2017-11-30

Development of Chimeric Cas9 Nucleases for Accurate and Flexible Genome Editing

Mehmet F. Bolukbasi
University of Massachusetts Medical School

DEVELOPMENT OF CHIMERIC CAS9 NUCLEASES FOR ACCURATE AND FLEXIBLE GENOME EDITING

A Dissertation Presented

By

MEHMET FATIH BOLUKBASI

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

November 30th, 2017

Biochemistry and Molecular Pharmacology

DEVELOPMENT OF CHIMERIC CAS9 NUCLEASES FOR ACCURATE AND FLEXIBLE GENOME EDITING

A Dissertation Presented

By

MEHMET FATIH BOLUKBASI

This work was undertaken in the Graduate School of Biomedical Sciences

Biochemistry and Molecular Pharmacology

Under the mentorship of

Scot Wolfe, Ph.D., Thesis Advisor

Paul Kaufman, Ph.D., Member of Committee

Craig Peterson, Ph.D., Member of Committee

Erik Sontheimer, Ph.D., Member of Committee

Charles Gersbach, Ph.D., External Member of Committee

Job Dekker, Ph.D., Chair of Committee

Anthony Carruthers, Ph.D.,

Dean of the Graduate School of Biomedical Sciences

November 30th, 2017

DEDICATION

To my family for their endless support:

My father and mother for their love and devotion,

My brother for his wisdom and guidance,

My parents-in-law for their encouragement,

And

My wife, for her love and being a life-long friend to share the joy and the sorrow.

ACKNOWLEDGEMENTS

There are many people that I would like to thank for having a significant share in my graduate school period. First of all, I would like to thank Dr. Scot Wolfe for mentoring my doctoral work over the years. Scot has always been there at times when I needed guidance. I particularly appreciate the freedom he provided me to follow new research directions, and the space he offered for my personal development. In addition of being a caring mentor and a great role model, Scot is a brilliant scientist and a hardworking colleague. I always feel lucky to be mentored by him.

I also would like to thank the members of my thesis research advisory and dissertation examination committees: Dr. Job Dekker, Dr. Craig Peterson, Dr. Paul Kaufman, and Dr. Erik Sontheimer, and Dr. Charles Gersbach. They have created the time to listen my work, research interests and provided valuable feedback, constructive criticism. I would like give another thank to Dr. Gersbach for taking his time for a trip to Worcester to join my thesis defense.

I would like to extend my acknowledgments to the all past and current members of Wolfe Lab. They created a great working environment to get things done while having fun. Particularly, I would like to thank to Ankit for his guidance during the early years of my work and to Pengpeng for all his contributions, which are critical to the success of our research. Beyond Wolfe Lab, I would like to thank all the people I collaborated with: Sarah and Mike from Brodsky Lab, Nadia and Erik from Sontheimer Lab, Chun-Qing and Wen from Xue Lab, Alan and Manuel from Garber Lab, and Julie Zhu.

Outside of the work, I would like to thank my friends in UMASS and in the USA, whom I felt like in a family with: Arda, Aylin, Cem, Fatma, Özgün, Özkan, Rasim, Seda, Yeliz, and Zeynep.

Last but not least, I would like to thank my family: My parents Yusuf and Şükran for their love, care and constant support, my brother Ahmet, who has been my life-long mentor and a friend, and my parents-in-law İbrahim and Serpil for their encouragement and support. Finally I would like to thank my wife Özge, who has always been there for me with love and care, even when we were on separate continents. I probably could not do this without the support of my family.

ABSTRACT

There has been tremendous amount of effort focused on the development and improvement of genome editing applications over the decades. Particularly, the development of programmable nucleases has revolutionized genome editing with regards to their improvements in mutagenesis efficacy and targeting feasibility. Programmable nucleases are competent for a variety of genome editing applications. There is growing interest in employing the programmable nucleases in therapeutic genome editing applications, such as correcting mutations in genetic disorders.

Type II CRISPR-Cas9 bacterial adaptive immunity systems have recently been engineered as RNA-guided programmable nucleases. Native CRISPR-Cas9 nucleases have two stages of sequence-specific target DNA recognition prior to cleavage: the intrinsic binding of the Cas9 nuclease to a short DNA element (the PAM) followed by testing target site complementarity with the programmable guide RNA. The ease of reprogramming CRISPR-Cas9 nucleases for new target sequences makes them favorable genome editing platform for many applications including gene therapy. However, wild-type Cas9 nucleases have limitations: (i) The PAM element requirement restricts the targeting range of Cas9; (ii) despite the presence of two stages of target recognition, wild-type Cas9 can cleave DNA at unintended sites, which is not desired for therapeutic purposes; and (iii) there is a lack of control over the mutagenic editing product that is procuded.

In this study, we developed and characterized chimeric Cas9 platforms to provide solutions to these limitations. In these platforms, the DNA-binding affinity of Cas9 protein from *S. pyogenes* is attenuated such that the target site binding is dependent on a fused

programmable DNA-targeting-unit that recognizes a neighboring DNA-sequence. This modification extends the range of usable PAM elements and substantially improves the targeting specify of wild type Cas9. Furthermore, one of the featured chimeric Cas9 variants developed in this study has both robust nuclease activity and ability to generate predictable uniform editing products. These superior properties of the chimeric Cas9 platforms make them favorable for various genome editing applications and bring programmable nucleases one step closer to therapeutic applications.

PREFACE

Parts of this dissertation appear in the publications listed below:

Bolukbasi, M. F., A. Gupta, S. Oikemus, A. G. Derr, M. Garber, M. H. Brodsky, L. J. Zhu, and S. A. Wolfe. 2015. 'DNA-binding-domain fusions enhance the targeting range and precision of Cas9', *Nat Methods*, 12: 1150-6.

Bolukbasi, M. F., A. Gupta, and S. A. Wolfe. 2016. 'Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery', *Nat Methods*, 13: 41-50.

Other publications from my graduate study that are not included in the thesis:

Yin, H., C. Q. Song, J. R. Dorkin, L. J. Zhu, Y. Li, Q. Wu, A. Park, J. Yang, S. Suresh, A. Bizhanova, A. Gupta, **M. F. Bolukbasi**, S. Walsh, R. L. Bogorad, G. Gao, Z. Weng, Y. Dong, V. Koteliansky, S. A. Wolfe, R. Langer, W. Xue, and D. G. Anderson. 2016. 'Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo', *Nat Biotechnol*, 34: 328-33.

Yin, H., C. Q. Song, S. Suresh, Q. Wu, S. Walsh, L. H. Rhym, E. Mintzer, **M. F. Bolukbasi**, L. J. Zhu, K. Kauffman, H. Mou, A. Oberholzer, J. Ding, S. Y. Kwan, R. L. Bogorad, T. Zatsepin, V. Koteliansky, S. A. Wolfe, W. Xue, R. Langer, and D. G. Anderson. 2017. 'Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing', *Nat Biotechnol*. Online on Nov 13th, 2017

LIST OF THIRD PARTY COPYRIGHTED MATERIAL

The below listed material is reprinted and adapted from the co-authored publications.

