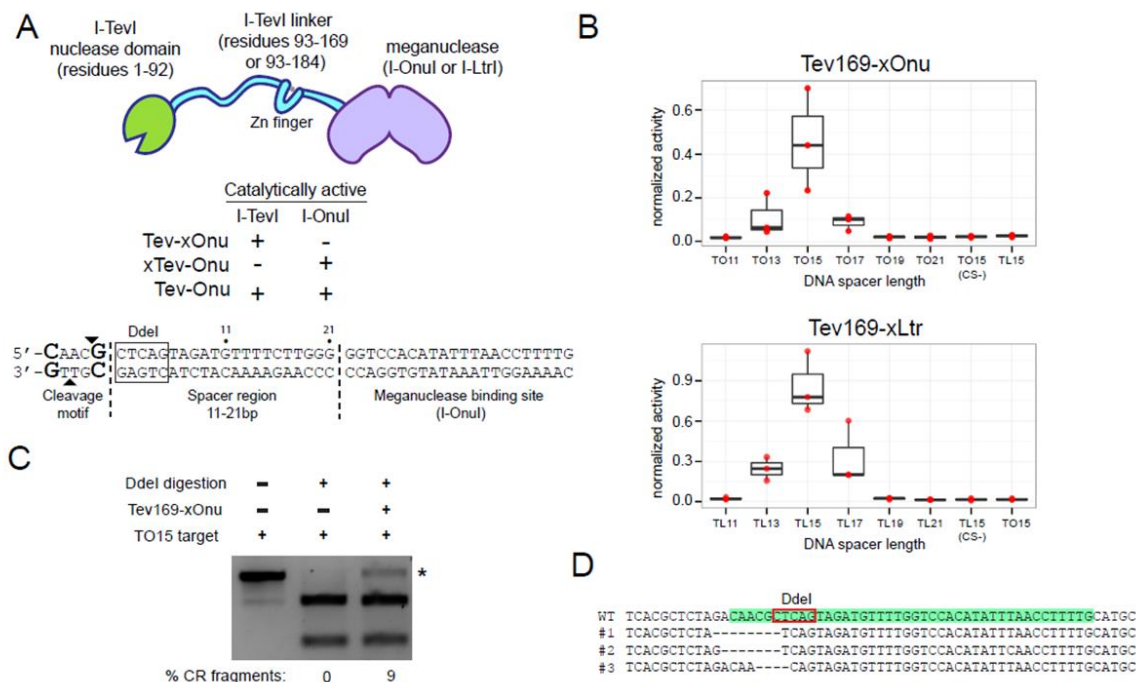


Figure 3.1 MegaTev activity in yeast and HEK 293T cells

(A) (*top*) Schematic for the modular MegaTev fusions, consisting of the I-TevI nuclease domain and various linker lengths fused to a meganuclease. Catalytically inactive R27A I-TevI and E22Q I-OnuI are denoted by xTev and xOnu, respectively. (*bottom*) The modular composition of the target site consisting of the 5'-CNNNG-3' (CAACG) cleavage motif and variable length spacer region (11-21bp) from the native I-TevI thymidylate synthase (*td*) gene located upstream of the meganuclease binding site. The arrows highlight the top and bottom I-TevI cleavage sites. Highlighted in the box is the DdeI site adjacent to the I-TevI cleavage site used to identify mutagenic events in HEK 293T cells. (B) Boxplot of the activity for Tev169-xOnu (*Top*) and Tev169-xLtr (*Bottom*) on various DNA spacer lengths in the yeast recombination assay. TO15CS(-), AAACA cleavage motif in the TO15 target site; TL15CS(-), AAACA cleavage motif in the TL15 target site; xLtr, a catalytically inactive E29Q I-LtrI. Activity was normalized to the homodimeric FokI-Zif268 fusion and data are plotted with SD for n=3. (C) DdeI digestion of PCR products to identify mutagenic events for an episomal target in HEK 293T cells. Shown is a representative agarose gel for DdeI digested PCR amplified target sites from plasmids isolated 48 hrs post-transfection. Indicated by the asterisk is the DdeI cleavage-resistant product, with percent product indicated below each lane. (D) Cleavage-resistant products from HEK 293T cells transfected with Tev169-xOnu were cloned and analyzed by sequencing. The TO15 target site is highlighted in green and in the red box is the DdeI site used to identify cleavage resistant products. Deletions encompass the I-TevI cleavage site with the length of deletion illustrated by the dashed lines.

meganuclease and not the I-TevI nuclease domain and linker, we tested the Tev-xOnu against the TL15 target site, and tested Tev-xLtr against the TO15 target site. No activity was observed for either fusion on the reciprocal substrates (Fig. 3.1B), showing that the I-TevI nuclease domain does not direct targeting of the MegaTevs.

To further demonstrate the genome-editing relevance of MegaTevs, we constructed a Tev169-xOnu fusion codon-optimized for expression in human cells, and co-transfected it and the TO15 target site cloned on a separate plasmid into HEK 293T cells. To monitor Tev-xOnu activity, we took advantage of a DdeI site that lies immediately downstream of the I-TevI CNNNG motif (Fig. 3.1A). Plasmid substrates cleaved *in vivo* by the MegaTev would be subject to mutagenic non-homologous end joining repair, destroying the DdeI site. Subsequent resistance of target sites to DdeI digestion after PCR amplification from genomic DNA reflects Tev-xOnu cleavage and mutagenic repair. Significantly, we observed 9% DdeI cleavage resistance for target sites amplified from unsorted HEK 293T cells co-transfected with Tev-xOnu and the TO15 target site (Fig 3.1C). In contrast, no DdeI resistant products were observed for cells transfected with the TO15 target site plasmid only. Cleavage resistant products were cloned and sequenced, revealing deletions that spanned the I-TevI cleavage site (Fig. 3.1D), demonstrating that MegaTevs function to induce mutagenic DNA repair in human cells.

3.3.2 Dual active site MegaTevs for highly efficient targeted deletions

A unique aspect of the MegaTevs is the fusion of two homing endonuclease active sites into a single polypeptide chain. Each active site is positioned such that the top-strand nicking sites are separated by ~30 bp on the TO15 DNA substrate. When both active sites are functional, this arrangement presents the opportunity to introduce two DSBs with different cohesive ends at a single site in a highly efficient and concerted process. As a proof of concept, we constructed and purified a Tev169-Onu dual nuclease where both the I-TevI and I-OnuI active sites are wild type (Tev169-Onu, Fig. 3.2A). Activity was tested *in vitro* utilizing an internally radiolabelled PCR product of 242-bp containing the TO15 target site. Cleavage by the dual nuclease would be evident by the release of 29-bp product corresponding to the internal sequence between the I-TevI and I-OnuI cleavage sites (Fig. 3.2A). As shown in Figure 3.2B, the dual active MegaTev efficiently produced three products after 5 mins of digestion, with the accumulation of the internal product

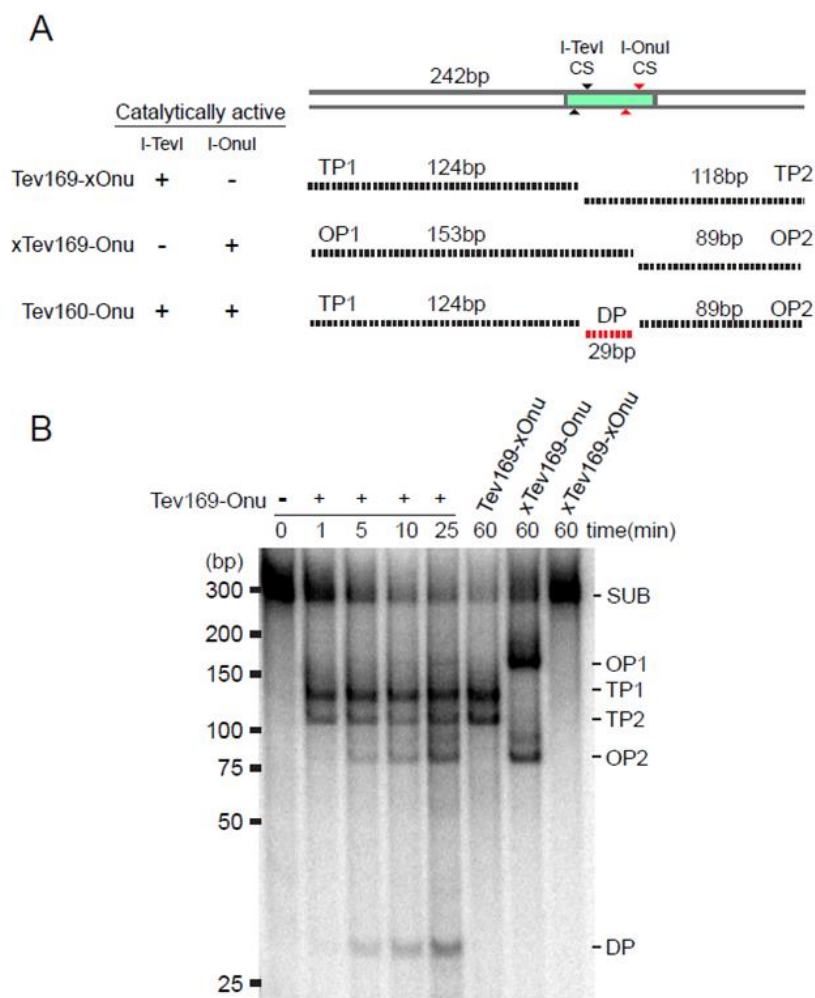


Figure 3.2: Dual active site MegaTev activity *in vitro*

(A) Schematic of substrate and cleavage products containing the TO15 target site indicated by a green-filled rectangle. The black and red arrows indicate the top and bottom cleavage sites (CS) for I-TevI and I-OnuI, respectively. Black dashed lines indicate I-TevI cleavage products (TP1 and TP2), I-OnuI cleavage products (OP1 and OP2), and the red dashed line is the internal dual product (DP) from both I-TevI and I-OnuI cleavage. Sizes of the substrate and products are indicated in base pairs (bp). (B) Polyacrylamide gel of internally labeled TO15 PCR substrate (SUB) incubated with (+) or without (-) purified dual active Tev169-Onu and single-active site variants. The MegaTevs were incubated with substrate for the indicated times, and cleavage products are indicated on the right side of the gel based on their predicted sizes from panel (A). The sizing standard (in bp) was cropped from the gel image.

(IP) after 25 mins. Interestingly, cleavage by the I-TevI nuclease domain precedes cleavage by I-OnuI, as I-TevI-specific products are detected at 1 min (TP1 and TP2), whereas I-OnuI products are detected after 5 mins (OP1 and OP2). Cleavage assays where I-TevI is active and I-OnuI is inactive (Tev169-xOnu) produced two products consistent with only I-TevI cleavage activity. Similarly, a catalytically inactive R27A I-TevI in the context an active I-OnuI fusion (xTev169-Onu) produced two products consistent with I-OnuI cleavage. No cleavage was observed for the dual dead nuclease (inactive I-TevI R27A and I-OnuI E22Q, xTev-xOnu) after 1 hr of incubation. We also constructed and purified an analogous Tev169-Ltr dual nuclease, and tested for activity on an internally labeled PCR product containing the TL15 site (Supplemental Fig. S3.1). I-TevI-specific cleavage products were observed before I-LtrI products, with the internal cleavage product visible after 5 mins of incubation.

To extend these results to an *in vivo* context, we stably integrated the TO15 target site into the genome of HEK 293 cells, and assayed for activity after independent transfections with the Tev169-Onu dual nuclease and single nuclease variants. The expression constructs included the MegaTev open-reading frame (ORF), followed by an in-frame T2A peptide and mCherry ORF, allowing us to monitor MegaTev expression by mCherry levels (Fig. 3.3A). After 48 hours of incubation, the TO15 target site was PCR amplified and subject to digestion by T7 endonuclease I (T7EI) to evaluate the extent of MegaTev activity. T7EI was used rather than DdeI digestion to allow us to more accurately detect I-OnuI-specific events with the xTev-Onu construct. As shown in Figure 3B, the dual active Tev169-Onu showed 24 +/- 6% indels, whereas undetectable levels of cleavage were observed with both of the single active site Tev169-Onu fusions (xTev169-Onu and Tev169-xOnu), in spite of similar expression levels (Fig. 3.3A). It is important to note that we did not co-express DNA end processing enzymes, such as the 3'-5' exonuclease Trex2, to enhance mutagenic repair at the TO15 cleavage site. Extremely low levels of cleavage by the xTev169-Onu variant, where I-OnuI E2 is active, is consistent with previous studies that required sorting for cells with high I-OnuI expression levels and multiple rounds of PCR enrichment to visualise I-OnuI E2 cleavage (43).

To confirm that cleavage by the dual MegaTev resulted in the deletion of the internal sequence between the I-TevI and I-OnuI cleavage sites, the PCR products that were resistant to cleavage by DdeI were cloned and sequenced. We recovered three sequences (M1, M2 and M3) that each

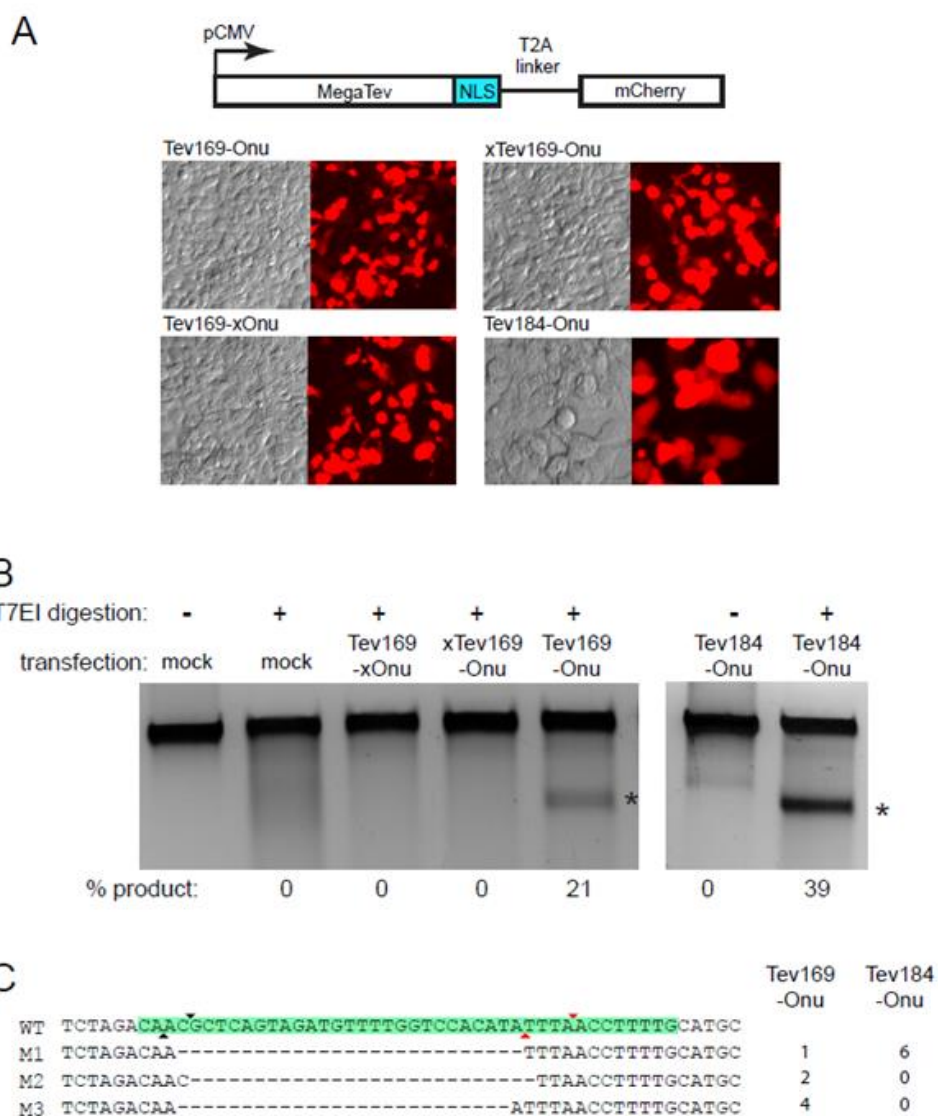


Figure 3.3: Dual active site MegaTev activity in HEK 293 cells

(A) Schematic of the expression construct, with the MegaTev ORF linked to mCherry through a T2A peptide. The CMV promoter was used to express the constructs. NLS, nuclear-localization signal. Representative contrast and fluorescence images for cells transfected with the indicated MegaTev fusions are shown below. (B) T7 endonuclease I (T7EI) digests of TO15 target site PCR products from HEK 293T cells transfected with the indicated MegaTevs. The extent of T7EI digestion is indicated below each lane. An asterisk indicates the T7EI cleavage products. (C) Summary of sequencing of PCR amplified TO15 target sites that were resistant to cleavage by DdeI. Three sequence types were found, indicated by M1, M2 and M3. The number of independent occurrences that each sequence type was identified from Tev169-Onu and Tev184-Onu transfections is indicated on the right.

occurred multiple times, and in each case the intervening sequence was deleted (Fig. 3.3C). We also transfected another dual active MegaTev construct containing the I-TevI 1-184 fragment fused to I-OnuI E2 (Tev184-Onu) into the HEK 293 cells, and observed a similar high level of activity (Fig. 3.3B). Sequencing of DdeI-resistant PCR fragments from the Tev184-Onu transfection revealed only the M1 sequence that was observed with the Tev184-Onu construct (Fig. 3.3C). One explanation for the recovered sequences is that the I-TevI 2-bp 3'-overhang is ligated directly to the I-OnuI 4-bp 3'-overhang with minimal processing and fill-in of the 2-bp gap.

Collectively, our data show that the fusion of two homing endonuclease active sites into a single polypeptide chain creates a dual nuclease that can introduce two DSBs at a single target site. In both an *in vitro* and *in vivo* context, the two DSBs excise a short internal fragment, thus increasing the efficiency of targeted deletion without expression of DNA end-processing enzymes.

3.3.3 Requirements of the I-TevI CNNNG cleavage motif *in vivo*

Efficient cleavage by the I-TevI nuclease domain requires that the 5'-CNNNG-3' cleavage motif be spaced 15 bp from the meganuclease binding site (Fig. 3.1B). The C and G of the motif are critical for cleavage, while the central three bases (the NNN triplet) exhibit a substantial degree of tolerance to substitution (44,45). To determine the tolerance of the central three bases to substitution in the context of the MegaTev fusion, we used a variation of a two-plasmid bacterial selection to rapidly assess survival of all 64 variants of the central triplet cloned into the toxic plasmid. In this assay (42), the 64 toxic plasmids were transformed into cells harbouring the Tev169-xOnu fusion, plated on non-selective plates, and then replica-gridded onto selective plates to induce expression of the *ccdB* gene on the toxic plasmid. Cells survive this challenge if Tev169-xOnu can cleave and promote elimination of the toxic plasmid. The Tev169-xOnu fusion was used for this experiment to ensure that survival was due to I-TevI activity and not I-OnuI activity. Three different morphologies were observed on the selective plates; no growth (dead), colonies that grew to the same diameter on both the selective and non-selective plates (strong survivors), and colonies that grew on selective plates but were smaller in diameter than on non-selective plates and often formed a cauliflower morphology (weak survivors) (Supplemental Fig. S3.4). Weak survivors were determined to be target sites that promoted < 1% percent survival when assayed individually. Survival was plotted in a heatmap format (Fig. 3.4A) revealing that C/G rich triplets generally inhibited Tev169-xOnu survival.

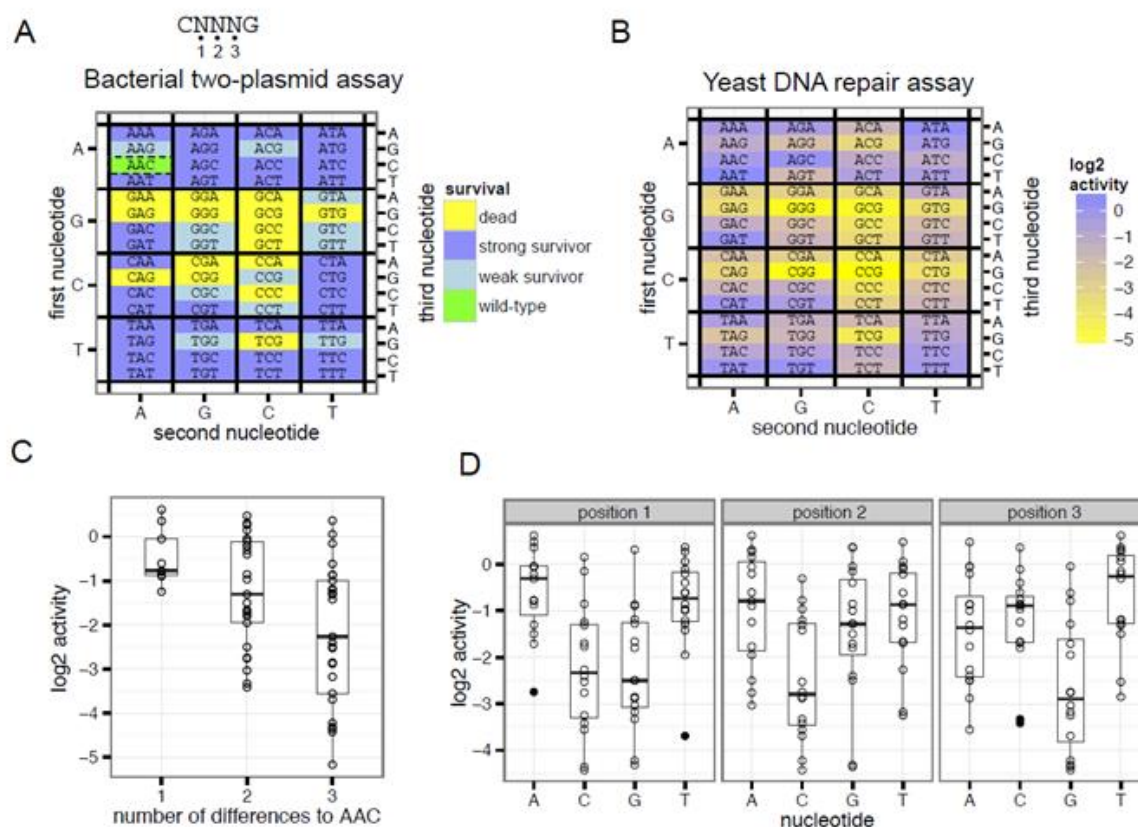


Figure 3.4: MegaTev nucleotide preference within the CNNNG cleavage motif

(A) The three nucleotides of the NNN triplet are labeled as 1, 2, and 3 from 5' to 3' within the CNNNG motif. Heat map of Tev169-xOnu activity on all 64 variants of the cleavage motif in the bacterial two-plasmid selection. The first nucleotide of the NNN triplet is plotted on the left axis, the second nucleotide of the NNN triplet on the bottom axis, and the third nucleotide on the right axis. (B) Heat map of the activity of Tev169-xOnu on all 64-nucleotide variants in the yeast β -galactosidase repair assay. The values are normalized to the wild-type sequence (AAC) and plotted on a log₂ scale. The heatmap is labeled as in panel (A). (C) Boxplots depicting the log₂ activity for NNN triplets that have 1, 2 or 3 nucleotide differences relative to the wild-type AAC sequence. For each boxplot, the upper and lower bounds of the boxes indicate the 25th and 75th percentile of the data, the solid horizontal bar indicates the median of the data, and the ends of the whiskers represent 1.5 times the interquartile range. Individual data points are shown as open circles, and data points that lie outside of the interquartile range (outliers) are shown as black points. (D) Boxplots of the influence of nucleotide identity at each of the three positions within the NNN motif on activity relative to the wild-type AAC sequence. The boxplots are labeled as in panel (C).

We also assayed all 64-triplet variants in the quantitative yeast-based *lacZ* repair assay, and plotted the activity for each triplet normalized to the activity of the wild-type AAC triplet on a log₂ scale (Fig. 3.4B). In general, more substitutions within the triplet resulted in lower activity, however some triplets with 2 or 3 substitutions were as active as the wild-type AAC triplet (Fig. 3.4C). A/T rich target sites displayed activity on par with the wild type sequence. As with the bacterial two-plasmid selection, triplets with a C and/or G in the first two positions supported lower β -galactosidase activity, and triplets with a G in the third position were less active than other nucleotides at this position (Fig. 3.4D).

3.3.4 MegaTevs selects for the appropriately spaced cleavage motif from a random substrate

To determine the optimal sequence and spacing of the cleavage motif, we generated a plasmid library where the CAACG cleavage motif and three downstream base pairs were completely randomized (Fig. 3.5A, the N8 library). A round of *in vitro* selection with the N8 plasmid library consisted of *in vitro* digestion with purified Tev169-xOnu, isolation of linearized plasmid, religation and transformation into *E. coli* for amplification. After three rounds of selection, the input library and the final round of selection were sequenced using the Ion Torrent platform. After data processing, we first determined the proportion of all 16 possible dinucleotide combinations (ANNNA, ANNNC, ANNNG, etc) regardless of position within the N8 randomized region to ascertain if the MegaTev displayed a preference for the CNNNG motif. As shown in Figure 5A, the CNNNG motif was greatly enriched by round 1 relative to the other dinucleotide combinations, and predominated by round 3, indicating that CNNNG is the preferred motif. Although minor enrichment relative to the input library was observed for other dinucleotide combinations (ANNNA and ANNNT), these combinations do not support activity in cell-based assays, and are not considered relevant.

We next analyzed the phasing of the CNNNG motif within the N8 randomized region for reads containing this motif. This analysis was undertaken as native I-TevI can cleave the wild-type CAACG motif that has been moved closer to the primary binding site (albeit with lower efficiency) (46). The statistical occurrence of the CNNNG motif is 1 in 15 bps, and would be expected to occur at the four possible positions within the input N8 library (C1:G5, C2:G6, C3:G7, and

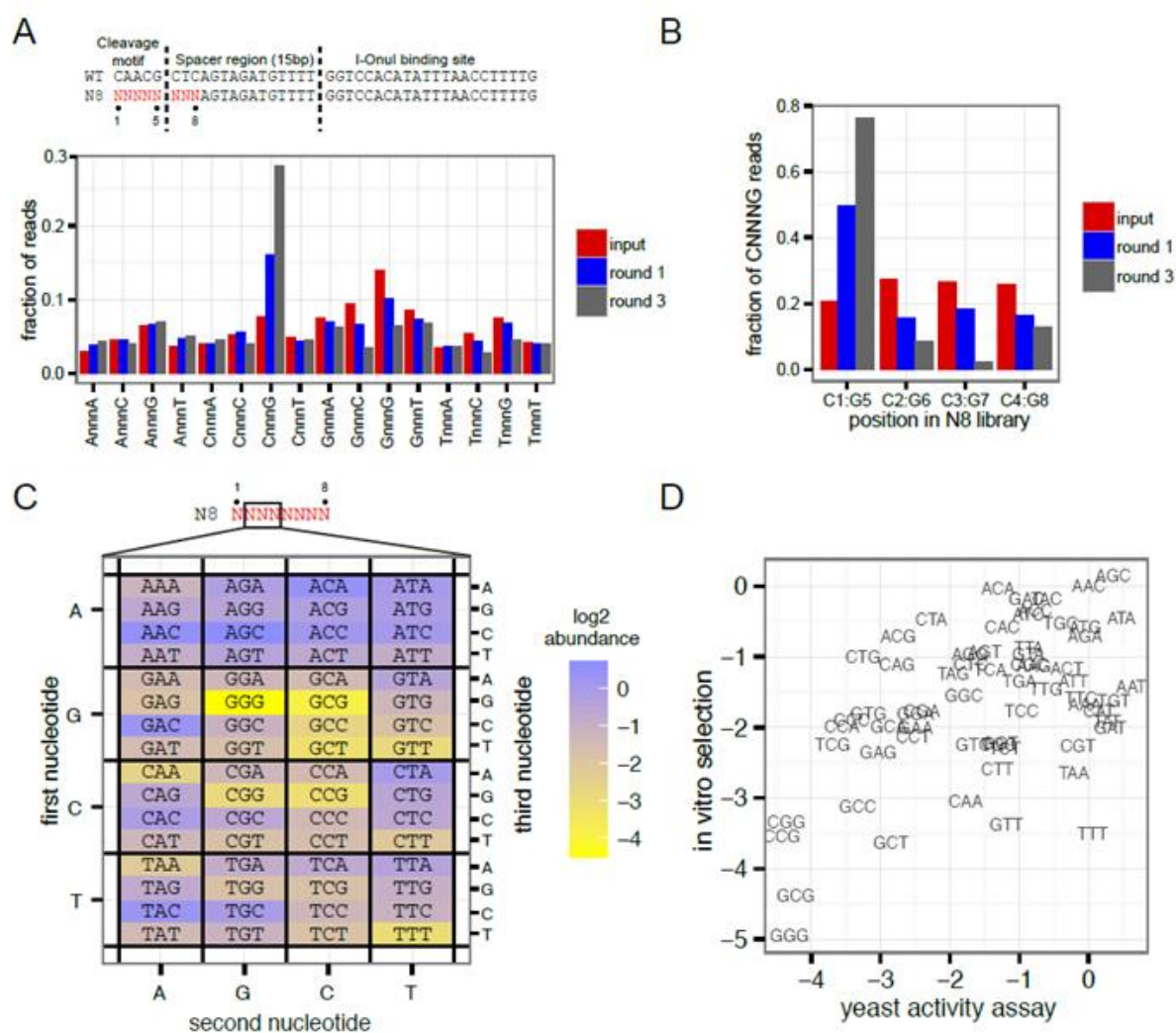


Figure 3.5: MegaTev selects for an appropriately spaced cleavage motif

(A) At top is a schematic of the TO15 substrate and N8 randomized library with the eight randomized positions colored in red. The randomized positions are labeled 1 to 8, with position 1 corresponding to the C of the CNNNG cleavage motif. Shown below is a bar graph showing the fraction of reads with each of the 16 possible dinucleotide combinations. (B) Bar graph showing the fraction of reads with the CNNNG cleavage motif at positions C1:G5, C2:G6, C3:G7, and C4:G8 for the input library and selection rounds 1 and 3. (C) Heat map plotted on a log₂ scale showing abundance of all 64 CNNNG motifs after three rounds of selection with the Tev169-xOnu fusion. The first, second and third nucleotide corresponds to the second, third and fourth nucleotide in the N8 library, respectively. (D) Correlation between the *in vitro* abundance from the selection experiments and *in vivo* activity in the yeast-based assay for each NNN triplet.

C4:G8). Indeed, we observed the CNNNG motif with approximately equal frequency at the four possible positions within the input N8 library (Fig. 3.5B, red bars). However, when the sequencing reads for selection rounds 1 and 3 were analyzed, we found an overwhelming preference for C and G at positions 1 and 5, respectively (Fig. 3.5B, blue and grey bars). This analysis confirms that the MegaTev is cleaving at correctly positioned CNNNG motifs within the N8 library, with the G of the motif positioned 15-bp from the LHE binding.

Using sequencing reads that possessed the CNNNG motif at positions 1 and 5 of the randomized region, we determined the abundance of each NNN triplet within the motif and plotted the log₂ abundance in a heatmap format (Fig. 3.5C). As anticipated, A/T rich triplets were preferred over G/C rich triplets. This analysis also facilitated a comparison to the activity of each NNN triplet in the yeast DNA repair assay (Fig. 3.4B). Plotting the normalized abundance of each NNN triplet from the sequencing data versus the activity in the yeast-based assay showed that A/T rich NNN triplets supported higher activity relative to G/C rich sequences in both datasets (Fig. 3.5D).

3.3.5 Nucleotide preference within the CNNNG motif and at flanking positions

We next analyzed nucleotide preferences at each position in the N8 library after the third round of selection. Nucleotide preferences were determined by calculating the proportional abundance of each nucleotide at each position for both the input library and round 3 selection, and then plotting the difference (enrichment) between round 3 and input as a heatmap (Fig. 3.6A). One advantage of this analysis is that it corrects for nucleotide bias in the input library. As shown in Figure 6A, apart from the expected C and G preference at positions 1 and 5, the strongest preference was observed at position 7, where T or A were selected for while a C or G were selected against. Interestingly, in four of the positions (3, 6, 7, and 8), the wild-type nucleotide was not preferred, implying that the native *td* target site of I-TevI is not the optimal substrate.

To provide an *in vivo* context for the nucleotide preferences observed in the flanking DNA sequence, we independently made point substitutions in the TO15 substrate at positions 6 and 7, and tested their activity in the yeast β -galactosidase assay (Fig. 3.6B). Substitutions to A or T at position 6 did not drastically reduce activity as compared to the TO15 substrate, while the C6G substitution showed a modest increase in activity. At position 7, the T7A substitution reduced

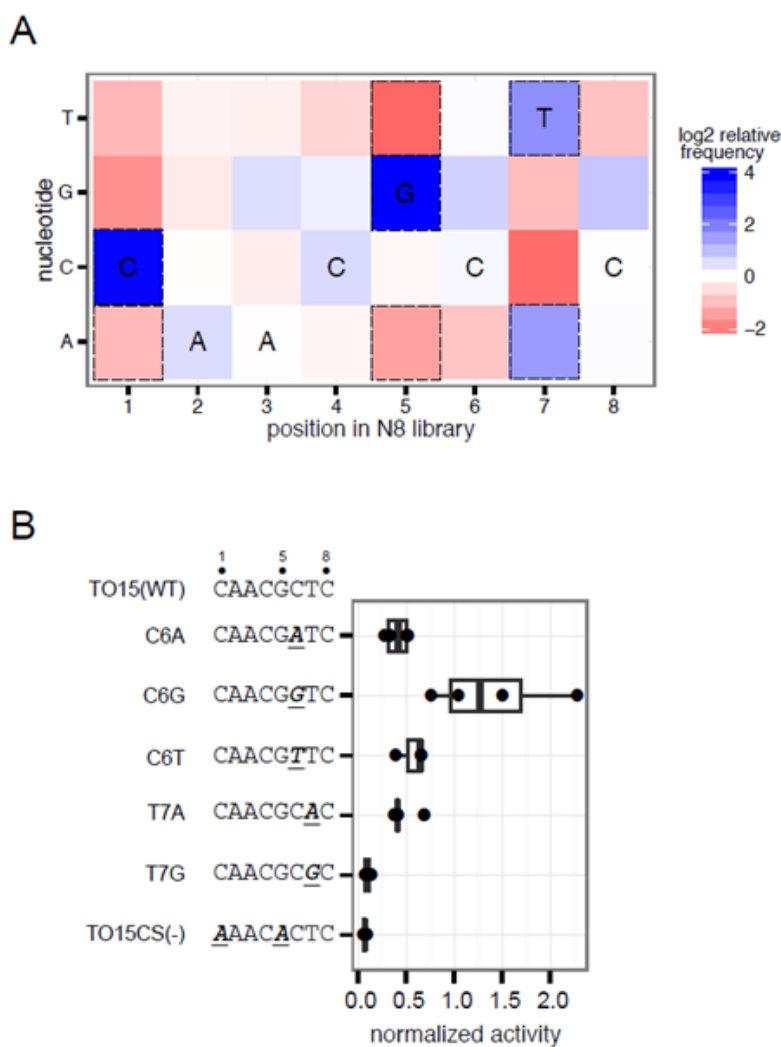


Figure 3.6: Nucleotide preference within the N8 randomized library

(A) Heatmap for nucleotide enrichment after three rounds of selection with the Tev169-xOnu fusion. Dashed boxes represent enrichment values that are greater than two standard deviations away from the mean of enrichment values at all positions. The wild-type sequence is indicated at each position. (B) Validation of nucleotide preferences within the N8 randomized region. Boxplot of Tev169-xOnu activity of on TO15 substrates with point mutants indicated by underlined, and italicized bold-type font. Activity is normalized to Tev169-xOnu on the TO15 substrate. Each mutant was assayed at least three times. Sequences are displayed on the left of the graph with the mutations bolded and underlined.

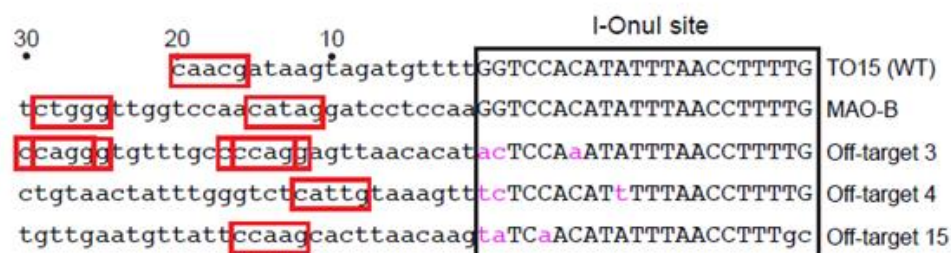
activity by half, while the T7G substitution reduced activity to background levels, supporting the enrichment preferences seen in the *in vitro* data.

Collectively, these data show that the MegaTev strongly prefers to cleave CNNNG motifs spaced 15-bp from the meganuclease-binding site, agreeing with previous studies on spacing of the cleavage motif (27). We also found that the I-TevI nuclease domain is tolerant of multiple substitutions within the CNNNG motif, with many variants cleaved better than the wild-type AAC triplet, but in general preferring A/T rich sequences. Defining a strict consensus sequence within the motif is complicated by the observation of different levels of activity for each of the 64-possible CNNNG variants.

3.3.6 Testing for off-target cleavage

Off-target cleavage is a significant concern for genome-editing endonucleases. To assess off-target cleavage of the MegaTev architecture, we took advantage of previously predicted off-targets for the I-OnuI E2 variant (the backbone meganuclease for our MegaTevs) (43). The I-OnuI E2 variant was optimized to cleave a site within the human MAO-B gene, and a CNNNG motif is positioned 11-bp upstream of the I-OnuI E2 binding site. A number of the top-ranked off-target sites also possess CNNNG motifs within 30-bp of the I-OnuI E2 binding site (Fig. 3.7A), suggesting that they might be substrates for the I-TevI nuclease domain. These sites differ in the spacing of the CNNNG motif from the I-OnuI E2 binding site, and also in the NNN triplet of the cleavage motif. We tested for cleavage at these sites using T7EI digestion of PCR products amplified from HEK 293 cells that were transfected with the dual active Tev169-Onu MegaTev. As shown in Figure 3.7B, we did not observe any T7EI digestion products at these sites, whereas robust T7EI digestion was seen at the integrated TO15 site. We attribute undetectable (or extremely low) activity at these sites to the sub-optimal spacing of the CNNNG motif from the I-OnuI site, and to the presence of NNN triplets that are weakly active as judged by the yeast DNA repair assay. While the sites tested represent a small number of the potential off-target sites, our data suggest that very low levels of off-target cleavage will be observed with the MegaTev at off-target sites where the CNNNG motif is not optimally spaced, or where the NNN central triplet does not support robust activity.

A



B

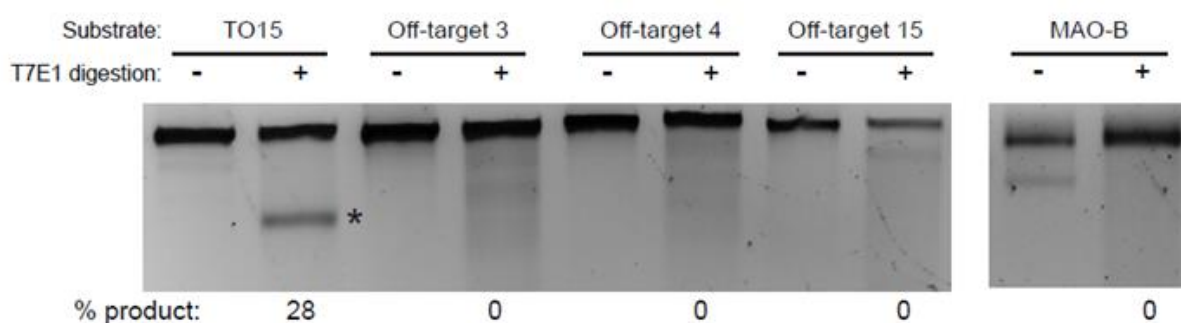


Figure 3.7: Surveyor assay for off-target cleavage.

(A) Schematic showing the TO15, MAO-B and three off-target sites. The I-OnuI E2 site is boxed, with the nucleotide differences in pink lower case font, while the upstream sequences are in lower case black font. Numbering is relative to the 5' end of the I-OnuI binding site. Potential CNNNG cleavage motifs are highlighted in red boxes. (B) T7 endonuclease I (T7E1) digests of PCR-products corresponding to the indicated target sites amplified from HEK 293 cells transfected the dual active site MegaTev. The extent of T7E1 cleavage is indicated below each lane.