CHAPTER I is reprinted and adapted from the following co-authored publication:

Bolukbasi, M. F., A. Gupta, and S. A. Wolfe. 2016. 'Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery', *Nat Methods*, 13: 41-50.

License #: permission is not required for the authors.

CHAPTER II is reprinted and adapted from the following co-authored publication:

Bolukbasi, M. F., A. Gupta, S. Oikemus, A. G. Derr, M. Garber, M. H. Brodsky, L. J. Zhu, and S. A. Wolfe. 2015. 'DNA-binding-domain fusions enhance the targeting range and precision of Cas9', *Nat Methods*, 12: 1150-6.

License #: permission is not required for the authors.

TABLE OF CONTENTS

TABLE OF FIGURES

ABBREVIATIONS

DSB	Double Strand Break
NHEJ	Non-Homologous End Joining
HDR	Homology-Directed Repair
Indel	Insertions and Deletions
pDBD	Programmable DNA-Binding Domain
ZFP	Zinc Finger Protein
ZFN	Zinc Finger Nuclease
TALE	Transcription Activator-Like Effectors
TALEN	Transcription Activator-Like Effector Nuclease
CRISPR	Clustered Regularly Interspersed Short Palindromic Repeats
Cas	CRISPR-associated
PAM	Protospacer Adjacent Motif
PID	PAM-interaction Domain
crRNA	CRISPR RNA
tracrRNA	Trans-Activating crRNA
sgRNA	Single Guide RNA
tru-sgRNA	Truncated Single Guide RNA
SpCas9	*Streptococcus pyogenes* Cas9
NmCas9	*Neisseria meningiditis* Cas9
SaCas9	*Staphylococcus aureus* Cas9
GUIDE-seq	Genome-Wide, Unbiased Identification of DSBs Enabled by Sequencing

CHAPTER I: Introduction

1.1 Introduction on understanding the genetic code

The information defining the heritable characteristics of living beings is stored in their genetic material, their genome. For the majority of organisms, the genome is constituted of deoxyribonucleic acid (DNA). The stored genetic information within DNA is encoded as an alphabet with four letters, bases: Adenine, Thymine, Guanine, and Cytosine. Throughout time, cellular machineries have evolved to compile, execute, restore, and reproduce the encoded genetic information.

Understanding the functions of genes—the basic unit of heredity—in the context of physiology or disease is a fundamental scientific interest. Since Mendel's first pea hybridization experiments, geneticists have been working hard to deconvolute the meaning of the genetic code. Early research into gene function has relied upon forward genetic approaches, where a gene is associated with an observed phenotype caused by a random loss of function mutation. Technological advancements such as development of DNA sequencing and polymerase chain reaction (PCR) led to the rise of reverse genetics to decrypt the genetic code. In the genomics age, with the combination of information technologies and modern genetics it is possible to read the whole genome of various organisms (including humans) fast and accurately. Researchers can now access large datasets to discover the function of genes and other regulatory genomic regions, associate the variations and mutations with physiological processes and diseases. Two patients with a similar disease phenotype may possess different types of underlying mutations or gene

variations that alter how they respond to a conventional therapy. Accumulation of deeper datasets is shaping the future of medicine towards personalized treatment options.

1.2 Rise of genome engineering via programmable nucleases

As researchers have been deciphering the genetic code, scientific curiosity and the quests of geneticist began to shift into editing the genetic code. Gene targeting—the ancestor phrase of genome editing—allows sequence-specific alterations of a genomic locus via homologous recombination[1]. Such alterations are useful to study the function of a gene by generation of null mutants or insertion of reporters to regulate its expression spatially or temporally[2]. Beyond understanding gene function, gene targeting has the potential to revert mutations that are underlying genetic disorders. In complex genomes such as humans, homologous recombination is extremely rare event such that its frequency is less than one in a million somatic cells[1]. Presence of a double strand break at a particular genomic site stimulates the gene targeting for this site by orders of magnitude[3, 4]. Therefore, generation of intentional double strand breaks (DSB) at a defined DNA sequence became the interest genome engineering efforts.

In order to create a DSB at a defined site in genome, researchers use programmable nucleases. These nucleases are composed of a sequence-specific DNA targeting unit and a catalytic unit; the activity of the latter depends on success of former. There are four types of programmable nucleases: Meganucleases, zinc finger nucleases, TALENs, and CRISPR-Cas system[5]. The first three nucleases use protein-DNA interactions to achieve sequence-specific DNA targeting. Whereas, CRISPR-Cas system is an RNA-guided nuclease platform. The utility of genome editing via programmable nucleases is

revolutionizing biomedical research. Over the last two decade, there has been tremendous progress in the development of programmable nucleases and their creative applications. Programmable nucleases can be employed on research models (cell lines and model organisms) to study biological processes or disorders, economically important plants and livestock to improve the quality of products and choose desired traits, and therapeutic genome editing applications[6]. Among these various applications, therapeutic genome editing requires serious considerations, such as the efficacy, safety, and successful delivery of the nucleases[7, 8]. The center work of this thesis is to assess and improve the specificity (safety) of CRISPR-Cas9 nucleases for future therapeutic genome editing applications.

1.3 Mechanism of programmable nucleases-directed genome editing

In common to all programmable nucleases, the outcome of genome editing depends on cellular events. Once a nuclease docks on the genome and finds its target site, it generates a double-strand break. A genomic double strand break is a serious threat to cells' well-being. Therefore, cells respond to a DSB by triggering two primary types of DNA-repair pathways: Non-homologous end joining (NHEJ) and homology-directed repair (HDR)[9] (**Figure 1.1a)**. Mammalian cells predominantly respond to a DSB via canonical NHEJ. Depending on the conformation of the broken ends NHEJ machinery ligates the ends of broken DNA strands differently. For blunt ended DSBs, the ends are preferentially ligated precisely[10, 11]. Such ligation may restore the nuclease target site and leads to another round of nuclease cleavage and DSB formation. During the repair of DSBs with staggered ends, NHEJ machinery may insert or delete a few random bases at the site of a DSB (indels) prior to the ligation of the broken ends[10, 11]. Indel mutations may cause gene disruption due

to changes in the open reading frame. If two separate DSBs occur on the genome simultaneously, more complex genomic arrangements may occur: such as deletion or inversion of the intervening sequence if the breaks are on the same chromosome[12], or translocations if the breaks are on different chromosomes[13]. Homology-directed repair of a DSB requires a template DNA that has homology to the flanking sequences around the both ends of the broken DNA. The template DNA can either be another non-cut allele in the genome or an exogenously supplied donor, which can be double stranded or single stranded[8, 9, 14]. Double stranded donors utilize a Rad51-dependent homologous recombination pathway, whereas HDR with single stranded donors is Rad51 independent[15]. Both HDR pathways with exogenous donors lead to tailored site-specific alterations in the target genome including gene correction, insertion or deletion.

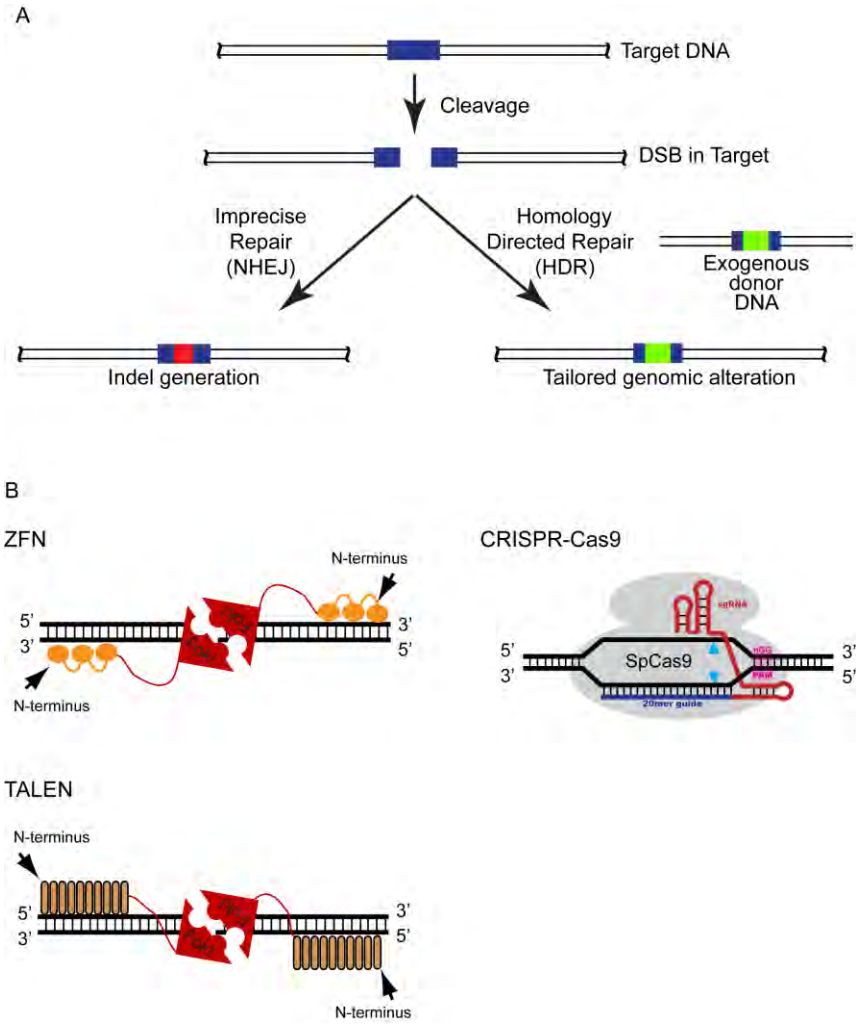


Figure 1. 1 Overview of programmable nuclease-directed genome editing
a) Schematic view of nuclease-directed genome editing. Once a programmed nuclease docks on the target site (blue rectangle), both strands of the DNA get cleaved. Typically, NHEJ or HDR machineries involve in the repair the DSB. NHEJ-mediated repairs may generate small indels at the target site. In the presence of a template DNA that has homology to the target site (such as exogenous donor DNA), HDR may occur. **b)** Schematic of three most commonly used programmable nuclease platforms: ZFNs, TALENs, CRISPR-Cas9. Both ZFNs and TALENs use programmable DNA-binding domains for sequence-specific DNA recognition. ZFNs dock on the antiparallel to the sequence; 5' end of the target sequence is with C-terminal of the protein and 3' end of the target sequence is with N-terminal of the protein. In contrast, TALENs dock on the DNA parallel to the sequence. Both ZFNs and TALENs use the nuclease domain of FokI. If two pairs ZFN and TALEN monomer binds on DNA in correct orientation and spacing, FokI domains of each monomer can dimerize and cleaves DNA. CRISPR-Cas9 is an RNA-guided nuclease platform. 5' end of the sgRNA is programmed for sequence-specific DNA recognition. Cas9 proteins have intrinsic DNA binding feature against a short motif downstream of the sgRNA target site.

1.4 Programmable DNA-binding domains

Both zinc finger nucleases and TALENs employ programmable DNA-binding domains (pDBD) to achieve sequence-specific DNA targeting. The pDBDs of zinc finger nucleases and TALENs are Cys2His2 zinc finger proteins and transcription activator-like effectors respectively.

1.4.1 Cys2His2 Zinc Finger Proteins (ZFP)

Cys2His2 ZFPs are the largest class of transcription factors found in metazoan genomes[16]. More than three decades ago, the first member of Cys2His2 ZFP family was identified as the tandem repeats of about 30 residues within the TFIIIA protein sequence from *Xenopus*[17]. As additional studies characterized more and more zinc finger protein sequences, their alignment reveal a consensus motif: (F/Y)-X-C-X_{2-5}-C-X_3-(F/Y)-X-X-X-X-X-Ψ-X-X-H-X_{3-5}-H, where X is any amino acid and Ψ is a hydrophobic residue[18]. Nuclear magnetic resonance structures reported that in the presence of zinc these residues fold into a ββα motif and coordinate zinc tetrahedrally[19, 20]. This zinc coordinated, compact, conserved ββα motif is referred as "finger". The crystal structure of three-finger Zif268 bound to DNA provided insight on its DNA recognition[21-23]. These studies featured Zif268 as model system to study ZFP-DNA interaction. ZFPs fit into the major groove of the target DNA in modular fashion meaning that the positioning of each finger on the DNA is similar. The docking of ZFP to the target DNA is anti-parallel such that the N-terminal finger docks along the 3' end and the C-terminal finger is with 5' end of the primary DNA strand that is recognized. The surface residues of the α-helix of each finger contact bases in the major groove of the target DNA. The amino acids at positions -1, 3, 6 of the α-helix of each finger

contact three consecutive bases on the same DNA strand (primary strand). The amino acid at position 2 of the α-helix may contact to a base on the complementary strand[21-23]. Therefore, in canonical form, each finger is considered have specificity against a triplet DNA sequence. Mutations in the DNA contact residues of the recognition helix can alter (reprogram) the finger specificity against a target triplet DNA sequence[24-27]. Over the last two decades, numerous studies aimed to select and identify novel ZFP-DNA binding patterns[28-38]. In a single finger format, all 64 possible triplet DNA sequences can be targeted[37]. However, in a ZFP array, addition of a consecutive finger does not always mean concatenation of the recognition sequence of the added finger due to context dependencies. The incompatibility of the residues at the finger-finger interface or recognition overlap of the adjacent fingers can account for the context dependency[36].