3.4 Discussion

A number of characteristics of the I-TevI nuclease domain and meganucleases make them well suited for genome-editing applications. Of relevance to the current study is the fact that cleavage by each enzyme generates different length non-cohesive overhangs (2-nt 3' overhangs for I-TevI, and 4-nt 3' overhangs for meganucleases) (47-49). I-TevI remains bound to its cleavage products, and protects DNA ends from exonucleolytic resection, affecting the extent and directionality of DNA repair events (50). Kinetic studies have also shown that product release by meganucleases is rate limiting (51), and denaturation of meganuclease-DNA complexes is often required to resolve cleavage products in agarose gels (52). Thus, the efficiency of the dual active MegaTevs in HEK 293 cells can in part be attributed to the non-compatible cohesive ends that are sequestered from the NHEJ repair machinery. Sequencing of cleaved targets sites largely supports this hypothesis, as the DSBs are repaired with minimal processing without evidence of deletions up- or downstream of the I-TevI and I-OnuI cleavage sites. End-sequestration by the single-chain MegaTev would also minimize DNA rearrangements such as translocations, inversions or duplications, as is observed with simultaneous expression of multiple genome-editing nucleases in multiplexing experiments.

The MegaTevs used in this study were targeted to model DNA substrates. To be generally useful, the MegaTev platform must be able to target a range of sequences. Engineering meganuclease specificity has been greatly accelerated by a detailed understanding of protein-DNA contacts through crystallographic, computational, and biochemical studies, and by improvements in screening methodologies (43,53-56). Recent efforts suggest that 1 in 300 bp can be targeted by the current set of meganucleases for which detailed protein-DNA contacts maps are available. The targeting range of the MegaTev platform can also be increased by fusing the I-TevI domain to different meganucleases, as shown with the I-LtrI fusions. A series of tailored enzymes that can target sites with a variety of nucleotide compositions could also be achieved by fusing different GIY-YIG nuclease domains with distinct cleavage preferences to the same meganuclease backbone. Our preliminary studies in this regard have generated active enzymes by fusion of the GIY-YIG domain from the I-TuIa homing endonuclease to I-OnuI (J.M.W and D.R.E, unpublished results) (57).

Precise targeting and prediction of off-target sites will also require a detailed understanding of the nucleotide requirements of the GIY-YIG nuclease domain. In the case of I-TevI, past studies revealed that the nuclease domain required a 5'-CNNNG-3' cleavage motif (45,46), and that the linker was tolerant of multiple substitutions within the DNA spacer that separates the cleavage motif from binding site (44). More recently, the I-TevI cleavage motif was defined as CDDHGS (D=A,G,T;H=A,C,T;S=G or C) in the context of a monomeric TALEN architecture (cTALENs) (28). This consensus sequence differs from our observed nucleotide preferences, and one reason may be that the cTALEN study examined cleavage preference within a CNNNG motif positioned 7-bp from the TALE binding site. Our screen of DNA spacer length variation, in contrast, shows a clear spacing preference of 15-bp with MegaTevs derived from both I-OnuI and I-LtrI (Fig. 1B). Thus, the cTALEN study may in fact be describing nucleotide preference within the DNA spacer region that is contacted by the I-TevI linker. Our results also revealed a wide range of tolerance to substitution within the CNNNG motif, with some NNN triplets cleaved more efficiently than the wild-type AAC sequence. We anticipate that both the spacing and sequence requirement of the CNNNG motif will “de-toxify” off-target cleavage, for the simple reason that not all off-target sites will have a permissive motif positioned appropriately from the meganuclease binding site. Indeed, our data, while representing a small number of potential off-target sites, revealed no detectable cleavage by T7EI assays.

In summary, the MegaTevs represent a novel fusion of two different active sites to generate a dual nuclease with a high efficiency of gene disruption without the need to overexpress DNA end-processing enzymes. The compact size of the MegaTev, coupled with the high fidelity imparted by the specificity of the I-TevI and meganuclease domains, makes them suitable for genome-engineering applications where minimizing off-target effects is paramount.

3.5 References

1. Gaj, T., Gersbach, C.A. and Barbas, C.F., 3rd. (2013) ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends Biotechnol*, 31, 397-405.
2. Scharenberg, A.M., Duchateau, P. and Smith, J. (2013) Genome engineering with TAL-effector nucleases and alternative modular nuclease technologies. *Curr Gene Ther*, 13, 291-303.

3. Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. and Voytas, D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, 186, 757-761.
4. Li, T., Huang, S., Zhao, X., Wright, D.A., Carpenter, S., Spalding, M.H., Weeks, D.P. and Yang, B. (2011) Modularly assembled designer TAL effector nucleases for targeted gene knockout and gene replacement in eukaryotes. *Nucleic Acids Res*, 39, 6315-6325.
5. Kim, Y.G., Shi, Y., Berg, J.M. and Chandrasegaran, S. (1997) Site-specific cleavage of DNA-RNA hybrids by zinc finger/FokI cleavage domain fusions. *Gene*, 203, 43-49.
6. Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y. et al. (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods*, 8, 67-69.
7. Bitinaite, J., Wah, D.A., Aggarwal, A.K. and Schildkraut, I. (1998) FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A*, 95, 10570-10575.
8. Vanamee, E.S., Santagata, S. and Aggarwal, A.K. (2001) FokI requires two specific DNA sites for cleavage. *J Mol Biol*, 309, 69-78.
9. Smith, J., Bibikova, M., Whitby, F.G., Reddy, A.R., Chandrasegaran, S. and Carroll, D. (2000) Requirements for double-strand cleavage by chimeric restriction enzymes with zinc finger DNA-recognition domains. *Nucleic Acids Res*, 28, 3361-3369.
10. Gupta, A., Meng, X., Zhu, L.J., Lawson, N.D. and Wolfe, S.A. (2011) Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Res*, 39, 381-392.
11. Pattanayak, V., Ramirez, C.L., Joung, J.K. and Liu, D.R. (2011) Revealing off-target cleavage specificities of zinc-finger nucleases by in vitro selection. *Nat Methods*, 8, 765-770.
12. Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppl, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G. et al. (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol*, 29, 816-823.
13. Burgess, D.J. (2013) Technology: a CRISPR genome-editing tool. *Nat Rev Genet*, 14, 80.
14. Pennisi, E. (2013) The CRISPR craze. *Science*, 341, 833-836.
15. Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F. and Nureki, O. (2014) Crystal Structure of Cas9 in Complex with Guide RNA and Target DNA. *Cell*, 156, 935-949.

16. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. and Charpentier, E. (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science*, 337, 816-821.
17. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. and Doudna, J.A. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*.
18. Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S. et al. (2014) Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. *Science*.
19. Stoddard, B.L. (2011) Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure*, 19, 7-15.
20. Silva, G., Poirot, L., Galetto, R., Smith, J., Montoya, G., Duchateau, P. and Paques, F. (2011) Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther*, 11, 11-27.
21. Gaj, T., Mercer, A.C., Sirk, S.J., Smith, H.L. and Barbas, C.F., 3rd. (2013) A comprehensive approach to zinc-finger recombinase customization enables genomic targeting in human cells. *Nucleic Acids Res*, 41, 3937-3946.
22. Gordley, R.M., Gersbach, C.A. and Barbas, C.F., 3rd. (2009) Synthesis of programmable integrases. *Proc Natl Acad Sci U S A*, 106, 5053-5058.
23. Fonfara, I., Curth, U., Pingoud, A. and Wende, W. (2011) Creating highly specific nucleases by fusion of active restriction endonucleases and catalytically inactive homing endonucleases. *Nucleic Acids Res*, 40, 847-860.
24. Schierling, B., Dannemann, N., Gabsalilow, L., Wende, W., Cathomen, T. and Pingoud, A. (2011) A novel zinc-finger nuclease platform with a sequence-specific cleavage module. *Nucleic Acids Res*, 6, 2623-38.
25. Yanik, M., Alzubi, J., Lahaye, T., Cathomen, T., Pingoud, A. and Wende, W. (2013) TALE-PvuII Fusion Proteins - Novel Tools for Gene Targeting. *PLoS One*, 8, e82539.
26. Boissel, S., Jarjour, J., Astrakhan, A., Adey, A., Gouble, A., Duchateau, P., Shendure, J., Stoddard, B.L., Certo, M.T., Baker, D. et al. (2013) MegaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res*. 4, 2591-2601.
27. Kleinstiver, B.P., Wolfs, J.M., Kolaczyk, T., Roberts, A.K., Hu, S.X. and Edgell, D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proc Natl Acad Sci U S A*, 109, 8061-8066.

28. Beurdeley, M., Bietz, F., Li, J., Thomas, S., Stoddard, T., Juillerat, A., Zhang, F., Voytas, D.F., Duchateau, P. and Silva, G.H. (2013) Compact designer TALENs for efficient genome engineering. *Nat Commun*, 4, 1762.
29. Lieber, M.R. (2011) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem*, 79, 181-211.
30. McVey, M. and Lee, S.E. (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet*, 24, 529-538.
31. Shrivastav, M., De Haro, L.P. and Nickoloff, J.A. (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Res*, 18, 134-147.
32. Certo, M.T., Gwiazda, K.S., Kuhar, R., Sather, B., Curinga, G., Mandt, T., Brault, M., Lambert, A.R., Baxter, S.K., Jacoby, K. et al. (2012) Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nat Methods*, 9, 973-975.
33. Delacote, F., Perez, C., Guyot, V., Duhamel, M., Rochon, C., Ollivier, N., Macmaster, R., Silva, G.H., Paques, F., Daboussi, F. et al. (2013) High frequency targeted mutagenesis using engineered endonucleases and DNA-end processing enzymes. *PLoS One*, 8, e53217.
34. Sollu, C., Pars, K., Cornu, T.I., Thibodeau-Beganny, S., Maeder, M.L., Joung, J.K., Heilbronn, R. and Cathomen, T. (2010) Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res*, 38, 8269-8276.
35. Ran, F.A., Hsu, P.D., Lin, C.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D.A., Inoue, A., Matoba, S., Zhang, Y. et al. (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell*, 154, 1380-1389.
36. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. et al. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339, 819-823.
37. Kleinstiver, B.P., Fernandes, A., Gloor, G.B. and Edgell, D.R. (2010) A unified genetic, computational, and experimental framework identifies non-conserved residues as critical for function of the homing endonuclease I-BmoI. *Nucleic Acids Res*, 38, 2411-2427.
38. Townsend, J.A., Wright, D.A., Winfrey, R.J., Fu, F., Maeder, M.L., Joung, J.K. and Voytas, D.F. (2009) High-frequency modification of plant genes using engineered zinc-finger nucleases. *Nature*, 459, 442-445.

39. Edgell, D.R., Stanger, M.J. and Belfort, M. (2003) Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Curr Biol*, 13, 973-978.
40. Wickham, H. (2009) *ggplot2: elegant graphics for data analysis*. Springer, New York.
41. R, T.C. (2013). R Foundation for Statistical Computing, Vienna, Austria.
42. Wolfs, J.M., Kleinstiver, B.P. and Edgell, D.R. (2014) Rapid screening of endonuclease target site preference using a modified bacterial two-plasmid selection. *Methods Mol Biol*, 1123, 97-104.
43. Takeuchi, R., Lambert, A.R., Mak, A.N., Jacoby, K., Dickson, R.J., Gloor, G.B., Scharenberg, A.M., Edgell, D.R. and Stoddard, B.L. (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc Natl Acad Sci U S A*, 108, 13077-13082.
44. Bryk, M., Quirk, S.M., Mueller, J.E., Loizos, N., Lawrence, C. and Belfort, M. (1993) The td intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *EMBO J*, 12, 2141-2149.
45. Edgell, D.R., Stanger, M.J. and Belfort, M. (2004) Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *J Mol Biol*, 343, 1231-1241.
46. Bryk, M., Belisle, M., Mueller, J.E. and Belfort, M. (1995) Selection of a remote cleavage site by I-TevI, the td intron-encoded endonuclease. *J Mol Biol*, 247, 197-210.
47. Bell-Pedersen, D., Quirk, S., Clyman, J. and Belfort, M. (1990) Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. *Nucleic Acids Res*, 18, 3763-3770.
48. Gimble, F.S. and Thorner, J. (1992) Homing of a DNA endonuclease gene by meiotic gene conversion in *Saccharomyces cerevisiae*. *Nature*, 357, 301-306.
49. Stoddard, B.L. (2005) Homing endonuclease structure and function. *Q Rev Biophys*, 38, 49-95.
50. Mueller, J.E., Smith, D. and Belfort, M. (1996) Exon coconversion biases accompanying intron homing: battle of the nucleases. *Genes Dev*, 10, 2158-2166.

51. Chevalier, B., Sussman, D., Otis, C., Noel, A.J., Turmel, M., Lemieux, C., Stephens, K., Monnat, R.J., Jr. and Stoddard, B.L. (2004) Metal-dependent DNA cleavage mechanism of the I-CreI LAGLIDADG homing endonuclease. *Biochemistry*, 43, 14015-14026.
52. Argast, G.M., Stephens, K.M., Emond, M.J. and Monnat, R.J., Jr. (1998) I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. *J Mol Biol*, 280, 345-353.
53. Jarjour, J., West-Foyle, H., Certo, M.T., Hubert, C.G., Doyle, L., Getz, M.M., Stoddard, B.L. and Scharenberg, A.M. (2009) High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. *Nucleic Acids Res*, 37, 6871-6880.
54. Thyme, S.B., Jarjour, J., Takeuchi, R., Havranek, J.J., Ashworth, J., Scharenberg, A.M., Stoddard, B.L. and Baker, D. (2009) Exploitation of binding energy for catalysis and design. *Nature*, 461, 1300-1304.
55. Arnould, S., Chames, P., Perez, C., Lacroix, E., Duclert, A., Epinat, J.C., Stricher, F., Petit, A.S., Patin, A., Guillier, S. et al. (2006) Engineering of large numbers of highly specific homing endonucleases that induce recombination on novel DNA targets. *J Mol Biol*, 355, 443-458.
56. Smith, J., Grizot, S., Arnould, S., Duclert, A., Epinat, J.C., Chames, P., Prieto, J., Redondo, P., Blanco, F.J., Bravo, J. et al. (2006) A combinatorial approach to create artificial homing endonucleases cleaving chosen sequences. *Nucleic Acids Res*, 34, e149.
57. Sandegren, L. and Sjöberg, B.-M. (2004) Distribution, sequence homology, and homing of group I introns among T-even-like bacteriophages: evidence for recent transfer of old introns. *J Biol Chem*, 279, 22218-22227.

Chapter 4

4 Biasing genome-editing events towards precise length deletions with RNA-guided TevCas9 dual nuclease (Part 1: I-TevI variants)

The work presented in this chapter is reproduced (with permission, Appendix S1) from:

Wolfs, J.M., Hamilton, T.A., Lant, J.T., Laforet, M., Zhang, J., Salemi, L., Gloor, G.B., Schild-Poulter, C. and Edgell, D.R. (2016) Biasing genome-editing events towards precise length deletions with RNA-guided TevCas9 dual nuclease. *PNAS* (pii: 201616343. [Epub ahead of print])

4.1 Introduction

Genome editing with engineered nucleases has revolutionized the targeted manipulation of the genomes of organisms ranging from bacteria to mammals ¹. Zinc finger nucleases (ZFNs) ², LADLIDADG homing endonucleases (also called LHEs or meganucleases) ³, transcription-like effector nucleases (TALENs) ⁴, and nucleases based on the clustered regularly interspersed short palindromic repeat (CRISPR) systems all represent programmable genome-editing nucleases that have successfully been used to introduce targeted changes in genomes ⁵. One of the most common applications of genome-editing nucleases is gene knockouts that are performed in the absence of an exogenously added repair template. The nuclease-introduced break is repaired by the mutagenic non-homologous end-joining pathway (mNHEJ) ⁶⁻⁸, resulting in small insertion or deletions (indels) that, if targeted to an exon, can change the reading frame to produce a non-functional protein. The efficiency of gene knockouts is dependent on multiple factors, including the chromatin context of the target site ^{9,10}, the cell cycle stage ¹¹, and the nature of the DNA overhangs generated by the nuclease ¹². In the case of the commonly used Cas9 (CRISPR-associated protein 9), robust activity coupled with blunt ends generated upon DNA cleavage often leads to a persistent cycle of re-ligation of the DNA ends and regeneration of the target site that will continue until sufficient modification of the target site by mNHEJ prevents cleavage, creating a heterogeneous population of modified sites ¹³. Although more commonly used for homology-directed repair applications ¹⁴, paired Cas9 nickase variants can also be used to create gene knockouts.

We previously created a modular dual-active site nuclease that introduces two DSBs with non-compatible DNA ends at a target site to bias repair events towards deletion of the intervening DNA segment between the two cleavage sites, escaping the cycles of persistent cleavage and re-ligation

cycle¹⁵. We fused the monomeric nuclease and linker domains of the sequence-tolerant GIY-YIG homing endonuclease I-TevI to the N-terminus of the I-OnuI LAGLIDADG homing endonuclease (or meganuclease), which acts as both a targeting and nuclease domain, creating MegaTevs (Fig. 4.1)¹⁶⁻¹⁸. The I-TevI nuclease domain generates a 2-nt 3' overhang¹⁹, whereas the meganuclease generates a non-compatible 4-nt 3' overhang¹⁷. Using model DNA substrates derived from the native I-TevI thymidylate synthase (*td*) and meganuclease target sites, we showed that MegaTevs efficiently excises a 30-bp fragment from the target site in HEK 293 cells with minimal exonucleolytic processing¹⁵. However, the utility of the MegaTev platform is limited by re-engineering of meganuclease specificity on a case-by-base basis, and by uncharacterized interactions of the I-TevI linker domain with the DNA spacer region that are critical for positioning of the nuclease domain at the 5'-CNNNG-3' cleavage site^{20,21} (Fig. 4.1).

Here, we address these limitations by profiling the sequence requirements of the I-TevI linker-DNA spacer interactions, identifying nucleotide positions critical for activity and delineating a putative linker-DNA code. Directed evolution experiments isolated I-TevI linker variants with activity on DNA spacers not cleaved by the wild-type linker, generating a set of variants with defined DNA spacer preferences. We created MegaTev and Tev-mTALENs fusions with the new I-TevI linker variants to demonstrate that the variants maintain identified target specificity across different DNA-binding architectures. In addition, we used the I-TevI DNA code to develop a simple prediction module for novel cleavage site identification.

4.2 Material and Methods

4.2.1 Bacterial strains and plasmid construction

Escherichia coli DH5 α (New England Biolabs) was used for plasmid manipulations, and *E. coli* ER2566 (New England Biolabs) was used for protein expression. Unless stated, *E. coli* strains were grown in Luria-Bertani (LB) media. *E. coli* BW25141(λ DE3) was used for all two-plasmid selection assays (Supplementary Table S1)⁵⁰. MegaTev constructs that included residues 1-169 or 1-184 of I-TevI were cloned into pACYCDuet-1 using the NcoI and XhoI sites as previously described¹⁸. For the bacterial two-plasmid selection, MegaTev target sites were cloned into pTox using the XbaI and SphI cloning sites. The low mutational density MegaTev linker library

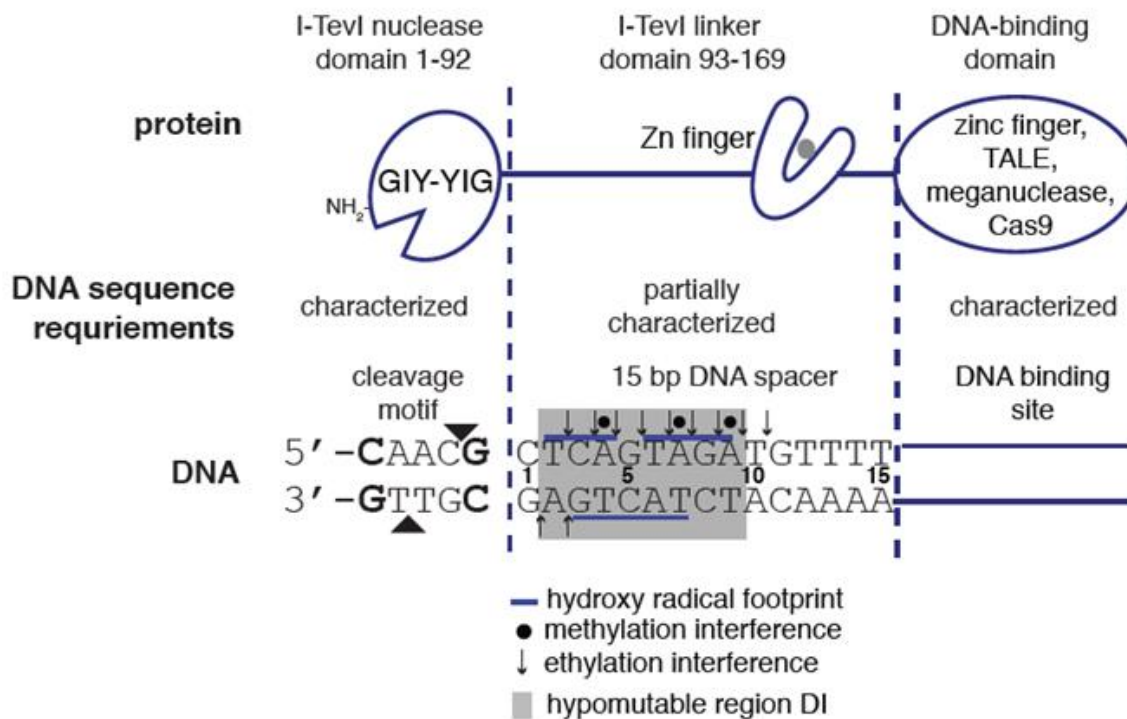


Figure 4.1: Schematic of modular genome-editing nucleases based on the I-TevI nuclease and linker domains.

The corresponding interactions of each of the protein modules with DNA substrate is shown below. The cleavage motif and DNA spacer region are derived from bases -27 to -9 of the native I-TevI *td* substrate, but are renumbered here with $+1$ indicating the first base (C) of the top strand of the DNA spacer. Known interactions of I-TevI with the DNA spacer region based on genetic and biochemical data is summarized.

encompassing amino acids 104-164 of I-TevI were constructed using Mutazyme II Polymerase (Agilent Technologies) with an estimated complexity of 1.4×10^5 and a mutation frequency of 1-3 base pairs per clone. A library with a higher mutational density was constructed with a mutation frequency of 3-10 base pairs per clone using mutagenic dNTPs (TriLink Biotechnologies) with an estimated library complexity of 5.7×10^4 . Each linker library pool was cloned into pACYCDuet-1 harbouring a MegaTev open reading frame using the PciI and BamHI sites located in the I-TevI linker region. To create TevCas9 fusions, a human codon optimized version of I-TevI (amino acids 1-169) with a GGSGGS peptide at the C-terminal end was fused to the N-terminus of Cas9 using SOEing PCR, and cloned into the AgeI and BglII sites of pX458 (Addgene). The TevCas9 gRNAs were cloned into the BbsI site of pX458, or into the XbaI and SphI sites of pTox. All constructs were confirmed by sequencing.

4.2.2 *In vitro* enrichment assay

Target site oligonucleotides containing a randomized 12 bp spacer in the context of a *td* target site were cloned into the pSP72 vector as previously described to create pSP72-N12³⁹. MegaTev purification was performed as previously described¹⁸. Cleavage assays were performed with 360 nM or 1100 nM purified MegaTev and 10 nM pSP72-N15 plasmid in cleavage buffer (50 mM Tris-HCl pH 7.9, 300 mM NaCl, 10 mM MgCl₂ and 1 mM dithiothreitol (DTT)) at 37°C for 5 min. Subsequent rounds of enrichment was performed as previously described³⁹. Samples were prepared for Illumina sequencing at the London Regional Genomics Centre by PCR amplification of target site from both the input and round 3 enrichment samples using Phusion polymerase (New England Biolabs) with barcoded primers. Custom Perl scripts were used to separate the Illumina reads by barcode and to extract the randomized region. Enrichment was determined by using the center logarithmic transformation of the nucleotide counts⁵¹, and by taking the difference between the enriched and input values per nucleotide per position.

4.2.3 Bacterial two-plasmid selection

The bacterial two-plasmid selection was performed with single MegaTev constructs or with MegaTev linker libraries as previously described^{15,18}. Briefly, 50 ng of MegaTev pEndo vector was transformed into *E. coli* BW25141 cells harbouring the pTox plasmid containing the target site of choice. Cells were recovered at 37 °C for 1.5 hrs and aliquots plated on selective (LB media

plus 25 µg/ml chloramphenicol and 10 mM L-(+)-arabinose) and non-selective plates (LB media plus 25 µg/ml chloramphenicol, 50 µg/ml kanamycin and 0.2% glucose). Survival was expressed as the ratio of the number of colonies on the selective plate versus the number of colonies on the non-selective plate. The MegaTev mutant linker libraries were screened either on solid media, or in liquid LB media under selective and non-selective conditions. Identified MegaTev linker variants were re-cloned into pEndo and the selection repeated on the appropriate target sites.

4.3 Results

4.3.1 Identifying and validating I-TevI nucleotide preference in the DNA spacer

To determine nucleotide preference of the I-TevI linker for the DNA spacer region, we performed sequential *in vitro* enrichment with the Tev169-xOnu MegaTev nuclease (hereafter MegaTev) and a plasmid library containing a randomized 12-bp DNA spacer (Fig. 4.2A). In this construct, I-OnuI cleavage is inactivated by an E22Q mutation (represented by an xOnu), meaning that cleavage can only occur if the I-TevI linker interacts with the DNA spacer to correctly position the I-TevI nuclease domain on substrate. Supercoiled plasmid was incubated with purified MegaTev protein (consisting of I-TevI residues 1-169) under two different protein:DNA ratios. Substrates that were linearized by the MegaTev were isolated from an agarose gel, re-circularized by ligation and amplified in *E. coli*. After two rounds of enrichment, the DNA spacer region was PCR amplified and the nucleotide content at each position relative to the input DNA library was readout by Illumina sequencing. As shown in Fig. 4.2B, similar nucleotide enrichment was observed at each position under both protein:DNA conditions, differing only in the higher enrichment values for the more stringent protein:DNA conditions (~30:1 ratio MegaTev to substrate). We also tested a MegaTev construct with a longer I-TevI linker fragment encompassing residues 1-184, and found similar nucleotide enrichment in the DNA spacer (Supplementary Figure S4.1). In general, I-TevI tolerated a number of nucleotide substitutions at each position, with the exception of position 8 where only G was enriched relative to the input pool.

To extend these findings to an *in vivo* environment, we used a well-characterized two-plasmid selection system in *E. coli* where survival is promoted by the MegaTev endonuclease (expressed on pEndo) cleaving the target site cloned on a second plasmid that also encodes a DNA gyrase toxin (pTox) (Fig. 4.2C)²³. We made single substitutions in critical positions of the *td* wild-type

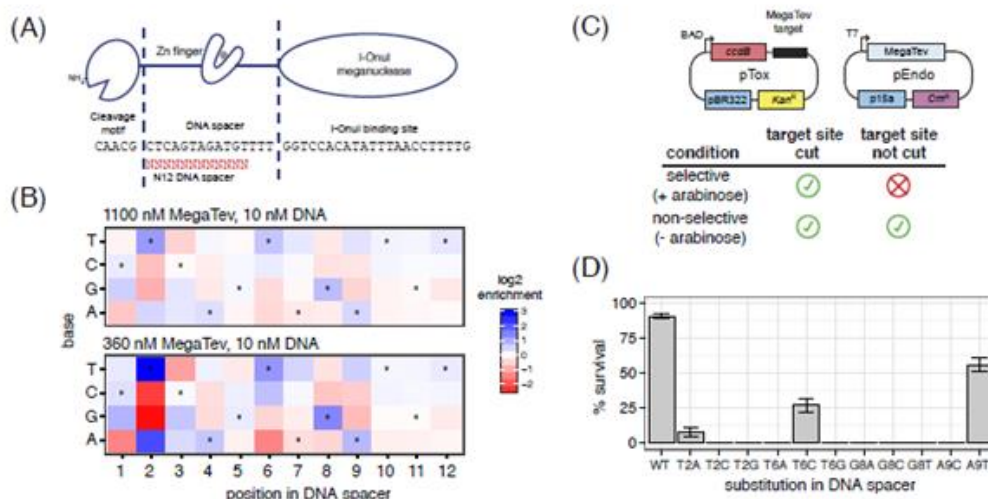


Figure 4.2: Determination of I-TevI nucleotide preference in the DNA spacer

(A) Schematic of the MegaTev DNA substrate that contains the I-TevI cleavage motif, DNA spacer, and meganuclease (I-OnuI) binding site. The randomized bases of the DNA spacer are represented by N. (B) Profiling of nucleotide preference in the DNA spacer using sequential enrichment of a randomized DNA spacer substrate. Two separate enrichment experiments were performed with different protein-DNA ratios. Nucleotide preference in the DNA spacer region is displayed as proportional log₂ enrichment of each nucleotide at each position relative to the abundance in the input pool. Black dots indicate the wild-type nucleotide in the *td* DNA spacer at each position. (C) Schematic of the two-plasmid survival assay, and conditions that promote growth (circled green check mark) or no growth (circled red X). (D) Survival in the two-plasmid assay of the wild-type MegaTev against DNA substrates with single substitutions in the DNA spacer. Data is presented as mean of three independent experiments with error reported as standard deviation.

DNA spacer in the pTox plasmid, and individually tested these against the MegaTev (Fig. 4.2D). In this experiment, MegaTev activity is readout out as the ratio of colonies on non-selective versus selective plates, typically expressed as a percentage. The results of these experiments were in agreement with the *in vitro* enrichment data, as we observed survival on substrates with T2A, T6C or A9T substitutions. No survival was observed on any of the position 8 substitutions, agreeing with the singular preference for G at this position.

4.3.2 Position G8 of the DNA spacer acts as an anchor for the linker

The above data implicate G at position 8 of the DNA spacer as a critical determinant for linker interactions, possibly by acting as an anchor point for interactions with the DNA spacer such that the linker can correctly position the nuclease domain at the 5'-CNNNG-3' cleavage motif, as is observed with the related enzyme I-BmoI. One prediction of the linker-anchor model is that the distance between the anchor G8 and the 5'-CNNNG-3' cleavage motif should remain constant (8 base pairs, measured from G8 to the G of the cleavage motif) independent of the position of the G8 anchor (Fig. 4.3). We tested this model by making substrates with a wild-type DNA spacer length of 15 bp, as well as two substrates that differed by +/- 1 base pairs, creating 14-bp and 16-bp DNA spacers, respectively (Fig. 4.3). To determine if the position of the anchor G8 residue influenced selection of a CNNNG cleavage motif by the nuclease domain, the substrates were randomized in the predicted cleavage site region, allowing us to readout the distance (or phasing) between the G of observed CNNNG motifs and the anchor G8 residue by Illumina sequencing after *in vitro* enrichment. We found that regardless of the DNA substrate, the preferred spacing between the G8 anchor base and the cleavage motif was 8 base pairs (Fig. 4.3) as opposed to the random distribution of spacing seen in the input pools (Supplemental Fig. S4.2). Illumina sequencing also allowed us to determine nucleotide enrichment in the randomized region of each substrate, and we observed strikingly similar patterns between the substrates (Fig. 4.3). In particular, we noted a strong preference for A or T two base pairs downstream of the G of the CNNNG motif, corresponding to position 2 of the DNA spacer region (Fig. 4.3). This finding indicates that sequence preference at the critical position 2 of the DNA spacer is maintained regardless of the position of the cleavage motif.

Collectively, these data show that I-TevI linker interactions with the DNA spacer include both sequence-preference and sequence-spacing requirements. Position G8 is critical for linker

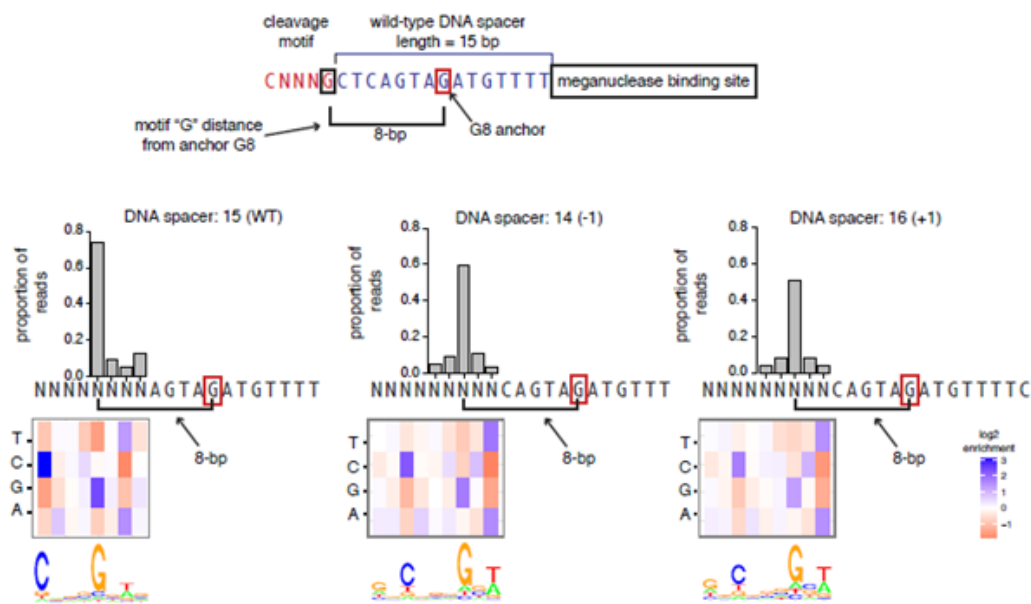


Figure 4.3: Determination of I-TevI distance preference from G8 position in the DNA spacer

(Top) Schematic of substrates used to determine the effect of positioning of the G8 anchor on cleavage motif spacing. (Bottom) For each substrate examined, the position of the G of the C N N N G cleavage motif relative to the G8 anchor nucleotide (highlighted by red rectangle) is plotted as the proportion of Illumina reads. Note that the maximum distance of the G of the cleavage motif from the G8 position is determined by the length of the randomized region. Below each substrate is a heat map and sequence logos plot showing nucleotide enrichment at each position in the randomized region relative to the starting input. Position 6 of the randomized region is equivalent to position 2 of the DNA spacer in Figure 4.2.

interactions, possibly by acting as an anchor for the linker such that the nuclease domain is positioned 8 base pairs upstream. Our data also agree with previous footprinting, modification interference assays, and low-resolution enrichment studies showing that positions 2-10 of the DNA spacer are important for I-TevI-DNA interactions²⁴⁻²⁶

4.3.3 Isolation of I-TevI linker variants that can cleave substrates with substitutions in critical positions

The identification of nucleotide positions in the DNA spacer critical for I-TevI linker interactions prompted us to screen for I-TevI linker variants that could promote cleavage of substrates with substitutions in these positions (summarized in Fig. 4.4A). We used two strategies to isolate I-TevI linker variants. First, we created a mutagenic library encompassing amino acid residues 104 to 169 of the I-TevI linker with on average 1-3 amino acids substitutions per clone. This library was screened using the two-plasmid genetic selection system on DNA substrates with single nucleotide substitutions in positions 2, 6, 8 and 9. This experiment resulted in the identification of the K135N linker variant on the T6G substrate, and the S134G variant on the A9G substrate. No viable clones were recovered on the other substrates. A library with a higher substitution rate (3-10 amino acid substitutions per clone), was also screened against the singly substituted substrates, resulting in the identification of the linker variants K135R/N140S, S134G/N140S, K135R/N140S/Q158R, K118R/K135R/N140S. All of these variants were recovered on the G8A substrate, but subsequent testing revealed survival on G8T substrates for some of the variants (Fig. 4.4B).

Our second strategy used a hybrid non-native DNA spacer derived from the human monoamine oxidase B gene (MAO-B)¹⁷. This substrate was used to determine if nucleotide identity at positions 2, 6, 8 and 9 was sufficient for activity of the wild-type I-TevI linker, or if activity was influenced by context-dependent effects of the surrounding DNA spacer sequence. The MAO-B substrate possessed the wild-type I-TevI linker-DNA spacer nucleotides at positions 2 (A), 6 (C), and 9 (A), and an A8G substitution was made so that the substrate conformed to the I-TevI wild-type preferences (Fig. 4.4A). We found that the wild-type I-TevI linker did not promote cleavage of this substrate in the bacterial two-plasmid assay. Subsequent screening of the mutagenic libraries on this substrate identified the V117F/D127G linker variant.

We de-convoluted the functional importance of the identified amino acid substitutions by

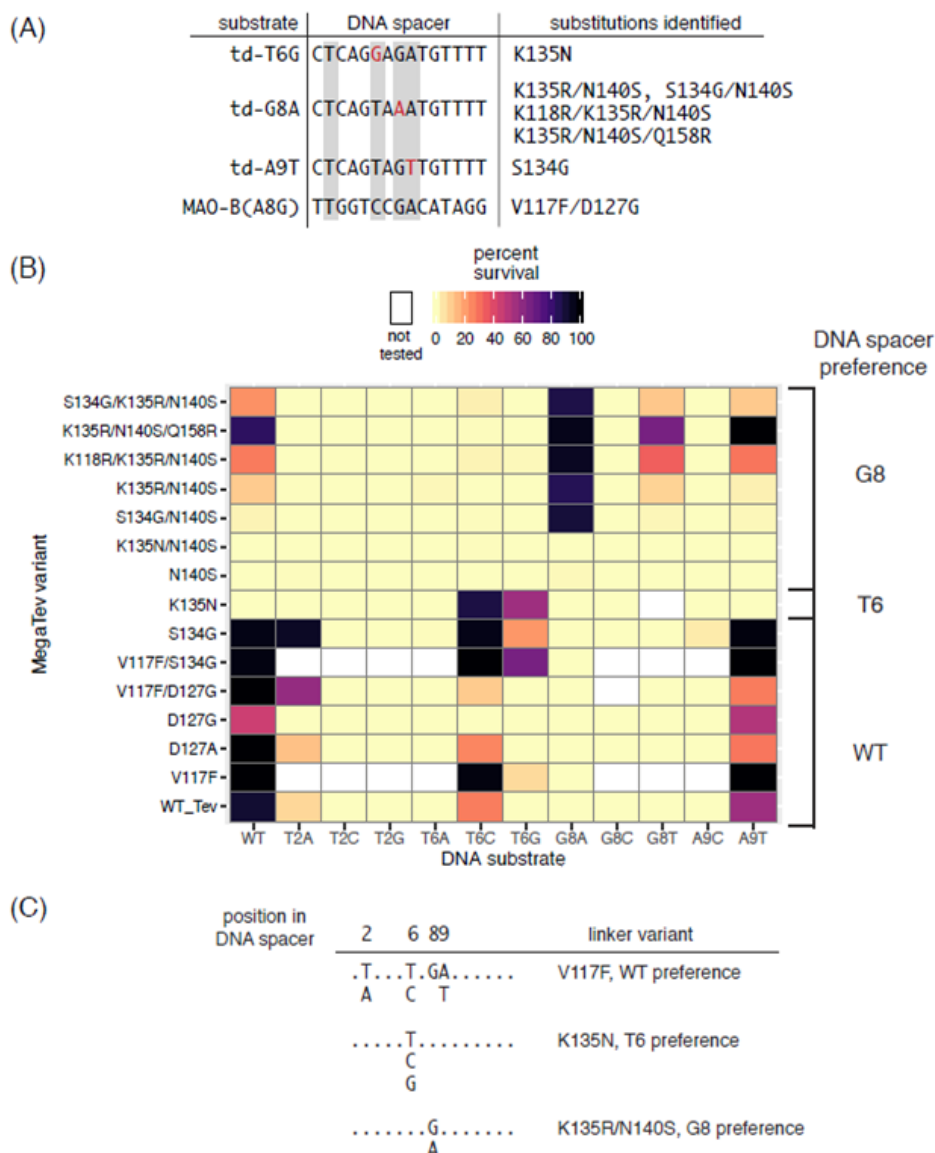


Figure 4.4: Identification and characterization of I-TevI linker variants with altered DNA preferences.

(A) Summary of directed evolution experiments to isolate I-TevI linker variants with altered specificity against the indicated DNA spacers. For the three *td* substrates, nucleotide substitutions relative to the wild-type sequence are in red type font. Positions 2, 6, 8, and 9 of the DNA spacer that are important for I-TevI activity are highlighted by grey rectangles. (B) De-convolution of identified amino acid substitutions on MegaTev survival against singly substituted DNA spacer substrates in the bacterial two-plasmid survival assay. Mean values from three independent biological replicates are reported in the heat map, with white boxes representing linker variant-substrate combinations that were not tested. MegaTev variants fall into three general classes; those with wild-type (WT) preference, those with preference on G8 substituted substrates (G8), and those with preference on T6 substituted substrates (T6). (C) Summary of I-TevI linker variant nucleotide preferences in the DNA spacer.

screening linker variants with individual substitutions, or with combinations of double or triple substitutions, against the singly substituted DNA spacer substrates (Fig. 4.4B). Based on these results, we classified the I-TevI linker variants into three types. The first type, including variants with the V117F substitution, increased activity on DNA spacers on which the wild-type linker was active (WT preference). The second type of variants were those active on DNA spacers with T6C or T6G substitutions (T6 preference). These variants possessed a K135N substitution, or a S134G substitution. The third type of variant was active on substrates with a G8A substitution, and weakly active on G8T substrates (G8 preference). These variants include K135R/N140S. Interestingly, both the second and third type of linker variants possessed substitutions in residue K135, and were not active on the WT DNA spacer substrate suggesting a changed specificity. We noted that the N140S substitution was only isolated in the context of a S134G or K135R substitution, suggestive of a cooperative effect on linker-DNA spacer interactions. When N140S was combined with the K135N substitution, the double mutant K135N/N140S lost activity on all DNA spacers tested, indicative of a negative epistatic interaction.