1.4.2 Transcription activator like (TAL) effectors

In nature, bacterial pathogen *Xanthomonas* secrete transcription activator like (TAL) effectors to infect plant cells. Once injected into the host cell, TAL effectors bind to promoters of target plant genes and activate their expression to facilitate the spreading of the infection[39, 40]. Typically, a TAL effector is composed of three major domains: the N-terminal domain contains signaling motif required for secretion, the central domain is a tandem array of multiple highly conserved repeat modules, and the C-terminal domain carries nuclear localization sequences (NLS) and transcriptional activation domains[41]. The central repeat modules account for sequence specific DNA recognition[42]. Each repeat module is 34 amino acid long, and its sequence composition is highly conserved except the positions 12^{th} and 13^{th}, which are termed the repeat variable diresidue (RVD)[43]. The

number of repeat modules and the composition of RVDs are distinct for each TAL effector. Computational and empirical approaches decrypted sequence-specific DNA recognition code of the TAL effectors. Each repeat module involves in the recognition of a single base pair whose specificity is determined by the composition of the RVD. These studies suggested the use of four primary RVDs for construction of novel TAL effectors for each base: NI for A, HD for C, NN for R (G or A), and NG for T[44, 45].

Structural studies indicated that a TAL effector dock on major grove of B form target DNA and wrap around in a right-handed superhelical architecture. In contrast to ZFPs, TAL effectors bind on the target DNA strand in parallel to the sequence such that N-terminal repeat module recognizes 5' end and C-terminal repeat module recognizes 3' end of the target DNA sequence. Each 34 amino acid long repeat module forms a two-helix bundle structure, where the RVD locate within the loop connecting the helices. In each repeat module, the 12th residue stabilizes the 13th to make the contact with the corresponding DNA base[46, 47]. In addition to the compatible RVD composition, TAL-effectors also require the presence of a 5' Thymine base (T) at the N0 position of a target DNA sequence[44, 45]. Crystal structure also confirmed the recognition of the N0 T by the N-terminal domain[46, 47]. Structure-guided directed evolution of the N-terminal domain eliminated requirement N0 Thymidine at the target DNA site[48]. Therefore, theoretically it is possible to design and construct a TAL-effector array against any given DNA sequence.

1.5 FokI endonuclease

FokI is a type IIs restriction endonuclease naturally found in *Flavobacterium okeanokoites* bacteria. FokI binds to 5´-GGATC-3´ sequence and cleaves the duplex DNA

at 9 and 13 nucleotides downstream of its recognition sequence by generating 4 nucleotides 5' overhang[49]. Proteolytic and structural studies revealed that FokI consist of two domains: N-terminal DNA-binding domain, and C-terminal non-sequence specific nuclease domain[50, 51]. Researchers harnessed the modular nature of FokI to substitute the DNA-binding domain with other natural DNA-binding domains[52-56]. Mutational analysis shows that FokI utilizes a single catalytic center to cleave a target DNA[57]. Although FokI is present as monomer in a solution or docked on a target DNA[58, 59], biochemical and structural studies reported that the nuclease domain of FokI requires homo-dimerization to cleave the substrate DNA[60, 61]. Either two FokI monomers bound on the same DNA molecule or—at very high concentrations—one DNA bound FokI monomer and another one in the solution must dimerize to produce a DSB[60, 61]. *In vitro* DNA cleavage experiments point out that on a target DNA molecule with two FokI site some distance from each other, two FokI monomers can dimerize by forming a loop architecture[62].

1.6 Zinc Finger Nucleases (ZFNs)

Zinc finger nucleases (ZFNs) are the chimera of a Cys2His2 zinc finger protein (ZFP), and the nuclease domain of the FokI type IIs restriction endonuclease. ZFNs function as a pair where each monomer of ZFN binds to 9-18 base pairs of DNA. When two ZFN monomers recognize nearby DNA sequence in correct orientation and spacing (**Figure 1.1b**), FokI from each ZFN can dimerize and cleave both strands of the DNA[63]. In a random genome of 3 billion base pairs, about a 16 base pair-long DNA sequence should be unique. However, site-specific targeting within a complex genome is a challenging task. The target site must be distinguished from a number of similar sequences—including

paralogues and pseudogenes—effectively and reproducibly. To reduce off-target cleaves of ZFNs, multiple approaches focused on different components of the ZFNs are described. One of the approaches is to select from a library of engineered of ZFPs with different DNA recognition residues for those better target site specificity[28-38]. For example, a recent study described a reporter system to distinctively select target but not off-target site specific ZFP arrays on therapeutically relevant genomic sites[64]. However, there is no easy formula to construct high fidelity ZFNs without site-specific selection. Increasing the number of fingers in a zinc finger array can bring more specificity but also create a mismatch between the helical periodicity of the ZFP and the DNA[18, 65]. When a ZFP docks on the target DNA, the inter finger linker generates rigid conformation for the two adjacent fingers to contact successive DNA triplets[21]. Zif268-templated engineered ZFPs usually contain evolutionary conserved canonical 5 amino acid (TG[E/Q]KP) linker[21, 66]. Mutations in the residues of the canonical linker reduce the DNA binding affinity of the ZFP[67, 68]. Addition of non-canonical linkers in between every two or three fingers brings the helical periodicities of the ZFP and the DNA to a match and allows construction of longer ZFPs[69-71].

Another level of targeting specificity can be provided to ZFNs by engineering of the FokI nuclease domain. Structural and biochemical studies revealed the dimer interface between two FokI monomers[60, 61]. Harnessing the structural information, three studies reported the generation of obligatory heterodimeric FokI variants. ZFNs constructed with these heterodimeric FokI variants displayed reduced the levels of off-target activity-associated toxicity in comparison to canonical ZFNs[72-74]. Another strategy is to use ZFNickases. This approach also uses obligatory FokI heterodimers, where one of the

monomers is catalytically inactive due to a point mutation (D450A)[75, 76]. ZFNickases cleave one of the strands (nick) at its target site and compatible with homology-directed repair. Although ZFNickases display lower activity, reduction of random indels at the target site can be considered as improved specificity. After all these years of engineering efforts, now it is possible to produce highly active and specific ZFNs. Yet, the construction of ZFNs still requires technical expertise.

1.6.1 Applications of ZFNs

Following the identification of the modular architecture of FokI, a stream of proof-of-principle studies reported the development of functional chimeric restriction endonucleases, where the nuclease domain of FokI is fused to naturally occurring DNA-binding domains[52-54]. One of these studies reported the first functional ZFN as the fusion of CP-QDR or Sp1-QNR zinc finger proteins to the nuclease domain of FokI[53]. In early 2000s, ZFNs were used in living cells. Carroll and Chandrasegaran laboratories applied ZFNs to induce homologous recombination in *Xenopus laevis* oocytes[77]. A number of studies followed this and reported ZFN-mediated genome editing in cell lines[78-80], plants[81, 82], and model organisms including fruit fly[83, 84], zebrafish[34], and mice[85] (reviewed in Urnov *et al*.[63]). These studies not only suggested the broad utility of programmable nucleases but also helped researchers to answer basic biological questions. Beyond the biomedical laboratory usage, ZFNs also edited the genomes of non-model systems for development of biotechnology applications in agriculture[81, 86] and livestock[87-89].