4.3.4 I-TevI linker variants cleave target sites derived from human genes

I-TevI linker variants with the strongest phenotypes were chosen for testing on non-native targets derived from human genes. We first tested the V117F linker variant that enhanced the wild-type preference in the MegaTev architecture. To identify potential DNA target sites, we searched 19-21 bp windows of the human gene tuberous sclerosis 1 (*TSC1*) coding region for sites that contained a 5'-CNNNG-3' cleavage motif followed by a DNA spacer of 14, 15, or 16 bps. Of the 964 sites with DNA spacer lengths of 15 bp, we arbitrarily chose two sites, TSC1.2125 and TSC1.5054, for testing in the two-plasmid selection system in *E. coli*. (Fig 4.5). The MegaTev-V117F variant showed robust survival (~100%) on both sites, whereas the MegaTev-WT protein had significantly lower levels of survival (2% and 5%). We also showed that the MegaTev-V117F variant promoted 100% survival on a target site from the human breast cancer 1 gene (BRCA1.2541), whereas the MegaTev-WT showed no survival (Fig 4.5).

Encouraged by these results with the MegaTev-V117F variant, we next tested the I-TevI linker variants with altered DNA spacer preferences at position 6 (K135N) and positions 8 and 9 (K135R/N140S). Potential target sites were chosen from human genes based on the presence and

A

	Gene abbreviation	Position	CNNNG activity	CNNNG	Spacer sequence
1	XPC	1277	0.86	CTAAG	GTGGTCCTTCCCAGA
2	Ku70	14915	1.29	CTGTG	CTCAGGTGATCCTCC
3	Hprt	5740	1.04	CTTTG	GAAGGTAATTACAGT
4	pkd1	2398	0.96	CAGAG	GTGGGCAATGGCGTG
5	VEGFA	2522	0.86	CATTG	CTGTGCTTTGGGGAT
6	TSC1 site 1	2125	0.96	CAGAG	GAAGATGGTGTGCC
7	TSC1 site 2	5054	1.29	CTGTG	GTCTGGTGTTCAGC
8	TSC1 site 3	6374	1.29	CTGTG	CAGCATGGAATTCCT
9	Wilms	684	0.76	CTGCG	GAGCCCAATACAGAA
10	APC	1440	0.86	CATTG	CAGAATTATTGCAAG
11	BRCA1 site 1	2541	0.86	CATTG	GTACCTGGTACTGAT
12	BRCA1 site 2	3999	0.96	CAGAG	GAGAATTTATTATCA

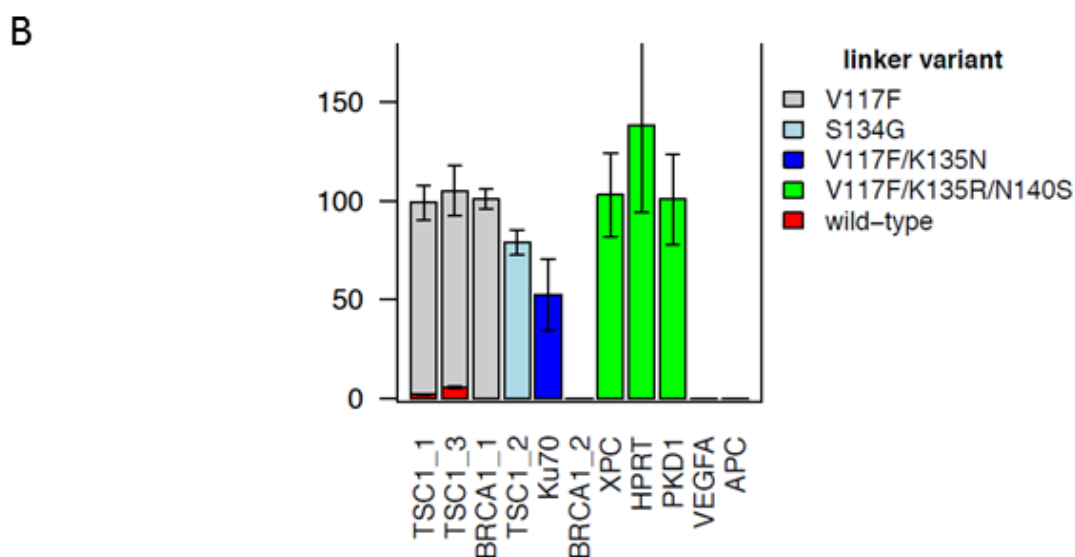


Figure 4.5: I-TevI linker variants cleavage activity for various human gene sequences.

(A) List of genes and target sequences for various human genes. Gene targets are color coded based on I-TevI variant used to cleave the site; V117F (V) in purple, V117F/K135N (VK) in red and V117F/K135R/N140S (VKN) in yellow (B) Survival percentage on various human target sites listed in (A).

predicted activity of the CNNNG cleavage motif, a DNA spacer of 15-bps, and nucleotide identity at positions 2, 6, 8 and 9 that correlated with preferences of the linker variants (Fig. 4.7). Because the potential DNA target sites had a T or A at position 2, and T or C at position 6, the V117F mutation (which enhances activity at these positions) was included in all MegaTev variants with the K135N and K135R/N140S substitutions. As shown in Figure 4.5B, we observed robust survival of the MegaTev variants on 5 of the 8 sites. No survival was seen with the MegaTev-WT protein.

To further investigate new I-TevI variant specificity in alternative DNA-binding platforms we created Tev-mTALEN fusions and tested them using a yeast based recombination assay. TALEs were assembled using golden gate assembly into vectors containing various I-TevI variants (Fig. 4.6A). Not surprisingly, I-TevI linker variants enhanced activity compared to wild-type I-TevI consistent with MegaTev fusions (Fig. 4.6B). No cleavage was observed at other Tev-mTALEN sites with new I-TevI variants demonstrating that the TALE domain still guided targeting (Fig. 4.7C). To further demonstrate the consistency between MegaTev and Tev-mTALENs, we created nucleotide substitutions in the Ku70 Tev-mTALEN target site. We made critical G8A substitution alone and full spacer sequence swap for XPC spacer sequence that is predicted to effect VK cleavage as well as a C12T that should not have detrimental effects (Fig 4.6D). We observed that a G8A substitution abolished activity and C12T had little effect (Fig 4.6E). To cleave the XPC spacer sequence swap, we used a VKN-Ku70 Tev-mTALEN and observed robust cleavage on the XPC spacer and not the Ku70 spacer sequence (Fig 4.6E). These results demonstrated that targeting rules identified in the MegaTev architecture was portable to the Tev-mTALEN architecture.

4.4 Discussion

Homing endonucleases are rare-cleaving enzymes with many properties that are suitable for genome-editing applications, including high fidelity, small size, and generation of 3' DNA overhangs that enhance homology-directed repair³¹. Of the six families of homing endonucleases, the most success in re-programming specificity has come with the LAGLIDADG family (LHEs or meganucleases), typically involving prior crystallographic knowledge to direct mutagenesis of the network of protein residues involved in DNA contacts by cycles of design and optimization

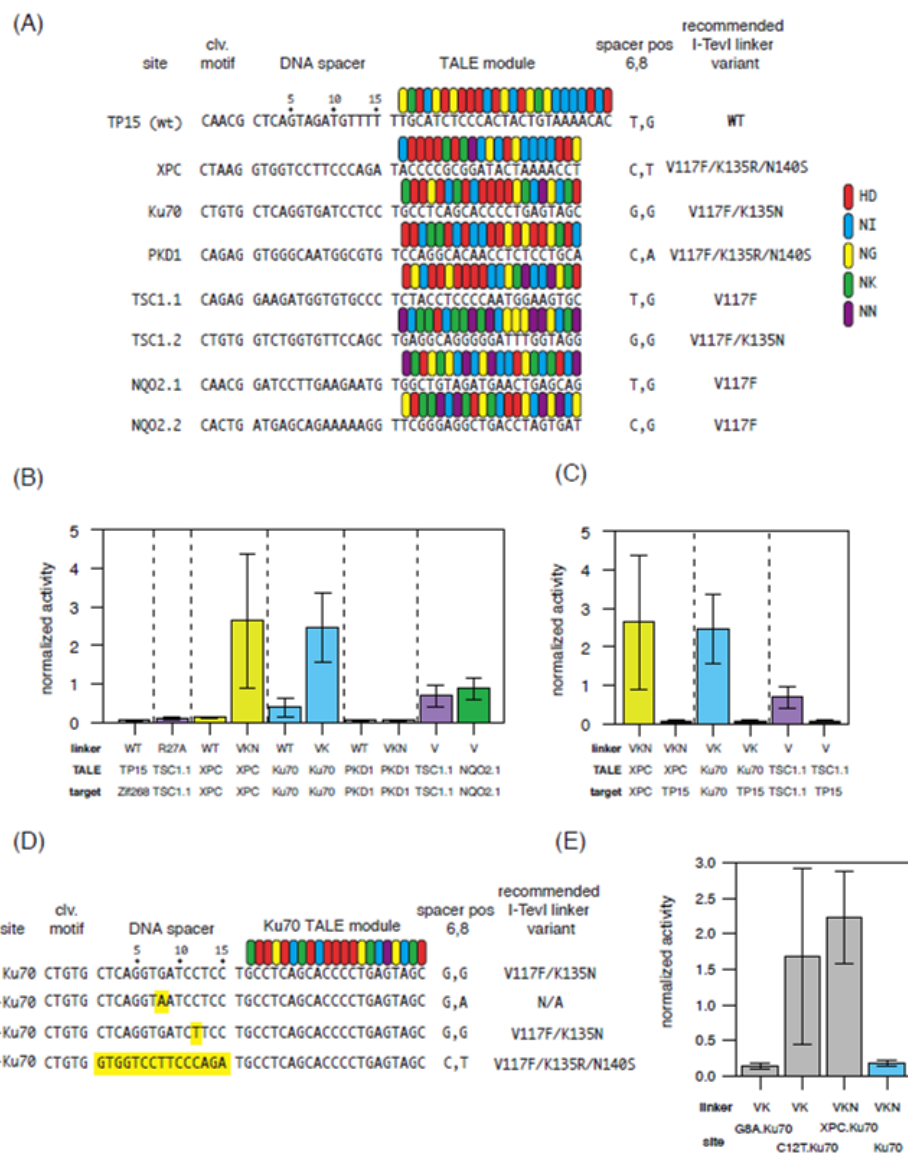


Figure 4.6: I-TevI linker variants function in the context of Tev-mTALENs.

(A) Schematic of the target sites, TALE modules, and recommended I-TevI linker variants based on nucleotide identity at positions 6 and 8 of the DNA spacer. (B) Barplots of activity of the Tev-mTALEN variants on the different target sites (bars are colored according to target site identity). The data represent activity values normalized to the activity of control Tev169-PthXO1 Tev-mTALEN on the TP15 substrate. (C) Barplots showing that activity of Tev-mTALENs is a function of the TALE domain. (D) and (E) Testing the effects of single nucleotide substitutions in the DNA spacer region of the Ku70 target sites. The different substrates are indicated by the substitution in the DNA spacer, and are highlighted in yellow. For panels (B), (C), and (E), data are averages of at least three independent experiments with error bars representing standard deviation.

for each desired target site^{32,33}. Here, we have explored an alternative strategy to using homing endonucleases in genome-editing applications by taking advantage of the sequence-tolerant binding properties of the modular GIY-YIG family to design reagents that can be used without further engineering steps²⁴.

I-TevI is the prototypical member of the GIY-YIG family of homing endonucleases that are characterized by a distinctive modularity, consisting of a N-terminal nuclease domain, a flexible linker domain, and a C-terminal DNA-binding domain^{26,34,35}. Known I-TevI-substrate interactions, as deduced from crystallographic and biochemical studies^{24,26,36}, are a mixture of discriminatory contacts to bases and sequence-tolerant interactions. In the native I-TevI protein, the linker domain acts as a communication device between the N-terminal nuclease and C-terminal DNA-binding domains to position the nuclease domain at the correct distance from the DNA-binding domain such that it cleaves at the appropriate 5'-CNNNG-3' cleavage motif²⁰. This stringent ruler-like behaviour likely evolved as a mechanism to prevent cleavage at a 5'-CCGTG-3' cleavage site that lies upstream of the I-TevI autoregulatory binding site in the phage T4 genome³⁷. Linker function is maintained in the chimeric I-TevI fusions, as substrates with suboptimal spacing of the cleavage site are poor substrates relative to those with optimal spacing of the cleavage site. Interestingly, for both the native and chimeric enzymes, cleavage activity correlates with a ~10-bp periodicity of the cleavage motif^{15,18,27}. One interpretation of this periodicity is that it reflects the molecular distance between where the linker contacts the DNA spacer and where the nuclease is domain positioned on DNA by the linker.

A critical question thus becomes where does the linker contact the DNA spacer, and what residue(s) in the linker are involved in this contact? Our data strongly implicates position G8 of the DNA spacer as the anchor point for the linker, and linker residues 134-140 as mediating this contact, either directly or indirectly. The distance between G8 and the G of the CNNNG cleavage motif is 8-bps, closely matching the ~10-bp periodicity relationship seen between activity and motif spacing. The 8-bp distance is also maintained when the G8 nucleotide is moved, further emphasizing the ruler-like relationship between the linker and DNA spacer. Position G8 corresponds to the first nucleotide of a universally conserved Asp codon (GAY) in thymidylate synthase genes^{38,39}. Homing endonucleases typically use conserved nucleotides such as this as cleavage determinants because these positions are mutationally constrained by the fact that they

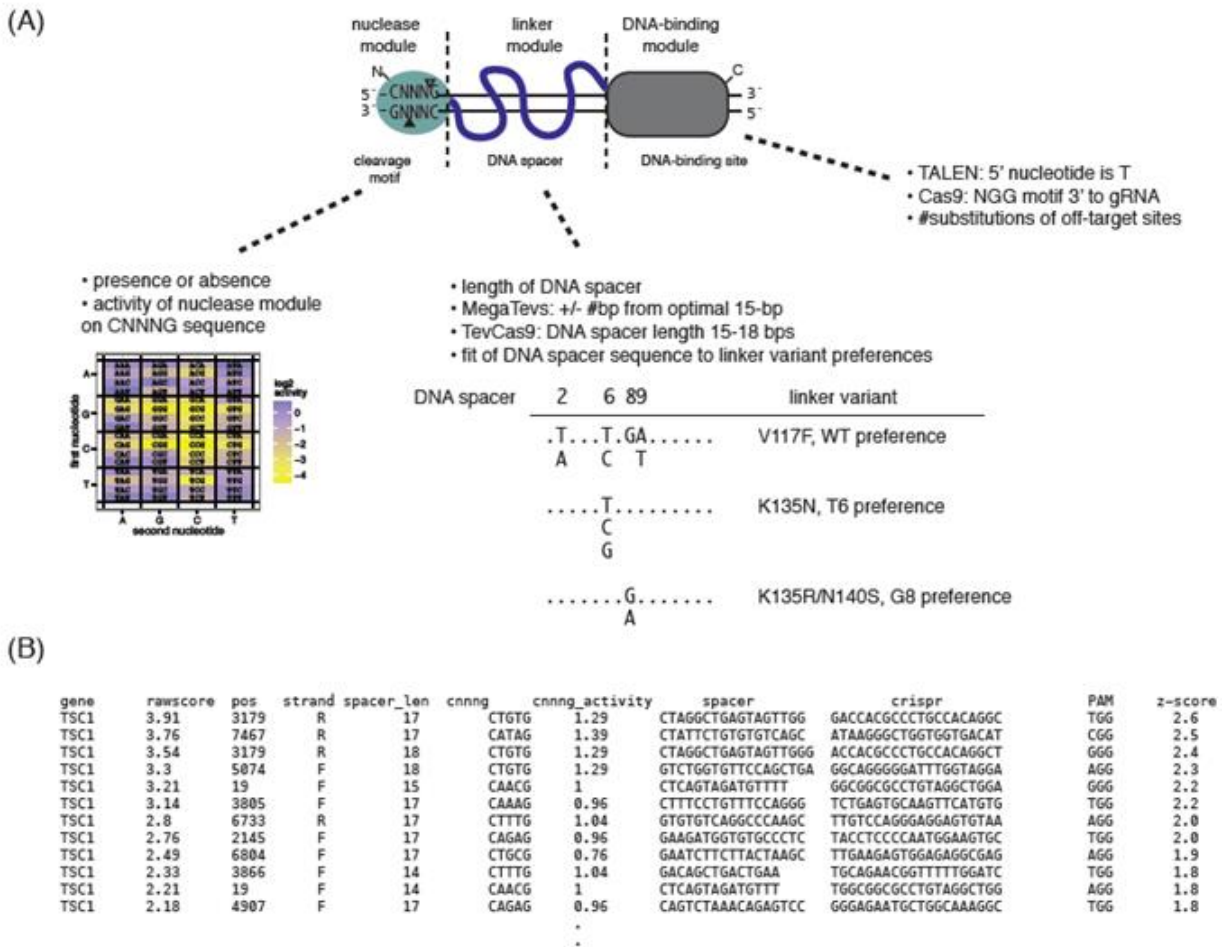


Figure 4.7: Overview of TevCas9 targeting model

(A) Schematic of chimeric I-TevI-based nucleases, with the individual components of the binding model highlighted for each domain. The heatmap of relative cleavage CNNNG activity is derived from reference (15). (B) Example of TevCas9 site predictions in the human TSC1 coding region ranked by Z-score.

encode functionally critical residues of the protein (the Asp residue is found near the thymidylate synthase active site⁴⁰), and thus represent evolutionary stable recognition determinants. The related enzyme I-BmoI also uses a homologous G residue in its thyA target site as a critical cleavage determinant⁴¹. Thus, at least for I-TevI and possibly for other GIY-YIG endonucleases, the linker-DNA code can be partly deciphered by identifying nucleotides that correspond to conserved nucleotides of the target genes.

Residues 123-148 of the I-TevI linker are predicted to form a β -strand²¹, and it is possible that this region of the linker interacts with DNA substrate between the G of the cleavage motif and G8 in the DNA spacer. Our data implicate residue K135 as important for linker-DNA interactions at positions T6 and G8 of the DNA spacer. Interestingly, the phenotypes of the two substitutions we isolated at this position, K135N and K135R, have opposing effects and are modulated by substitutions at neighboring positions, notably S134 and N140, although we do not yet know how S134 or N140 modulate activity. More generally, our data imply that I-TevI linker libraries with a high mutational density focused on residues 128-148 may reveal variants that are tolerant of a broader range of nucleotide substitutions at positions T6 and G8, or in other positions of the DNA spacer we identified as important, notably position T2.

In summary, we have shown that profiling the DNA sequence requirement of the sequence-tolerant I-TevI endonuclease facilitates genome-editing applications. That new I-TevI variants are portable to various DNA-binding domains and target sequences can be predicted to identify target sites within the genome (fig. 4.7).

4.5 References

- 1 Doudna, J. A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096, doi:10.1126/science.1258096 (2014).
- 2 Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics* **11**, 636-646 (2010).
- 3 Stoddard, B. L. Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mobile DNA* **5**, 7, doi:10.1186/1759-8753-5-7 (2014).
- 4 Bogdanove, A. J. & Voytas, D. F. TAL effectors: customizable proteins for DNA targeting. *Science* **333**, 1843-1846 (2011).

- 5 Sternberg, S. H. & Doudna, J. A. Expanding the Biologist's Toolkit with CRISPR-Cas9. *Molecular cell* **58**, 568-574, doi:10.1016/j.molcel.2015.02.032 (2015).
- 6 Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* **79**, 181-211 (2011).
- 7 McVey, M. & Lee, S. E. MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* **24**, 529-538, doi:10.1016/j.tig.2008.08.007 (2008).
- 8 Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. Regulation of DNA double-strand break repair pathway choice. *Cell Res* **18**, 134-147, doi:10.1038/cr.2007.111 (2008).
- 9 Chen, X. *et al.* Probing the impact of chromatin conformation on genome editing tools. *Nucleic acids research* **44**, 6482-6492, doi:10.1093/nar/gkw524 (2016).
- 10 Daboussi, F. *et al.* Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. *Nucleic acids research* **40**, 6367-6379, doi:10.1093/nar/gks268 (2012).
- 11 Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766, doi:10.7554/eLife.04766 (2014).
- 12 Certo, M. T. *et al.* Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nat Methods* **9**, 973-975, doi:10.1038/nmeth.2177 (2012).
- 13 Richardson, C. D., Ray, G. J., DeWitt, M. A., Curie, G. L. & Corn, J. E. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology* **34**, 339-344, doi:10.1038/nbt.3481 (2016).
- 14 Ran, F. A. *et al.* Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389, doi:10.1016/j.cell.2013.08.021 (2013).
- 15 Wolfs, J. A. *et al.* MegaTevs: Single chain dual nucleases for efficient gene disruptions. *Nucleic acids research* **42**, 8816-8829, doi:10.1093/nar/gku573 (2014).
- 16 Sethuraman, J., Majer, A., Friedrich, N. C., Edgell, D. R. & Hausner, G. Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of Ophiostoma and related taxa. *Molecular biology and evolution* **26**, 2299-2315 (2009).
- 17 Takeuchi, R. *et al.* Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 13077-13082 (2011).

- 18 Kleinstiver, B. P. *et al.* Monomeric site-specific nucleases for genome editing. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 8061-8066, doi:10.1073/pnas.1117984109 (2012).
- 19 Bell-Pedersen, D., Quirk, S., Clyman, J. & Belfort, M. Intron mobility in phage T4 is dependent upon a distinctive class of endonucleases and independent of DNA sequences encoding the intron core: mechanistic and evolutionary implications. *Nucleic acids research* **18**, 3763-3770 (1990).
- 20 Dean, A. B. *et al.* Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 8554-8561 (2002).
- 21 Liu, Q. *et al.* Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *Journal of molecular biology* **379**, 1094-1106 (2008).
- 22 Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821, doi:10.1126/science.1225829 (2012).
- 23 Chen, Z. & Zhao, H. A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic acids research* **33**, e154 (2005).
- 24 Bryk, M. *et al.* The *td* intron endonuclease I-TevI makes extensive sequence-tolerant contacts across the minor groove of its DNA target. *The EMBO journal* **12**, 2141-2149 (1993).
- 25 Mueller, J. E., Smith, D., Bryk, M. & Belfort, M. Intron-encoded endonuclease I-TevI binds as a monomer to effect sequential cleavage via conformational changes in the *td* homing site. *The EMBO journal* **14**, 5724-5735 (1995).
- 26 Derbyshire, V., Kowalski, J. C., Dansereau, J. T., Hauer, C. R. & Belfort, M. Two-domain structure of the *td* intron-encoded endonuclease I-TevI correlates with the two-domain configuration of the homing site. *Journal of molecular biology* **265**, 494-506 (1997).
- 27 Kleinstiver, B. P. *et al.* The I-TevI Nuclease and Linker Domains Contribute to the Specificity of Monomeric TALENs. *G3 (Bethesda)* **4**, 1155-1165, doi:10.1534/g3.114.011445 (2014).
- 28 Qiu, P. *et al.* Mutation detection using Surveyor nuclease. *Biotechniques* **36**, 702-707 (2004).
- 29 Guschin, D. Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* **649**, 247-256, doi:10.1007/978-1-60761-753-2_15 (2010).

- 30 Mueller, J. E., Smith, D. & Belfort, M. Exon coconversion biases accompanying intron homing: battle of the nucleases. *Genes & development* **10**, 2158-2166 (1996).
- 31 Stoddard, B. L. Homing endonucleases: from microbial genetic invaders to reagents for targeted DNA modification. *Structure* **19**, 7-15, doi:10.1016/j.str.2010.12.003 (2011).
- 32 Takeuchi, R., Choi, M. & Stoddard, B. L. Redesign of extensive protein-DNA interfaces of meganucleases using iterative cycles of in vitro compartmentalization. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 4061-4066, doi:10.1073/pnas.1321030111 (2014).
- 33 Silva, G. *et al.* Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther* **11**, 11-27 (2011).
- 34 Van Roey, P. & Derbyshire, V. in *Homing Endonucleases and Inteins* (ed Stoddard BL Belfort M, Wood DW, Derbyshire V) 67-83 (Springer-Verlag, 2005).
- 35 Dunin-Horkawicz, S., Feder, M. & Bujnicki, J. M. Phylogenomic analysis of the GIY-YIG nuclease superfamily. *BMC Genomics* **7**, 98 (2006).
- 36 Van Roey, P., Waddling, C. A., Fox, K. M., Belfort, M. & Derbyshire, V. Intertwined structure of the DNA-binding domain of intron endonuclease I-TevI with its substrate. *The EMBO journal* **20**, 3631-3637 (2001).
- 37 Edgell, D. R. *et al.* Intron-encoded homing endonuclease I-TevI also functions as a transcriptional autorepressor. *Nature structural & molecular biology* **11**, 936-944, doi:10.1038/nsmb823 (2004).
- 38 Edgell, D. R., Stanger, M. J. & Belfort, M. Coincidence of cleavage sites of intron endonuclease I-TevI and critical sequences of the host thymidylate synthase gene. *Journal of molecular biology* **343**, 1231-1241, doi:10.1016/j.jmb.2004.09.005 (2004).
- 39 Edgell, D. R., Stanger, M. J. & Belfort, M. Importance of a single base pair for discrimination between intron-containing and intronless alleles by endonuclease I-BmoI. *Current biology : CB* **13**, 973-978 (2003).
- 40 Carreras, C. W. & Santi, D. V. The catalytic mechanism and structure of thymidylate synthase. *Annual review of biochemistry* **64**, 721-762 (1995).
- 41 Kleinstiver, B. P., Wolfs, J. M. & Edgell, D. R. The monomeric GIY-YIG homing endonuclease I-BmoI uses a molecular anchor and a flexible tether to sequentially nick DNA. *Nucleic acids research* **41**, 5413-5427, doi:10.1093/nar/gkt186 (2013).
- 42 Roy, A. C., Wilson, G. G. & Edgell, D. R. Perpetuating the homing endonuclease life cycle: identification of mutations that modulate and change I-TevI cleavage preference. *Nucleic acids research* **44**, 7350-7359, doi:10.1093/nar/gkw614 (2016).

Chapter 5

5 Biasing genome-editing events towards precise length deletions with RNA-guided TevCas9 dual nuclease (Part 2: TevCas9 Fusions)

The work presented in this chapter is reproduced (with permission, Appendix S1) from:

Wolfs, J.M., Hamilton, T.A., Lant, J.T., Laforet, M., Zhang, J., Salemi, L., Gloor, G.B., Schild-Poulter, C. and Edgell, D.R. (2016) Biasing genome-editing events towards precise length deletions with RNA-guided TevCas9 dual nuclease. *PNAS* (pii: 201616343. [Epub ahead of print])

5.1 Introduction

Genome editing with engineered nucleases has revolutionized the targeted manipulation of the genomes of organisms ranging from bacteria to mammals (1). Zinc finger nucleases (ZFNs) (2), transcription-like effector nucleases (TALENs) (3), MegaTALs (fusion of a LAGLIDADG homing endonuclease and TALE domain) (4-6), and nucleases based on the clustered regularly interspersed short palindromic repeat (CRISPR)-associated protein 9 (Cas9) all represent programmable genome-editing nucleases that have successfully been used to introduce targeted changes in genomes (7-11). One of the most common applications of genome-editing nucleases is gene knockouts that are performed in the absence of an exogenously added repair template (12). In the case of Cas9, the blunt DNA ends introduced at DNA cleavage are substrates for error-free repair by the classical NHEJ pathway (c-NHEJ) (13), regenerating the target site for re-cleavage by the nuclease. This cycle of cleavage, ligation and target site regeneration is perturbed when the DSB is sufficiently modified by exonucleolytic processing by c-NHEJ, or by the alternative NHEJ pathway (alt-NHEJ), to prevent cleavage by the nuclease (14-18). Imprecise repair by either of the NHEJ pathways generates the characteristic spectrum of heterogeneous length insertions or deletions (indels) centered around the break site (19, 20). The heterogeneous distribution of indels, and the fact that not all indels generate gene knockouts, means that downstream screening and confirmation of knockout genotypes is required. In addition, the chromatin context of the target site (21, 22), the cell cycle stage and cell type (23), and the nature of the DNA overhangs generated by genome-editing nucleases influence the types of indels and efficiency of gene knockouts (24). Coupling the expression of DNA end processing enzymes and genome-editing nucleases can bias

gene knockouts by enhancing exonucleolytic end-processing before ligation by NHEJ, with variable rates of success depending on the end-processing enzyme employed (24, 25). Although more commonly used for homology-directed repair applications (26), paired Cas9 nickase variants can be used to generate gene knockouts, but also generate heterogeneous length deletions.

Here, we provide a simple and robust solution to bypass the persistent cycle of cleavage and re-ligation and shift genome-editing events towards deletions of defined length. We created a dual nuclease that introduces two non-compatible DNA breaks at a target site such that the majority of the target site is deleted, preventing regeneration of the target site and continued cleavage. We fused the monomeric nuclease and linker domains from the I-TevI homing endonuclease to Cas9 (27, 28), creating an RNA-guided TevCas9 dual nuclease that functions robustly in HEK 293 cells at endogenous target sites. TevCas9 generates defined length deletions of 33-36 base pairs at high frequencies with minimal DNA end processing compared to Cas9. We envision that the non-compatible and directional nature of the TevCas9 overhangs will enhance applications such as site-directed mutagenesis by oligonucleotide ligation, deletion of binding sites, or epitope-tagging.

5.2 Material and Methods

5.2.1 Bacterial strains and plasmid construction

See chapter 4, 4.2.1 Bacterial strains and plasmid construction for TevCas9 construction.

5.2.2 Protein purification

For bacterial expression and purification, TevCas9 or Cas9 variants were cloned in pACYC-Duet1 expression plasmids under control of a T7-regulated promoter. The RARA gRNA was cloned in the BmsI sites of the pACYC plasmid. The expression plasmids were transformed into *E. coli* ER2566, and grown in 1L LB broth until an A_{600} reading of 0.5, induced with isopropyl-D-thiogalactopyranoside (IPTG) at 1 mM final concentration, and grown overnight at 16°C. Cells were harvested, resuspended in binding buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 10 mM imidazole, 1 mM dithiothreitol (DTT)), lysed and centrifuged to clarify cell debris. The lysate was loaded onto a 1 ml Ni-NTA affinity column (GE Healthcare), washed with 5 column volumes binding buffer and 5 column volumes of binding buffer with 25 mM imidazole, before step elution with binding buffer containing 100, 200, 300 or 400 mM imidazole. Fractions were dialyzed into

storage buffer (20 mM Tris-HCl pH 8, 500 mM NaCl, 1 mM DTT, 5% glycerol) and frozen at -80°C.

5.2.3 gRNA visualization

To confirm gRNA co-purification with TevCas9 or Cas9, purified protein (4.4 µg) was treated with 20 µg of proteinase K in NEB Buffer 3 at 37°C for 30 mins. The sample was split, and half treated with RNase A for 30 mins at 37°C. Samples were heated to 95°C for 5 mins before loading on a 12% denaturing urea-polyacrylamide gel. RNA was visualized by staining with ethidium bromide.

5.2.4 TevCas9 endonuclease assays

The RARA target site was amplified from pTox-RARA by PCR using primers DE1544 and DE1191. Cleavage assays were performed with a 90:1 molar ratio of protein:DNA in a pooled reaction in cleavage buffer (20 nM DNA substrate, 50 mM Tris-HCl pH 8, 125 mM NaCl, 10 mM MgCl₂, 1 mM DTT). Aliquots were taken at the appropriate time points, and stopped in stop buffer (200 mM ethylenediaminetetraacetic acid (EDTA), 30% glycerol, 1% sodium dodecyl sulphate, and xylene cyanol). Reactions were treated with proteinase K and RNase A before electrophoresis on an 8% polyacrylamide gel in Tris-borate EDTA buffer.

5.2.5 HEK293 transfections and genomic extractions

HEK293 cells were cultured in Dulbecco's modified Eagle's medium supplemented with 10 % fetal bovine serum and grown at 37 °C with 5 % CO₂. One day before transfection, approximately 1 x 10⁵ cells were seeded into 24-well plates. Transfections were performed using JetPRIME® (Polyplus transfection) with 500 ng of pX458 plasmid as recommended in the JetPRIME® protocol. Transfection reagent was removed 4 hrs after transfection and replaced with fresh media and incubated for 48 hrs at 37 °C with 5 % CO₂. Cells were harvested and resuspended in 100µl TE pH 8.0. Cells were lysed in 300µl of extraction buffer (100 mM NaCl, 10 mM Tris-HCl pH 8.0, 25 mM Ethylenediaminetetraacetic acid (EDTA), 0.5 % sodium dodecyl sulfate (SDS) and 0.1 mg/ml proteinase K) for 16 hrs at 50 °C. Cell lysates were treated with 1 mg/ml RNaseA for 1 hr at 37 °C and genomic DNA was purified using phenol/chloroform extraction and ethanol precipitation. Genomic DNA was resuspended to a final volume of 70 µl in TE pH 8.0.

5.2.6 T7 endonuclease I mismatch repair assays and restriction endonuclease assays

Genomic TevCas9 target sites were PCR amplified with Taq polymerase (Thermo Fisher) using nested PCR. For T7 endonuclease 1 (T7E1) mismatch assays^{28,29}, PCR products were boiled at 95 °C for 5 min, slowly cooled to 50 °C before flash freezing at -20 °C for 2 min. Approximately 250 ng of PCR product was incubated with 2 U of T7E1 (New England Biolabs) with NEBuffer 2 for 15 mins at 37 °C. Cleavage products were separated on a 1.5% agarose gel stained with ethidium bromide and analyzed using AlphaImager™ (Alpha Innotech). For restriction endonuclease assays, 250ng of PCR product was digested with BamHI or PvuII for 1 hr at 37 °C. Digests were electrophoresed on a 1.5% agarose gel containing ethidium bromide and analyzed using AlphaImager™ (Alpha Innotech).

5.2.7 Binding models for MegaTev and TevCas9

The binding models for MegaTevs and Cas9 included two components essential for I-TevI activity: a 5'-CNNNG-3' cleavage motif, and a DNA spacer. Putative MegaTev binding sites were found using custom Perl scripts that windowed across human coding regions for 5'-CNNNG-3 motifs followed by a DNA spacer of 14-16 bps. The raw score of the putative binding site was calculated by: (1) the log₂ value of the activity of the 5'-CNNNG-3' cleavage motif (relative to the 5'-CAACG-3' native motif); (2) the log₂ score of the DNA spacer length, and given scores 18 bp = 0.6; 17 bp = 1; 16 bp = 0.6; 15 bp = 0.4; 14 bp = 0 (3) the log₂ score of the fit of the DNA spacer sequence to the position weight matrix (PWM) derived from the DNA spacer *in vitro* enrichment data. The predicted binding sites were subsequently ranked by their Z score. The gRNA portion of highly ranked TevCas9 sites were individually examined using the XXX server, and TevCas9 sites that included poor gRNAs were excluded.

5.2.8 Next generation DNA sequencing

Genomic DNA from HEK 293 cells corresponding to three independent transfections of TevCas9, Cas9, and one mock transfection, was used to PCR amplify the NQO2, RARA, TSC1-2125 or TSC1-5054 target sites with barcoded primers. The PCR products for each site were combined into pools, and a single-end 250-bp run performed on an Illumina Mi-Seq platform at the London Regional Genomics Centre. Reads were trimmed to 200 nts and then parsed for each barcode using custom Perl scripts. We first analyzed the reads for length differences (insertions and deletions)

relative to the unmodified site by extracting sequence between anchor sequences that flanked the TevCas9 and Cas9 target sites. The proportion of insertions or deletions for each length variant was the ratio of the number of reads with length N to the total number of extracted reads, excluding reads with less than 1000 counts. To map the positions of deletions within each target site, we used the UBLAST function of USEARCH⁵², with the following parameters: -evaluate 0.01, -mismatch 5, -gapopen 20IQ/*LQ/*RQ, -gapext 0IQ, -strand plus, -maxhits 2. Searches were exported using the -userout option, with the following fields: query, target, id, alnlen, mism, opens, qlo, qhi, tlo, thi, thi, qrow. The output files were parsed for deletion start and deletion end point, and length, and each deletion plotted as a proportion relative to all mapped deletions using the graphing functions of the R project for statistical computing.

5.3 Results

5.3.1 Fusion of I-TevI to Cas9 creates an RNA-guided dual nuclease

To create a TevCas9 RNA-guided dual nuclease, we fused the I-TevI nuclease domain (residues 1-92) and a portion of the linker domain (residues 92-169) to the N-terminus of SpCas9 nuclease derived from the *Streptococcus pyogenes* Type II CRISPR system (27, 28) (Fig. 5.1A). A TevCas9 target site includes the I-TevI 5'-CNNNG cleavage motif, a DNA spacer that interacts with the I-TevI linker and that separates the cleavage motif from the binding site, the DNA-binding site complementary to the gRNA, and the downstream NGG protospacer-associated motif (PAM) sequence. Previous studies revealed that the length of the DNA spacer is crucial for I-TevI cleavage function, with lengths of 14-19 base pairs (bps) supporting activity in various chimeric contexts (27, 29, 30). We co-expressed TevCas9 with a C-terminal histidine tag and a gRNA targeting a site in the retinoic acid receptor alpha gene (RARA.233, numbered according to start of target site in the RARA cDNA) from the same plasmid in *E. coli*. The RARA.233 target site was predicted according to a binding model that utilized data from *in vitro* profiling of I-TevI linker-DNA spacer nucleotide preference (Fig. 4.1, 4.2 and 4.3), and the activity of the I-TevI nuclease domain on all possible 64 CNNNG variants (30). The TevCas9 variant used in this experiment (hereafter, TC-V) contains an activity-enhancing V117F substitution in the I-TevI linker.

We purified the TC-V by metal-affinity chromatography, and showed that TC-V was complexed with an RNA species of the size predicted for the transcribed gRNA (indicated by an asterisk, Fig.

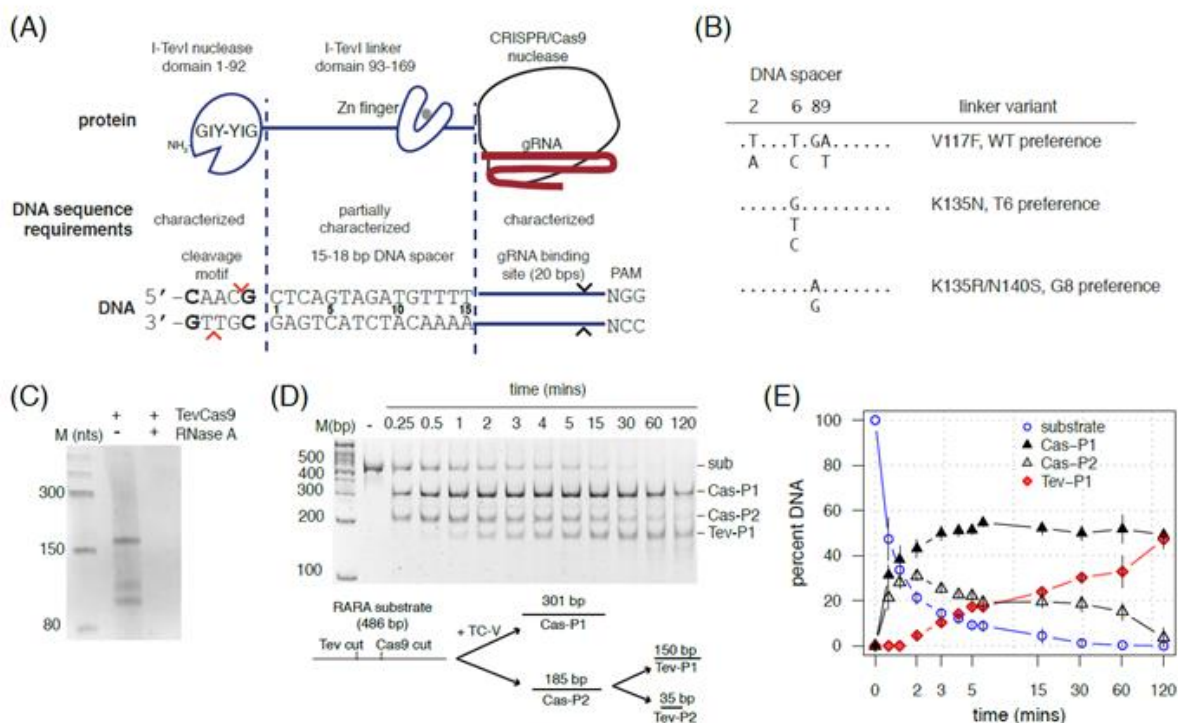


Figure 5.1: Purification and characterization of a TevCas9 dual nuclease

A) Schematic of the TevCas9 nuclease, and general organization of the DNA substrate. (B) Summary of in vitro profiling of I-TevI-linker DNA spacer nucleotide preferences, and directed evolution experiments that identified I-TevI linker variants with activity on DNA spacers with substitutions in positions T6 and G8. Linker variants used in the TevCas9 nuclease are indicated on the right, with DNA spacer preference indicated. (C) TevCas9 co purifies with an RNA species of the size predicted for the transcribed gRNA. Shown is an 12% urea-polyacrylamide gel of TevCas9 protein samples treated with proteinase K, and then with (+) or without (-) RNase A. The marker is an RNA ladder with sizes in nucleotides indicated. (D) (top) Representative time course cleavage assay (in minutes) with TevCas9-V117F (TC-V) and a PCR substrate containing the RARA target site. The substrate and cleavage products are indicated on the right side of the gel. The gel image is inverted. (bottom) Reaction pathway as deduced from the appearance of products in the cleavage assay. Note that the 35-bp Tev-P2 is not shown on the gel image. (E) Plot of reaction progress in minutes versus percent DNA for the TevCas9 time course cleavage assay with the RARA substrate. Data points are mean values of four independent experiments, with vertical bars representing standard deviation.

1C). To determine if the TC-V/gRNA ribonucleoprotein complex (RNP) introduced two DSBs, we examined the cleavage profile on a PCR substrate that contained the RARA.233 target site. Cleavage was consistent with two sequential DSBs, first by the Cas9 nuclease, and second by the I-TevI nuclease (Fig. 5.1D and 5.1E). The predicted 33-bp product corresponding to the fragment excised between the I-TevI and Cas9 cleavage sites was difficult to reproducibly visualize, and not used to quantitate reaction progress. Interestingly, cleavage by Cas9 was ~30-fold faster than cleavage by I-TevI (Fig. 5.1E, as judged by time required for 50% product formation).

5.3.2 TevCas9 functions robustly in HEK 293 cells

TevCas9 is targeted by the Cas9-associated gRNA, a property that allowed us to directly compare TevCas9 and Cas9 activity at the same sites in HEK 293 cells. To do so, we used T7 endonuclease 1 (T7E1) mismatch cleavage assays to measure on-target modification at the RARA.232 site and at two other sites, one in the tuberous sclerosis 1 gene (TSC1.2125) and one in the quinone reductase 2 gene (NQO2.54) (Fig. 5.2A) (31, 32). With TC-V transfections, we observed modification rates ranging from 16-23%, versus 12–29% modification rates at the same sites when cells were transfected with Cas9. We also targeted TevCas9 to a second site in the TSC1 gene (TSC1.5054). The TevCas9 variant (TC-VK) used in this experiment contained V117F and K135N substitutions in the I-TevI linker domain that enhance activity on DNA spacers with a T->C substitution in position 6 of the DNA spacer (Fig. 4.2). We observed very similar levels of modification with both TC-VK (mean = 15%) and Cas9 (mean = 13%) at the TSC1.5054 site. No activity at the NQO2.54, TSC1.2125 or TSC1.5054 sites was observed in cells transfected with TC-V programmed with the RARA.233 gRNA, showing that the I-TevI domains do not influence gRNA-mediated targeting of Cas9. We also predicted off-targets sites based on the NQO2.54 gRNA (33), and examined cleavage at these sites by T7E1 mismatch assays (Fig. S4.4). No cleavage was observed at these sites, three of which had a correctly spaced CNNNG motif from the binding site for the gRNA. These experiments show that the I-TevI nuclease and variant linker domains can be directed by Cas9 to cleave target sites with diverse cleavage motifs and DNA spacers, suggesting that the I-TevI sequence requirements are not limiting for TevCas9 targeting.

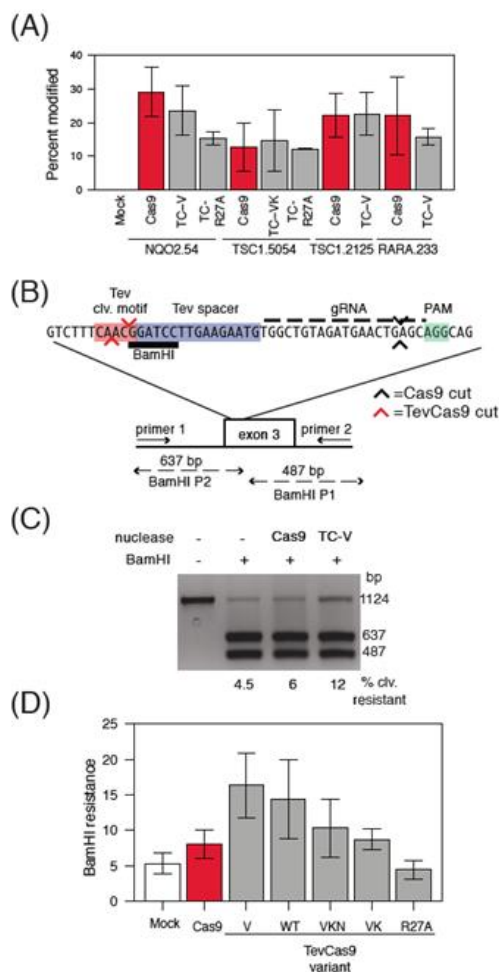


Figure 5.2: Activity and specificity of TevCas9 in HEK 293 cells

(A) T7E1 mismatch cleavage assays of PCR fragments amplified from the indicated target sites after transfection with Cas9 or TevCas9. TC-V, TevCas9-V117F; TC-VK, TevCas9-V117F/K135N; TC-R27A, TevR27ACas9 (R27A inactivates I-TevI cleavage activity). (B) Detailed schematic of the target site in exon 6 of the human *NQO2* gene, positions of PCR primers used for amplification, and sizes of BamHI cleavage products. The I-TevI cleavage motif and DNA spacers are highlighted by red and blue rectangles, and the PAM motif by a green rectangle. Positions of I-TevI and Cas9 cleavage sites are represented by red and black arrows, respectively. (C) Representative agarose gel of BamHI cleavage assays on PCR products amplified from the *NQO2* locus 48 hrs after transfection with Cas9 or with TC-V. Sizes of the substrate (1124 bp) and two BamHI cleavage products (637 bp and 487 bp) are indicated on the right. The percent of substrate resistant to cleavage by BamHI is indicated below each lane. (D) Activity of different TevCas9 variants at the *NQO2* site measured by BamHI resistance. TevCas9 variants labelled as in panel (A) except for VKN – TevCas9-V117F/K135R/N140S. In panels (A) and (D), barplots are mean values of at least three independent experiments, with vertical bars representing standard deviation.

5.3.3 TevCas9 generates defined length deletions in HEK 293 cells

T7E1 mismatch cleavage assays report on the overall nuclease modification rate at a particular site, but cannot distinguish the relative contributions of the I-TevI and Cas9 active sites. To do so, we used a BamHI restriction site in the NQO2.54 target site that overlaps the predicted I-TevI CAACG cleavage motif (Fig. 5.2B). After transfections, TevCas9 or Cas9 activity was estimated from the fraction of PCR products amplified from the NQO2 site that are BamHI resistant as a result of nuclease cleavage and mNHEJ repair (Fig. 5.2C and 5.2D). In transfections with the TC-V and TC-WT (wild-type I-TevI) nucleases, we observed 14 – 16% BamHI cleavage resistance. This result implied that cleavage by the I-TevI nuclease domain at the CAACG motif knocked out the BamHI site. In contrast, transfections with Cas9 alone resulted in 5% BamHI cleavage resistance, implying that only a small fraction of Cas9-induced mNHEJ repair events destroy the BamHI site (Fig. 5.2D). A similar cleavage resistance assay was performed at the TSC1.5054 site, and showed that TC-VK cleavage generated a higher level of PvuII resistance than in transfections with Cas9 alone (18% versus 6%) (Fig. S4.3), consistent with deletion of the sequence between the I-TevI and Cas9 cleavage sites. We also took advantage of two adjacent NGG PAMs in the TSC1-5054 site to determine the effect of DNA spacer length on TevCas9 activity by simply changing the position of the gRNA (18_gRNA and 14_gRNA, Fig. 8A) (Fig. S4.3). We found that TC-VK programmed with the 18_gRNA produced ~4.5-fold more PvuII resistance than with TC-VK programmed with the 14_gRNA (Fig. S4.3), suggesting that the optimal DNA spacer length for the TevCas9 is between 15-18 bps.

5.3.4 Deep sequencing reveals minimal processing of TevCas9 deletions

To more accurately assess indels introduced by TevCas9 and Cas9 in HEK 293 cells, we used Illumina sequencing of PCR products amplified from the on-target sites to readout the type and length of indels. We first examined the reads for the length of deletions or insertions relative to the unmodified target site, and found that a high proportion of reads in the TevCas9 experiments were deletions that corresponded in length to the distance between the I-TevI and Cas9 cleavage sites (Fig. 5.3A, 5.3D, 5.3G, 5.3J). For the Cas9 experiments, a range of insertion and deletion lengths were observed, including a high proportion of +1 insertions, particularly at the RARA.233 target site. The overall modification rates at each target site, estimated from indel-containing reads, was

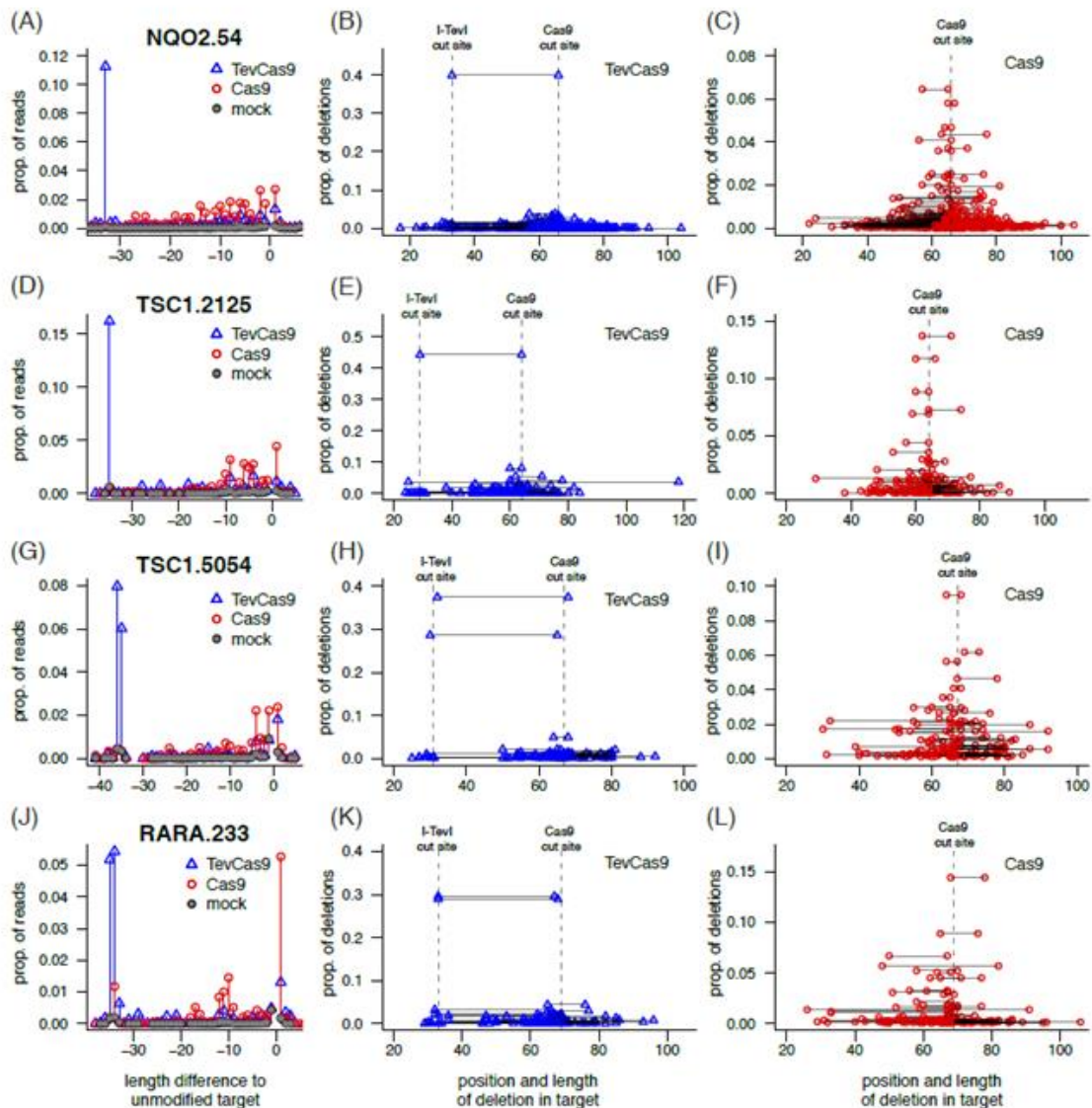


Figure 5.3: TevCas9 generates deletions of precise lengths in HEK 293 cells

Results from Illumina sequencing of PCR-amplified fragments for (A-C) NQO2, (D-F) TSC1-2125, (G-I) TSC1-5054, and (J-L) RARA target sites. (A,D,G,J) Proportion of reads with length differences relative to the unmodified target, with blue triangles representing TevCas9, red open circles representing Cas9, and grey dots representing mock transfection. (B,D,F,H) Proportion of TevCas9 reads with deletions mapped to the position in the target site. (C,F,I,L) Proportion of Cas9 reads with deletions mapped to the position in the target site. Dotted vertical lines indicate I-TevI and Cas9 cleavage sites.

very similar for TevCas9 and Cas9, agreeing with the T7E1 mismatch cleavage assays (Fig. S4.4). Strikingly, when we mapped the position of the deletions, the predominant deletions in the TevCas9 experiments were precise excisions of the DNA segment between the two cut sites (Fig. 5.3B, 5.3E, 5.3H, 5.3K). In contrast, we found that no single deletion product dominated the sequencing reads for the Cas9 experiment, that the deletion lengths were shorter and heterogeneous in length, and that the deletions were centered around the Cas9 cleavage site (Fig. 5.3C, 5.3F, 5.3I, 5.3L). These results demonstrate that TevCas9 cleavage generates defined length deletions at a high frequency in HEK 293 cells, and that the two non-compatible DNA ends are repaired with minimal DNA processing (Fig. 5.4).

5.3.5 Positioning of the I-TevI cleavage motif and gRNA biases in-frame to out-of-frame deletions

The relative position of the I-TevI and Cas9 cleavage sites at a given target site is determined by length of the DNA spacer that separates the I-TevI CNNNG cleavage motif from the gRNA binding site. At the NQO2.54 target, the distance between the I-TevI and Cas9 cleavage sites is 33-bps (measured from the I-TevI top strand cleavage site), which would generate an in-frame deletion of 11 amino acids. We examined the proportion of events at the NQO2.54 site and found that 59% of deletions are in-frame versus 41% that are out-of-frame (Fig. 5.4). In contrast, 20% of events generated by Cas9 are in-frame, and 70% out-of-frame. Conversely, TevCas9 events can be biased towards out-of-frame events, as seen at the TSC1.2125 and RARA.233 sites that have an even number of bases between the I-TevI and Cas9 cut sites. In both cases, TevCas9 generates a higher percentage of out-of-frame deletions than does Cas9 (Fig. S4.5).

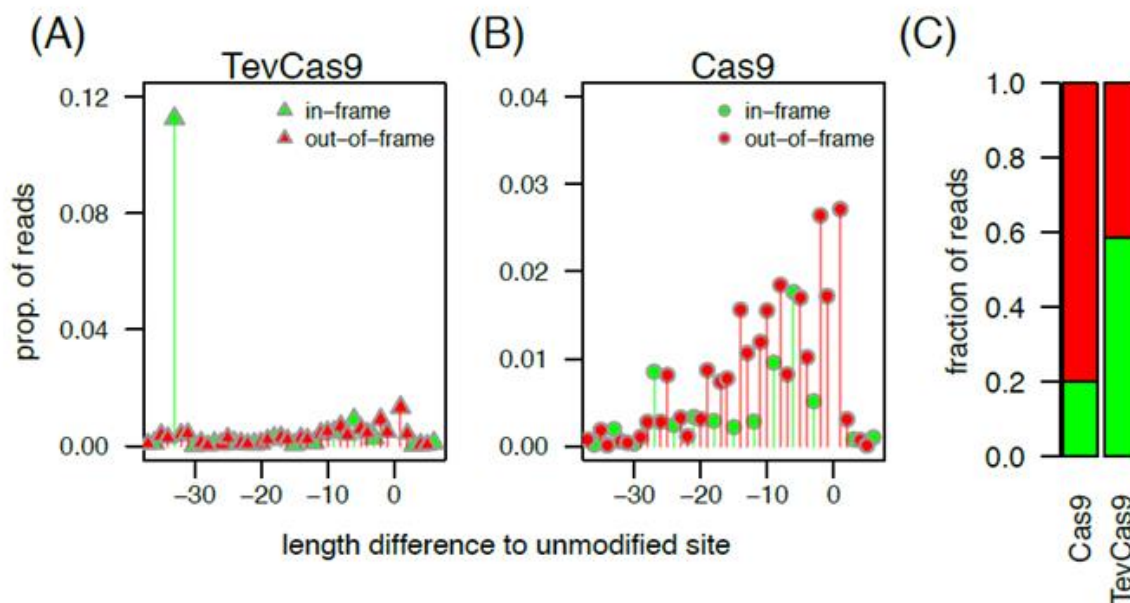


Figure 5.4: TevCas9 can bias the proportion of in-frame to out-of-frame indels

Illumina read data for the NQO2 target site is plotted as the proportion of reads that are in-frame (green) and out-of-frame (red) for (A) TevCas9 (triangles) and (C) Cas9 (open circles). (C) Fraction of reads that are in-frame or out-of-frame for TevCas9 and Cas9.

5.4 Discussion

Motivated by the observation that the majority of genome-editing nucleases generate a DSB with compatible cohesive ends, we hypothesized that a single-chain dual nuclease could bias genome-editing outcomes towards defined length deletions by generating two non-compatible DNA breaks at a target site. Subsequent repair would effectively delete the majority of the target site such that it would not be a substrate for further nuclease activity. Our first attempt at creating a dual-active site nuclease involved fusing the I-TevI nuclease and linker domains to the N-terminus of the I-OnuI LAGLIDADG homing endonuclease (or meganuclease), which acts as both a targeting and nuclease domain, creating MegaTevs (27, 34, 35). Using model DNA substrates, we showed that MegaTevs efficiently excises a 30-bp fragment from the target site in HEK 293 cells (30). However, the utility of the MegaTev platform is limited by re-engineering of meganuclease specificity to each desired target site, and by uncharacterized interactions of the I-TevI linker domain with the DNA spacer region that are critical for positioning of the nuclease domain at the 5'-CNNNG-3' cleavage site (36, 37).

Here, we have overcome limitations of the MegaTev platform, in two ways. First, we profiled the sequence requirements of the I-TevI linker-DNA spacer interactions, identifying nucleotide positions critical for activity and delineating a putative linker-DNA code that enabled development of a targeting model. Directed evolution experiments isolated I-TevI linker variants with activity on DNA spacers not cleaved by the wild-type linker, generating a set of variants with defined DNA spacer preferences. Second, we created an easier-to-target dual nuclease by fusing the I-TevI nuclease and linker domains to the N-terminus of Cas9 (28), generating an RNA-guided TevCas9 nuclease with two active sites. TevCas9 can be purified as an RNP that could be used for direct transfection of cell lines, as shown with Cas9 (38).

The I-TevI nuclease and linker domains have successfully been fused to four different DNA-binding architectures that are used in genome editing; zinc fingers (27), TALEs (29), meganucleases (30), and, as reported here, Cas9. An on-going debate in the genome-editing field centers on the ease of use (targeting range) versus specificity (off-target effects) of the various reagents, particularly for common laboratory manipulations of cell lines or model organisms. Of the I-TevI-based genome-editing nucleases, the TevCas9 fusions are likely to be the most user-friendly given the ease of programming Cas9-substrate interactions. Similarly, the identification

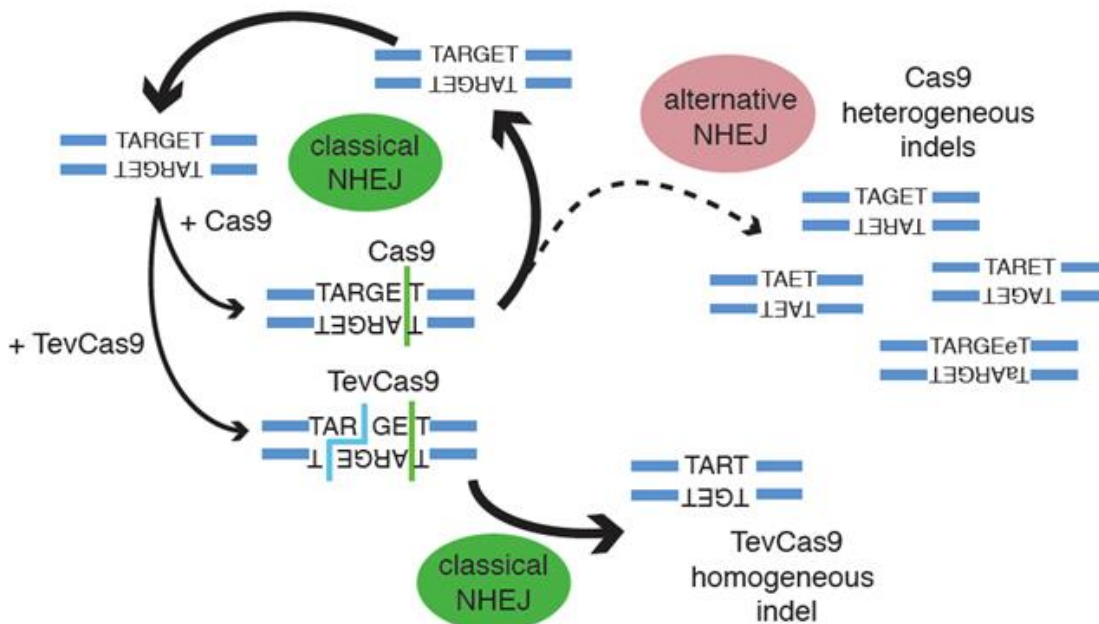


Figure 5.5: Model of how TevCas9 biases DNA repair outcomes

TevCas9 or Cas9 recognize and cleave a target site. The non-compatible DNA ends and 33-36 base pair deletion generated by TevCas9 prevents regeneration of the target site. Compatible DNA ends generated by Cas9 are repaired by NHEJ, regenerating the target site and inducing a cycle of cleavage and ligation. At a lower rate (dashed line), some Cas9 events undergo repair by the alternative NHEJ pathway, getting heterogeneous length indels.

and characterization of I-TevI linker variants makes the TevCas9 reagent easy-to-use in that targeting of a desired site does not require extensive cycles of engineering and optimization of I-TevI-DNA specificity. For TevCas9, we predicted a potential target site on average every 14 bps for the NQO2, TSC1 and RARA genes, with CG-rich regions having a higher number of potential sites due to the nature of the NGG PAM and I-TevI 5'-CNNNG-3' cleavage motif. The targeting requirements of TevCas9 (two cleavage sites separated by a defined length spacer) are similar to the parameters for paired Cas9 nickases and FokI-dCas9 nucleases that function as dimers (two independent gRNAs separated by a defined length) (9-11, 26, 33). While future refinements of the TevCas9 binding model will improve target site selection, especially for the I-TevI linker variants with preferences for substitutions in positions T6 or G8 of the DNA spacer, it is encouraging that the high density of predicted sites will not severely limit targeting of TevCas9. Moreover, our recent demonstration that I-TevI cleavage specificity can be modulated to include different cleavage motifs (notably 5'-NNNNG-3') could also increase targeting potential by alleviating the requirement for a CNNNG motif (39). I-TevI fusions to Cas9 variants with altered PAM specificities (40, 41), or fusion to other Cas9 homologs (42), could similarly broaden targeting range. It remains to be seen if the I-TevI nuclease and linker domains mitigate Cas9 off-target effects, as observed with the chimeric ZF-Cas9 and FokI-dCas9 fusions (9, 11, 43). The TevCas9 fusions described here further expand the repertoire of chimeric fusions that are possible with Cas9. Notably, the TevCas9 fusion is similar to the FokI-dCas9 fusion in that both I-TevI and FokI are nuclease domains.

One striking result from our study was that TevCas9 generates defined length deletions of 33-36 base pairs in HEK 293 with minimal DNA end processing. Sequences of the TevCas9 deletion products are consistent with repair by the c-NHEJ pathway whereby the protruding I-TevI 3' overhang was directly ligated to the blunt end generated by Cas9, with subsequent fill-in of the 2-nt gap on the opposite strand (Fig. 6). The nature of the repair products must mean that TevCas9 sequesters the DNA ends from exonucleolytic processing associated with NHEJ pathways after cleavage, possibly because product release by I-TevI and Cas9 (at least in the context of the chimeric TevCas9) is slow (44, 45). In contrast, the heterogeneous length indels observed at Cas9 events are typical of repair by the alternative NHEJ pathway. Although we did not examine the efficiency or types of indels generated by paired Cas9 nickases, published data demonstrates that Cas9 nickases introduce heterogeneous length indels (fig S4.5) (46, 47). Interestingly, our *in vitro*

data show that cleavage by I-TevI is ~30-fold slower than Cas9, which is also supported by the deep sequencing data showing that approximately half of all TevCas9 indel events are consistent with cleavage only by Cas9. This observed difference in cleavage rates could reflect a sub-optimal fusion point between I-TevI and Cas9 that hinders I-TevI activity, or inherently faster cleavage by Cas9. Improving the rate of I-TevI cleavage in the context of the chimeric TevCas9 nuclease by optimization of the fusion could conceivably bias events even more towards deletions of defined length than is observed now.

Cas9 is commonly used to create gene knockouts for downstream phenotypic studies (33). However, in-frame deletions could generate different phenotypes than knockouts if the deletions encompass functionally important regions. The TevCas9 nuclease can generate in-frame deletions of 11 amino acids efficiently, because the length of the deletion is determined by the distance between the I-TevI and Cas9 cleavage sites. This property of TevCas9, coupled with the directional nature of the TevCas9 deletion (a 3' 2-nt overhang by I-TevI and a blunt DNA end by Cas9), could enhance ligation of oligonucleotides, enabling applications such as site-directed mutagenesis, swapping of functional domains, or epitope-tagging (48). Conversely, TevCas9 could be used to tile across promoter regions to delineate functional elements (49), or protein-DNA interaction sites in introns or other non-coding regions.

In summary, we have shown that the TevCas9 dual nuclease provides a strategy to bias genome-editing events towards deletions of defined lengths, escaping the persistent cycle of cleavage and ligation of compatible DSBs that results in heterogeneous length deletions observed with other genome-editing nucleases.

5.5 References

- 1 Doudna, J. A. & Charpentier, E. (2014). Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096.
- 2 Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D (2010). Genome editing with engineered zinc finger nucleases. *Nature reviews. Genetics* **11**, 636-646.
- 3 Bogdanove, A. J. & Voytas, D. F. (2011) TAL effectors: customizable proteins for DNA targeting. *Science* **333**, 1843-1846.
- 4 Boissel S, *et al.* (2014) megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic acids research* 42(4):2591-2601.

5. Sather BD, *et al.* (2015) Efficient modification of CCR5 in primary human hematopoietic cells using a megaTAL nuclease and AAV donor template. *Sci Transl Med* 7(307):307ra156.
- 6 Stoddard, B. L. (2014) Homing endonucleases from mobile group I introns: discovery to genome engineering. *Mobile DNA* 5, 7.
- 7 Sternberg, S. H. & Doudna, J. A. (2015). Expanding the Biologist's Toolkit with CRISPR-Cas9. *Molecular cell* 58, 568-574.
- 8 Carroll D (2014) Genome engineering with targetable nucleases. *Annual review of biochemistry* 83:409-439.
- 9 Guilinger JP, Thompson DB, & Liu DR (2014) Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nature biotechnology* 32(6):577-582.
- 10 Aouida M, *et al.* (2015) Efficient fdCas9 Synthetic Endonuclease with Improved Specificity for Precise Genome Engineering. *PloS one* 10(7):e0133373.
- 11 Tsai SQ, *et al.* (2014) Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nature biotechnology* 32(6):569-576.
- 12 Bibikova M, Golic M, Golic KG, & Carroll D (2002) Targeted chromosomal cleavage and mutagenesis in *Drosophila* using zinc-finger nucleases. *Genetics* 161(3):1169-1175.
- 13 Mladenov E & Iliakis G (2011) Induction and repair of DNA double strand breaks: the increasing spectrum of non-homologous end joining pathways. *Mutat Res* 711(1-2):61-72.
- 14 Lieber, M. R. (2011) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annual review of biochemistry* 79, 181-211.
- 15 McVey, M. & Lee, S. E. (2008) MMEJ repair of double-strand breaks (director's cut): deleted sequences and alternative endings. *Trends Genet* 24, 529-538.
- 16 Shrivastav, M., De Haro, L. P. & Nickoloff, J. A. (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Res* 18, 134-147.
- 17 Deriano L & Roth DB (2013) Modernizing the nonhomologous end-joining repertoire: alternative and classical NHEJ share the stage. *Annu Rev Genet* 47:433-455.
- 18 Richardson CD, Ray GJ, Bray NL, & Corn JE (2016) Non-homologous DNA increases gene disruption efficiency by altering DNA repair outcomes. *Nat Commun* 7:12463.

- 19 Beumer KJ, *et al.* (2013) Comparing zinc finger nucleases and transcription activator-like effector nucleases for gene targeting in *Drosophila*. *G3 (Bethesda)* 3(10):1717-1725.
- 20 Chen S, *et al.* (2013) A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly. *Nucleic acids research* 41(4):2769-2778.
- 21 Chen, X. *et al.* (2016) Probing the impact of chromatin conformation on genome editing tools. *Nucleic acids research* **44**, 6482-6492.
- 22 Daboussi, F. *et al.* (2012) Chromosomal context and epigenetic mechanisms control the efficacy of genome editing by rare-cutting designer endonucleases. *Nucleic acids research* **40**, 6367-6379.
- 23 Lin, S., Staahl, B. T., Alla, R. K. & Doudna, J. A. (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *Elife* **3**, e04766.
- 24 Certo, M. T. *et al.* (2012) Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nat Methods* **9**, 973-975.
- 25 Delacote F, *et al.* (2013) High frequency targeted mutagenesis using engineered endonucleases and DNA-end processing enzymes. *PLoS one* 8(1):e53217.
- 26 Ran, F. A. *et al.* (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389.
- 27 Kleinstiver BP, *et al.* (2012) Monomeric site-specific nucleases for genome editing. *Proceedings of the National Academy of Sciences of the United States of America* 109(21):8061-8066.
- 28 Jinek M, *et al.* (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337(6096):816-821.
- 29 Kleinstiver BP, *et al.* (2014) The I-TevI Nuclease and Linker Domains Contribute to the Specificity of Monomeric TALENs. *G3 (Bethesda)* 4(6):1155-1165.
- 30 Wolfs JM, *et al.* (2014) MegaTevs: Single chain dual nucleases for efficient gene disruptions. *Nucleic acids research* 42(13):8816-8829.
- 31 Qiu P, *et al.* (2004) Mutation detection using Surveyor nuclease. *Biotechniques* 36(4):702-707.
- 32 Guschin DY, *et al.* (2010) A rapid and general assay for monitoring endogenous gene modification. *Methods Mol Biol* 649:247-256.

- 33 Ran FA, *et al.* (2013) Genome engineering using the CRISPR-Cas9 system. *Nature protocols* 8(11):2281-2308.
- 34 Sethuraman J, Majer A, Friedrich NC, Edgell DR, & Hausner G (2009) Genes within genes: multiple LAGLIDADG homing endonucleases target the ribosomal protein S3 gene encoded within an rnl group I intron of Ophiostoma and related taxa. *Molecular biology and evolution* 26(10):2299-2315.
- 35 Takeuchi R, *et al.* (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proceedings of the National Academy of Sciences of the United States of America* 108(32):13077-13082.
- 36 Dean AB, *et al.* (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proceedings of the National Academy of Sciences of the United States of America* 99(13):8554-8561.
- 37 Liu Q, *et al.* (2008) Role of the interdomain linker in distance determination for remote cleavage by homing endonuclease I-TevI. *Journal of molecular biology* 379(5):1094-1106.
- 38 Kim S, Kim D, Cho SW, Kim J, & Kim JS (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* 24(6):1012-1019.
- 39 Roy AC, Wilson GG, & Edgell DR (2016) Perpetuating the homing endonuclease life cycle: identification of mutations that modulate and change I-TevI cleavage preference. *Nucleic acids research* 44(15):7350-7359.
- 40 Kleinstiver BP, *et al.* (2015) Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nature biotechnology* 33(12):1293-1298.
- 41 Kleinstiver BP, *et al.* (2015) Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* 523(7561):481-485.
- 42 Esvelt KM, *et al.* (2013) Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* 10(11):1116-1121.
- 43 Bolukbasi MF, *et al.* (2015) DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat Methods* 12(12):1150-1156.
- 44 Mueller JE, Smith D, & Belfort M (1996) Exon coconversion biases accompanying intron homing: battle of the nucleases. *Genes & development* 10(17):2158-2166.
- 45 Richardson CD, Ray GJ, DeWitt MA, Curie GL, & Corn JE (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nature biotechnology*. 34 (3):339-345.

- 46 Karimova M, *et al.* (2015) CRISPR/Cas9 nickase-mediated disruption of hepatitis B virus open reading frame S and X. *Scientific reports* 5:13734.
- 47 Fu Y, Sander JD, Reyon D, Cascio VM, & Joung JK (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nature biotechnology*.32 (3)279-285.
- 48 Lackner DH, *et al.* (2015) A generic strategy for CRISPR-Cas9-mediated gene tagging. *Nat Commun* 6:10237.
- 49 Sanjana NE, *et al.* (2016) High-resolution interrogation of functional elements in the noncoding genome. *Science* 353(6307):1545-1549.
- 50 Edgar RC (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* 26(19):2460-2461.

Chapter 6

6 Discussion

The desire to manipulate model organisms or correct human disease has led to a huge development in engineered nucleases. Genome engineering was named Method of the Year in 2011 by Nature Methods with “CRISPR” publications continuing to grow at an exponential rate (1). A number of genome-editing reagents have been created over the years from engineered LAGLIDADG homing endonucleases to ZFNs and TALENs, a set of FokI derived reagents, and finally the most recent CRISPR-Cas9 system. Each reagent has its own set of advantages and disadvantages that make up the genome-editing “toolbox”, with no single reagent being ideal for every situation. One similarity of all four reagents is the persistence of a cleavage and re-ligation cycling event (2). Here, we describe another potential genome-editing reagent using the GIY-HE I-TevI catalytic and linker domain fused to the LHE (MegaTev) or Cas9 (TevCas9) nuclease, creating a dual-active engineered reagent capable of making two DSBs at a single target site. We demonstrate that these dual-active endonucleases efficiently excise a short fragment from their target site and the remaining ends are ligated together *in vivo*. In addition, we demonstrate that I-TevI is a portable domain that can be fused to all four DNA-binding platforms: LAGLIDADG, ZFs, TALEs, and Cas9. Utilizing the MegaTev fusion, I investigated the nuclease and linker specificity and identified key DNA sequence preferences required to promote robust I-TevI cleavage. Using a bacterial selection system, I was able to isolate I-TevI mutants with alternative DNA specificity, expanding the number of sites available for I-TevI reagents. Adapting native I-TevI to new target sequences was a critical step in uncovering its potential in the genome-editing field. I demonstrate the new I-TevI variants are portable to three different DNA-binding platforms and maintain their identified specificity. Ultimately, we fused I-TevI to Cas9 to create TevCas9, an adaptable reagent and demonstrated for the first time that I-TevI can target novel genomic locus. Distinctly, TevCas9 fusions are a programmable dual-endonuclease that can excise short fragments at a single target site which could provide alternative applications compared to conventional genome-editing reagents.

6.1 Considering off-target cleavage of genome editing reagents

One of the most important aspects when optimizing genome editing reagents is minimizing off-target cleavage events. Expressing a non-specific reagent can be detrimental to the cell, as off-target DSBs can lead to chromosomal rearrangements, inversion, deletions and substitutions within the genome (3–5). A nuclease that has only a couple off-target cleavage events can still be detrimental to the cell as changes on the genomic level in the germline are passed to future generations. This presents a common conundrum, how do you create a reagent that is highly adaptable but also highly-specific. We proposed that the inherent specificity of the I-TevI catalytic domain and linker region would reduce off-target cleavage as seen with other genome-editing reagents (6). We demonstrate in every platform that I-TevI cleavage is directed by the DNA-binding platform and not the I-TevI nuclease domain, which suggests minimal additional off-target cleavage events. This correlates well with previous studies of I-TevI, as the nuclease domain has low affinity for DNA substrates (7). We did explore some potential off-target sites for MegaTev and TevCas9 at predicted target sites and saw no evidence that I-TevI increased off-target cleavage. However a more exhaustive approach is required to better analyze the potential off-target effects of I-TevI-based reagents. Surveying for off-target events is a very technical and expensive challenge that should be performed with optimized reagents ready for agricultural or clinical applications.