The goal of therapeutic genome editing existed since the early days of programmable nucleases. As researchers gained insight on how these nucleases edit

genomes and developed improved nuclease architectures, more and more applications are being reported. Any nuclease-based therapeutic genome editing strategy has to utilize cellular DSB repair pathways such as NHEJ or HDR[8]. Disruption of a therapeutically relevant gene is relatively easier since NHEJ is the predominant cellular response to a DSB, yet applicable disease contexts are limited. For example, individuals carrying naturally occurring 32bp deletion in HIV-1 entry co-receptor *CCR5* gene (*CCR5del32*) display resistance to HIV-1 infection with no serious health issues[90, 91]. Harnessing this phenomenon, ZFNs were used to inactivate the *CCR5* gene in $CD4^+$ T cells and $CD34^+$ hematopoietic stem and progenitor cells (HSPCs)[92, 93]. Clinical trials are ongoing to evaluate the therapeutic potential of the autologous transplantation of the nuclease treated cells against HIV-1 infection[94]. In addition to offering treatment for infectious diseases, nuclease based genome editing should be useful for monogenic disorders. In sickle cell disease (SCD), the underlying cause is a point mutation in the adult β-globin gene, which causes the polymerization of hemoglobin in erythrocytes with a sickled conformation[95]. One of the proposed therapeutic approaches is to induce the expression of fetal γ-globin gene in red blood cells[95, 96]. ZFN-directed disruption of the erythroid lineage specific enhancer element in *BCL11A* gene[97, 98] in $CD34^+$ HSPCs induces the expression of fetal γ-globin gene in differentiated red blood cells[99].

Therapeutic genome editing via HDR requires an exogenous donor template that has homology to the nuclease target site. Nuclease directed HDR is compatible for both gene correction and gene replacement applications. These methods are more challenging than gene disruption strategies since HDR rates are typically lower than NHEJ, which can

introduce other undesired mutations at the target site and confound the outcome[8, 100]. An example of a gene correction approach is to revert the β-globin point mutation that causes SCD. In $CD34^+$ HSPCs, ZFN induced DSBs in the presence of donor DNA corrected the mutation within fraction of the treated population[101]. Such direct gene correction is not feasible for wide range of monogenic diseases, where there can be multiple types of the underlying mutations and each of them require customized nuclease and donor construction. Gene replacement approaches provide an alternate solution to this problem. This strategy aims to integrate a repair cassette containing the corrected open reading frame of the gene of interest (cDNA) into either the non-coding region of the same gene (e.g. introns) or a safe harbor locus in human genome such as *AAVS1* intron 1[102, 103]. In *ex vivo* therapeutics setting, ZFN driven gene replacement strategy has been applied on patient isolated $CD34^+$ HSPCs for two X-linked diseases: chronic granulomatous disease (X-CGD)[104] and severe combined immunodeficiency (SCID-X1)[105, 106]. AAV6 (adeno associated virus serotype 6) donor template carrying functional cDNA of *CYBB* gene with homology arms for *AAVS1* site restored the genetic deficiency in X-CGD[104]. For a prospective SCID-X1 treatment, ZFNs drove site-specific integration of IDLV (integration deficient lentivirus) donors carrying either full length *IL2RG* cDNA with homology arms against either *AAVS1* or partial *IL2RG* cDNA with homology arms against the intron of the endogenous gene[105, 106]. In common to all of studies editing $CD34^+$ HSPCs, the rate of HDR is lower for long term progenitors—hematopoietic stem cells—than the bulk population. This is possibly due to the quiescent nature of these cells. In addition to these *ex vi*vo approaches, ZFNs have also been used for *in vivo* gene replacement therapeutics. In

hemophilia B, the blood coagulation factor IX deficiency mutations are located in *F9* gene. In mouse hemophilia B model, tail vein injection of AAV vectors carrying the ZFN expression cassettes and the donor (promoterless functional factor IX cDNA) leads donor cDNA incorporation into *F9* or *Albumin* locus in the liver cells[107-109].

1.7 Transcription activator-like effector nucleases (TALENs)

Similar to ZFNs, transcription activator-like effector nucleases (TALENs) are also a chimera of a programmable DNA binding domain (TAL effectors) and FokI nuclease domain (**Figure 1.1b**). Shortly after the decryption of sequence specific DNA recognition via TAL-effectors[44, 45], three studies reported the development of TALENs[55, 56, 110]. Importantly, these studies identified critical functional parameters such as required domain architectures and optimal spacing between each TALEN monomer target site[55, 110]. In contrast to its simplicity of reprogramming, the repetitive nature of TAL effectors presents challenges with regards to the construction of the designer TALENs. Most commonly a golden gate cloning system, which takes advantage of type IIs restriction enzymes, is used to build TALENs within few days[111, 112]. Another approach is to stitch together modules with longer homology arms in ligation independent manner[113]. For higher throughput assemblies, solid support-based module ligation generates functional TALENs within a day[114]. Repetitive TALEN sequences impact the stabilities of the lentiviral expression constructs of these nucleases[115]. Re-coding of the TAL effectors to minimize sequence repetition provided a solution to this problem[116]. Typically, designer TALENs use 15-20 base pair DNA recognition provide sufficient site specificity. However, several studies reported that TALENs display off-target activity in human genome at relatively low

levels[117-121]. Altering the RVD compositions or reducing the excess positive charge in non-sequence specific DNA contacts provide improved specificity and on target activity of TALENs[120, 121]. Another approach is to truncate the C-terminal domain to restrict spacing requirement between two monomer binding sites[110, 121, 122]. Beyond the engineering of the DNA binding domain, usage of the heterodimeric FokI architectures can provide an additional level of targeting accuracy[72-74, 120]. Overall, TALENs can provide accurate and efficient targeted genome editing with reduced designing efforts relative to ZFNs.

1.7.1 Applications of TALENs

Over many years, applications of ZFNs determined the ground work for development and application of other programmable nucleases. Because of this existing knowledge and the simplicity of the TAL effector programming, many researchers interested in developing and applying TALENs for various biomedical questions. Immediately after deciphering the TAL effector DNA recognition code[44, 45], number of studies reported TALEN-directed successful genome editing in cell lines[110, 117] and organisms including yeast[55, 56], worms[123], zebrafish[124, 125], rat[126], fruit fly[127], and more (reviewed by D. Carroll[6]). Some studies suggest that TALENs can edit genomes in more complex manner beyond small indels; such as complex chromosomal deletions, inversions and translocations[12, 13, 128]. High-throughput production of designer TALENs opened up novel applications to study gene function[129]. These include construction of TALEN libraries for a small set of human genes involve in cancer/epigenetic regulation or targeting 18740 protein-coding genes[114, 122]. In addition to protein-coding genes, a library targeting human miRNAs has also been reported[130]. In combination with induced pluripotent stem

cells (hIPSCs), TALENs can be used to develop cellular disease models[131-133]. Applications of TALENs extend beyond basic science. Currently, TALENs are often being used in agriculture and livestock to improve the quality of products and select desirable traits[89, 134-137].

Similar to ZFNs, therapeutic genome editing via TALENs also involve gene disruption or correction or replacement strategies. TALENs can disrupt genes to provide protection against infectious diseases; such as targeting *CCR5* gene for HIV-1 infection resistance or depletion of hepatitis B virus from human genome[110, 138-140]. In monogenic disorders context, TALENs have also been used to perturb the erythroid-lineage specific enhancer element in *BCL11A* gene[95-98] within $CD34^+$ HSPCs[141]. TALEN induced gene correction reverted the disease-causing mutations in Douchenne muscular dystrophy[142], epidermolyisis bullosa[118], and sickle cell anemia[143]. Finally, TALENs have been utilized to incorporate gene replacement donors for β-Thalassemia[144] and X-linked chronic granulomatous disease[145]. Beyond the utility of programmable nucleases, TAL-effectors have been fused with other effector domains to generate site specific transcription regulation, and epigenetic modifications[146-150].

1.8 Clustered Regularly Interspersed Short Palindromic Repeats (CRISPR) and CRISPR-associated (Cas) systems

CRISPR-Cas systems provide RNA-guided adaptive immunity in bacteria and archaea against infectious genetic material (*e.g*. phages, plasmids *etc*.). To date, two classes and six types of CRISPR-Cas systems have been identified. Classification of CRISPR-Cas systems is done according to the organization of the CRISPR loci and the presence and

composition of Cas (CRISPR-associated) proteins[151, 152]. The CRISPR locus is an array of multiple short repetitive sequences interspersed by non-repetitive spacer sequences in a repeat-spacer-repeat format. The sequence of a spacer corresponds to the genetic material of a species infectious to the host organism[153-155]. Therefore, the CRISPR loci represent the immunization record of the host to previously encountered infectious nucleic acids. The composition of Cas genes varies among different CRISPR-Cas system. In *class 1* systems (types I, III and IV), multiple proteins are employed to form an effector complex, whereas in *class 2* systems (types II, V, and VI), a single multidomain effector protein is sufficient for immune response[151, 152, 156, 157]. Even though *class 1* systems account for vast majority of the all known CRISPR systems[157], the simplicity of *class 2* systems makes them favorable for development of biotechnology applications.

In nature, CRISPR-Cas mediated adaptive immunity occurs in three stages: adaptation (acquisition), expression and processing, and interference[158]. In the adaptation stage, Cas1-Cas2 protein complex incorporates new pieces of infectious foreign genetic material (protospacer) into the CRISPR loci[159-161]. The presence of 2 to 4 nucleotide long protospacer adjacent motif (PAM) is crucial in protospacer acquisition to distinguish self versus foreign genetic material[162-164]. The next stage is expression and processing of the CRISPR RNA (crRNA) from the transcribed locus. The CRISPR locus of a host genome, which contains several spacers in between the repeat sequences, is typically transcribed as a long RNA molecule, named precursory CRISPR RNA (pre-crRNA). In order to generate an active crRNA-Cas effector ribonucleoprotein (RNP) complex, the pre-crRNA needs to be processed into a mature crRNA form[156, 165]. The crRNA processing in *class 2* systems

can either be done by the cognate Cas effector[166, 167], or require the involvement of trans-activating crRNA (tracrRNA) and RNase III[168]. In the final interference stage, functional crRNA loaded Cas effectors locate and cleave the target nucleic acid site-specifically[155, 169-171].

Class 2 CRISPR-Cas systems have been repurposed for various biotechnological applications[5, 172]. Particularly, the type II (Cas9) and type V (Cas12-a, Cpf1) effectors are revolutionizing the genome editing applications[173, 174]. Recently, type VI effector (Cas13-a, C2c2) has been described as an RNA-guided RNA targeting effector that can knockdown or edit the target RNA[175, 176]. It is reasonable to anticipate the development of novel applications as new Cas effectors are identified. The next section focuses on the Cas9 effectors (nucleases) in the context of defining and improving their DNA targeting specificity.

1.9 Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery

The simplicity of site-specific genome targeting by type II clustered, regularly interspaced, short palindromic repeat (CRISPR)-Cas9 nucleases, along with their robust activity profile, has changed the landscape of genome editing. These favorable properties have made the CRISPR-Cas9 system the technology of choice for sequence-specific modifications in vertebrate systems. For many applications, whether the focus is on basic science investigations or therapeutic efficacy, activity and precision are important considerations when one is choosing a nuclease platform, target site and delivery method. Here we review recent methods for increasing the activity and accuracy of Cas9 and assessing the extent of off-target cleavage events.

1.9.1 Overview on CRISPR-Cas9 systems and their applications

Type II CRISPR-Cas9 systems are found in only a small fraction of bacterial species[152, 177]. These adaptive defense systems employ a single, large multisubunit endonuclease (Cas9) and a pair of RNAs that as a complex facilitate sequence-specific target cleavage of foreign DNA[158]. In 2012, the critical components required to program the *Streptococcus pyogenes* ribonucleoprotein complex to cleave a specific DNA sequence were defined and streamlined by engineering a single guide RNA (sgRNA) that could subsume all the RNA-based structural and recognition functions[170]. Based on lessons from other artificial nuclease platforms[43, 63], the insights provided by this *in vitro* study were rapidly translated in early 2013 to Cas9-sgRNA expression and delivery systems that facilitated sequence-specific genome editing in mammalian cell lines[173, 178-180] and model organisms[181, 182]. These and other investigations probed and improved aspects of nuclear localization, and sgRNA design and expression[173, 178, 180, 183] that rapidly culminated in a reliable *S. pyogenes* Cas9-based nuclease system that is revolutionizing biological studies and holds promise for therapeutic applications[7, 184, 185].

The Cas9-sgRNA nuclease system (**Figure 1.1b** and **Figure 1.2**) is more straightforward to program than prior artificial nuclease platforms because sequence-specific targeting resides primarily within the associated sgRNA, which is simple to recode. This attribute has permitted a proliferation of Cas9-based genome editing applications that harness the generation of a site-specific double-strand break (DSB) for targeted gene inactivation or modification. The ease of expressing many distinct sgRNAs within a cell (sgRNA multiplexing) facilitates the simultaneous targeting of multiple

genes[173, 181, 186-191] or the creation of larger genomic alterations between pairs of DSBs, such as segmental deletions[173, 192, 193], inversions[193, 194], and translocations[195-198]. Cell lines or organisms can be "preloaded" with Cas9[199-203] such that editing requires only the delivery of the desired sgRNAs. Lentiviral delivery of libraries of sgRNAs into Cas9 expressing cells permits genome-wide gene-inactivation screens that can identify participating genes in a cellular process of interest[199-202]. Similarly, a Cre-dependent Cas9 knock-in mouse model combined with viral-mediated sgRNA and Cre delivery facilitates tissue or cell-type specific gene modification[203], such as the analysis of regulatory networks in primary immune cells[204].