Many approaches have been used to study undesired modifications created by genome-editing reagents. Our initial screens for MegaTev and TevCas9 fusions involved only screening predicted sites with the fewest number of mismatches to the on-target. The challenge in developing a whole-genome screening method is it requires high sensitivity techniques capable of detecting low frequency mutations. There are a number of thorough approaches that have been described to date, but like genome-editing reagents, each technique has its advantages and disadvantages. A large number of bioinformatics approaches to predict off-target cleavage were developed for a number of reagents but many fell short at predicting all potential off-target sites (8, 9). Researchers turned to an *in vitro* SELEX (system evolution of ligands exponential enrichment) that allowed them to cover a large number of sites (10^{12}) but it failed to mimic a cellular environment (10). A similar approach, known as digenome-seq, used cell free genomic DNA to analyze Cas9 off-target cleavage and resulted in an overestimation of the number of target sites similar to SELEX (11). ChIP-seq (Chromatin immunoprecipitation sequencing) was another technique that attempted to capture all

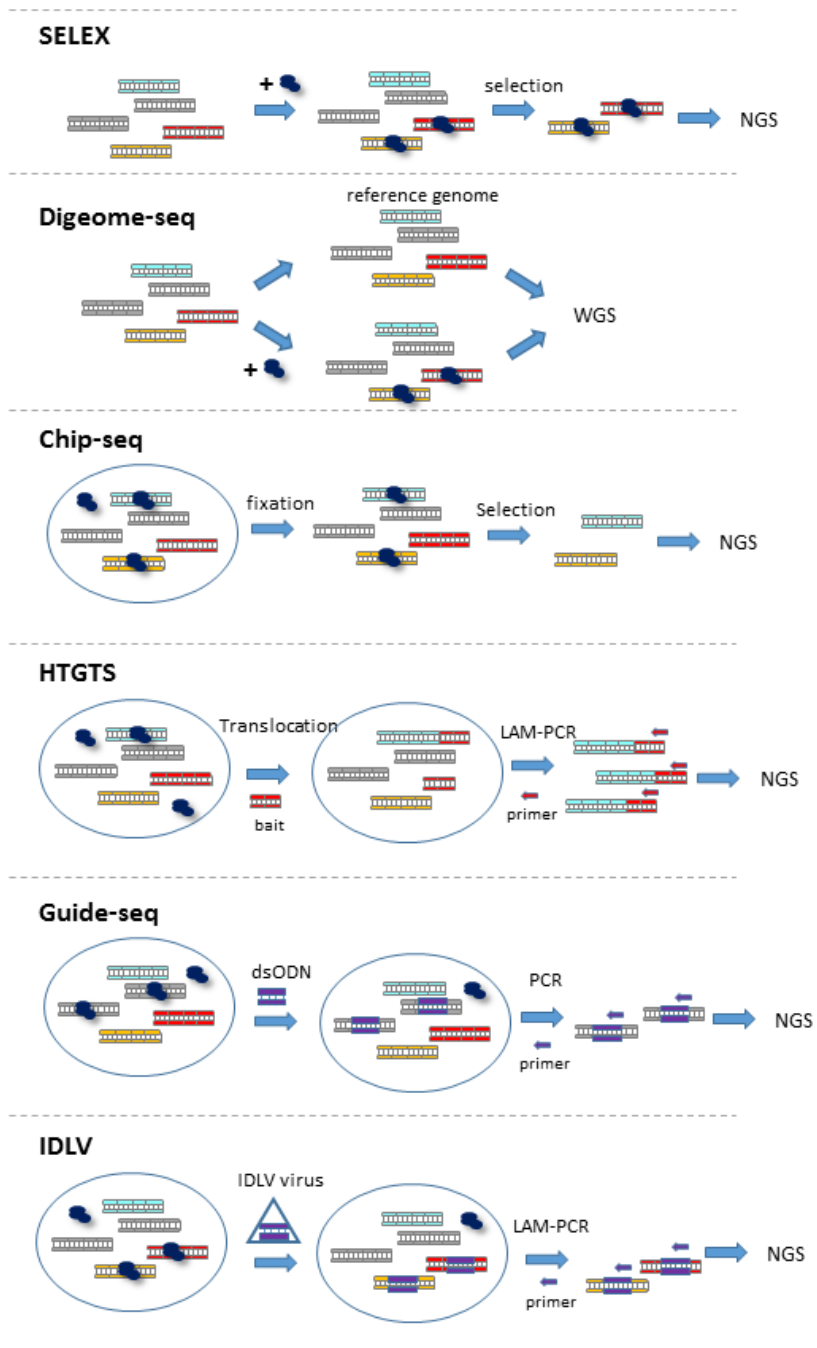


Figure 6.1: Schematic of off-target screening techniques using next-generation sequencing

Various methods used to determine off-target cleavage frequencies of genome-editing reagents. SELEX (system evolution of ligands by exponential enrichment), ChIP-seq (Chromatin immunoprecipitation sequencing), HTGTS (high-throughput genome-wide translocation sequencing), IDLV capture (integrase-defective lentiviral vectors), dsODN (double-stranded oligonucleotide), NGS (next-generation sequencing), WGS (whole-genome sequencing)

the off-target sites for many nucleases, but the high background noise and capture of binding sites once again did not properly represent off-target cleavage events in a cell (12–14). Some studies report the number of off-target binding sites being greater than 1200, however not all bound sites are cleaved suggesting binding is more promiscuous than cleavage activity (1, 13). Two approaches, IDLV capture (integrase-defective lentiviral vectors) and guide-seq, used a DNA integration approach to capture off-target events of a linear IDLV genome or a small double-strand oligonucleotide, respectively (15, 16). The main limitations of these techniques is that they depend on DNA integration to capture the off-target events which may be less efficient than indel formation. For IDLV, the major limitation is the sensitivity, as it can only detect mutation frequency of greater than 1% (15). Similarly, guide-seq failed to identify 7 well-defined off-target sites confirmed using chip-seq and cleavage by indel generation (13). The last method called HTGTS (high-throughput genome-wide translocation sequencing) detects DSBs based on translocation events to other DNA sequences (17). The most systematic comparison of these approaches was carried out using spCas9 at a VEGFA gene target with the same gRNA in all four studies ChiP-seq (5), Guide-seq (16), HTGTS (17) and digenome-seq (11) identifying 55, 21, 38 and 87 off-target sites, respectively. Note the number of off-target sites varies greatly with each technique and no single method identified all the same off-targets sites. Since each technique identified novel off-target sites, it poses the question, how many off-target sites are there truly and how safe are engineered nucleases? The difficulty in detecting off-target cleavage should be a sign to proceed with caution using these reagents (18). In recent years, concerns have risen after the failed attempts to modify human embryos (18–20).

Indeed, as technology advances our ability to survey for off-target cleavage will improve and engineering a precise reagent may become a reality. The limitation in current technology has not deterred researchers from attempting to develop precise reagents. A large number of studies have focused on improving genome-editing reagents using various techniques. Some of the first off-target cleavage events observed by ZFNs was a result of ZFNs binding at “half-sites” that caused dimerization and cleavage from FokI at unintended target sites. This was improved by eliminating homo-dimerization of FokI to create an obligate hetero-dimerization to reduce complex formation at “half-sites” (21). The discovery of TALENs also helped to de-toxify FokI based fusions as longer recognition domain could be assembled. Similarly, by fusing the TALE domain to LHEs, a reduction in off-target cleavage was observed and also resulted in an increase in on-target

efficiency (6). Off-target cleavage is no stranger to the new CRISPR-Cas9 system as the initial off-target analysis of spCas9 identified a large number of off-target sites (16, 5, 22). Initial studies demonstrated Cas9 could tolerate multiple mismatches in the target site with specificity varying based on Cas9 ortholog, gRNA architecture, target sequence, PAM, and relative concentration and duration of treatment (16, 23–26). Identified off-targets were poorly related to the on-target site making them difficult to predict bioinformatically (16, 22). However, as demonstrated with previous genome-editing reagents, modifications to the CRISPR-Cas9 architecture created more specific versions. One study truncated the gRNA to reduce the overall binding affinity of the DNA/RNA complex to reduce off-target cleavage (16, 27). Similarly, new spCas9 variants with mutations knocking-out the non-specific DNA interactions of spCas9 greatly reduced off-target cleavage (28, 29). Most recently, identification of new CRISPR-Cas Cpf1 nucleases have remarkable specificity only detecting off-target sites for 17 of 20 sites with AsCpf1 and 12 of 20 with LbCpf1 (30, 31). Using the new high-fidelity Cas9 variants or Cpf1 Cas proteins would help reduce any potential off-targets observed for TevCas9 fusions. In addition, eliminating non-specific contacts could be applied to I-TevI to further reduce the binding affinity to ensure only the on-target is cleaved.

6.2 Implications of I-TevI biology for genome engineering

I-TevI based genome engineering should proceed with caution as more thorough off-target profiling needs to be performed with these reagents. As discovered with FokI-based genome-editing reagents, DNA-binding platforms that guide nucleases to various sites in the genome bind a number of promiscuous sites resulting in off-target cleavage. FokI requirement of dimerization for cleavage was not sufficient in constructing a precise nuclease. However, this observation can also be attributed to its non-specific nature, resulting in cleavage at every dimerization complex. We approached this problem from the opposite direction, presenting an endonuclease, I-TevI which possesses inherent specificity and distance requirements for efficient DSB formation. At first glance, I-TevI seemed too restrictive to be a useful genome-editing reagent, but through further characterizing the nuclease and linker domain specificity, we were able to identify important base-pairs required for cleavage. This allowed us to take a directed evolution approach similar to engineering LAGLIDADG proteins, mutating specific modules that may be contacting these regions. Unfortunately, no crystal structures of the I-TevI nuclease domain and linker region

contacting DNA have been solved, so mutated regions were large to encompass potential DNA-binding residues. We identified I-TevI linker variants with altered DNA specificity that could be utilized to increase the number of potential target sites available for I-TevI-derived reagents. Site-specific nuclease fusions have been previously demonstrated to increase efficiency and specificity as described with the MegaTAL platform (6). In terms of available target sites in the genome, I-TevI is more malleable than LHEs but more restrictive than the CRISPR-Cas9 system and on par with TALENs. Nevertheless, one must select the appropriate genome-editing reagent depending on your desired application.

One important aspect of a good genome-editing reagent is its ability to be applied readily, without further protein engineering or understanding. The ultimate goal is to reach into a genetic “toolbox” and pick a reagent to manipulate a genome of interest, in the same fashion restriction endonucleases are used to manipulate plasmids. Therefore, a desirable reagent would check all three requirements of being highly efficient, specific and “ready-to-use”. CRISPR-Cas9 system is the closest reagent that checks all three of the boxes with exception of specificity that will require more studies. I-TevI-derived engineering reagents, specifically the TevCas9 fusion, is very close to checking all three requirements. With further characterization of the various I-TevI variants and profiling their specificity, a target site predictive software could be developed using a position-specific scoring matrix (PSSM) to rank various sites in the genome. Similar to software developed for other reagents like CRISPR-Cas9, an online tool would be available to insert your gene of interest and would rank potential TevCas9 target sites for the user. Therefore, this would transition TevCas9 into a more “ready-to-use” reagent where you order the correct I-TevI variant and gRNA specified from the software predictions.

Our ability to manipulate I-TevI to target various sites within the human genome was a large accomplishment in proving it could have a role in the genome-editing field. However, there are many hurdles that still must be overcome, as inherent sequence specificity is still restrictive and a thorough off-target profiling has yet to be performed. Understanding the basic biology of the I-TevI nuclease and linker domain may aid in overcoming these future hurdles. Cross-linking studies would help to identify key residues within the protein that are making base-specific contacts and permit a more directed protein engineering approach. Identifying base-specific contacts could help

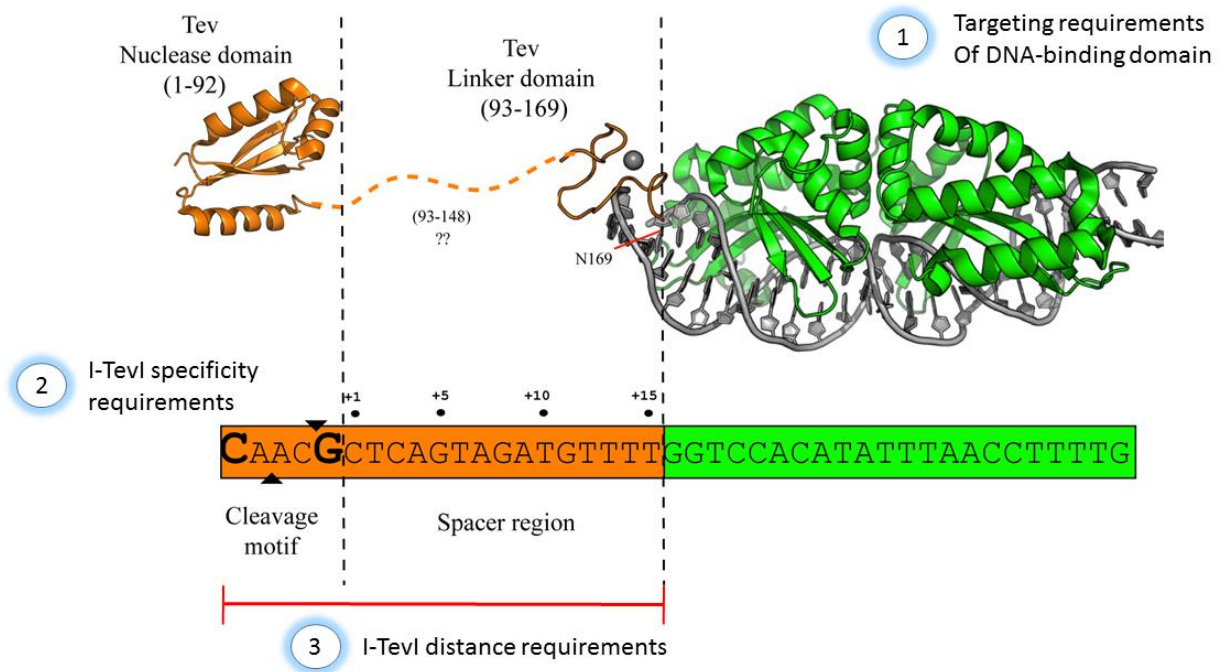


Figure 6.2: Considerations when targeting I-TevI derived nucleases

further engineer I-TevI variants with alternative specificity without reducing overall specificity of the nuclease domain. In addition, studying the mechanism could help reduce possible off-target sites by potentially making an I-TevI nickase variants. Whatever the implication, understanding the basic biology of I-TevI will aid in applying the nuclease domain in an effective manner.

6.3 Potential for dual-active GIY-YIG endonucleases

Current genome editing reagents only introduce a single DSB at their target site to induce mutagenic NHEJ repair to create gene knockouts. This may lead to persistent cleavage at their target sites with a cyclic cleavage and ligation reaction resulting in an unproductive gene modification. Although, a lot of reagents are efficient enough to overcome the classical NHEJ pathway and produce a mutagenic repair, the repair outcomes are unpredictable. This means only 2/3 of outcomes will result in an out-of-frame mutation resulting in gene knockout. One potential application of the dual-active TevCas9 fusion is the ability to bias the outcome through excising a short fragment from its target site. With careful design, excising a fragment not divisible by 3 (complete codons), would bias repair events to produce gene knockouts more frequently than 66%. As a result, dual-active TevCas9 would be more efficient genome-editing reagent in producing gene knockouts over Cas9 alone. However, expression of two active nuclease may result in increased off-target cleavage, but this has yet to be investigated. Alternatively, dual-active TevCas9 could be used to bias repair events toward in-frame deletions. This may be beneficial as the excised fragment is large enough to eliminate an alpha-helix from your protein of interest or a binding site from a promoter. Sequence deletions are not a new concept, multiplexing genome-editing reagents such as Cas9 could be used to delete intervening sequences (32). However, these studies produce more variable repair products compared to TevCas9 likely attributed to TevCas9 ability to sequester the excised fragment from cellular repair proteins more effectively than two independent nucleases. In addition, expressing two different gRNAs has been demonstrated to increase off-target events.

Another aspect of genome-editing is the ability to bias cellular DNA-repair pathways from mutagenic NHEJ to a gene correction homologous recombination (HR) repair pathway. Gene correction events using HR (1-8%) still trouble the genome-editing field as the efficiency is orders of magnitude lower than mutagenic NHEJ (15-80%) (33–35, 23, 36). Since HR and NHEJ are in competition the challenge is inducing HR without any NHEJ background (5, 37). The efficiency

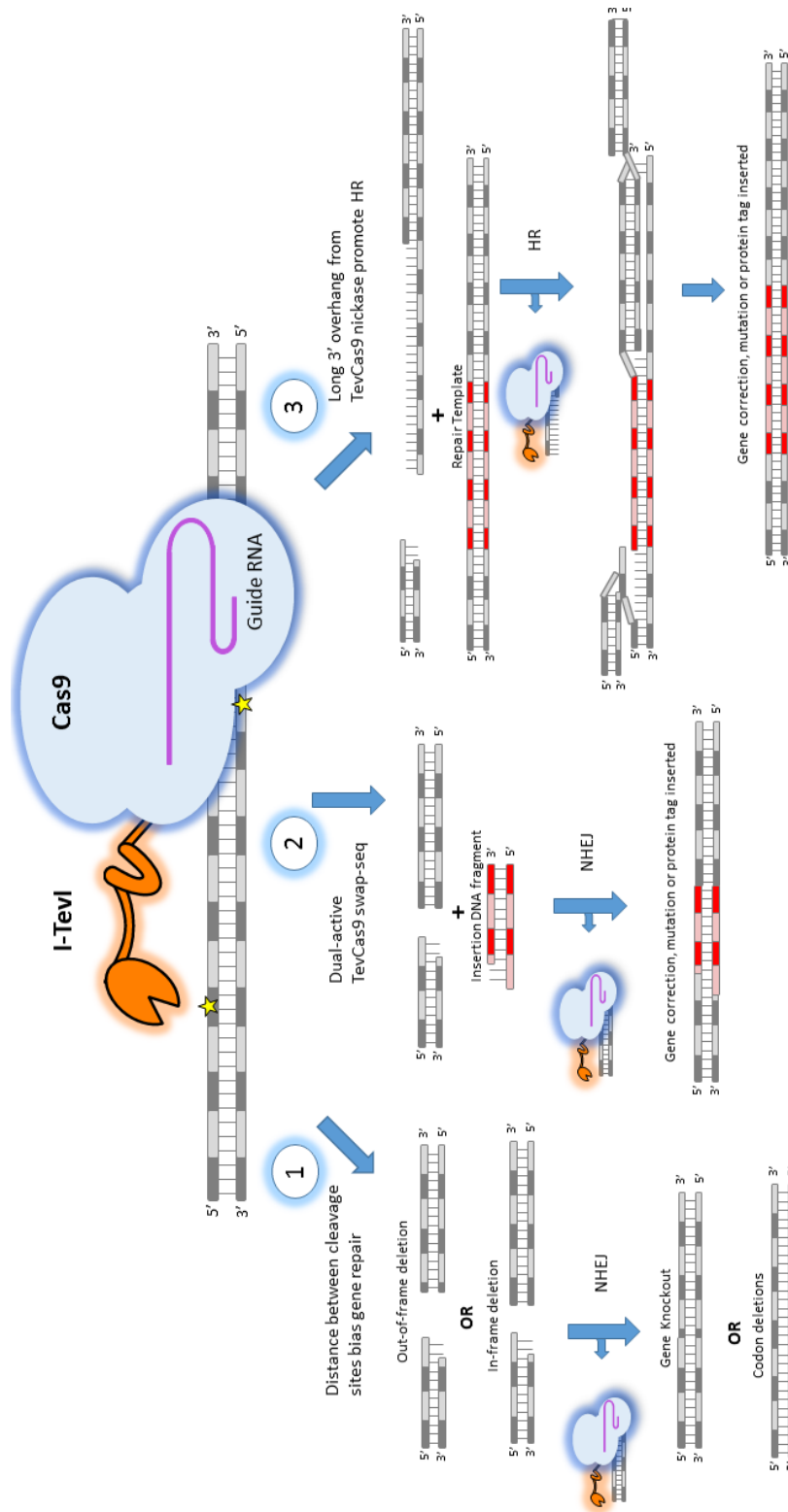


Figure 6.3: Potential genome engineering outcomes using TevCas9 dual-active nucleases

of both HR and NHEJ are highly dependent on chromosomal location, nuclease platform, cell cycle, cell type and topology of repair template (36, 38). Some efforts to increase HR and reduce NHEJ were to create nickases, this would still promote HR and yet not activate NHEJ by only cleaving one strand of DNA (39, 40). However, this did successfully reduce NHEJ but did not significantly increase the rate of HR. Other attempts to increase HR have been to use chemical inhibitors to inhibit NHEJ repair proteins and promote HR. One study utilized small molecules and only observed a 2-fold increase in knock-in efficiency ranging from 0.8% to 3.5% with only a 2-fold increase (41). Only moderate increases in HR were observed, but still not the efficiency required for widespread use. Recently, a study reported HR frequencies of ~60% through simple rules for designing a single stranded DNA donor template in conjunction with Cas9, but this has to be further studied at other genomic sites (42). Resection creating a free 3' overhang is the first step in HR repair, which could be established using a TevCas9 mutant variant (Figure 6.3). It would be intriguing to determine if a TevCas9 fusion containing a Cas9 nickase mutant could be utilized to produce a long 3' overhang to help stimulate HR repair. The long 3' overhang may direct cellular repair away from NHEJ and towards HR increasing the efficiency. This is unique from past attempts of just reducing NHEJ repair instead of stimulating HR repair.

The final potential application for the dual-active TevCas9 is to capitalize on its ability to excise a fragment from the genome and replace it with a novel sequence. Similar to off-target profiling approaches guide-seq and IDLV capture, a short double-stranded oligo or linear IDLV fragment are integrated into the genome. Similar to these two approaches, the excised TevCas9 fragment can be replaced with an alternative sequence of choice. This would circumvent the problem of low HR frequency as sequences could be readily swapped out of genomes using TevCas9, a protocol I termed "swap-seq". This has the potential to be used to correct gene mutations, add alternative mutations, or insert protein tags on endogenous genes. However, the inherent specificity of I-TevI may restrict the available sites that can be modified using swap-seq. Overall, the unique dual-active function and applicability of the I-TevI based genome-editing reagent TevCas9 presents some interesting new avenues for genome modifications.

6.4 References

1. Sternberg, S.H. and Doudna, J.A. (2015) Expanding the Biologist's Toolkit with CRISPR-Cas9. *Mol. Cell*, **58**, 568–574.

2. Certo, M.T., Gwiazda, K.S., Kuhar, R., Sather, B., Curinga, G., Mandt, T., Brault, M., Lambert, A.R., Baxter, S.K., Jacoby, K., *et al.* (2012) Coupling endonucleases with DNA end-processing enzymes to drive gene disruption. *Nat. Methods*, **9**, 973–5.
3. Cannan, W.J. and Pederson, D.S. (2016) Mechanisms and Consequences of Double-Strand DNA Break Formation in Chromatin. *J. Cell. Physiol.*, **231**, 3–14.
4. Dumitrache, L.C., Hu, L., Son, M.Y., Li, H., Wesevich, A., Scully, R., Stark, J. and Hasty, P. (2011) Trex2 enables spontaneous sister chromatid exchanges without facilitating DNA double-strand break repair. *Genetics*, **188**, 787–797.
5. Fu, Y., Foden, J. a, Khayter, C., Maeder, M.L., Reyon, D., Joung, J.K. and Sander, J.D. (2013) High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat. Biotechnol.*, **31**, 822–6.
6. Boissel, S., Jarjour, J., Astrakhan, A., Adey, A., Gouble, A., Duchateau, P., Shendure, J., Stoddard, B.L., Certo, M.T., Baker, D., *et al.* (2014) MegaTALs: A rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res.*, **42**, 2591–2601.
7. Dean, A.B., Stanger, M.J., Dansereau, J.T., Van Roey, P., Derbyshire, V. and Belfort, M. (2002) Zinc finger as distance determinant in the flexible linker of intron endonuclease I-TevI. *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 8554–61.
8. Hsu, P.D., Lander, E.S. and Zhang, F. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, **157**, 1262–1278.
9. Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., Vandyk, J.K. and Bogdanove, A.J. (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: Tools for TAL effector design and target prediction. *Nucleic Acids Res.*, **40**, 117–122.
10. Gabriel, R., Lombardo, A., Arens, A., Miller, J.C., Genovese, P., Kaeppl, C., Nowrouzi, A., Bartholomae, C.C., Wang, J., Friedman, G., *et al.* (2011) An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat. Biotechnol.*, **29**, 816–823.
11. Kim, D., Bae, S., Park, J., Kim, E., Kim, S., Yu, H.R., Hwang, J., Kim, J.I. and Kim, J.S. (2015) Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods*, **12**, 237–43.
12. Cencic, R., Miura, H., Malina, A., Robert, F., Ethier, S., Schmeing, T.M., Dostie, J. and Pelletier, J. (2014) Protospacer adjacent motif (PAM)-distal sequences engage CRISPR Cas9 DNA target cleavage. *PLoS One*, **9**, 1–13.
13. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. and Adli, M. (2014) Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol*, **32**, 677–683.
14. Wu, X., Scott, D.A., Kriz, A.J., Chiu, A.C., Hsu, P.D., Dadon, D.B., Cheng, A.W., Trevino, A.E., Konermann, S., Chen, S., *et al.* (2014) Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol*, **32**, 670–676.
15. Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R. and Yee, J. (2015) Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.*, **33**, 175–178.
16. Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V. V, Thapar, V., Wyvekens, N., Khayter, C.,

- lafrate,A.J., Le,L.P., *et al.* (2015) GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol*, **33**, 187–197.
17. Frock,R.L., Hu,J., Meyers,R.M., Ho,Y.-J., Kii,E. and Alt,F.W. (2015) Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat. Biotechnol.*, **33**, 179–86.
 18. Baltimore,D., Berg,P., Botchan,M., Carroll,D., Charo,R.A., Church,G., Corn,J.E., Daley,G.Q., Doudna,J. a, Fenner,M., *et al.* (2015) A prudent path forward for genomic engineering and germline gene modification. *Science*, **348**, 36.
 19. Cyranoski,D. and Reardon,S. (2015) Embryo Editing Sparks Epic Debate. *Nature*, **520**, 593–594.
 20. Liang,P., Xu,Y., Zhang,X., Ding,C., Huang,R., Zhang,Z., Lv,J., Xie,X., Chen,Y., Li,Y., *et al.* (2015) CRISPR/Cas9-mediated gene editing in human tripronuclear zygotes. *Protein Cell*, **6**, 363–372.
 21. Miller,J.C., Holmes,M.C., Wang,J., Guschin,D.Y., Lee,Y.-L., Rupniewski,I., Beausejour,C.M., Waite,A.J., Wang,N.S., Kim,K. a, *et al.* (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat. Biotechnol.*, **25**, 778–85.
 22. Guilinger,J.P., Pattanayak,V., Reyon,D., Tsai,S.Q., Sander,J.D., Joung,J.K. and Liu,D.R. (2014) Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat. Methods*, **11**, 429–35.
 23. Maruyama,T., Dougan,S.K., Truttmann,M.C., Bilate,A.M., Ingram,J.R. and Ploegh,H.L. (2015) Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.*, **33**, 538–42.
 24. Ran,F.A., Cong,L., Yan,W.X., Scott,D. a., Gootenberg,J.S., Kriz,A.J., Zetsche,B., Shalem,O., Wu,X., Makarova,K.S., *et al.* (2015) In vivo genome editing using *Staphylococcus aureus* Cas9. *Nature*, **520**, 186–190.
 25. Veres,A., Gosis,B.S., Ding,Q., Collins,R., Ragavendran,A., Brand,H., Erdin,S., Talkowski,M.E. and Musunuru,K. (2014) Low incidence of Off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. *Cell Stem Cell*, **15**, 27–30.
 26. Mali,P., Yang,L., Esvelt,K.M., Aach,J., Guell,M., DiCarlo,J.E., Norville,J.E. and Church,G.M. (2013) RNA-guided human genome engineering via Cas9. *Science*, **339**, 823–6.
 27. Fu,Y., Sander,J.D., Reyon,D., Cascio,V.M. and Joung,J.K. (2014) Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.*, **32**, 279–84.
 28. Kleinstiver,B.P., Pattanayak,V., Prew,M.S., Tsai,S.Q., Nguyen,N.T., Zheng,Z. and Keith Joung,J. (2016) High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, **529**, 490–495.
 29. Slaymaker,I.M., Gao,L., Zetsche,B., Scott,D.A., Yan,W.X. and Zhang,F. (2015) Rationally engineered Cas9 nucleases with improved specificity. *Science*, **351**, 84–88.
 30. Zetsche,B., Gootenberg,J.S., Abudayyeh,O.O., Slaymaker,I.M., Makarova,K.S., Essletzbichler,P., Volz,S.E., Joung,J., Van Der Oost,J., Regev,A., *et al.* (2015) Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, **163**, 759–771.
 31. Kleinstiver,B.P., Tsai,S.Q., Prew,M.S., Nguyen,N.T., Welch,M.M., Lopez,J.M., McCaw,Z.R., Aryee,M.J., Kleinstiver,B.P., Keith Joung,J., *et al.* (2016) Genome-wide specificities of CRISPR-Cas Cpf1

- nucleases in human cells. *Nat. Biotechnol.*, 10.1038/nbt.3620.
32. Cong,L., Ran,F.A., Cox,D., Lin,S., Barretto,R., Habib,N., Hsu,P.D., Wu,X., Jiang,W., Marraffini,L.A., *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–23.
 33. Mali,P., Esvelt,K.M. and Church,G.M. (2013) Cas9 as a versatile tool for engineering biology. *Nat Methods*, **10**, 957–963.
 34. Urnov,F.D., Miller,J.C., Lee,Y.L., Beausejour,C.M., Rock,J.M., Augustus,S., Jamieson, a C., Porteus,M.H., Gregory,P.D. and Holmes,M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646–651.
 35. Ramirez,C.L., Certo,M.T., Mussolino,C., Goodwin,M.J., Cradick,T.J., McCaffrey,A.P., Cathomen,T., Scharenberg,A.M. and Joung,J.K. (2012) Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res.*, **40**, 5560–5568.
 36. Miyaoka,Y., Berman,J.R., Cooper,S.B., Mayerl,S.J., Chan,A.H., Zhang,B., Karlin-Neumann,G.A. and Conklin,B.R. (2016) Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci. Rep.*, **6**, 23549.
 37. Hsu,P.D., Scott,D.A., Weinstein,J.A., Ran,F.A., Konermann,S., Agarwala,V., Li,Y., Fine,E.J., Wu,X., Shalem,O., *et al.* (2013) DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.*, **31**, 827–32.
 38. Cox,D.B.T., Platt,R.J. and Zhang,F. (2015) Therapeutic genome editing: prospects and challenges. *Nat. Med.*, **21**, 121–131.
 39. Ran,F.A., Hsu,P.D., Lin,C.Y., Gootenberg,J.S., Konermann,S., Trevino,A.E., Scott,D.A., Inoue,A., Matoba,S., Zhang,Y., *et al.* (2013) Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*, **154**, 1380–1389.
 40. Ran,F.A., Hsu,P.D., Wright,J., Agarwala,V., Scott,D.A. and Zhang,F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, **8**, 2281–2308.
 41. Yu,C., Liu,Y., Ma,T., Liu,K., Xu,S., Zhang,Y., Liu,H., La Russa,M., Xie,M., Ding,S., *et al.* (2015) Small molecules enhance crispr genome editing in pluripotent stem cells. *Cell Stem Cell*, **16**, 142–147.
 42. Richardson,C.D., Ray,G.J., DeWitt,M.A., Curie,G.L. and Corn,J.E. (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat. Biotechnol.*, **34**, 339–44.

Appendices

Appendix S1: Copyright permissions for Chapters 2-5

S1.2 Permission for Chapter 2



PNAS Permissions <PNASPermissions@nas.edu>

Mon 11/7, 11:14 AM

Jason Wolfs



Reply all | v



Action Items



Thank you for your message. I can address your query about the 2012 article. However, you'll receive a separate email from PNAS Production about your current manuscript under review (PNAS cannot grant permission for its use as it has not been accepted or published).

Regarding this article only:

Kleinstiver, B.P., Wolfs, J.M., Kolaczyk, T., Roberts, A.K., Hu, S.X. and Edgell, D.R. (2012) Monomeric site-specific nucleases for genome editing. *Proc. Natl. Acad. Sci. U. S. A.*, **109**, 8061–6

Authors do not need to obtain permission for the following uses of material they have published in PNAS: (1) to use their original figures or tables in their future works; (2) to make copies of their papers for their own personal use, including classroom use, or for the personal use of colleagues, provided those copies are not for sale and are not distributed in a systematic way; (3) to include their papers as part of their dissertations (this is the applicable situation); or (4) to use all or part of their articles in printed compilations of their own works.

Please cite the original PNAS article in full when re-using the article in your thesis. Because this material published after 2008, a copyright note is not needed. Feel free to contact us with any additional questions you might have.

Best regards,
Kay McLaughlin for
Diane Sullenberger
Executive Editor
PNAS

S1.3 Permission for Chapter 3



JOURNALS PERMISSIONS <Journals.Permissions@oup.com>

Thu 11/3, 10:04 AM

Jason Wolfs

Reply all |

Dear Jason Wolfs,

RE. Jason M. Wolfs et al. MegaTevs: Single-Chain Dual Nucleases For Efficient Gene Disruption. *Nucleic Acids Research* (2014) 42 (13): 8816-8829

Thank you for your recent email requesting permission to reuse all or part of your article in a thesis/dissertation.

As part of your copyright agreement with Oxford University Press you have retained the right, after publication, to use all or part of the article and abstract, in the preparation of derivative works, extension of the article into a booklength work, in a thesis/dissertation, or in another works collection, provided that a full acknowledgement is made to the original publication in the journal. As a result, you should not require direct permission from Oxford University Press to reuse your article.

Authors of *Oxford Open* articles are entitled to deposit their **original version or the version of record** in institutional and/or centrally organized repositories and can make this publicly available immediately upon publication, provided that the journal and OUP are attributed as the original place of publication and that correct citation details are given. Authors should also deposit the URL of their published article, in addition to the PDF version.

The journal strongly encourages *Oxford Open* authors to deposit the version of record instead of the original version. This will guarantee that the definitive version is readily available to those accessing your article from such repositories, and means that your article is more likely to be cited correctly.

For full details of our publication and rights policy please see the attached link to our website:
<http://www.oxfordjournals.org/en/access-purchase/rights-and-permissions/self-archiving-policy.html>

If you have any other queries, please feel free to contact us.

Kind regards,



| Permissions Assistant | Rights Department
Academic and Journals Divisions | Global Business Development

S1.4 Permission for Chapter 4 and 5

Permission to Use Copyrighted Material in a Doctoral Thesis

PP PNAS Permissions <PNASPermissions@nas.edu>
Tue 12/13, 4:50 PM
Jason Wolfs

Inbox

Action Items

Thank you for your message. Authors do not need to obtain permission for the following uses of material they have published in PNAS: (1) to use their original figures or tables in their future works; (2) to make copies of their papers for their own personal use, including classroom use, or for the personal use of colleagues, provided those copies are not for sale and are not distributed in a systematic way; (3) to include their papers as part of their dissertations; or (4) to use all or part of their articles in printed compilations of their own works.

Please cite the original PNAS article in full when re-using the material. Because this material published after 2008, a copyright note is not needed. Feel free to contact us with any additional questions you might have.

Best regards,

[Redacted Signature]

Executive Editor
PNAS

JW Jason Wolfs
Thu 12/13, 10:04 AM

Dear PNAS,

I am a University of Western Ontario graduate student completing my Doctoral thesis. My thesis will be available in full-text on the internet for reference, study and / or copy. Except in situations where a thesis is under embargo or restriction, the electronic version will be accessible through the Western Libraries web pages, the Library's web catalogue, and also through web search engines. I will also be granting Library and Archives Canada and ProQuest/UMI a non-exclusive license to reproduce, loan, distribute, or sell single copies of my thesis by any means and in any form or format. These rights will in no way restrict republication of the material in any other form by you or by others authorized by you.

I would like permission to allow inclusion of the following material in my thesis: Wolf, J.M., Hamilton, T.A., Laut, J.T., Laforet, M., Zhang, J., Salemi, L., Gloor, G.B., Schib-Poulsen, C. and Edgell, D.R. (2016) Biasing genome-editing events towards precise length deletions with RNA-guided Cas9 dual nuclease. *Proceedings of the National Academy of Sciences USA* (recently accepted)

The material will be attributed through a citation.

Appendix S2: Supplemental information for Chapter 2

S2 Supplemental material and methods

Bacterial strains and plasmid construction. *Escherichia coli* strains DH5 α and ER2566 (New England Biolabs) were used for plasmid manipulations and protein expression, respectively. *E.coli* strain BW25141(λ DE3) was used for genetic selection assays [1]. A complete description of all plasmids used in this study are listed in Table S3, and oligonucleotides are listed in Table S4. The ryA and ryB zinc-finger genes were synthesized by Integrated DNA Technologies with 5'-BamHI and 3'-XhoI sites and a C-terminal 6-histidine tag, and cloned into pACYCDuet-1 to generate pACYCryAZf+H and pACYCryBZF+H, respectively. A stop codon was introduced at the 3' end of the ryAZf gene using Quikchange (Stratagene) to generate pACYCryAZf. To create GIY-ZFEs, the I-TevI and I-BmoI GIY-YIG domains were PCR amplified from bacteriophage T4 gDNA and pACYCIBmoI, respectively, and cloned into the NcoI/BamHI sites of pACYCryAZf+H, pACYCryAZf, and pACYCryBZF+H. TevN201-ryA and TevN201R27A were subcloned into the XbaI and EcoRV sites of pTAL3 to generate the expression plasmids for the yeast reporter assay (pYTZN201 and pYTZN201R27A). To generate Tev-LHEs, the I-OnuI E1 gene was amplified with BamHI and Sall ends to clone into the BamHI and XhoI sites of pACYCDuet-1(Pcil) to create pACYCOnuE1(+H). This vector was subsequently Quikchanged to introduce an E22Q mutation in I-OnuI E1 to create pACYCOnuE1E22Q (+H). I-TevI catalytic domains were amplified as above and cloned into Pcil/BamHI of pACYCOnuE1E22Q (+H). The R27A mutants of Tev-ZFEs and Tev-LHEs were generated using Quikchange mutagenesis. Hybrid GIY-ZFE and Tev-LHE target sites (Fig. 1B and 4B, Table S1 and S2) were cloned into the toxic plasmid p11-lacY-wtx1 to generate reporter plasmids for the bacterial selection. Tev-ryA and Bmo-ryA target sites were cloned into pSP72 for *in vitro* cleavage assays. The Tev-ryA site hybrid homing site was also cloned into LITMUS28i using BamHI and XhoI to generate pTZHS1.35. The two-site Tev-ZF plasmids were created by sub-cloning the PvuII/HpaI fragment from pTZHS1.35 into the Swal site of pTZHS1.35 to generate pTZHS2.35 and pTZHS3.35 (with the second TZHS in either orientation). The G5A or C1A/G5A mutations were introduced into pToxTZ and pTZHS plasmids by Quikchange mutagenesis. To generate the target plasmids for the yeast reporter assay, the TZ-ryA target sites from toxic plasmids containing TZ1.33, TZ1.33G5A, and TZ1.33C1A/G5A were amplified and cloned into the BglII and SpeI sites of pCS753. All constructs were verified by sequencing, and the amino acid sequences of all GIY-ZFEs and Tev-LHEs constructed are provided in Figure S5.

Two-plasmid genetic selection. The two plasmid genetic selection was performed as described [1], with toxic (reporter) plasmids containing hybrid TZ-ryA, TZ-ryB, BZ-ryA, TO target sites (Table S1 and S2),

mutant target sites (with G5A or C1A/G5A substitutions), or plasmids lacking a target site (p11-lacY-wtx1). Survival percentage was calculated by dividing the number of colonies observed on selective by those observed on non-selective plates.

Yeast reporter assay. Transformants of *S.cerevisiae* YPH500(α) with Tev-ZFE constructs and YPH499(a) with target constructs were grown overnight (~230 rpm) at 30°C in synthetic complete medium lacking histidine (Tev-ZFEs) or lacking tryptophan and uracil (targets). Tev-ZFEs and targets were mated by adding equal densities (~400 μ l) of overnight culture to 1 ml YPD and left stationary for 5-6 hours at 30°C. Cells harvested by centrifugation were washed in 1 ml and resuspended in 4 ml of synthetic medium lacking histidine and tryptophan prior to shaking overnight at 30°C. Cells were harvested by centrifugation, washed in 1 ml Z buffer (60 mM Na₂PO₄, 40 mM NaH₂PO₄, 10 mM KCl, 1 mM MgSO₄, pH 7.0), and suspended in 250 μ l Z buffer. The suspension was diluted 20-fold into 1 ml Z buffer containing 0.27% β -mercaptoethanol, and 75 μ l CHCl₃ and 45 μ l 0.1% SDS were added prior to vortexing. Lysates were pre-incubated at 30°C prior to the addition of 100 μ l 4 mg/ml ONPG. Reactions proceeded until a yellow colour developed whereby progress was stopped by the addition of 300 μ l 1M Na₂CO₃. Stopped reactions were pelleted and the absorbance of the supernatant was analyzed at 420 nm and 550 nm.

Protein purification. Cultures overexpressing either TevN201-ZFE or BmoN221-ZFE were grown at 37°C to an OD₆₀₀~0.5 and expression induced by 0.5 mM IPTG (Bio Basic Inc.) overnight at 15°C. Cells were harvested by centrifugation at 8983 x *g* for 12 minutes, re-suspended in binding buffer (20 mM Tris-HCl pH 8.0, 500 mM NaCl, 10 mM imidazole, 5% glycerol, and 1 mM DDT), and lysed by homogenization at 15,000 psi. The cell lysate was clarified by centrifugation at 20400 x *g*, followed by sonication for 30 seconds, and centrifugation at 20400 x *g* for 15 minutes. The clarified lysate was loaded onto a 1 mL HisTrap-HP column (GE Healthcare), washed with 15 mL binding buffer and then 10 mL wash buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 50 mM imidazole, 5% glycerol, and 1 mM DDT). Bound proteins were eluted in 1.5 mL fractions in four 5 mL step elutions with increasing concentrations of imidazole. Fractions containing GIY-ZFEs were dialyzed twice against 1L dialysis buffer (20 mM Tris-HCl (pH 8.0), 500 mM NaCl, 5% glycerol, and 1 mM DDT) prior to storage at -80°C.

Cleavage assays. Single time-point cleavage assays to determine the EC_{0.5max} of TevN201-ryA were performed in buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl, 10 mM MgCl₂, 5% glycerol, 1 mM DTT and 10 nM pTZHS1.33. Reactions were incubated for 3 minutes at 37°C, stopped with 5 μ l stop solution (100 mM EDTA, 40% glycerol, and bromophenol blue), and electrophoresed on a 1% agarose gel

prior to staining with ethidium bromide and analysis on an Alphascreen™3400 (Alpha Innotech). The $EC_{0.5max}$ was determined by fitting the data to the equation

$$f_{([endo])} = \frac{f_{max} * [endo]^H}{EC_{0.5max} + [endo]^H}$$

where $f_{([endo])}$ is the fraction of substrate cleaved at concentration of TevN201-ryA [endo], f_{max} is the maximal fraction cleavage, with 1 being the highest value, and H is the Hill constant that was set to 1. The initial reaction velocity was determined using supercoiled plasmid substrate with varying concentrations of TevN201-ryA (0.7 nM to 47 nM) and buffer as above. Aliquots were removed at various times, stopped and analyzed as above. The data for product appearance was fitted to the equation

$$P = A(1 - e^{-k_1 t}) + k_2 t$$

where P is product (in nM), A is the magnitude of the initial burst, k_1 is the rate constant (s^{-1}) of the initial burst phase and k_2 is the steady state rate constant (s^{-1}). The two-site plasmid cleavage assays were conducted as above, using 10 nM pTZHS2.33 or pTZHS3.33 as substrates, and ~90 nM purified TevN201-ryA. The k_{obs} rate constants were calculated from the decay of supercoiled substrate by fitting to the equation

$$[C] = [C_0] \exp(-k_1 t)$$

where $[C]$ is the concentration (nM) of supercoiled plasmid at time t, $[C_0]$ is the initial concentration of supercoiled substrate (nM), and k_1 is the first order rate constant (in s^{-1}). At least 3 independent trials were conducted for each data set.