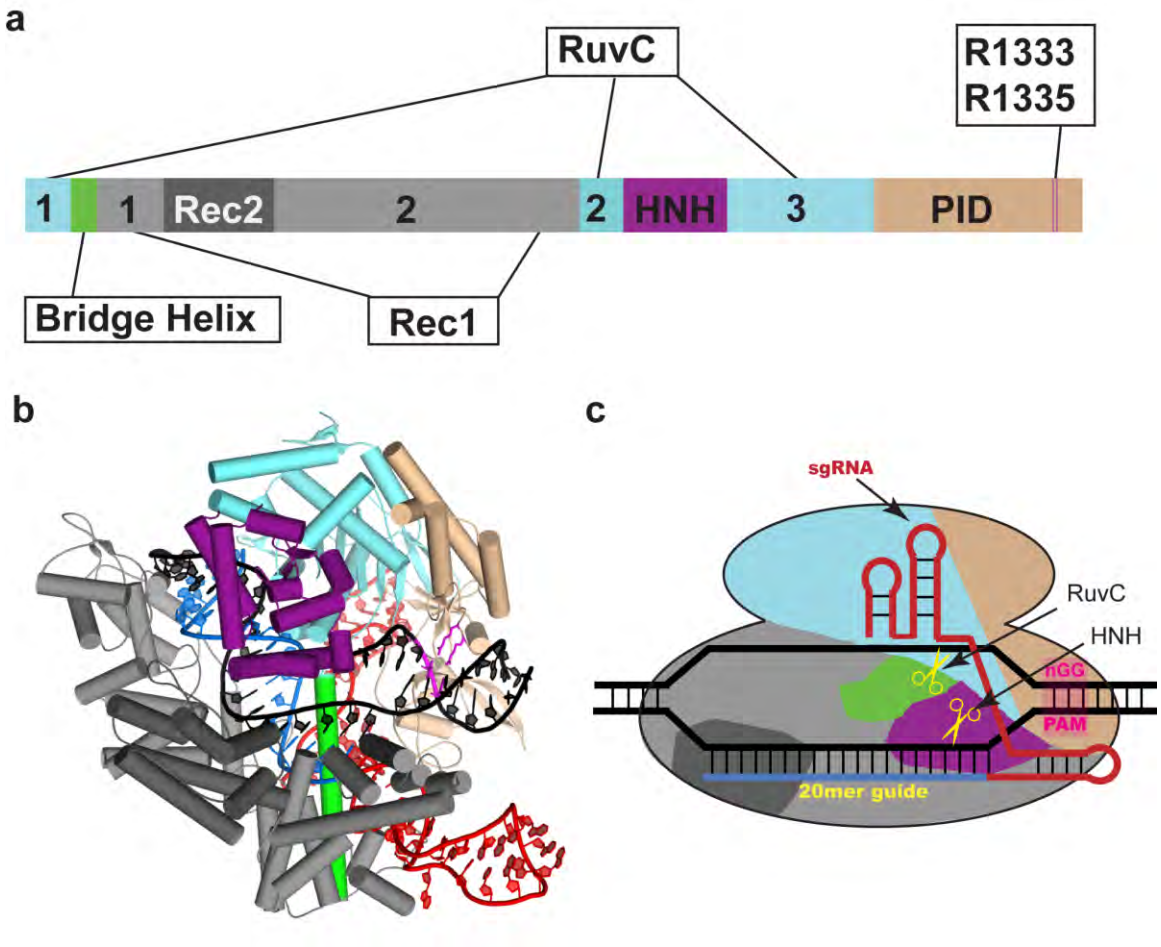


Figure 1. 2 Structural overview of SpCas9

a) Schematic view of the domain organization of SpCas9, where domains that are formed from discontinuous sequence elements are numbered. **b**) X-ray crystal structure of SpCas9 with sgRNA (blue - guide and red – constant sequence) and the target DNA (black)[205]. In this structure, sgRNA:DNA heteroduplex divides the Cas9 protein into two lobes connected by the bridge helix (green): the recognition lobe consist of the Rec1 (light grey) and Rec2 (dark grey) domains; and the nuclease lobe with RuvC (cyan), HNH (purple), and the PAM-interacting domain (wheat). Two Arginines (R1333 and R1335; magenta) within the PID make the primary contacts to the guanines within the PAM. **c**) Schematic view of the Cas9-sgRNA complex, upon R-loop formation cleavage of the DNA occurs 3 bp from the edge of the PAM element.

The Cas9-sgRNA system has proven to be similarly effective in promoting the targeted insertion of a desired DNA sequence through homology-directed repair (HDR) utilizing either exogenously supplied duplex DNA templates[173, 178, 192, 206], single stranded oligonucleotides[116, 181, 192, 207-213] or viral encoded templates[203, 214-217]. The efficiency of HDR events is cell-type dependent and typically occurs at a lower frequency than the rate of insertions or deletions (indels) created via non-homologous end joining (NHEJ)-mediated imprecise repair of the Cas9-generated DSB[218]. However, HDR rates can be increased in some cell types through cell cycle synchronization[209] or small molecules or proteins that interfere with alternate DNA repair pathways[219, 220]. In sum, Cas9-based technological advances have dramatically simplified the creation of modified vertebrate cell lines and animal models for the investigation of gene function during development and disease progression[194, 207, 221-232]. Similar revolution is being realized by the use of nuclease-dead Cas9 (dCas9) to deliver effector domains to perturb gene expression and chromatin modification states [233].

Therapeutic applications for CRISPR-Cas9 are also being actively explored. Cas9-mediated *ex vivo* gene editing in primary cells[189, 191, 208, 210, 213, 217] or induced pluripotent stem cells[145, 234, 235] provides a potential pathway for the creation of autologous cell-based therapies. *In vivo* correction of dystrophin gene mutations that are associated with Douchne muscular dystrophy holds therapeutic promise[236-238]. Another therapeutic level of Cas9-mediated *in vivo* gene correction was recently demonstrated in a mouse model of fumarylacetoacetate hydrolase (Fah) deficiency[207, 239]. In parallel, viral delivery systems, such as adeno-associated virus (AAV), for the tissue-specific delivery of Cas9-sgRNA are

being developed[240]. The limited cargo capacity of AAV (~4.7 kb)[241] creates challenges for packaging expression cassettes for *S. pyogenes* Cas9 (SpCas9, ~4.2 kb) and its sgRNA in a single vector[187]. More compact CRISPR-Cas9 systems—such as Cas9 orthologues from *Neisseria meningiditis* (NmCas9), *Staphylococcus aureus* (SaCas9), and *Campylobacter jejuni* (CjCas9)—have been developed[242-245] to overcome this problem. One such system, SaCas9[244], achieved efficient gene editing within the mouse liver when delivered via a single AAV construct. These proof-of-principle experiments highlight the promise of Type II Cas9 nucleases for therapeutic gene correction or gene replacement if compatible delivery systems can be realized and collateral damage to the genome can be minimized.