Cleavage mapping. Mapping of cleavage sites was performed as described [2]. Briefly, primers were individually end-labeled with γ - ^{32}P ATP, and used in PCR reactions with pTox or pSP72 plasmids carrying TZ-ryA or BZ-ryA target sites to generate strand-specific substrates. The substrates were incubated with purified protein as above, and electrophoresed in 8% denaturing gels alongside sequencing ladders generated by cycle sequencing with the same end-labeled primers (US Biologicals).

Bioinformatics. The distribution of the CNNNG cleavage motif was examined in a 35-bp window of ZFN sites predicted for *Dania rerio* chromosome 1 (Ensembl release 51) using a custom Perl script [6]. Briefly, 40 bp flanking the downstream region of each predicted ZFN was extracted from the corresponding zebrafish chromosome 1 cDNA, and searched for position $i = C$ and position $i+4 = G$, with the occurrences of each CNNNG reported at position i (the C of the motif). The number of CNNNG motifs for unique ZFN

sites was fit to a binomial distribution, and plotted in R.

S2 Supplemental figures

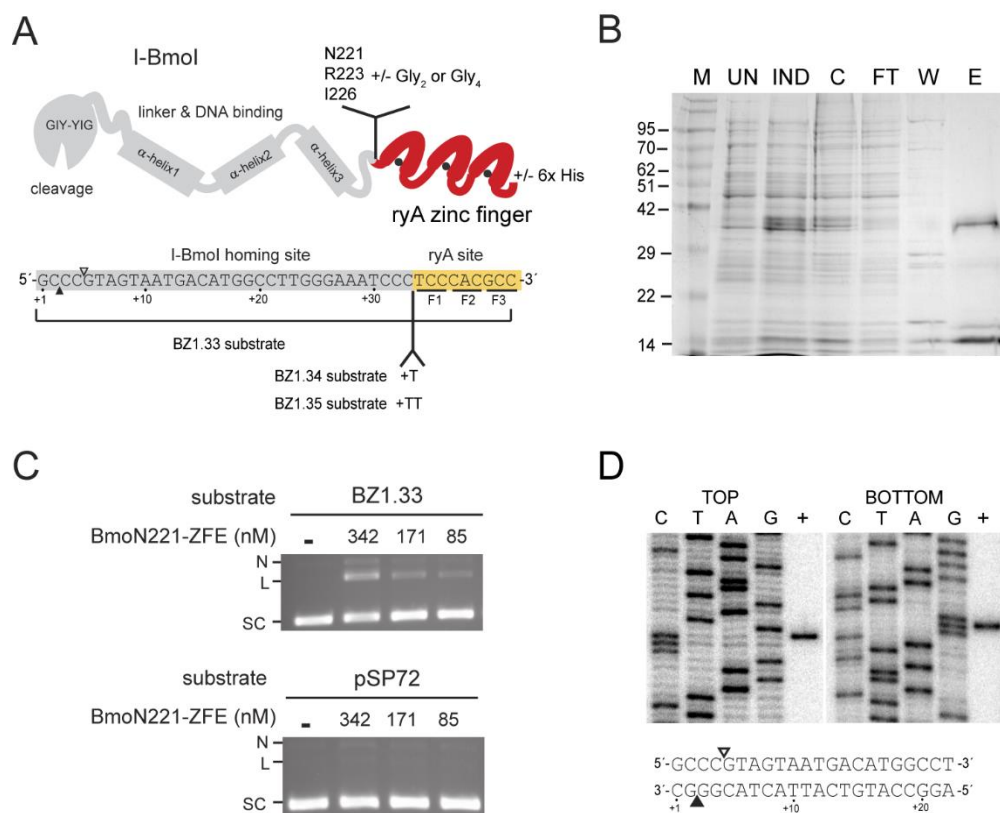


Figure S5.1: Design and functionality of Bmo-ZFEs.

(A) Schematic of Bmo-ZFE constructs, with I-BmoI protein and substrate shown in grey, and the ryA protein and binding site shown in red and yellow, respectively. *Top panel*, the fusion points for each of the Bmo-ZFEs are indicated as the last I-BmoI amino acid, with or without a 2xGlycine or 4xGlycine linker. Constructs were made with a 6xHis tag on the C-terminal end. *Bottom panel*, the substrate shown consists of 33-nts of the top strand I-BmoI *thyA* target site (BZ1.33), fused to the 5' end of the ryA binding site. Substrates tested differ by the insertion of one or two T nucleotides at the junction of the *thyA*/ryA sites. (B) Purification of His-tagged BmoN221-ryA. Shown is a representative SDS-PAGE gel; M, marker with molecular weights in kDa indicated on the left; UN, uninduced culture; IND, induced culture; C, crude lysate; FT, flow-through from metal-affinity column; W, wash; E, elution. (C) BmoN221-ryA cleavage specificity. Shown are representative agarose gels of cleavage assays with 10 nM pBZ1.33 or pSP72 (no target site) substrates and the indicated concentrations of BmoN221-ryA under standard assay conditions for 10 minutes. BmoN221-ryA cleaves the BZ-ryA target site plasmid (BZ1.33) but not the control plasmid (pSP72) lacking the target site. N, nicked; L, linear; SC, supercoiled. (D) Mapping of BmoN221-ryA cleavage sites on the BZ1.33 substrate, with top and bottom cleavage sites indicated by open and closed triangles, respectively.

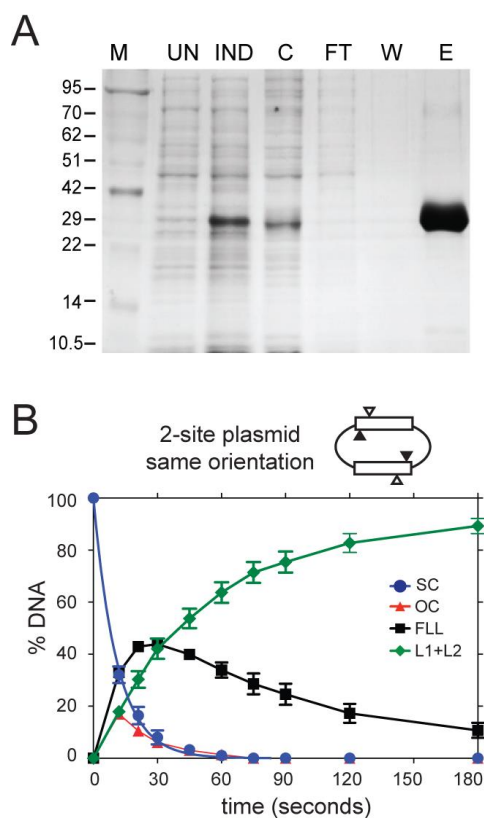


Figure S2.2: TevN201-ryA purification for *in vitro* experiments.

(A) Purification of TevN201-ryA. Shown is a representative SDS-PAGE gel; M, marker with molecular weights in kDa indicated on the left; UN, uninduced culture; IND, induced culture; C, crude lysate; FT, flow-through from metal-affinity column; W, wash; E, elution. (B) Graphical representation of cleavage assays with 90 nM TevN201-ZFE and 10 nM two-site pTZ1.33 plasmid with target sites in the same orientation. Data are plotted as averages of three independent replicates with standard deviations; SC, supercoiled; OC, open-circle (nicked); FLL, full-length linear; L1+L2, linear products.

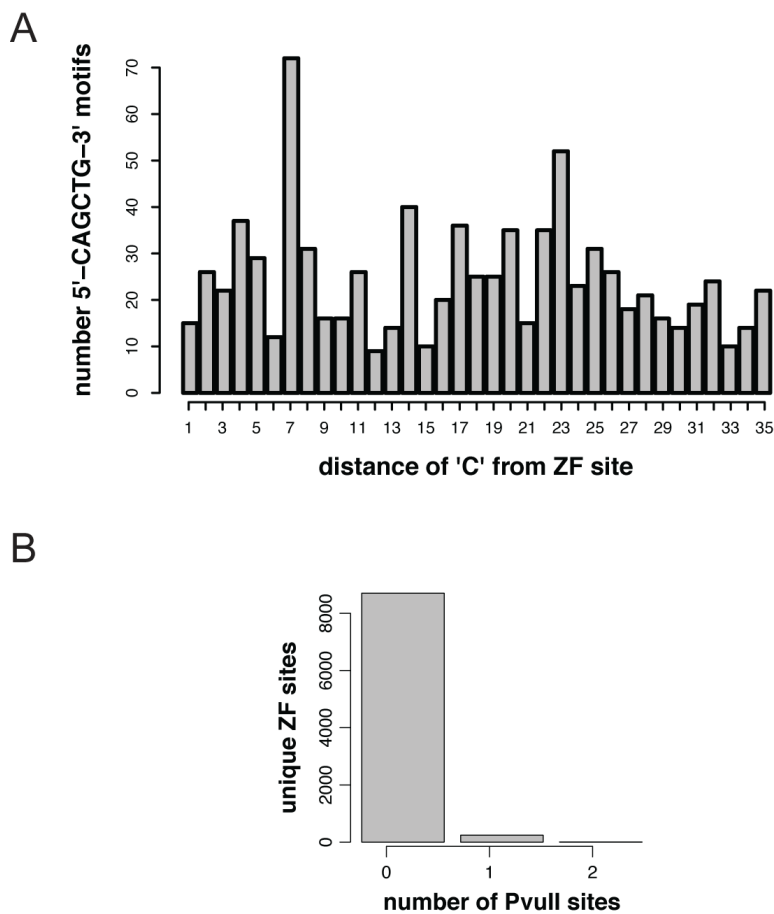


Figure S2.3: PvuII site analysis.

(A) Shown is the distribution of the 5'-CAGCTG-3' motif in a 35-bp window flanking 8,829 predicted ZFN sites on zebrafish chromosome 1. The number of occurrences of the 'C' of the motif at each distance is indicated. (C) Unique ZFN sites were grouped according to the number of occurrences of the 5'-CAGCTG-3' motif in the 35-bp window.

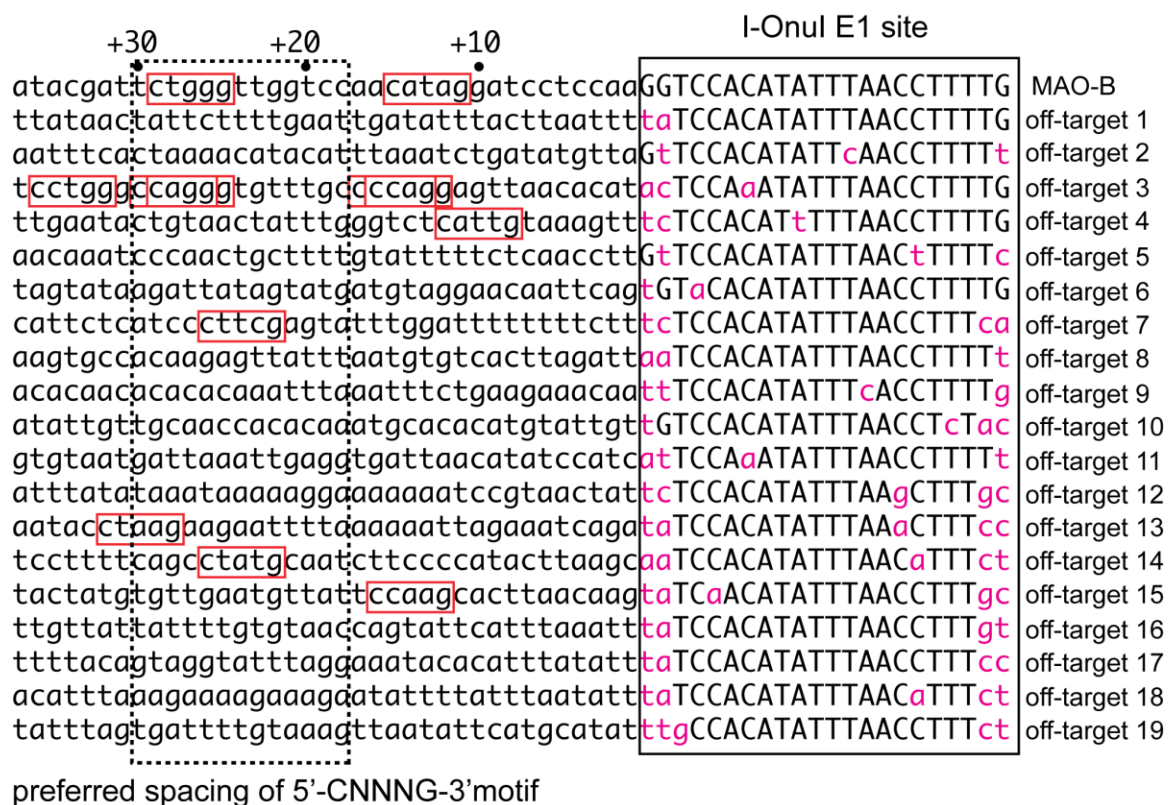


Figure S2.4: Occurrence of the 5'-CNNNG-3' motif upstream of I-OnuI E1 off-target sites.

Shown is 37-nt of upstream sequence adjacent to the 22-nt I-OnuI E1 MAO-B target site, along with 19 predicted off-target sites [3]. CNNNG motifs are highlighted in red, with only 3 of 19 predicted I-OnuI E1 off target sites containing a CNNNG motif at a targetable distance by Tev-LHE fusions. Nucleotide differences of the off-target sites to the I-OnuI E1 site are indicated in magenta lower case font.

S2 Supplemental tables

Table S2.1: Tev-ZFE selection data.

GIY-ZFE	Toxic plasmid								
	pToxTZ1.30		pToxTZ1.32		pToxTZ1.33			pToxTZ1.34	
	WT	WT	WT	G5A	C1A/G5A	WT	G5A	C1A/G5A	
TevN201	0	4 ± 1.2 (3)	86.8 ± 5.9 (6)	0	0	59.9 ± 9.5 (6)	0	0.2 ± 0.1 (3)	
TevN201G ₂	ND	ND	72.7 ± 10.7 (6)	0	0	56.9 ± 11.2 (6)	0	0	
TevN201G ₄	ND	ND	83.7 ± 15.2 (4)	0	0	42.8 ± 12.6 (6)	0	0	
TevN201R27A	ND	ND	0	0	0	0	0	0	
TevK203	0	5.4 ± 1.4 (3)	86.8 ± 7.1 (6)	0	0	50.7 ± 9.5 (6)	0	0	
TevK203G ₂	ND	ND	88 ± 13.9 (6)	0	0	53.7 ± 10.4 (6)	0	0	
TevK203G ₄	ND	ND	80.7 ± 7.9 (4)	0.2 ± 0.2 (3)	0.4 ± 0.3 (3)	43.6 ± 13.0 (6)	0	0	
TevK203R27A	ND	ND	0	0	0	0	0	0	
TevS206	0	0	86.6 ± 6.9 (6)	0	0	47.1 ± 8.6 (6)	0	0	
TevS206G ₂	ND	ND	70.7 ± 8.7 (4)	0	0	27.8 ± 7.4 (6)	0	0	
TevS206R27A	ND	ND	0	0	0	0	0	0	
TevN201-ryB	ND	ND	0	ND	ND	ND	ND	ND	

GIY-ZFE	pToxTZ1.35			pToxTZ1.36	pToxTZ1.38	p11lac ₁ wtx1	<u>pTox ryB</u>
	WT	G5A	C1A/G5A	WT	WT		WT
TevN201	49.8 ± 9.8 (6)	0	0	0	0	0	0
TevN201G ₂	38.6 ± 10.3 (4)	0	0	ND	ND	0	ND
TevN201G ₄	36.3 ± 7.1 (4)	0	0	ND	ND	0	ND
TevN201R27A	0	0	0	ND	ND	0	ND
TevK203	51.0 ± 6.6 (5)	0	0	1.6 ± 0.3 (3)	0	0	ND
TevK203G ₂	46.5 ± 10.9 (5)	0	0	ND	ND	0	ND
TevK203G ₄	48.0 ± 6.1 (4)	0	0	ND	ND	0	ND
TevK203R27A	0	0	0	ND	ND	0	ND
TevS206	62.3 ± 12.4 (4)	0	0	26.1 ± 1.8 (3)	0	0	ND
TevS206G ₂	44.2 ± 16.4 (4)	0	0	ND	ND	0	ND
TevS206R27A	0	0	0	ND	ND	0	ND

*ND, not determined; ±, standard deviation of replicates indicated in brackets

Table S2.2: Tev-LHE selection data.

Tev-LHE	Toxic plasmid							
	pToxTO1.12	pToxTO1.14	pToxTO1.16	pToxTO1.18		pToxTO1.20		pToxTO1.22
	WT	WT	WT	WT	C-1A/G-5A	WT	C-1A/G-5A	WT
TevS114	0.2 (2)	0	0	16.5 ± 6.2 (3)	0	11.6 ± 3.7 (3)	0	0
TevD127	0	0	0.1 (2)	31.8 ± 11.3 (3)	0	12.6 ± 3.2 (3)	0	0
TevN140	0	0	0	40.2 ± 3.8 (3)	0	24.6 ± 0.4 (3)		0
TevN169	0	0	0	45.5 ± 11.5 (3)	0	94.5 ± 2.1 (3)	0	0
TevD184G ₂	0	0.1 (2)	0	25 ± 6.9 (3)	0	98.3 ± 7.1 (3)	0	11.9 ± 1.8 (3)
TevN201G ₄	0	0	0.1 (2)	29.5 ± 1.2 (3)	0	8.3 ± 0.4 (3)	0	0
TevN203	0.8 (2)	0	0.1 (2)	24.2 ± 2.0 (3)	0	7.7 ± 3.1 (3)	0	0
Tev-LHE	pToxTO1.24	pToxTO1.26		pToxTO1.28	pToxTO1.30			pToxTZ1.33
Tev-LHE	WT	WT	C-1A/G-5A	WT	WT	G-5A	C-1A/G-5A	WT
TevS114	0	65.2 ± 9 (3)	0	0	0	ND	ND	ND
TevD127	0	83.2 ± 7.9 (3)	0	0	0	ND	ND	ND
TevN140	0	74.9 ± 4.2 (3)	0	0	0	ND	ND	ND
TevN169	0	102.2 ± 10.4 (3)	0	0.4 ± 0.3 (3)	0	ND	ND	ND
TevD184G ₂	0	87.0 ± 4.7 (3)	0	0.0	0	ND	ND	ND
TevN201G ₄	0	76.0 ± 12.3 (3)	0	87.1 ± 9.5 (3)	96.9 ± 2.5 (3)	0	0	0
TevN203	0	97.5 ± 5.3 (3)	0	99.7 ± 12 (3)	87.2 ± 0.8 (3)	ND	ND	ND
TevR27A(N201G ₄)- OnuE1 (E22Q)	ND	ND	ND	ND	0	0	ND	ND
TevR27A(N201G ₄)- OnuE1	ND	ND	ND	ND	101.6 ± 4.1 (3)	93.3 ± 4.6 (3)	ND	ND

*ND, not determined; ±, standard deviation of replicates indicated in brackets.

Supplemental Table S2.3: Strains	Strains and plasmids used in this study. Description	Source
<i>E.coli</i> - DH5 α	F ⁻ , ϕ 80 <i>dlacZ</i> Δ M15, Δ (<i>lacZYA-argF</i>)U169, <i>deoR</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> (<i>rk</i> ⁻ , <i>mk</i> ⁺), <i>phoA</i> , <i>supE44</i> , λ ⁻ , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i>	Invitrogen
<i>E.coli</i> - ER2566	F ⁻ λ - <i>fhuA2</i> [<i>lon</i>] <i>ompT lacZ::T7 gene 1 gal sulA11</i> Δ (<i>mcrC-mrr</i>)114:: <i>IS10 R(mcr-73::miniTn10-TetS)2 R(zgb-210::Tn10)(TetS)</i> <i>endA1</i> [<i>dcm</i>]	N.E.B.
<i>E.coli</i> - BW25141(λ DE3)	F ⁻ <i>lacI</i> ^q <i>rrnB</i> _{T14} <i>DlacZ</i> _{WJ16} <i>DphoBR580 hsdR514 DaraBAD</i> _{AH33} <i>DrhaBAD</i> _{LD78} <i>galU95 endA</i> _{BT333} <i>uidA(DMluI)::pir+</i> <i>recA1</i> , λ DE3 lysogen	Ref [1],[4]
<i>S.cerevisiae</i> - YPH499	MATa <i>ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1</i>	Dr. Adam Bogdanove
<i>S.cerevisiae</i> - YPH500	MAT α <i>ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1</i>	Dr. Adam Bogdanove

Plasmids	Description	Source
pACYCDuet-1	<i>ori</i> _{p15A} , <i>cm</i>	Novagen
pACYCDuet-1(PciI)	<i>ori</i> _{p15A} , <i>cm</i> , pACYCDuet-1 with a PciI site substituted for the NcoI site	Novagen
p11-lacY-wtx1	<i>ori</i> _{pBR322} , <i>amp</i>	Ref [4]
pSP72	<i>ori</i> _{pBR322} , <i>amp</i>	Promega
LITMUS28i	<i>ori</i> _{pMB1} , <i>amp</i>	N.E.B.
pACYCIBmoI	pACYCDuet-1, containing the 798bp codon optimized I-BmoI gene in the NdeI and XhoI sites	Ref [1]
pryAzf	<i>ori</i> _{pUC} , <i>kan</i>	I.D.T.
pTAL3	<i>ori</i> _{pBR322} , <i>amp</i>	Dr. Adam Bogdanove
pCP5.1	pCP5 derivative, <i>amp</i> , with ColE1 origin from pBluescript IKS(-), reporter plasmid for the yeast based recombination assay, see [5]	Dr. Adam Bogdanove
pOnuE1	<i>ori</i> _{pBR322} , <i>amp</i>	Ref [3]
pACYCryAZf+H	pACYCDuet-1, containing the ryA zinc-finger gene with a c-terminal 6-histidine tag cloned into the BamHI and XhoI sites	This study
pACYCryAZf	pACYCDuet-1, containing the ryA zinc-finger gene cloned into the BamHI and XhoI sites	This study

pACYCryBzf+H	pACYCDuet-1, containing the ryB zinc-finger gene with a c-terminal 6-histidine tag cloned into the BamHI and XhoI sites	This study
pTevN201-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N201 of I-TevI (DE832/840) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevN201-ryB (or +H)	pACYCryBzf (or +H), with residues 1-N201 of I-TevI (DE832/840) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevN201G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N201 of I-TevI (DE833/840) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevN201G ₄ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N201 of I-TevI (DE834/840) + 4 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevK203-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-K203 of I-TevI (DE835/840) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevK203G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-K203 of I-TevI (DE836/840) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevK203G ₄ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-K203 of I-TevI (DE837/840) + 4 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevS206-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-S206 of I-TevI (DE838/840) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevS206G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-S206 of I-TevI (DE839/840) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoN221-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N221 of I-BmoI (DE841/849) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoN221G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N221 of I-BmoI (DE842/849) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoN221G ₄ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-N221 of I-BmoI (DE843/849) + 4 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoR223-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-R223 of I-BmoI (DE844/849) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study

pBmoR223G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-R223 of I-BmoI (DE845/849) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoR223G ₄ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-R223 of I-BmoI (DE846/849) + 4 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoI226-ZFE (or +H)	pACYCryAZf (or +H), with residues 1-I226 of I-BmoI (DE847/849) cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pBmoI226G ₂ -ZFE (or +H)	pACYCryAZf (or +H), with residues 1-I226 of I-BmoI (DE848/849) + 2 glycine residues cloned into the NcoI and BamHI sites (+/- 6xHis)	This study
pTevN201R27A	Similar to pTevN201-ZFE, with an R27A mutation	This study
pTevN201G ₂ R27A	Similar to pTevN201G ₂ -ZFE, with an R27A mutation	This study
pTevN201G ₄ R27A	Similar to pTevN201G ₄ -ZFE, with an R27A mutation	This study
pTevK203R27A	Similar to pTevK203-ZFE, with an R27A mutation	This study
pTevK203G ₂ R27A	Similar to pTevK203G ₂ -ZFE, with an R27A mutation	This study
pTevK203G ₄ R27A	Similar to pTevK203G ₄ -ZFE, with an R27A mutation	This study
pTevS206R27A	Similar to pTevS206-ZFE, with an R27A mutation	This study
pTevS206G ₂ R27A	Similar to pTevS206G ₂ -ZFE, with an R27A mutation	This study
pToxTZ1.35	p11-lacY-wtx1, that contains a 44-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +8 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites (DE824/825)	This study
pToxBZ1.35	p11-lacY-wtx1, that contains a 44-bp hybrid I-BmoI/ryA zinc-finger homing site (<i>thyA</i> bases -6 to +27 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites (DE826/827)	This study
pSP-TZHS1.35	pSP72, that contains a 44-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +8 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites (DE824/825)	This study
pTZHS1.35	LITMUS28i, with the 44-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +8 fused to the 9-bp ryAZf site) sub-cloned from pSP-TZHS1.35 into the BamHI and XhoI sites	This study
pBZHS1.35	pSP72, that contains a 44-bp hybrid I-BmoI/ryA zinc-finger homing site (<i>thyA</i> bases +6 to -27 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites (DE826/827)	This study
pTZHS2.35	Similar to pTZHS1.35, with a second Tev-ZFE1.35 target site sub-cloned from pSP-TZHS1.35 (using PvuII/HpaI) into the SmaI site	This study
pTZHS3.35	Similar to pTZHS2.35, with the second Tev-ZFE1.35 target site in the alternate orientation	This study

pToxTZ1.35G5A	Similar to pToxTZ1.35, with a G5A substitution (DE917/918)	This study
pToxTZ1.35C1A/G5A	Similar to pToxTZ1.35, with C1A and G5A substitutions (DE919/920)	This study
pTZHS1.35G5A	Similar to pTZHS1.35, with a G5A substitution	This study
pTZHS1.35C1A/G5A	Similar to pTZHS1.35, with C1A and G5A substitutions	This study
pToxTZ1.34	p11-lacY-wtx1, that contains a 43-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +7 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites	This study
pToxTZ1.34G5A	Similar to pToxTZ1.34, with G5A substitution	This study
pToxTZ1.34C1A/G5A	Similar to pToxTZ1.34, with C1A and G5A substitutions	This study
pToxTZ1.33	p11-lacY-wtx1, that contains a 42-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites	This study
pToxTZ1.33G5A	Similar to pToxTZ1.33, with G5A substitution	This study
pToxTZ1.33C1A/G5A	Similar to pToxTZ1.33, with C1A and G5A substitutions	This study
pToxTZ1.33-ryB	p11-lacY-wtx1, that contains a 42-bp hybrid I-TevI/ryB zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryBZf site) cloned into the XbaI and SphI sites	This study
pToxBZ1.34	p11-lacY-wtx1, that contains a 43-bp hybrid I-BmoI/ryA zinc-finger homing site (<i>thyA</i> bases -6 to +26 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites (DE826/827)	This study
pToxBZ1.33	p11-lacY-wtx1, that contains a 42-bp hybrid I-BmoI/ryA zinc-finger homing site (<i>thyA</i> bases -6 to +25 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites (DE826/827)	This study
pTZHS1.34	Similar to pTZHS1.35, with a 43-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +7 fused to the 9-bp ryAZf site)	This study
pTZHS1.33	Similar to pTZHS1.35, with a 42-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryAZf site)	This study
pBZHS1.34	Similar to pBZHS1.35, with a 43-bp hybrid I-BmoI/ryA zinc-finger homing site (<i>thyA</i> bases +6 to -26 fused to the 9-bp ryAZf site)	This study
pBZHS1.33	Similar to pBZHS1.35, with a 42-bp hybrid I-BmoI/ryA zinc-finger homing site (<i>thyA</i> bases +6 to -25 fused to the 9-bp ryAZf site)	This study
pTZHS2.34	Similar to pTZHS2.35, with both Tev-ZFE target sites as 43-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +7 fused to the 9-bp ryAZf site)	This study

pTZHS3.34	Similar to pTZHS3.35, with both Tev-ZFE target sites as 43-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +7 fused to the 9-bp ryAZf site)	This study
pTZHS2.33	Similar to pTZHS2.35, with both Tev-ZFE target sites as 42-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryAZf site)	This study
pTZHS3.33	Similar to pTZHS3.35, with both Tev-ZFE target sites as 42-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryAZf site)	This study
pTZHS1.34G5A	Similar to pTZHS1.34, with a G5A substitution	This study
pTZHS1.33G5A	Similar to pTZHS1.33, with a G5A substitution	This study
pTZHS1.34C1A/G5A	Similar to pTZHS1.34, with C1A and G5A substitutions	This study
pTZHS1.33C1A/G5A	Similar to pTZHS1.33, with C1A and G5A substitutions	This study
pToxTZ1.30	p11-lacY-wtx1, that contains a 39-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +3 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites (DE1085/1086)	This study
pToxTZ1.32	p11-lacY-wtx1, that contains a 41-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +5 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites (DE1087/1088)	This study
pToxTZ1.36	p11-lacY-wtx1, that contains a 45-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +9 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites (DE1134/1135)	This study
pToxTZ1.38	p11-lacY-wtx1, that contains a 47-bp hybrid I-TevI/ryA zinc-finger homing site (<i>td</i> bases -27 to +11 fused to the 9-bp ryAZf site) cloned into the XbaI and SphI sites (DE1136/1137)	This study
pToxTZ1.33ryB	p11-lacY-wtx1, that contains a 42-bp hybrid I-TevI/ryB zinc-finger homing site (<i>td</i> bases -27 to +6 fused to the 9-bp ryBZf site) cloned into the XbaI and SphI sites	This study
pToxTO1.12	p11-lacY-wtx1, that contains a 34-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to -16 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites (DE1072/1073)	This study
pToxTO1.14	p11-lacY-wtx1, that contains a 36-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to -14 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites (DE1070/1071)	This study
pToxTO1.16	p11-lacY-wtx1, that contains a 38-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to -12 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites (DE1068/1069)	This study

pToxTO1.18	p11-lacY-wtx1, that contains a 40-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to -10 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites(DE1066/1067)	This study
pToxTO1.20	p11-lacY-wtx1, that contains a 42-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to -8 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites(DE1064/1065)	This study
pToxTO1.22	p11-lacY-wtx1, that contains a 44-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to -6 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites(DE1062/1063)	This study
pToxTO1.24	p11-lacY-wtx1, that contains a 46-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to -4 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites(DE1060/1061)	This study
pToxTO1.26	p11-lacY-wtx1, that contains a 48-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to -2 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites(DE1058/1059)	This study
pToxTO1.28	p11-lacY-wtx1, that contains a 50-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to +1 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites(DE1056/1057)	This study
pToxTO1.30	p11-lacY-wtx1, that contains a 52-bp hybrid I-TevI/I-OnuI E1 homing site (<i>td</i> bases -27 to +3 fused to the I-OnuI E1 site) cloned into the XbaI and SphI sites(DE976/977)	This study
pToxTO1.18C1A/G5A	Similar to pToxTO18, with C1A and G5A substitution(DE1154/1155)	This study
pToxTO1.20 C1A/G5A	Similar to pToxTO20, with C1A and G5A substitution(DE1156/1157)	This study
pToxTO1.26 C1A/G5A	Similar to pToxTO26, with C1A and G5A substitution(DE1158/1159)	This study
pToxTO1.30 G5A	Similar to pToxTO30, with a G5A substitution	This study
pToxTO1.30 C1A/G5A	Similar to pToxTO30, with C1A and G5A substitution	This study
pACYCOnuE1(+H)	pACYCDuet-1(PciI), containing the I-OnuI E1 gene cloned into the BamHI and XhoI sites	This study
pACYCOnuE1(E22Q)(+H)	pACYCDuet-1(PciI), containing the I-OnuI E1 gene with a E22Q mutation cloned into the BamHI and XhoI sites	This study
pTevK203-OnuE1(E22Q)	pACYCOnuE1(E22Q)(+H), with residues 1-K203 of I-TevI (DE) cloned into the PciI and BamHI sites	This study
pTevN201G ₄ -OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), with residues 1-N201 of I-TevI + 4 glycine residues (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevD184G ₂ -OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), with residues 1-D184G ₂ of I-TevI + 2 glycine residues (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevN169-OnuE1(E22Q)(+H)	pACYCOnuE1(E22Q)(+H), with residues 1-N169 of I-TevI (DE) cloned into the PciI and BamHI sites (+6xHis)	This study

pTevN140-OnuE1(E22Q) (+H)	pACYCOnuE1(E22Q)(+H), with residues 1-N140 of I-TevI (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevD127-OnuE1(E22Q) (+H)	pACYCOnuE1(E22Q)(+H), with residues 1- D127 of I-TevI (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevS114-OnuE1(E22Q) (+H)	pACYCOnuE1(E22Q)(+H), with residues 1- S114 of I-TevI (DE) cloned into the PciI and BamHI sites (+6xHis)	This study
pTevN201G4R27A - OnuE1(+H)	pACYCOnuE1(+H), with residues 1-N201 of I-TevI + 4 glycine residues (DE) with a R27A mutation in I-TevI (+6xHis)	This study
pTevN201G4R27A-OnuE1(E22Q) (+H)	pACYCOnuE1(E22Q)(+H), similar to pTevN201G4-OnuE1(E22Q) (+H) with an R27A mutation in I-TevI	This study
pZif268	pTAL3 expression vector containing homodimeric FokI-Zif268 ZFN	Dr. Adam Bogdanove
pZif268target	pCS753 reporter plasmid with a homodimeric FokI-Zif268 target site interrupting a partially duplicated <i>lacZ</i> gene	Dr. Adam Bogdanove
pYTZN201	pTAL3, with TevN201-ryA cloned in XbaI/EcoRV using DE1121/1128	This study
pYTZN201R27A	Similar to pYTZN201, with an R27A I-TevI mutation	This study
pTZYHS1.33	pCS753, with the 42-bp hybrid I-TevI/ryA zinc-finger homing site cloned into the BglII/SpeI sites using DE1117/1118	This study
pTZYHS1.33G5A	Similar to PTZYHS1.33, with G5A substitution	This study
pTZYHS1.33C1A/G5A	Similar to PTZYHS1.33, with C1A and G5A substitutions	This study

Table S2.4: Oligonucleotides used in this study

Name	Sequence (5'-3')	Notes
DE410	GGAAGAAGTGGCTGATCTCAGC	Forward primer to generate all cycle-seq products for target sites cloned into pTox
DE411	CAGACCGCTTCTGCGTTCTG	Reverse primer to generate all cycle-seq products for target sites cloned into pTox
DE613	GCTAAAGATTTTGAAAAGGCATGGA AGAAGCATTTTAAAG	Forward quikchange primer to create R27A Tev-ZFEs
DE614	CTTTAAAATGCTTCTTCCATGCCTTTT CAAAATCTTTAGC	Reverse quikchange primer to create R27A Tev-ZFEs

DE824	<u>CTAGACAACGCTCAGTAGATGTTTTCTTGGTCTACCGTTTCCCACGCCGCA</u> <u>TG</u>	Top-strand oligo to clone the hybrid 35-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE825	<u>CGGCGTGGGAAACGGTAGACCCAAG</u> <u>AAAACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 35-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE826	<u>CTAGAGCCCGTAGTAATGACATGGCC</u> <u>TTGGGAAATCCCTTTCCCACGCCGCA</u> <u>TG</u>	Top-strand oligo to clone the hybrid 35-bp I-BmoI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE827	<u>CGGCGTGGGAAAGGGATTTCCAAG</u> <u>GCCATGTCATTACTACGGGCT</u>	Bottom-strand oligo to clone the hybrid 35-bp I-BmoI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE832	<u>CCGCGGATCCATTACTAGGCTTTTTA</u> <u>CC</u>	Reverse primer for TevN201-ZFE cloning, BamHI site underlined
DE833	<u>CCGCGGATCCACCACCATTACTAGGC</u> <u>TTTTTACC</u>	Reverse primer for TevN201G ₂ -ZFE cloning, BamHI site underlined
DE834	<u>CCGCGGATCCACCACCACCACCTTA</u> <u>CTAGGCTTTTTACC</u>	Reverse primer for TevN201G ₄ -ZFE cloning, BamHI site underlined
DE835	<u>CCGCGGATCCTTTAATATTACTAGGC</u> <u>TTTTTACC</u>	Reverse primer for TevK203-ZFE cloning, BamHI site underlined
DE836	<u>CCGCGGATCCACCACCTTTAATATTA</u> <u>CTAGGCTTTTTACC</u>	Reverse primer for TevK203G ₂ -ZFE cloning, BamHI site underlined
DE837	<u>CCGCGGATCCACCACCACCACCTTTA</u> <u>ATATTACTAGGCTTTTTACC</u>	Reverse primer for TevK203G ₄ -ZFE cloning, BamHI site underlined
DE838	<u>CCGCGGATCCTGAAATCTTTTTAATA</u> <u>TTACTAGGC</u>	Reverse primer for TevS206-ZFE cloning, BamHI site underlined
DE839	<u>CCGCGGATCCACCACCTGAAATCTTT</u> <u>TTAATATTACTAGGC</u>	Reverse primer for TevS206G ₂ -ZFE cloning, BamHI site underlined
DE840	<u>GCCGCCATGGGTAAAAGCGGAATTT</u> <u>ATCAGATT</u>	Forward primer for Tev-ZFE cloning, NcoI site underlined
DE841	<u>CCGCGGATCCGTTTTTCGGTTTACGA</u> <u>CC</u>	Reverse primer for BmoN221-ZFE cloning, BamHI site underlined
DE842	<u>CCGCGGATCCACCACCGTTTTTCGGT</u> <u>TTACGACC</u>	Reverse primer for BmoN221G ₂ -ZFE cloning, BamHI site underlined

DE843	CCGCGGATCCACCACCACCACCGTTT TTCGGTTTACGACC	Reverse primer for BmoN221G ₄ -ZFE cloning, BamHI site underlined
DE844	CCGCGGATCCACGAGAGTTTTTCGGT TTACG	Reverse primer for BmoR223-ZFE cloning, BamHI site underlined
DE845	CCGCGGATCCACCACCACGAGAGTTT TTCGGTTTACG	Reverse primer for BmoR223G ₂ -ZFE cloning, BamHI site underlined
DE846	CCGCGGATCCACCACCACCACCACG AGAGTTTTTCGGTTTACG	Reverse primer for BmoR223G ₄ -ZFE cloning, BamHI site underlined
DE847	CCGCGGATCCGATAACCGGACGAGA GTTTTTCGG	Reverse primer for BmoI226-ZFE cloning, BamHI site underlined
DE848	CCGCGGATCCACCACCGATAACCGG ACGAGAGTTTTTCGG	Reverse primer for BmoI226G ₂ -ZFE cloning, BamHI site underlined
DE849	GCCGCCATGGGTAAATCTGGTGTTTA CAAATC	Forward primer for Bmo-ZFE cloning, NcoI site underlined
DE850	CTTGGGTCTACCGTCCCACGCCGCA TG	Forward quikchange primer to make the 1.34 I-TevI/ryA <i>zinc-finger</i> target site
DE851	CATGCGGCGTGGGAACGGTAGACCC AAG	Reverse quikchange primer to make the 1.34 I-TevI/ryA <i>zinc-finger</i> target site
DE852	CTTGGGTCTACCGTCCCACGCCGCAT G	Forward quikchange primer to make the 1.33 I-TevI/ryA <i>zinc-finger</i> target site
DE853	CATGCGGCGTGGGACGGTAGACCCA AG	Reverse quikchange primer to make the 1.33 I-TevI/ryA <i>zinc-finger</i> target site
DE854	GCCTTGGGAAATCCCTCCCACGCCG CATG	Forward quikchange primer to make the 1.34 I-BmoI/ryA <i>zinc-finger</i> target site
DE855	CATGCGGCGTGGGAAGGGATTTCCCA AGGC	Reverse quikchange primer to make the 1.34 I-BmoI/ryA <i>zinc-finger</i> target site
DE856	GCCTTGGGAAATCCCTCCCACGCCG ATG	Forward quikchange primer to make the 1.33 I-BmoI/ryA <i>zinc-finger</i> target site
DE857	CATGCGGCGTGGGAGGGATTTCCCAA GGC	Reverse quikchange primer to make the 1.33 I-BmoI/ryA <i>zinc-finger</i> target site
DE858	CAGAAACAGCTGGTTTAATAACATCA TCACCACTAACTCG	Forward quikchange primer to add stops to the 3'-end of the ryA <i>zinc-finger</i>
DE859	CGAGTTAGTGGTGATGATGTTATTAA ACCAGCTGTTTCTG	Reverse quikchange primer to add stops to the 3'-end of the ryA <i>zinc-finger</i>

DE917	CTAGACAAC <u>ACT</u> CTAGTAGATGTTTTTC TTGGGTCTACCGTTT <u>CCCACGCCGCA</u> TG	Top strand oligo similar to DE824 with <u>G-23A</u> substitution
DE918	CGGCGTGGGAAACGGTAGACCCAAG AAAACATCTACTGAGT <u>GTTGT</u>	Bottom strand oligo similar to DE825 with <u>C-23T</u> substitution
DE919	CTAGAA <u>AA</u> CA <u>CT</u> CTAGTAGATGTTTTTC TTGGGTCTACCGTTT <u>CCCACGCCGCA</u> TG	Top strand oligo similar to DE824 with <u>G-23A</u> and <u>C-27A</u> substitutions
DE920	CGGCGTGGGAAACGGTAGACCCAAG AAAACATCTACTGAGT <u>GTTTT</u>	Bottom strand oligo similar to DE825 with <u>C-23T</u> and <u>G-27T</u> substitutions
DE973	<u>CTAGACAACGCT</u> CTAGTAGATGTTTTTC TTGGGTCTACCGTTAG <u>CTACTACGCAT</u> <u>G</u>	Top-strand oligo to clone the hybrid 33-bp I-TevI/9-bp <i>ryBZf</i> using <u>XbaI</u> and <u>SphI</u>
DE974	<u>CGTAGTAGCTA</u> ACGGTAGACCCAAGA AAACATCTACTGAGCGT <u>TGT</u>	Bottom-strand oligo to clone the hybrid 33-bp I-TevI/9-bp <i>ryBZf</i> using <u>XbaI</u> and <u>SphI</u>
DE976	<u>CTAGACAACGCT</u> CTAGTAGATGTTTTTC TTGGGTCTAGG <u>TCCACATATTTAACCTT</u> <u>TTGCATG</u>	Top strand oligo to clone the hybrid 30-bp I-TevI/22-bp <i>I-OnuI E1</i> using <u>XbaI</u> and <u>SphI</u>
DE977	<u>CAAAAGGT</u> TAAATATGTGGACCTAGAC CCAAGAAAACATCTACTGAGCGT <u>TGT</u>	Bottom strand oligo to clone the hybrid 30-bp I-TevI/22-bp <i>I-OnuI E1</i> using <u>XbaI</u> and <u>SphI</u>
DE978	<u>CTAGACAACACT</u> CTAGTAGATGTTTTTC TTGGGTCTAGG <u>TCCACATATTTAACCTT</u> <u>TTGCATG</u>	Top strand oligo similar to DE976 with <u>G5A</u> substitution
DE979	<u>CAAAAGGT</u> TAAATATGTGGACCTAG ACCAAGAAAACATCTACTGAGTGT <u>GT</u>	Bottom strand oligo similar to DE977 with <u>G5A</u> substitution
DE980	<u>CTAGAAAACACT</u> CTAGTAGATGTTTTTC TTGGGTCTAGG <u>TCCACATATTTAACCTT</u> <u>TTGCATG</u>	Top strand oligo similar to DE976 with <u>C1A</u> and <u>G5A</u> substitution
DE981	<u>CAAAAGGT</u> TAAATATGTGGACCTAGAC CCAAGAAAACATCTACTGAGTGT <u>TTTT</u>	Bottom strand oligo similar to DE977 with <u>C1A</u> and <u>G5A</u> substitution
DE982	CGGTTTTCGCCGACGCGCAAGGCTCCT TTTTGCTGCG	Forward quikchange primer to create E22Q I-OnuI E1 mutant

DE983	CGCAGCAAAAAGGAGCCTTGCGCGT CGGCGAAACCG	Reverse quikchange primer to create E22Q I-OnuI E1 mutant
DE991	CGCGT <u>CGACT</u> TAGAATACTCTGCCCT TGTTTC	Reverse primer to amplify the I-OnuI E1 gene with a 3' Sall site
DE1017	CGCGGAT <u>CC</u> ACCACCGTCTGAATGCT TATGATTAAG	Reverse primer for TevD184G ₂ cloning, BamHI site underlined
DE1040	CGCGGAT <u>CC</u> CAGAACGTTTCTTAATAA TTTC	Reverse primer for TevS114 cloning, BamHI site underlined
DE1042	CGCGGAT <u>CC</u> ATCAGGTCCAAGTTTAA GC	Reverse primer for TevD127 cloning, BamHI site underlined
DE1044	CGCGGAT <u>CC</u> GTTTTTACTTCCGGGTT TAC	Reverse primer for TevN140 cloning, BamHI site underlined
DE1045	CGCGGAT <u>CC</u> ATTTCTGCATTTACTAC AAG	Reverse primer for TevN169 cloning, BamHI site underlined
DE1056	<u>CTAGACAACGCTCAGTAGATGTTTTTC</u> <u>TTGGGTCCGTCACATATTTAACCTTTT</u> <u>GCATG</u>	Top-strand oligo to clone the hybrid 28- bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1057	<u>CAAAAGGTAAATATGTGGACCGACCC</u> <u>AAGAAAACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 28-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1058	<u>CTAGACAACGCTCAGTAGATGTTTTTC</u> <u>TTGGGGTCCACATATTTAACCTTTTGC</u> <u>ATG</u>	Top-strand oligo to clone the hybrid 26- bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1059	<u>CAAAAGGTAAATATGTGGACCCCAA</u> <u>GAAAACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 26-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1060	<u>CTAGACAACGCTCAGTAGATGTTTTTC</u> <u>TTGGGTCCACATATTTAACCTTTTGCAT</u> <u>G</u>	Top-strand oligo to clone the hybrid 24- bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1061	<u>CAAAAGGTAAATATGTGGACCCAAGA</u> <u>AAACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 24-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1062	<u>CTAGACAACGCTCAGTAGATGTTTTTC</u> <u>TGGTCCACATATTTAACCTTTTGCATG</u>	Top-strand oligo to clone the hybrid 22- bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>

DE1063	<u>C</u> AAAAGGTTAAATATGTGGACCAGAAA ACATCTACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 22-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1064	CTAG <u>A</u> CAACGCTCAGTAGATGTTTTG GTCCACATATTTAACCTTTTGCATG	Top-strand oligo to clone the hybrid 20-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1065	<u>C</u> AAAAGGTTAAATATGTGGACCAAAC ATCTACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 20-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1066	CTAG <u>A</u> CAACGCTCAGTAGATGTTGGT CCACATATTTAACCTTTTGCATG	Top-strand oligo to clone the hybrid 18-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1067	<u>C</u> AAAAGGTTAAATATGTGGACCAACAT CTACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 18-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1068	CTAG <u>A</u> CAACGCTCAGTAGATGGGTCC ACATATTTAACCTTTTGCATG	Top-strand oligo to clone the hybrid 16-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1069	<u>C</u> AAAAGGTTAAATATGTGGACCCATCT ACTGAGCGTTGT	Bottom-strand oligo to clone the hybrid 16-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1070	CTAG <u>A</u> CAACGCTCAGTAGAGGTCCAC ATATTTAACCTTTTGCATG	Top-strand oligo to clone the hybrid 14-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1071	<u>C</u> AAAAGGTTAAATATGTGGACCTCTACT GAGCGTTGT	Bottom-strand oligo to clone the hybrid 14-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1072	CTAG <u>A</u> CAACGCTCAGTAGGTCCACAT ATTTAACCTTTTGCATG	Top-strand oligo to clone the hybrid 12-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1073	<u>C</u> AAAAGGTTAAATATGTGGACCTACTG AGCGTTGT	Bottom-strand oligo to clone the hybrid 12-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1074	CGCGT <u>C</u> GA <u>C</u> TTAGTGGTGATGATGGT GATGGAATACTCTGCCCTTGTTCC	Reverse primer to amplify the I-OnuI E1 gene with a 3' <i>his-tag</i> and <u>Sall</u> site

DE1082	<u>GCGAGATCTGGTTCCGCCTATATGTC</u> CCG	Forward primer to amplify the I-OnuI E1 gene with a 5' BamHI site
DE1085	<u>CTAGACAACGCTCAGTAGATGTTTTTC</u> <u>TTGGGTCTACCTCCCACGCCGCATG</u>	Top-strand oligo to clone the hybrid 32-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE1086	<u>CGGCGTGGGAGGTAGACCCAAGAAA</u> <u>ACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 32-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE1087	<u>CTAGACAACGCTCAGTAGATGTTTTTC</u> <u>TTGGGTCTATCCCACGCCGCATG</u>	Top-strand oligo to clone the hybrid 30-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE1088	<u>CGGCGTGGGATAGACCCAAGAAAAC</u> <u>ATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 30-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE1117	<u>GCAGATCTCTAGCATTACGCTAGGG</u>	Forward primer to amplify target site in pTox with 5' <u>BglII</u> site to clone into pCS753
DE1118	<u>GCACTAGTCTTCTCTCATCCGCC</u>	Reverse primer to amplify target site in pTox with 3' <u>SpeI</u> site to clone into pCS753
DE1121	<u>CGTCTAGAATGAAAAGCGGAATTTAT</u> CAGATT	Forward primer to amplify I-TevI with 5' <u>XbaI</u> site to clone into pTAL3
DE1128	<u>CGGATATCTTATTAACCAGCTGTTT</u> CTGACGCAGG	Reverse primer to amplify <i>ryAZF</i> with 3' <u>EcoRV</u> site to clone into pTAL3
DE1134	<u>CTAGACAACGCTCAGTAGATGTTTTTC</u> <u>TTGGGTCTACCGTTTTCCCACGCCGC</u> <u>ATG</u>	Top-strand oligo to clone the hybrid 36-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE1135	<u>CGGCGTGGGAAAACGGTAGACCCAA</u> <u>GAAAACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 36-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE1136	<u>CTAGACAACGCTCAGTAGATGTTTTTC</u> <u>TTGGGTCTACCGTTTTCCCACGCCGCA</u> <u>TG</u>	Top-strand oligo to clone the hybrid 38-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>
DE1137	<u>CGGCGTGGGAAAACGGTAGACCCAAAG</u> <u>AAAACATCTACTGAGCGTTGT</u>	Bottom-strand oligo to clone the hybrid 38-bp I-TevI/9-bp <i>ryAZf</i> using <u>XbaI</u> and <u>SphI</u>

DE1154	<u>CTAGAAAACACTCAGTAGATGTTTTCTTGGGGGTCCACATATTTAACCTTTTGCATG</u>	Top-strand oligo to clone the hybrid 26-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1155	<u>CAAAAGGTAAATATGTGGACCCCAA</u> <u>GAAAACATCTACTGAGTGTTTT</u>	Bottom-strand oligo to clone the hybrid 26-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1156	<u>CTAGAAAACACTCAGTAGATGTTTTGTCCACATATTTAACCTTTTGCATG</u>	Top-strand oligo to clone the hybrid 20-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1157	<u>CAAAAGGTAAATATGTGGACCAAAC</u> <u>ATCTACTGAGTGTTTT</u>	Bottom-strand oligo to clone the hybrid 20-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1158	<u>CTAGAAAACACTCAGTAGATGTTGGTCCACATATTTAACCTTTTGCATG</u>	Top-strand oligo to clone the hybrid 18-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>
DE1159	<u>CAAAAGGTAAATATGTGGACCAACATCTACTGAGTGTTTT</u>	Bottom-strand oligo to clone the hybrid 18-bp I-TevI/22-bp I-OnuI E1 using <u>XbaI</u> and <u>SphI</u>

S2 Supplemental References

1. Kleinstiver BP, Fernandes A, Gloor GB, Edgell DR (2010) A unified genetic, computational, and experimental framework identifies non-conserved residues as critical for function of the homing endonuclease I-BmoI. *Nucleic Acids Res* 38: 2411-2427.
2. Edgell DR, Shub DA (2001) Related homing endonucleases I-*BmoI* and I-*TevI* use different strategies to cleave homologous recognition sites. *Proc Natl Acad Sci U S A* 98: 7898-7903.
3. Takeuchi R, Lambert AR, Mak AN, Jacoby K, Dickson RJ, et al. (2011) Tapping natural reservoirs of homing endonucleases for targeted gene modification. *Proc Natl Acad Sci U S A* 108: 13077-13082.
4. Chen Z, Zhao H (2005) A highly sensitive selection method for directed evolution of homing endonucleases. *Nucleic Acids Res* 33: e154.
5. Cermak T, Doyle EL, Christian M, Wang L, Zhang Y, et al. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* 39: e82.
6. Foley JE, *et al.* (2009) Rapid mutation of endogenous zebrafish genes using zinc finger nucleases made by Oligomerized Pool ENgineering (OPEN). *PLoS One* 4(2):e4348.

Appendix S3: Supplemental information for Chapter 3

S3 Supplemental figures

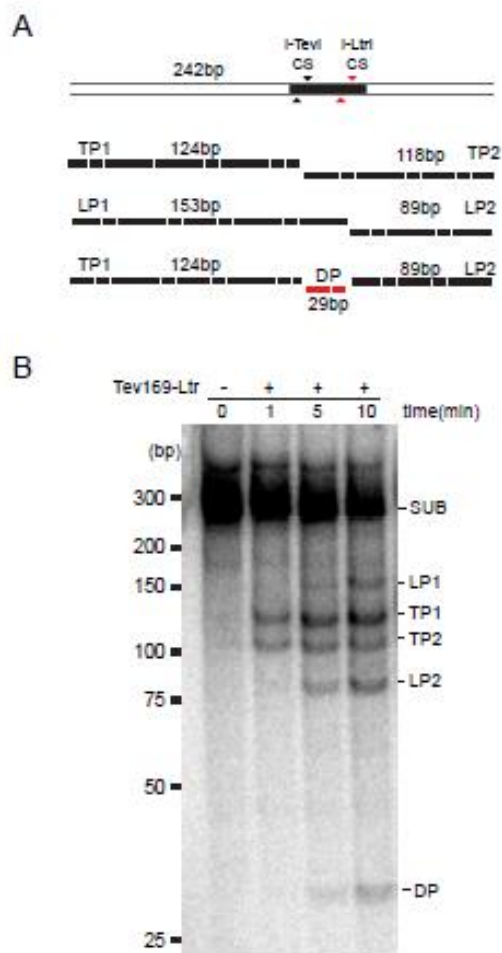


Figure S3.1: Dual active Tev169-Ltr activity *in vitro*

(A) Schematic of substrate and cleavage products containing the TL15 target site indicated by a black-filled rectangle. The black and red arrows indicate the top and bottom cleavage sites (CS) for I-TevI and I-LtrI, respectively. Black dashed lines indicate I-TevI cleavage products (TP1 and TP2), I-LtrI cleavage products (LP1 and LP2), and the red dashed line is the internal dual product (DP) from both I-TevI and I-LtrI cleavage. Sizes of the substrate and products are indicated in base pairs (bp). Mutant R27A I-TevI and E29Q I-LtrI are denoted by xTev and xLtr, respectively. (B) Polyacrylamide gel of internally labeled TL15 PCR product incubated with (+) or without (-) purified dual active Tev169-Ltr and single-active site variants. Tev169-Ltr was incubated with substrate for the indicated times, and cleavage products are indicated on the right side of the gel based on their predicted sizes from panel (A). The sizing standard (in bp) was cropped from the gel image.

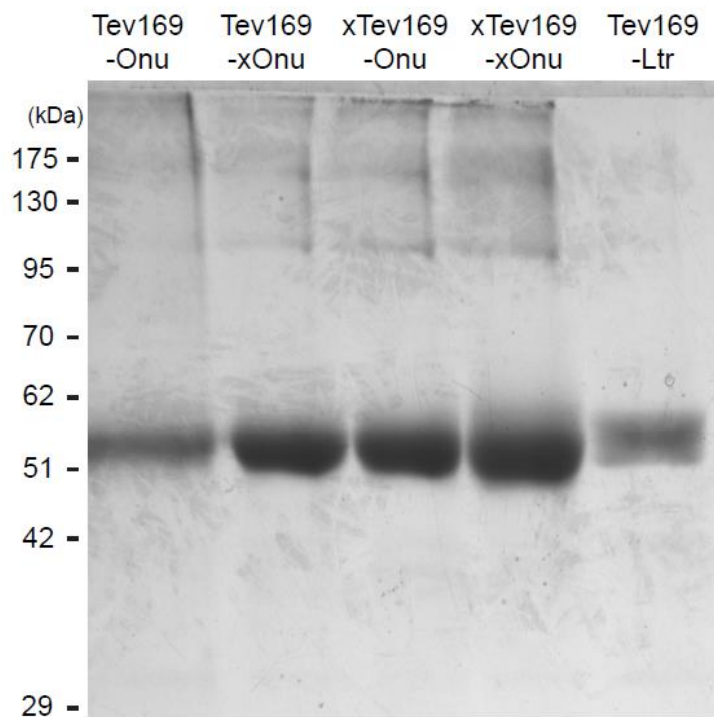


Figure S3.2: Purified MegaTev constructs on a Ni²⁺ column

SDS-Polyacrylamide gel of the purified MegaTev constructs used *in vitro*. Protein ladder was cropped from the image and sizes markers are indicated in kDa.

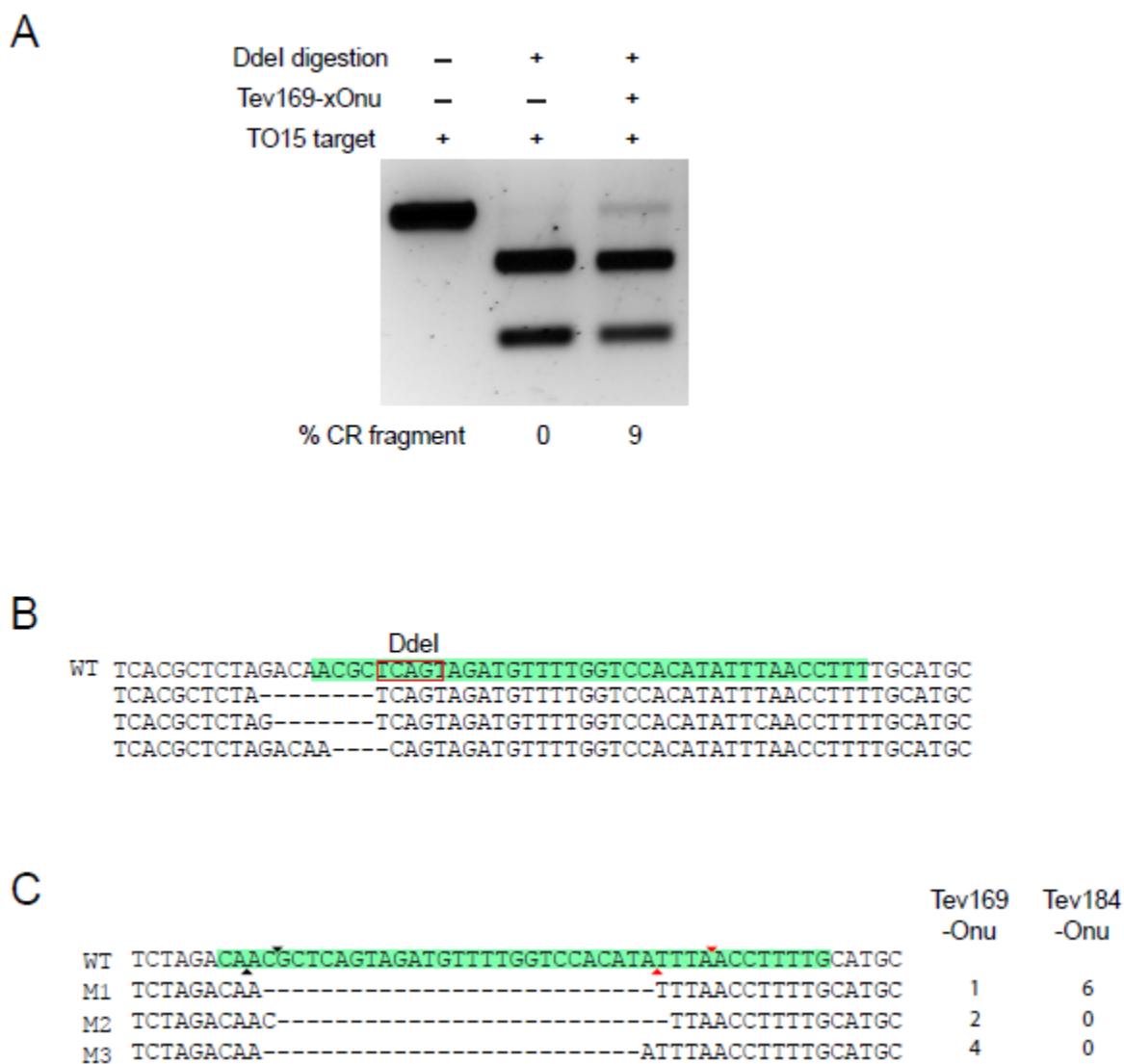


Figure S3.2: Representation of MegaTev cleavage

(A) Agarose gel of a DdeI digest for MegaTev with inactive I-OnuI. (B) Sanger sequencing for individual DdeI resistant sites treated with Tev169-xOnu (inactive I-OnuI). (C) Sanger sequencing for individual DdeI resistant sites treated with Tev169-Onu (MegaTev dual-active).

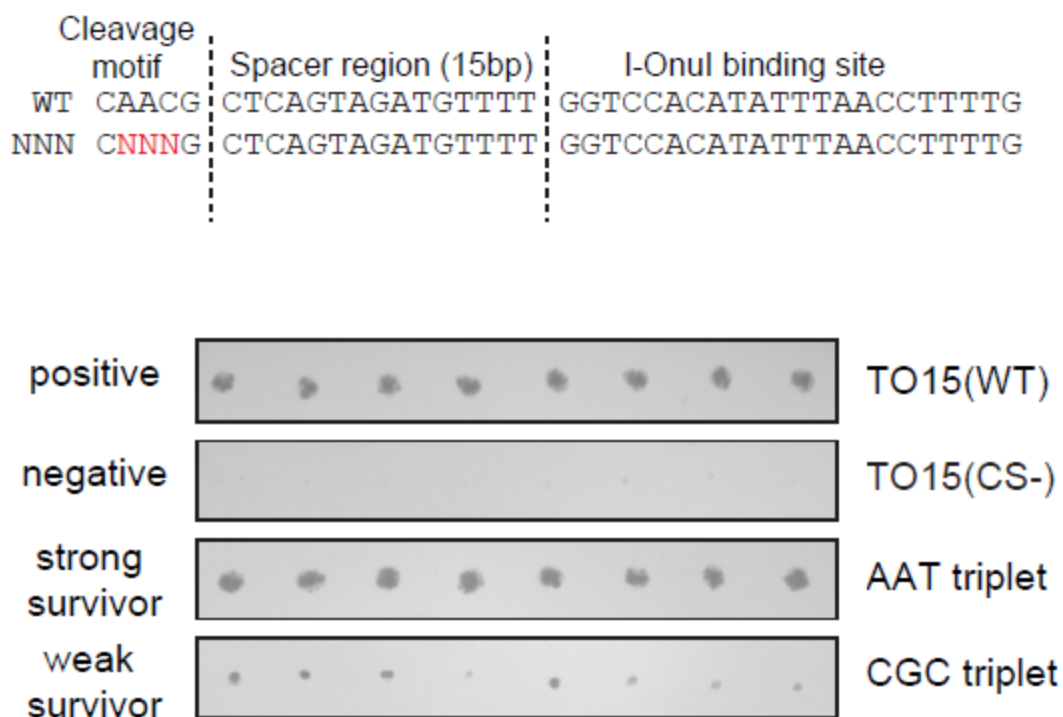


Figure S3.4: Examples of modified two-plasmid selection for CNNNG TO15 target sites

(A) Schematic for the TO15 target site highlighting in red the NNN-triplet within the cleavage motif. All 64-variants in the CNNNG cleavage motif were screened on selective plates. (B) Sample selective plate for the two control target sites and two NNN-triplets substrates consisting of a strong and weak survivors. Eight colonies were gridded for each target site with the identity displayed on the right.

S3: Supplemental Tables

Table S3.1: Strains used in this study

Construction	Description	Source
Strains		
DH5 α	F ⁻ , ϕ 80 <i>lacZ</i> Δ M15, Δ (<i>lacZYA-argF</i>)U169, <i>deoR</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> (rk ⁻ , mk ⁺), <i>phoA</i> , <i>supE44</i> , λ ⁻ , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i>	Invitrogen
ER2566	F- λ - <i>fhuA2</i> [lon] <i>ompT lacZ::T7 gene 1 gal sulA11</i> Δ (<i>mcrC-mrr</i>)114::IS10 R(<i>mcr-73::miniTn10-TetS</i>)2 R(<i>zgb-210::Tn10</i>)(TetS) <i>endA1</i> [dcm]	N.E.B.
BW25141 (λ DE3)	F ⁻ <i>lacI^q rrnB_{T14} DlacZ_{WJ16} DphoBR580 hsdR514 DaraBAD_{AH33} DrhaBAD_{LD78} galU95 endA_{BT333} uidA(DMluI)::pir⁺ recA1</i> , λ DE3 lysogen	
<i>S.cerevisiae</i> - YPH499	MATa <i>ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1</i>	Dr. Adam Bogdanove
<i>S.cerevisiae</i> - YPH500	MAT α <i>ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1</i>	Dr. Adam Bogdanove
HEK 293 and HEK293T	Human Embryonic kidney 293 cells	Dr. Schild-Poulter

Table S3.2: Oligonucleotides used in this study

Name	Sequence (5'-3')	Notes
DE-1058	CTAGACAACGCTCAGTAGATGTT TTCTTGGGGGTCCACATATTTAAC CTTTTGCATG	Top strand for T021 target site, XbaI/SphI
DE-1059	CAAAAGGTAAATATGTGGACCC CCAAGAAAACATCTACTGAGCGT TGT	Bottom strand for T021 target site, XbaI/SphI
DE-1060	CTAGACAACGCTCAGTAGATGTT TTCTTGGGTCCACATATTTAACCT TTTGCATG	Top strand for T019 target site, XbaI/SphI
DE-1061	CAAAAGGTAAATATGTGGACCC AAGAAAACATCTACTGAGCGTTG T	Bottom strand for T019 target site, XbaI/SphI

DE-1062	CTAGACAACGCTCAGTAGATGTT TTCTGGTCCACATATTTAACCTTT TGCATG	Top strand for T017 target site, XbaI/SphI
DE-1063	CAAAAGGTAAATATGTGGACCA GAAAACATCTACTGAGCGTTGT	Bottom strand for T017 target site, XbaI/SphI
DE-1064	CTAGACAACGCTCAGTAGATGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand for T015 target site, XbaI/SphI
DE-1065	CAAAAGGTAAATATGTGGACCA AAACATCTACTGAGCGTTGT	Bottom strand for T015 target site, XbaI/SphI
DE-1066	CTAGACAACGCTCAGTAGATGTT GGTCCACATATTTAACCTTTTGCA TG	Top strand for T013 target site, XbaI/SphI
DE-1067	CAAAAGGTAAATATGTGGACCA ACATCTACTGAGCGTTGT	Bottom strand for T013 target site, XbaI/SphI
DE-1068	CTAGACAACGCTCAGTAGATGGG TCCACATATTTAACCTTTTGATG	Top strand for T011 target site, XbaI/SphI
DE-1069	CAAAAGGTAAATATGTGGACCC ATCTACTGAGCGTTGT	Bottom strand for T011 target site, XbaI/SphI
DE-1156	CTAGAAAACACTCAGTAGATGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand for TO15(CS-) target site, XbaI/SphI
DE-1157	CAAAAGGTAAATATGTGGACCA AAACATCTACTGAGTGTTTT	Bottom strand for TO15(CS-) target site, XbaI/SphI
DE-1257	CCATCTCATCCCTGCGTGTCTCCG ACTCAGtatcgCCggcagatctgatcatcg	Forward barcoded primer to amplify the N8 library for Ion Torrent sequencing
DE1259	CCATCTCATCCCTGCGTGTCTCCG ACTCAGtgcacCCggcagatctgatcatcg	Forward barcoded primer to amplify the N8 round 3 selected sites for Ion Torrent sequencing
DE-1267	CCTCTCTATGGGCAGTCGGTGATg aactcgagcagctgaagc	Reverse primer to amplify the N8 target sites for Ion Torrent sequencing
DE-1338	CTAGACAACGCTCAGTAGATGTT TTAATGCTCCTATACGACGTTTAG CATG	Top strand for TL15 target site, XbaI/SphI

DE-1339	CTAAACGTCGTATAGGAGCATTA AAACATCTACTGAGCGTTGT	Bottom strand for TL15 target site, XbaI/SphI
DE-1340	CTAGAAAACACTCAGTAGATGTT TTAATGCTCCTATACGACGTTTAG CATG	Top strand for TL15(CS-) target site with an AAACA cleavage motif, XbaI/SphI
DE-1341	CTAAACGTCGTATAGGAGCATTA AAACATCTACTGAGTGTTTT	Bottom strand for TL15(CS-) target site with an AAACA cleavage motif, XbaI/SphI
DE-1379	CGGAATTCNNNNNNNNNAGTAGAT GTTTTGGTCCACATATTTAACCTT TTGCATG	Top strand oligo to create the N8 library with EcoRI site and SphI overhang
DE-1380	CAAAAGGTTAAATATGTGG	Bottom strand oligo to make the N8 library double-stranded
DE-1497	CTAGACAACGCGCAGTAGATGTT TTGGTCCACATATTTAACCTTTTG	Top strand TO15 target site with a T7G mutation, XbaI/EcoRI
DE-1498	AATTCAAAAGGTTAAATATGTGG ACCAAACATCTACTGCGCGTTG T	Bottom strand TO15 target site with a T7G mutation, XbaI/EcoRI
DE-1499	CTAGACAACGCACAGTAGATGTT TTGGTCCACATATTTAACCTTTTG	Top strand TO15 target site with a T7A mutation, XbaI/EcoRI
DE-1500	AATTCAAAAGGTTAAATATGTGG ACCAAACATCTACTGTGCGTTGT	Bottom strand TO15 target site with a T7A mutation, XbaI/EcoRI
DE-1501	CTAGACAACGCTCAGTAGATGTT TTGGTCCACATATTTAACCTTTTG	Top strand TO15 target site with a C6A mutation, XbaI/EcoRI
DE-1502	AATTCAAAAGGTTAAATATGTGG ACCAAACATCTACTGAGCGTTG T	Bottom strand TO15 target site with a C6A mutation, XbaI/EcoRI
DE-1503	CTAGACAACGGTCAGTAGATGTT TTGGTCCACATATTTAACCTTTTG	Top strand TO15 target site with a C6G mutation, XbaI/EcoRI
DE-1504	AATTCAAAAGGTTAAATATGTGG ACCAAACATCTACTGACCGTTG T	Bottom strand TO15 target site with a C6G mutation, XbaI/EcoRI
DE-1505	CTAGACAACGTTTCAGTAGATGTTT TTGGTCCACATATTTAACCTTTTG	Top strand TO15 target site with a C6T mutation, XbaI/EcoRI

DE-1506	AATTCAAAGGTTAAATATGTGG ACCAAACATCTACTGAACGTTG T	Bottom strand TO15 target site with a C6T mutation, XbaI/EcoRI
DE-1544	GCAGATCTCCAGTGTGCCGGTCTC CG	Forward primer to clone target sites from pKox into pCSTox using homology- directed repair
DE-1545	GCACTAGTCTTCTGAGTTCGGCAT GGG	Reverse primer to clone target sites from pKox into pCSTox using homology- directed repair
DE-1645	GCTCTGCAGAATGGCCAACCTTT AACG	Forward internal primer to DE1680/1682 to amplify the TO15 integrated target site
DE-1646	CTACCCGCTTCCATTGCTCAGCGG TGC	Reverse internal primer to DE1680/1682 to amplify the TO15 integrated target site
DE-1662	CTAGACAACGCTCAGTAGATGTT TTCTTGGGAATGCTCCTATACGAC GTTTAGCATG	Top strand for TL21 target site, XbaI/SphI
DE-1663	CTAAACGTCGTATAGGAGCATTC CCAAGAAAACATCTACTGAGCGT TGT	Bottom strand for TL21 target site, XbaI/SphI
DE-1664	CTAGACAACGCTCAGTAGATGTT TTCTTGAATGCTCCTATACGACGT TTAGCATG	Top strand for TL19 target site, XbaI/SphI
DE-1665	CTAAACGTCGTATAGGAGCATTC AAGAAAACATCTACTGAGCGTTG T	Bottom strand for TL19 target site, XbaI/SphI
DE-1666	CTAGACAACGCTCAGTAGATGTT TTCTAATGCTCCTATACGACGTTT AGCATG	Top strand for TL17 target site, XbaI/SphI
DE-1667	CTAAACGTCGTATAGGAGCATT GAAAACATCTACTGAGCGTTGT	Bottom strand for TL17 target site, XbaI/SphI
DE-1668	CTAGACAACGCTCAGTAGATGTT AATGCTCCTATACGACGTTTAGCA TG	Top strand for TL13 target site, XbaI/SphI
DE-1669	CTAAACGTCGTATAGGAGCATT ACATCTACTGAGCGTTGT	Bottom strand for TL13 target site, XbaI/SphI

DE-1670	CTAGACAACGCTCAGTAGATGAA TGCTCCTATACGACGTTTAGCATG	Top strand for TL11 target site, XbaI/SphI
DE-1671	CTAAACGTCGTATAGGAGCATTC ATCTACTGAGCGTTGT	Bottom strand for TL11 target site, XbaI/SphI
DE-1672	AATGTGCAGTCTAGGGCAGC	Forward primer to PCR amplify I-OnuI E1 site in the MAO-B gene
DE-1673	TGAGGGGCGAACTTTGAACA	Reverse primer to PCR amplify I-OnuI E1 site in the MAO-B gene
DE-1680	GCTAACTAGCTCTGTATCTGGCG	Forward primer to amplify the TO15 integrated target site
DE-1682	CTGACGAGTTCTGAACACCC	Reverse primer to amplify the TO15 integrated target site
DE-1853	TGAGGGGCGAACTTTGAACA	Internal forward primer to DE1702/1703 to PCR amplify I-OnuI E1 site in the MAO-B gene
DE-1854	CAAATCACATAAGAAGTGATCG	Internal reverse primer to DE1702/1703 to PCR amplify I-OnuI E1 site in the MAO-B gene
DE-1855	AGCAGAGAGGCATTCAACCC	Forward primer to PCR amplify I-OnuI E1 off-target #3
DE-1856	CTGGGGAGGAAGAGCTTTGG	Reverse primer to PCR amplify I-OnuI E1 off-target #3
DE-1857	CTGAGGGCAGTTAATGAGGC	Internal forward primer to DE1855/1856 to PCR amplify I-OnuI E1 off-target #3
DE-1858	GGAGTCACAAAGCAGGGGC	Internal reverse primer to DE1855/1856 to PCR amplify I-OnuI E1 off-target #3
DE-1859	AGAGTGAAGTCTGTGGTGG	Forward primer to PCR amplify I-OnuI E1 off-target #4
DE-1860	GGCTGATAAGGGAGACTGGC	Reverse primer to PCR amplify I-OnuI E1 off-target #4
DE-1861	CACACTGAACATCTAAACTCATG G	Internal forward primer to DE1859/1860 to PCR amplify I-OnuI E1 off-target #4
DE-1862	CAAAGAGTAGGATTGTGCTCC	Internal reverse primer to DE1859/1860 to PCR amplify I-OnuI E1 off-target #4

DE-1863	ATTGCCAAGAAGTCACCCG	Forward primer to PCR amplify I-OnuI E1 off-target #15
DE-1864	AAAGGTTTGCAAAGCAGCG	Reverse primer to PCR amplify I-OnuI E1 off-target #15
DE-1865	GAATACGCTACCACTCCTTACC	Internal forward primer to DE1863/1864 to PCR amplify I-OnuI E1 off-target #15
DE-1866	CTGGAAGGAACGTCTCCAAGC	Internal reverse primer to DE1863/1864 to PCR amplify I-OnuI E1 off-target #15

Appendix S4: Supplemental information for Chapter 4 and 5

S4 Supplemental Figures

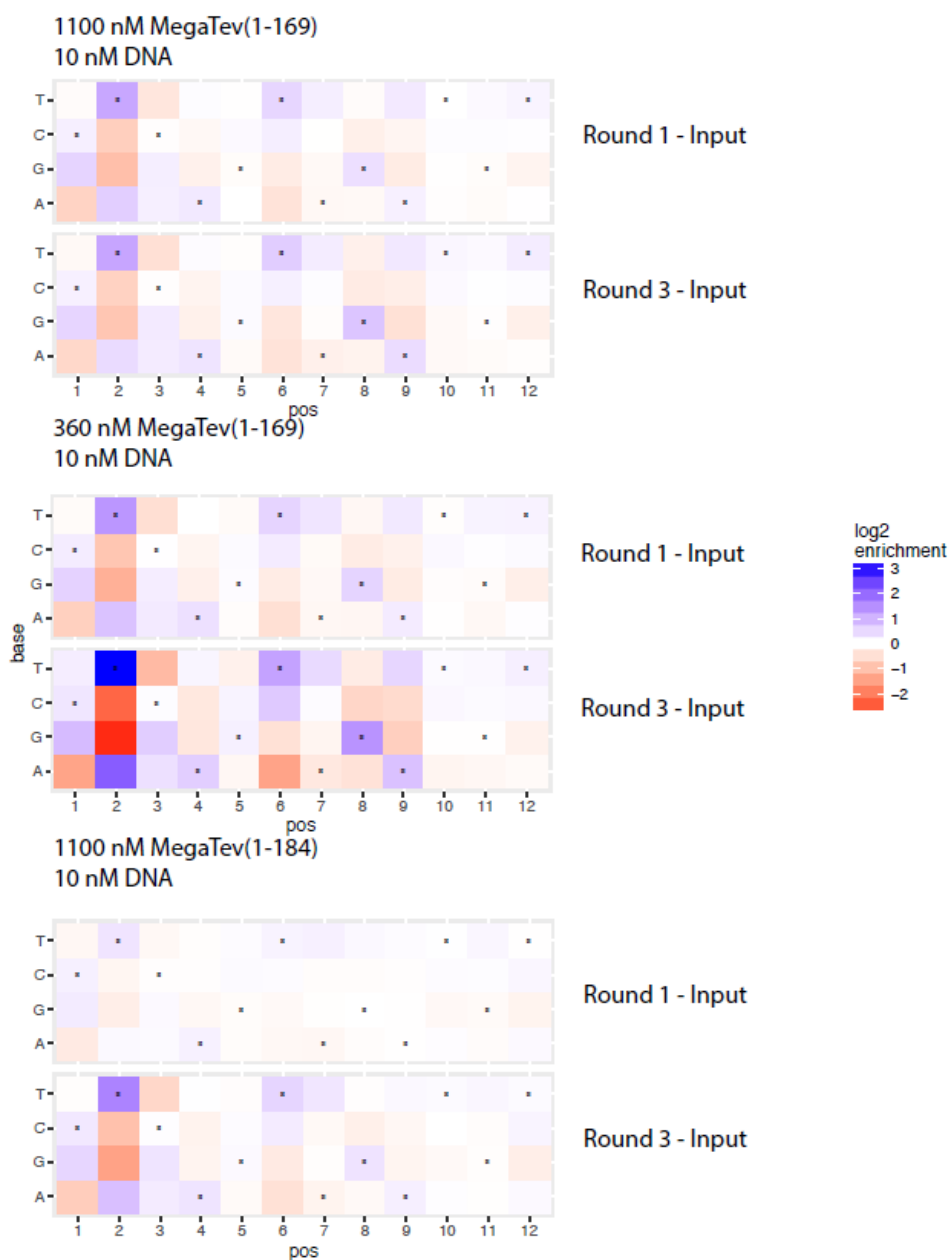


Figure S4.1: In vitro enrichment of N12 DNA spacer library with the MegaTev(1-169) or MegaTev(1-184) constructs under different reaction conditions.

Nucleotide preference in the DNA spacer region is displayed as proportional log₂ enrichment of each nucleotide at each position relative to the abundance in the input pool. Black dots indicate the wild-type nucleotide in the td DNA spacer at each position.

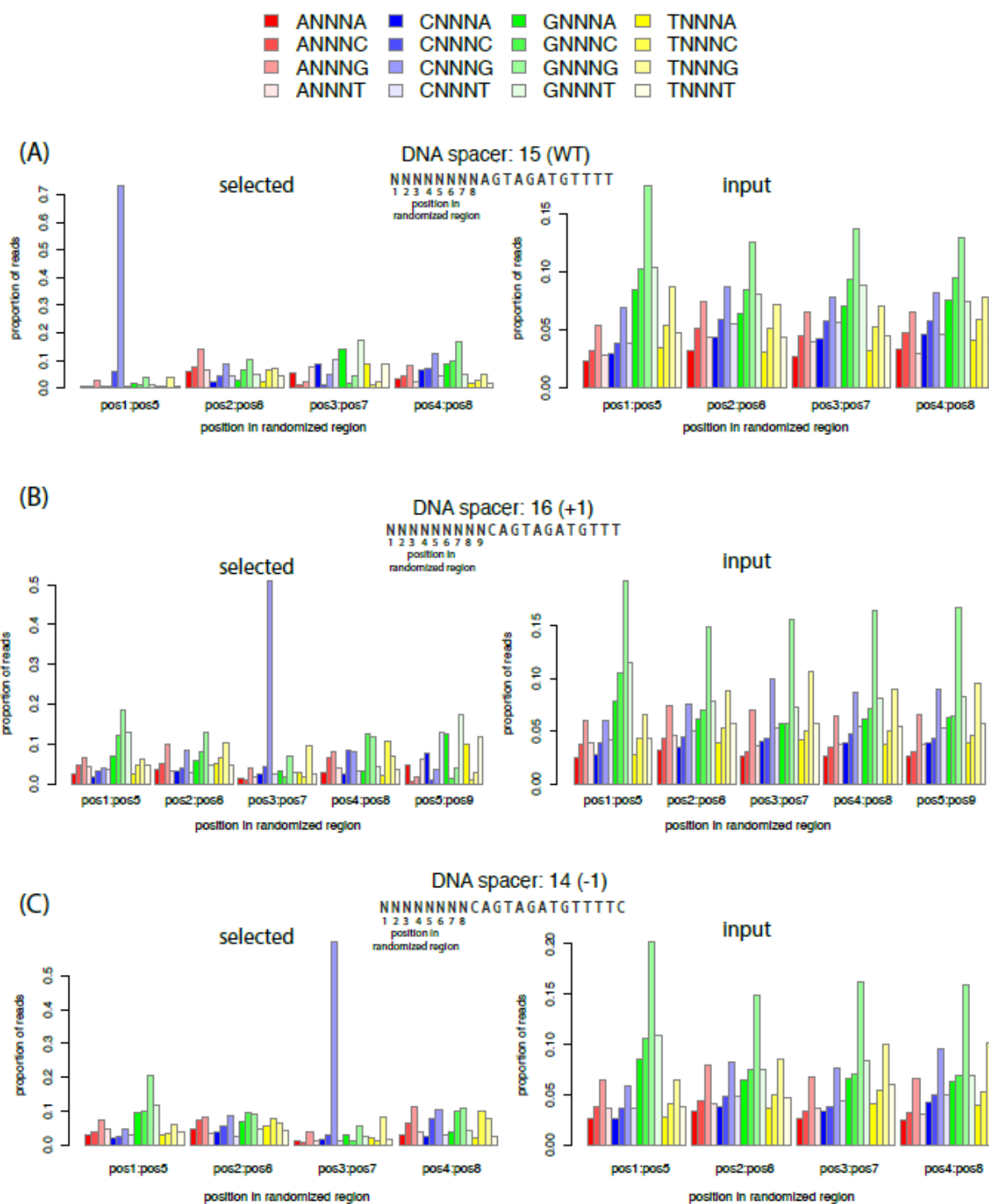


Figure S4.2: Frequency and phasing of all 16 NNNNN motifs in the three different DNA spacer substrates (panels A, B, and C).

The frequency of each NNNNN is the proportion of reads with the indicated NNNNN sequence relative to total reads at each position (ie. pos1:pos5). The left barplot in each panel is derived from the selected data, and the right panel in each plot is from the input library.

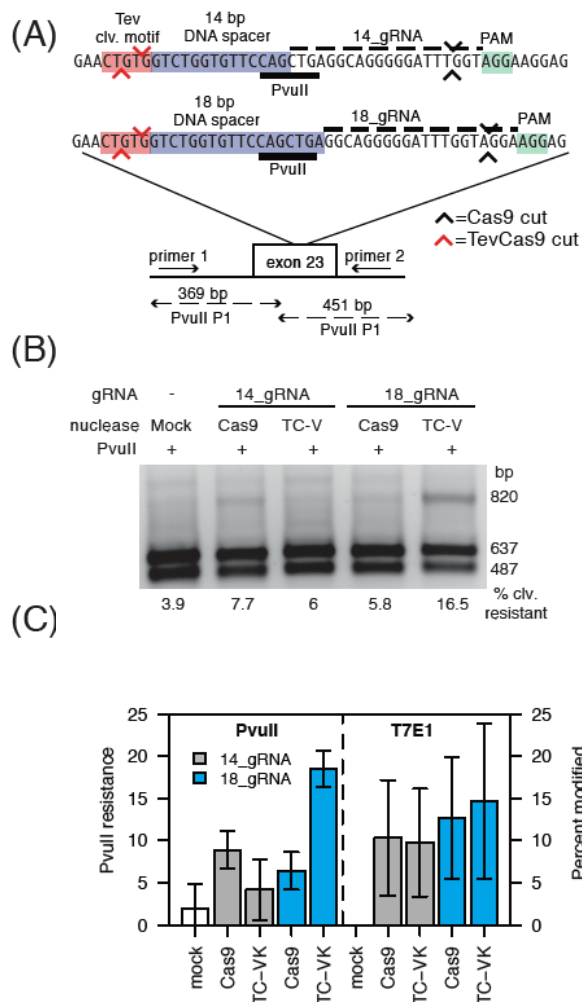
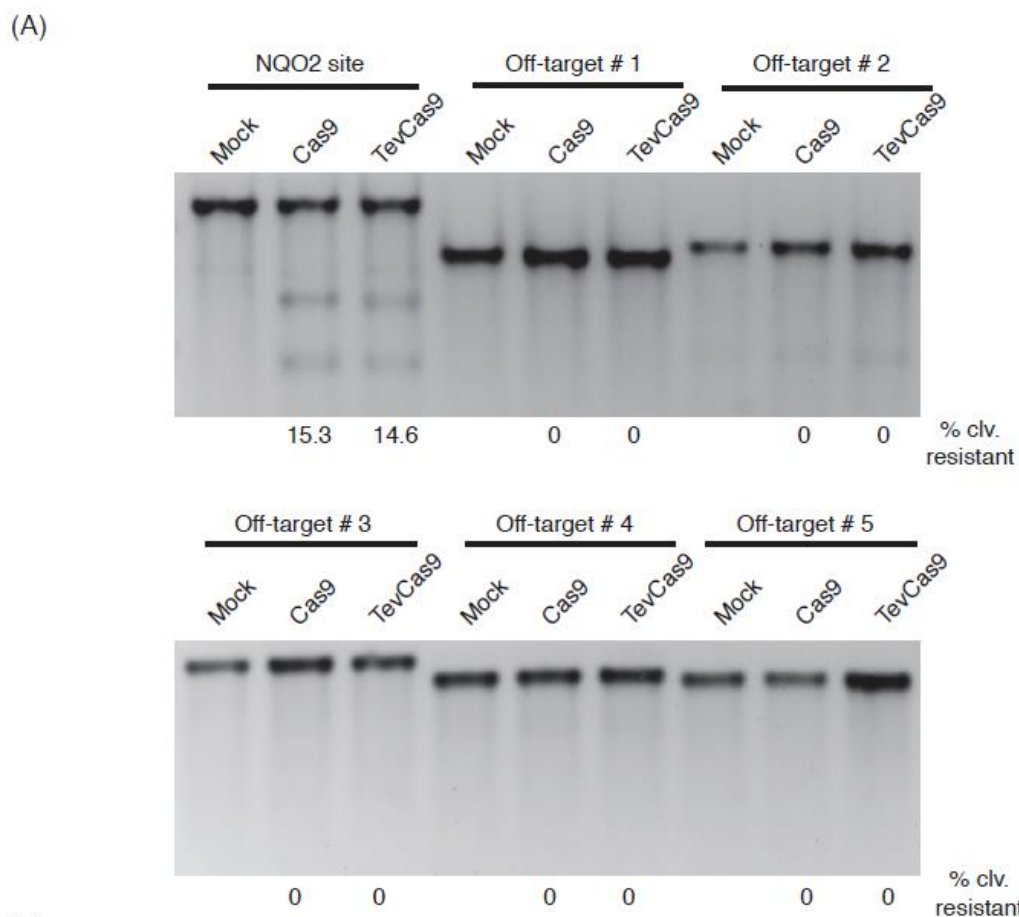


Figure S4.3: Activity and specificity of TevCas9 in HEK 293 cells

(A) Detailed schematic of the target site in exon 23 of the human *TSC1* gene, positions of PCR primers used for amplification, and sizes of PvuII cleavage products. The I-TevI cleavage motif and DNA spacers are highlighted by red and blue rectangles, and the PAM motif by a green rectangle. Positions of I-TevI and Cas9 cleavage sites are represented by red and black arrows, respectively. (B) Representative agarose gel of PvuII cleavage assays on PCR products amplified from the *TSC1* locus 48 hrs after transfection with Cas9 or with TC-VK. Sizes of the substrate (1124 bp) and two PvuII cleavage products are indicated on the right. The percent of substrate resistant to cleavage by PvuII is indicated below each lane. (C) Activity of TevCas9 TC-VK variant and Cas9 at the *TSC1* site measured by PvuII resistance and T7E1. Barplots are mean values of at least three independent experiments, with vertical bars representing standard deviation.



(B)

Site	25bp upstream sequence	gRNA sequence	Score	Mismatches (MMs)	CNNNG Motif
NQO2	tcttt caacg gatccttgaagaatg	TGGCTGTAGATGAACTGAGCAGG	—	No MMs	Yes
Off-target # 1	gtttggaagcacctcagcaggaca	GGGCTGTAGAGGAAGTGAAGTGG	0.9	3MMs [1:11:20]	No
Off-target # 2	tcatctacgtgcatccttgaagaac	TGGGTGTGGGTGAAGTGAAGCAGG	2.3	3MMs [4:8:10]	No
Off-target # 3	ctgga cgagg tggcgggccttgggg	AGGCAGTGGATGAAGTGAACCAG	0.5	4MMs [1:5:8:19]	Yes
Off-target # 4	acaggttcattctctctctgctgct	TGGCTAAAGATGAAGTGAAGCCTGG	0.4	3MMs [6:7:19]	No
Off-target # 5	cgggat ctggg gttcaccatcccc	TGGCTTCAAGGAAGTGAAGCTGG	0.3	4MMs [6:8:9:11]	Yes

Figure S4.4: Off-target profiling of TevCas9 and Cas9 programmed with the NQO2 gRNA.

The 5 off-target sites were predicted using the CRISPR design tools (<http://crispr.mit.edu/>). (A) Representative agarose gel images of T7E1 mismatch cleavage assays of PCR fragments amplified from HEK 293 cells transfected with Cas9 or TevCas9. (B) For each site, the gRNA target is in capital letters with mismatches to the NQO2 gRNA in red font. The lower case letters are the 25-bp upstream of the gRNA that would include the DNA spacer contacted by the I-TevI linker, and a CNNNG cleavage motif (if present) highlighted by a yellow box.

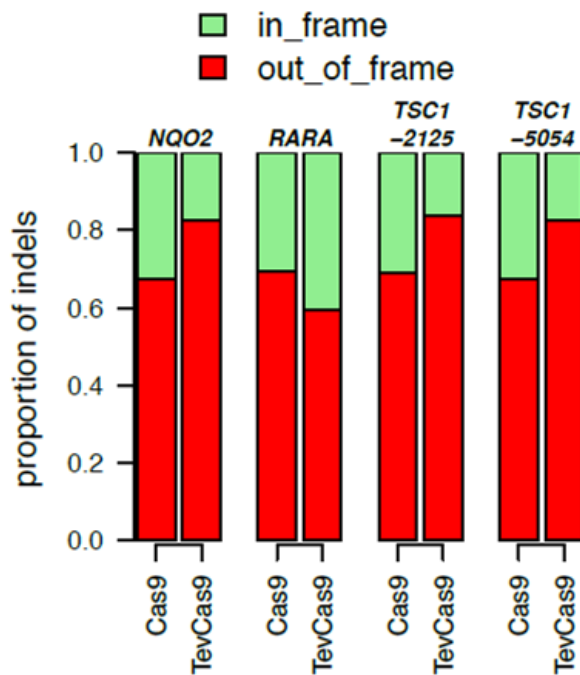


Figure S4.5: TevCas9 can bias the proportion of in-frame to out-of-frame indels

Illumina read data for the target site is plotted as the proportion of reads that are in-frame (green) and out-of-frame (red) for TevCas9 and Cas9.

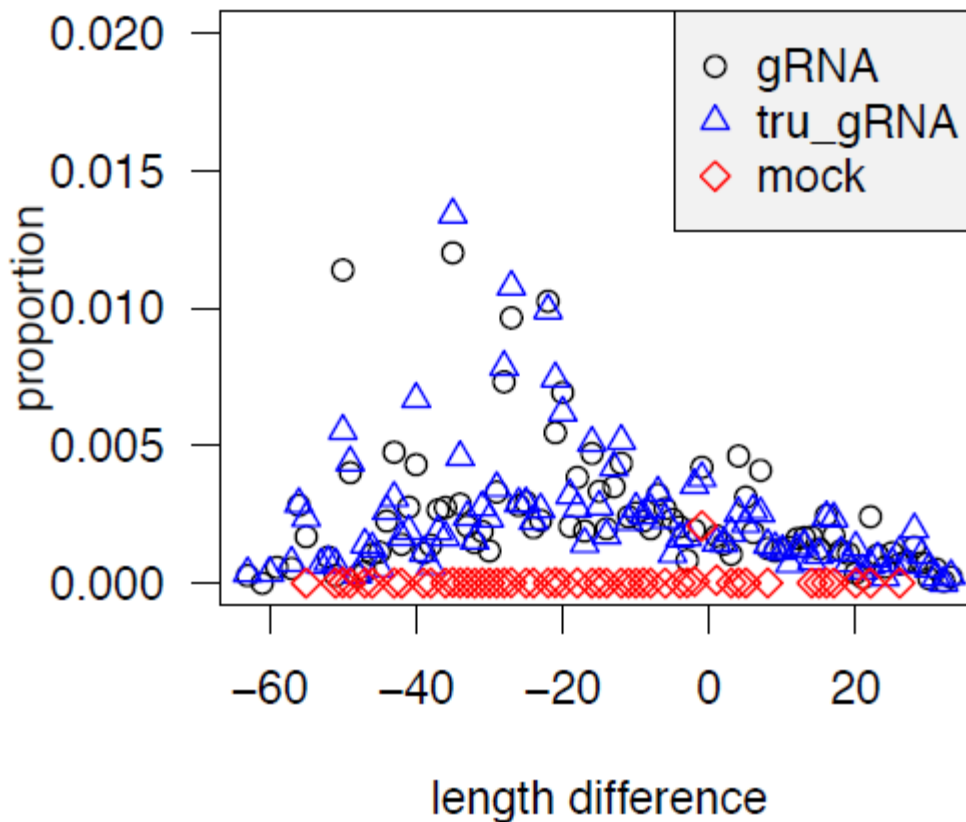


Figure S4.6: Indels generated at the VEGFA gene by Cas9 paired nickases

Illumina sequencing data was downloaded from the Sequencing Read Archive (SRP03315, experiments SRX380882 (control), SRX380880 (tru_gRNA), SRX380879 (gRNA)), and reads analyzed for length of insertion or deletion relative to the unmodified target. Data are plotted as length difference to unmodified site versus proportion of each indel relative to total reads.

S4 Supplemental Tables

Table S4.1: Strains used in this study

Construction	Description	Source
Strains		
DH5 α	F ⁻ , ϕ 80 <i>dlacZ</i> Δ M15, Δ (<i>lacZYA-argF</i>)U169, <i>deoR</i> , <i>recA1</i> , <i>endA1</i> , <i>hsdR17</i> (<i>rk⁻</i> , <i>mk⁺</i>), <i>phoA</i> , <i>supE44</i> , λ ⁻ , <i>thi-1</i> , <i>gyrA96</i> , <i>relA1</i>	Invitrogen
ER2566	F ⁻ λ - <i>fhuA2</i> [<i>lon</i>] <i>ompT lacZ::T7 gene 1 gal sulA11</i> Δ (<i>mcrC-mrr</i>)114:: <i>IS10</i> R(<i>mcr-73::miniTn10-TetS</i>)2 R(<i>zgb-210::Tn10</i>)(<i>TetS</i>) <i>endA1</i> [<i>dcm</i>]	N.E.B.
BW25141 (λ DE3)	F ⁻ <i>lacI^q</i> <i>rrnB_{T14}</i> <i>DlacZ_{WJ16}</i> <i>DphoBR580</i> <i>hsdR514</i> <i>DaraBAD_{AH33}</i> <i>DrhaBAD_{LD78}</i> <i>galU95</i> <i>endA_{BT333}</i> <i>uidA(DMlu1)::pir+</i> <i>recA1</i> , λ DE3 lysogen	
<i>S.cerevisiae</i> - YPH499	MATa <i>ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1</i>	Dr. Adam Bogdanove
<i>S.cerevisiae</i> - YPH500	MATa <i>ura3-52 lys2-801_amber ade2-101_ochre trp1-Δ63 his3-Δ200 leu2-Δ1</i>	Dr. Adam Bogdanove
HEK 293	Human Embryonic kidney 293 cells	Dr. Schild-Poulter

Table S4.2: Oligonucleotides used in this study

Name	Sequence (5'-3')	Notes
DE-1180	CGCATGCGGCGCNCNCAAAGGTT AAATATGTGGACCNNNNNNNNNN NNNNNCGTTGGCCGAATTCCG	Bottom strand for MegaTev TO15 N15 randomized spacer target site
DE-1181	CGGAATTCGGCCAAC	Forward primer to make DE1180 double-stranded DNA for cloning
DE-1064	CTAGACAACGCTCAGTAGATGTT TTGGTCCACATATTTAACCTTTTGCATG	Top strand oligonucleotide for TevOnu TO15 (15-bp spacer) target site
DE-1065	CAAAAGGTTAAATATGTGGACCA AAACATCTACTGAGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 (15-bp spacer) target site
DE-1412	CTAGACAACGCTCAGTACATGTTT TTGGTCCACATATTTAACCTTTTGCATG	Top strand oligonucleotide for TevOnu TO15 G8C target site
DE-1413	CAAAAGGTTAAATATGTGGACCA AAACATGTACTGAGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 G8C target site

DE-1414	CTAGACAACGCTCAGTAAATGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15 G8A target site
DE-1415	CAAAAGGTAAATATGTGGACCA AAACATTTACTGAGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 G8A target site
DE-1416	CTAGACAACGCTCAGTATATGTTT TGGTCCACATATTTAACCTTTTG ATG	Top strand oligonucleotide for TevOnu TO15 G8T target site
DE-1417	CAAAAGGTAAATATGTGGACCA AAACATAACTGAGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 G8T target site
DE-1418	CTAGACAACGCTCAGTAGTTGTTT TGGTCCACATATTTAACCTTTTG ATG	Top strand oligonucleotide for TevOnu TO15 A9T target site
DE-1419	CAAAAGGTAAATATGTGGACCA AAACAACTACTGAGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 A9T target site
DE-1420	CTAGACAACGCTCAGTAGGTGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15 A9G target site
DE-1421	CAAAAGGTAAATATGTGGACCA AAACACTACTGAGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 A9G target site
DE-1422	CTAGACAACGCTCAGTAGCTGTTT TGGTCCACATATTTAACCTTTTG ATG	Top strand oligonucleotide for TevOnu TO15 A9C target site
DE-1423	CAAAAGGTAAATATGTGGACCA AAACAGCTACTGAGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 A9C target site
DE-1209	CTAGACAACGCCAGTAGACATA GGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15 T2C (includes 10-15 bp spacer swapped for MAO-B sequence) target site
DE-1210	CAAAAGGTAAATATGTGGACCC CTATGTCTACTGGGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 T2C (includes 10-15 bp spacer swapped for MAO-B sequence) target site
DE-1497	CTAGACAACGCGCAGTAGATGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15 T2G target site

DE-1498	CAAAAGGTTAAATATGTGGACCA AAACATCTACTGCGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 T2G target site
DE-1499	CTAGACAACGCACAGTAGATGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15 T2A target site
DE-1500	CAAAAGGTTAAATATGTGGACCA AAACATCTACTGTGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 T2A target site
DE-1501	CTAGACAACGATCAGTAGATGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15 C1A target site
DE-1502	CAAAAGGTTAAATATGTGGACCA AAACATCTACTGATCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 C1A target site
DE-1503	CTAGACAACGCTCAGTAGATGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15 C1G target site
DE-1504	CAAAAGGTTAAATATGTGGACCA AAACATCTACTGACCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 C1G target site
DE-1505	CTAGACAACGTTTCAGTAGATGTTT TGGTCCACATATTTAACCTTTTGC ATG	Top strand oligonucleotide for TevOnu TO15 C1T target site
DE-1506	CAAAAGGTTAAATATGTGGACCA AAACATCTACTGAAACGTTGT	Bottom strand oligonucleotide for TevOnu TO15 C1T target site
DE-1958	CTAGACAACGCTCAGCAGATGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15 T6C target site
DE-1959	CAAAAGGTTAAATATGTGGACCA AAACATCTGCTGAGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 T6C target site
DE-1960	CTAGACAACGCTCAGGAGATGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15 T6G target site
DE-1961	CAAAAGGTTAAATATGTGGACCA AAACATCTCTGAGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 T6G target site
DE-1962	CTAGACAACGCTCAGAAGATGTT TTGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15 T6A target site

DE-1963	CAAAAGGTTAAATATGTGGACCA AAACATCT T CTGAGCGTTGT	Bottom strand oligonucleotide for TevOnu TO15 T6A target site
DE-2142	CTAGA CAGAGGTGGGCAATGGCG TG GGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15-PKD1 2386 target site
DE-2143	CAAAAGGTTAAATATGTGGACCC ACGCCATTGCCACCTCTGT	Bottom strand oligonucleotide for TevOnu TO15-PKD1 2386 target site
DE-2144	CTAGA CTTTGGAAGGTAATTACA GT GGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15-HPRT 5740 target site
DE-2145	CCAAAAGGTTAAATATGTGGACC ACTGTAATTACCTTCCAAAGT	Bottom strand oligonucleotide for TevOnu TO15- HPRT 5740 target site
DE-2146	CTAGA CTGTGCTCAGGTGATCCTC C GGTCCACATATTTAACCTTTTGC ATG	Top strand oligonucleotide for TevOnu TO15-Ku70 14915 target site
DE-2147	CAAAAGGTTAAATATGTGGACCG GAGGATCACCTGAGCACAGT	Bottom strand oligonucleotide for TevOnu TO15- Ku70 14915 target site
DE-2148	CTAGA CTAAGGTGGTCCTTCCCA G AGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15- XPC 1277 target site
DE-2149	CAAAAGGTTAAATATGTGGACCT CTGGGAAGGACCACCTTAGT	Bottom strand oligonucleotide for TevOnu TO15-XPC 1277 target site
DE-2206	CTAGA CATTGCTGTGCTTTGGGGA T GGTCCACATATTTAACCTTTTGC ATG	Top strand oligonucleotide for TevOnu TO15-VEGFA 2522 target site
DE-2207	CAAAAGGTTAAATATGTGGACCC TTGCAATAATTCTGCAATGT	Bottom strand oligonucleotide for TevOnu TO15-VEGFA 2522 target site
DE-2208	CTAGA CAGAGGAAGATGGTGTGC CC GGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15-TSC1-2125 target site
DE-2209	CAAAAGGTTAAATATGTGGACCG GGCACACCATCTTCTCTGT	Bottom strand oligonucleotide for TevOnu TO15-TSC1-2125 target site
DE-2210	CTAGA CTGTGGTCTGGTGTCCAG C GGTCCACATATTTAACCTTTTGC ATG	Top strand oligonucleotide for TevOnu TO15-TSC1-5054 target site

DE-2211	CAAAAGGTAAATATGTGGACCG CTGGAACACCAGACCACAGT	Bottom strand oligonucleotide for TevOnu TO15-TSC1-5054 target site
DE-2212	CTAGACTGTGCAGCATGGAATTC CTGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15-TSC1-6374 target site
DE-2213	CAAAAGGTAAATATGTGGACCA GGAATTCCATGCTGCACAGT	Bottom strand oligonucleotide for TevOnu TO15-TSC1-6374 target site
DE-2216	CTAGACATTGCAGAATTATTGCA AGGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15-APC 1440 target site
DE-2217	CAAAAGGTAAATATGTGGACCC TTGCAATAATTCTGCAATGT	Bottom strand oligonucleotide for TevOnu TO15-APC 1440 site
DE-2218	CTAGACATTGGTACCTGGTACTG ATGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15-BRAC1-2541 target site
DE-2219	CAAAAGGTAAATATGTGGACCA TCAGTACCAGGTACCAATGT	Bottom strand oligonucleotide for TevOnu TO15-BRAC1-2541 target site
DE-2220	CTAGACAGAGGCGAATTTATTAT CAGGTCCACATATTTAACCTTTTG CATG	Top strand oligonucleotide for TevOnu TO15-BRAC1-3999 target site
DE-2221	CAAAAGGTAAATATGTGGACCT GATAATAAATTCGCCTCTGT	Bottom strand oligonucleotide for TevOnu TO15-BRAC1-3999 target site
DE-2645	CACCGTGGCTGTAGATGAACTGA GC	Top strand oligonucleotide for Cas9 human NQO2 gRNA with BbsI overhangs
DE-2646	AAACGCTCAGTTCATCTACAGCC AC	Bottom strand oligonucleotide for Cas9 human NQO2 gRNA with BbsI overhangs
DE-2701	CACCGTTAGCATCTTGGGGAACA TG	Top strand oligonucleotide for Cas9 human RARA gRNA with BbsI overhangs
DE-2702	AAACCATGTTCCCAAGATGCTA AC	Bottom strand oligonucleotide for Cas9 human RARA gRNA with BbsI overhangs

DE-2703	CACCGTACCTCCCCAATGGAAGT GC	Top strand oligonucleotide for Cas9 human TSC1-1 gRNA with BbsI overhangs
DE-2704	AAACGCACTTCCATTGGGGAGGT AC	Bottom strand oligonucleotide for Cas9 human TSC1-1 gRNA with BbsI overhangs
DE-2705	CACCGCTGAGGCAGGGGGATTTG GT	Top strand oligonucleotide for Cas9 human TSC1-2 14bp gRNA with BbsI overhangs
DE-2706	AAACACCAAATCCCCCTGCCTCA GC	Bottom strand oligonucleotide for Cas9 human TSC1-2 14bp gRNA with BbsI overhangs
DE-2707	CACCGGGCAGGGGGATTTGGTAG GA	Top strand oligonucleotide for Cas9 human TSC1-2 gRNA 18bp with BbsI overhangs
DE-2708	AAACTCCTACCAAATCCCCCTGCC C	Bottom strand oligonucleotide for Cas9 human TSC1-2 18bp gRNA with BbsI overhangs
DE-2510	GTTCAAACACACATGCTCTGC	Forward primer for 1 st round amplification of NQ02 54 site human site
DE-2511	GCACTCCTGATGCTTCCTGTGGG	Reverse primer for 1 st round amplification of NQ02 54 site human site
DE-2512	CAATCATCACAGGGTCCTGAGGC	Forward primer for 2 nd round amplification of NQ02 54 site human site
DE-2513	GGAACCCCAGAAAATTGAGAAGC	Reverse primer for 2 nd round amplification of NQ02 54 human site
DE-2359	TGAGAGCCGTTTTCAACCCT	Forward primer for 1 st round amplification of TSC1 2125 human site
DE-2360	CCAGCCTTCTCTGTTCAGCA	Reverse primer for 1 st round amplification of TSC1 2125 human site
DE-2361	GGGATGTCAGGCCATCTGAA	Forward primer for 2 nd round amplification of TSC1 2125 human site
DE-2362	GGTGAATACCGACTGCCAT	Reverse primer for 2 nd round amplification of TSC1 2125 human site

DE-2779	ACGTCCATGGAGCTTCTAGC	Forward primer for 1 st round amplification of TSC1 5054 human site
DE2780	CCCGAGAAGCACTTGAGCAT	Reverse primer for 1 st round amplification of TSC1 5054 human site
DE2713	TTCAGTACCAACTGCCAGCC	Forward primer for 2 nd round amplification of TSC1 5054 human site
DE-2714	ACCCTTCTGTCTACTGGCCT	Reverse primer for 2 nd round amplification of TSC1 5054 human site
DE-2777	ACGTCCATGGAGCTTCTAGC	Forward primer for 1 st round amplification of RARA 233 human site
DE-2778	CCCGAGAAGCACTTGAGCAT	Reverse primer for 1 st round amplification of RARA 233 human site
DE-2711	TTCAGTACCAACTGCCAGCC	Forward primer for 2 nd round amplification of RARA 233 human site
DE-2712	ACCCTTCTGTCTACTGGCCT	Reverse primer for 2 nd round amplification of RARA 233 human site
DE-2783	TTCTCTCCACAGCTGTCCA	Forward primer for 1 st round amplification of NQ02 off-target #1
DE-2784	CACTAGGTGCAGACTCAGGC	Reverse primer for 1 st round amplification of NQ02 off-target #1
DE-2719	CAAGGCCCTCTCTCTTTTCG	Forward primer for 2 nd round amplification of NQ02 off-target #1
DE-2720	ATGGGTAGAAGCAGATGCCG	Reverse primer for 2 nd round amplification of NQ02 off-target #1
DE-2721	CCACTGCCAGATTTCTCCCC	Forward primer for 1 st round amplification of NQ02 off-target #2
DE-2722	GGGCATTCATTTGTCTGCACTT	Reverse primer for 1 st round amplification of NQ02 off-target #2
DE-2723	TCCTGCCAGGGAGTGATACA	Forward primer for 1 st round amplification of NQ02 off-target #3
DE-2724	CCAGGGTCGCGAACTAATGA	Reverse primer for 1 st round amplification of NQ02 off-target #3
DE-2785	TGCAACACCCTCTTTAATACTGA	Forward primer for 1 st round amplification of NQ02 off-target #4

DE-2786	AAAACGACCTCCGGTTTGTG	Reverse primer for 1 st round amplification of NQ02 off-target #4
DE-2725	AGCTGAATCCAGATGCCAGT	Forward primer for 2 nd round amplification of NQ02 off-target #4
DE-2726	GTGAAACTGGGTTTGCCCCT	Reverse primer for 2 nd round amplification of NQ02 off-target #4
DE-2727	GCCCCATTTTCCTGATGGGA	Forward primer for 1 st round amplification of NQ02 off-target #5
DE-2728	ATCCAGAGTGGTTCCATGCG	Reverse primer for 1 st round amplification of NQ02 off-target #5
DE-2787	CTTGGTGCTGTTGCACTCAT	Forward primer for 1 st round amplification of NQ02 off-target #6
DE-2788	GCCGCCTACTCCTCTTTTCTT	Reverse primer for 1 st round amplification of NQ02 off-target #6
DE-2729	GTGCCTCCTCAATGGTGACT	Forward primer for 2 nd round amplification of NQ02 off-target #6
DE-2730	GGTCAGGTGTGAGGGACTCT	Reverse primer for 2 nd round amplification of NQ02 off-target #6
DE-2835	ACACTCTTTCCTACACGACGCTC TTCCGATCTNNNNccaaggccTCTAT GCACACCAGG	Forward primer to amplify NQ02 54 for illumina sequencing
DE-2836	ACACTCTTTCCTACACGACGCTC TTCCGATCTNNNNcattaaggTCTATG CACACCAGG	Forward primer to amplify NQ02 54 for illumina sequencing
DE-2837	ACACTCTTTCCTACACGACGCTC TTCCGATCTNNNNataacgaaTCTATG CACACCAGG	Forward primer to amplify NQ02 54 for illumina sequencing
DE-2843	CGGTCTCGGCATTCCTGCTGAACC GCTCTTCCGATCTGGCTCAAGGTT CATGGC	Reverse primer to amplify NQ02 54 for illumina sequencing
DE-2844	ACACTCTTTCCTACACGACGCTC TTCCGATCTNNNNagttaaccGAGTGC CCCAGTCCC	Forward primer to amplify TSC1 2125 for illumina sequencing

DE-2845	ACACTCTTTCCCTACACGACGCTC TTCCGATCTNNNNcaggcttaGAGTG CCCCAGTCCC	Forward primer to amplify TSC1 2125 for illumina sequencing
DE-2846	ACACTCTTTCCCTACACGACGCTC TTCCGATCTNNNNtggcggctGAGTG CCCCAGTCCC	Forward primer to amplify TSC1 2125 for illumina sequencing
DE-2847	CGGTCTCGGCATTCCTGCTGAACC GCTCTTCCGATCTTGCCAAAGACA GCCC	Reverse primer to amplify TSC1 2125 for illumina sequencing
DE-2848	ACACTCTTTCCCTACACGACGCTC TTCCGATCTNNNNcttctctgTTCGTA CCGTGACAG	Forward primer to amplify TSC1 5054 for illumina sequencing
DE-2849	ACACTCTTTCCCTACACGACGCTC TTCCGATCTNNNNgtactcgTTCGTA CCGTGACAG	Forward primer to amplify TSC1 5054 for illumina sequencing
DE-2850	ACACTCTTTCCCTACACGACGCTC TTCCGATCTNNNNgctcattgTTCGTA CCGTGACAG	Forward primer to amplify TSC1 5054 for illumina sequencing
DE-2851	CGGTCTCGGCATTCCTGCTGAACC GCTCTTCCGATCTTTGGCAGTGGC AGAG	Reverse primer to amplify TSC1 5054 for illumina sequencing
DE-2852	ACACTCTTTCCCTACACGACGCTC TTCCGATCTNNNNgacactcgCGCTG CTGGAGGCGC	Forward primer to amplify RARA 233 for illumina sequencing
DE-2853	ACACTCTTTCCCTACACGACGCTC TTCCGATCTNNNNatccactaCGCTGC TGGAGGCGC	Forward primer to amplify RARA 233 for illumina sequencing
DE-2854	ACACTCTTTCCCTACACGACGCTC TTCCGATCTNNNNttatacagCGCTGC TGGAGGCGC	Forward primer to amplify RARA 233 for illumina sequencing
DE-2855	CGGTCTCGGCATTCCTGCTGAACC GCTCTTCCGATCTGGGCCAGGTGT CGGG	Reverse primer to amplify RARA 233 for illumina sequencing

JASON M. WOLFS

22112 Gibson rd, Mount Brydges,
Ont, Canada. N0L 1W0
519-319-4760
jwolfs@uwo.ca

Education:

PhD Candidate, Biochemistry, 2011-current
University of Western Ontario, London, Ontario

Honors Bachelor of Medical Science with Distinction, Biochemistry of infection and immunity, 2007-11
University of Western Ontario, London, Ontario

Publications:

Wolfs. J.M., Hamilton, T.A., Lant, J.T, Laforet, M., Zhang, J., Salemi, L., Gloor, G.B., Schild-Poulter,C. and Edgell,D.R. (2016) Biasing genome-editing events towards precise length deletions with RNA-guided TevCas9 dual nuclease. *PNAS* (pii: 201616343. [Epub ahead of print])

Wolfs. J.M., DaSilva M., Meister S.E., Wang X., Schild-Poulter C., Edgell, D.R. (2014) MegaTevs: Single chain dual nucleases for efficient gene disruption. *Nucleic Acids Res.*, **42**, 8816–8829.

Kleinstiver, B. P., Wang, L., **Wolfs. J.M.**, Kolaczyk, T., McDowell, B., Wang, X., Schild-Poulter,C. and Edgell,D.R. (2014) The I-TevI Nuclease and Linker Domains Contribute to the Specificity of Monomeric TALENs. *G3 (Bethesda)* **4**, 1155-1165.

Wolfs. J.M., Kleinstiver, B.P., Edgell, D.R. (2014) Rapid screening of endonuclease target site preference using a modified bacterial two-plasmid selection. *Methods Mol. Biol.* **1123**: 97-104

Kleinstiver, B.P., **Wolfs. J.M.**, Edgell, D.R. (2013) The monomeric GIY-YIG homing endonuclease I-BmoI uses a molecular anchor and a flexible tether to sequentially nick DNA. *Nucleic Acids Res.* **41**: 5413-5427

Kleinstiver, B.P., **Wolfs. J.M.**, Kolaczyk, T., Roberts, A.K., Hu, S.X., Edgell, D.R. (2012) Monomeric site-specific nucleases for genome editing. *PNAS* **109**: 8061-8066

Presentations:

Wolfs. J.M., DaSilva M., Meister S.E., Wang X., Schild-Poulter C., Edgell, D.R. (March, 2014, London, Ontario). MegaTevs: Single chain dual nucleases for efficient gene disruption. FASEB conference in Nassau, Bahamas. (abstract/poster)

Wolfs, J.M., DaSilva M., Meister S.E., Wang X., Schild-Poulter C., Edgell, D.R. (March, 2014, London, Ontario). MegaTevs: Single chain dual nucleases for efficient gene disruption. London Health Research Day. (abstract/poster)

Wolfs, J.M.*, Kleinstiver, B.P., DaSilva M., Edgell, D.R. (July, 2013, University of McMaster, Hamilton, Ontario). MegaTev for target gene disruption. Recombination, replication, repair, and chromosomal segregation conference. (Invited speaker)

Kleinstiver, B.P.*, **Wolfs, J.M.**, Roberts, A.K., Hu, S.X., Edgell, D.R (September, 2012, Lucca, Italy) Monomeric nucleases for genome engineering. Genome engineering: research and applications. (Work presented)

Wolfs, J.M.*, Kleinstiver, B.P., Edgell, D.R. (July, 2012, University of Toronto, Mississauga campus, Toronto, Ontario). Novel genome editing reagents: I-TevI/I-OnuI fusions. Recombination, replication, repair, and chromosomal segregation conference. (Invited speaker)

Wolfs, J.M.*, Kleinstiver, B.P., Kolaczyk, T., Edgell, D.R. (March, 2012, London, Ontario). Novel genome engineering reagents: creating chimeric endonucleases. London Health Research Day. (abstract/poster)

Kleinstiver, B.P.*, **Wolfs, J.M.**, Kolaczyk, T., Roberts, A.K., Hu, S.X., Edgell, D.R. (March, 2012, London, Ontario). Developing precise enzymes for genome editing applications. London Health Research Day. (abstract/poster)

Kleinstiver, B.P.*, **Wolfs, J.M.**, Roberts, A.K., Hu, S.X., Edgell, D.R. (November, 2011, University of Washington Medicine, Seattle, Washington). A monomeric GIY-YIG nuclease domain for genome engineering applications. Workshop on Genome Engineering. (abstract/poster)

Kleinstiver, B.P.*, **Wolfs, J.M.**, Edgell, D.R. (July, 2011, University of Western Ontario, London, Ontario). GIY-YIG homing endonucleases function as monomers. Recombination, replication, repair, and chromosomal segregation conference. (Work presented)

Awards and Honors:

<u>Award</u>	<u>Value (CDN)</u>	<u>Level</u>	<u>Location of Tenure</u>	<u>Period Held</u>
FASEB 2 nd place poster award	500	International	University of Western Ontario	2014
NSERC PGS Doctoral	63,000	National	University of Western Ontario	2014/01-2016/12
Ontario Graduate Scholarship	10,000	Provincial	University of Western Ontario	2013/05-2013/12
QEII Graduate Scholarship in Science and Technology (QEII GSST)	15,000	Institutional	University of Western Ontario	2012/05-2013/05
Western Research Graduate Scholarship	(tuition)	Institutional	University of Western Ontario	2011/08-2013/08
Schulich Graduate Scholarship	-	Institutional	University of Western Ontario	2011/08-2012/04

Biochemistry Undergraduate Summer Research Program	6,000	Institutional	University of Western Ontario	2010
Dean's Honour List	-	Institutional	University of Western Ontario	2008/09-2011/05

Experience:

Committees

Area Safety Committee, Biochemistry department, 2012-14
 Proctoring Committee, Biochemistry department, 2011-16
 Social Committee, Biochemistry department, 2011-14
 Society of Graduate Students, Biochemistry representative, 2012-13

Laboratory

Volunteer Research Program, University of Western Ontario, Department of Biochemistry 2009-10 (Dr. Ball)
 Biochemistry Undergraduate Summer Research Program (BUSRP), University of Western Ontario, Department of Biochemistry 2010 (Dr. Ball)
 Summer research program, University of Western Ontario, Department of Biochemistry 2011 (Dr. Edgell)
 Honours Research Thesis, University of Western Ontario, Department of Biochemistry 2010-11 (Dr. Edgell)

Teaching

2013 Biochemistry 2280A, University of Western Ontario
 2014 Biochemistry 2280A, University of Western Ontario
 2014 Biochemistry 3380G, University of Western Ontario
 2015 Biochemistry 2280A, University of Western Ontario
 2015 Biochemistry 3380G, University of Western Ontario