1.9.2 Licensing a DNA sequence for cleavage

Whether utilizing programmable nucleases for editing cell lines, model organisms or therapeutic applications, nuclease precision is an important consideration. Pioneering gene therapy efforts using gamma-retroviruses in hematopoietic stem cells were stymied by the insertional mutagenesis inherent with these gene delivery vectors, which led to oncogenic transformation in a subset of patients[246, 247]. Likewise, the generation of DNA breaks at unintended (off-target) sites by imprecise nucleases has the potential to alter gene expression and function through direct mutagenesis or the generation of genomic rearrangements. Understanding the series of events associated with the "licensing" of a DNA sequence for cleavage by the Cas9-sgRNA complex should permit the accurate prediction of potential cleavage sites and the development of Cas9-sgRNA variants with improved precision. A molecular and mechanistic description of target site recognition and cleavage has been provided by biochemical[170, 248-251] and structural studies[205, 252-255],

buttressed by nuclease activity assays in cell culture[173, 256-258] and data on the function of Cas9 in adaptive bacterial immunity[158]. Mechanistic analyses suggest that a DNA sequence is licensed for cleavage through two stages of recognition: 1) Cas9 binding to a favorable Protospacer Adjacent Motif (PAM) element, and 2) sgRNA-mediated interrogation of the neighboring DNA sequence via Watson-Crick base pairing to allosterically switch Cas9 into an active state (**Figure 1.3.a**).

Single molecule and bulk biochemical assays indicate that the residence time of a Cas9-sgRNA complex on a DNA sequence during target acquisition is dependent on the presence of a compatible PAM element[248]. This PAM-dependence provides kinetic control over the sampling of target sequences by restricting DNA-sgRNA heteroduplex (R-loop) formation[249]. This preference is observed in ChIP-seq experiments on dCas9, where the vast majority of occupied sequences contained an optimal PAM element[250, 259]. In the case of SpCas9, an NGG PAM element is strongly preferred[1704], although NAG and NGA PAMs are inefficiently recognized in some contexts[256, 260, 261].

sgRNA interrogation of the DNA sequence initiates next to the PAM element through Watson-Crick pairing in a 3′ to 5′ direction relative to the guide sequence (20 nucleotides [nt] for SpCas9)[248, 249]. Progressive R-loop formation is a passive process[248, 254, 262] in which mismatches within the seed region are more likely to abrogate nuclease activity[170, 171, 173, 248, 250, 259]. Data analysis from large-scale screens utilizing libraries of sgRNAs suggests that there is an optimal level of heteroduplex stability for efficient target cleavage[199, 263, 264] and that other sequence features in the guide[199, 258, 263-266] and neighboring the canonical PAM element[263, 264, 266] can influence nuclease efficiency. Using

these data, researchers have developed algorithms that predict Cas9-sgRNA activity as a function of these parameters[258, 264-267]. Chromatin architecture also influences the binding of Cas9-sgRNA complexes at suboptimal (non-cognate) sequences within the genome[250, 259], but the extent that it impacts nuclease activity remains uncertain[258, 268]. As more aspects of the mechanism of site licensing for cleavage by Cas9-sgRNA are defined, they should facilitate the creation of advanced predictive models describing sequence cleavage efficiency as a function of the guide sequence. These algorithms should enable the identification of efficient nuclease target sequences that have few compatible off-target sequences within the genome.

1.9.3 Cas9 promiscuity in cell lines, primary cells and model organisms

Precision is an enigma that has dogged the therapeutic application of all artificial nuclease platforms. There is abundant evidence that SpCas9-based nucleases in transformed cell lines can tolerate imperfections within the sgRNA-DNA heteroduplex[170, 173, 198, 256, 257, 261, 269, 270]. Analysis of the activity profiles of guides encoding one or more mismatches to a target site has shown that the number, position and type of base mismatches[256, 257, 261] can affect the activity level (for example, a G:U mismatch is better tolerated than a G:A mismatch)[256]. Heteroduplex sequence composition also influences tolerance to the number and position of the substitutions between the guide and target sequence[257, 261]. Reassuringly, deep sequencing of potential off-target sites (defined via computational prediction[198, 256, 257], ChIP-seq of dCas9 bound regions[250, 259] or *in vitro* library analysis[269]) within populations of nuclease-treated cells did not identify indels at the majority of these sequences. The absence of indels at most bound regions shown by

ChIP-seq [250, 259] indicates that nuclease activity is not merely a function of Cas9-sgRNA residence time. Notably a subset of off-target sites display nuclease activity[198, 250, 256, 257, 259, 269], with up to four or five mismatches between the guide and genomic sequence tolerated[257, 269]. This degree of potential promiscuity is a concern with respect to therapeutic applications, as thousands of sequences with four or five mismatches to a 20-nt guide sequence can typically be found in the human genome[198, 257]. In some instances Cas9-sgRNA complexes can also tolerate a 'bulge' between the guide and a noncognate sequence (a base flipped out of one component of the heteroduplex)[270], which further complicates the genome-wide prediction of nuclease activity as a function of guide sequence. These studies provided impetus for efforts to engineer Cas9-sgRNA systems with improved performance (described below).

Although studies of SpCas9-sgRNA-treated transformed cell lines have provided evidence of modest nuclease promiscuity, a more nuanced perception has emerged from the studies of Cas9-based editing in stem cells and model organisms. Sequence-capture characterization of a population of nuclease-treated $CD34^+$ cells at two to three dozen computationally predicted off-target sites detected significant indels at only a single off-target site for one of five active guides[189]. Importantly, stem cells facilitate the expansion of single nuclease-treated clones for direct comparison with the parental line via whole genome sequencing (WGS) in a nominally 'normal' cellular environment with regards to DNA repair. These comparative analyses revealed a large number of sequence alterations between the parental and nuclease-derived clones, but none could be ascribed to off-target cleavage by the nuclease[235, 271, 272]. Similar results were observed for exome sequencing in

a number of nuclease-treated clonal haploid cell lines[273]. These results, although promising must be interpreted cautiously, because WGS is most likely to detect only those active off-target sites that are cleaved with efficiency similar to that of the target site[273, 274]. Because lesions at most off-target sites are low-frequency events, the likelihood of a single clone containing an off-target lesion at a specific site is low. Similarly, analysis of Cas9-sgRNA-treated vertebrate embryos have occasionally identified a small number of active off-target events in the offspring of treated founder animals[192, 275-277]. These off-target sites typically have only one or two mismatches to the guide sequence, implying that they might have activity similar to that of the target site. The lower rate of Cas9-induced off-target lesions in stem cells and model organisms relative to transformed cell lines might be due in part to the higher propensity for imprecise DNA repair in transformed lines. In sum, these studies show that Cas9-edited stem cell lines or model organisms can be obtained without unwanted nuclease-based genomic modification, but the propensity for Cas9 off-target activity in transformed cell lines highlights the need for the evaluation of nuclease precision in therapeutic applications.

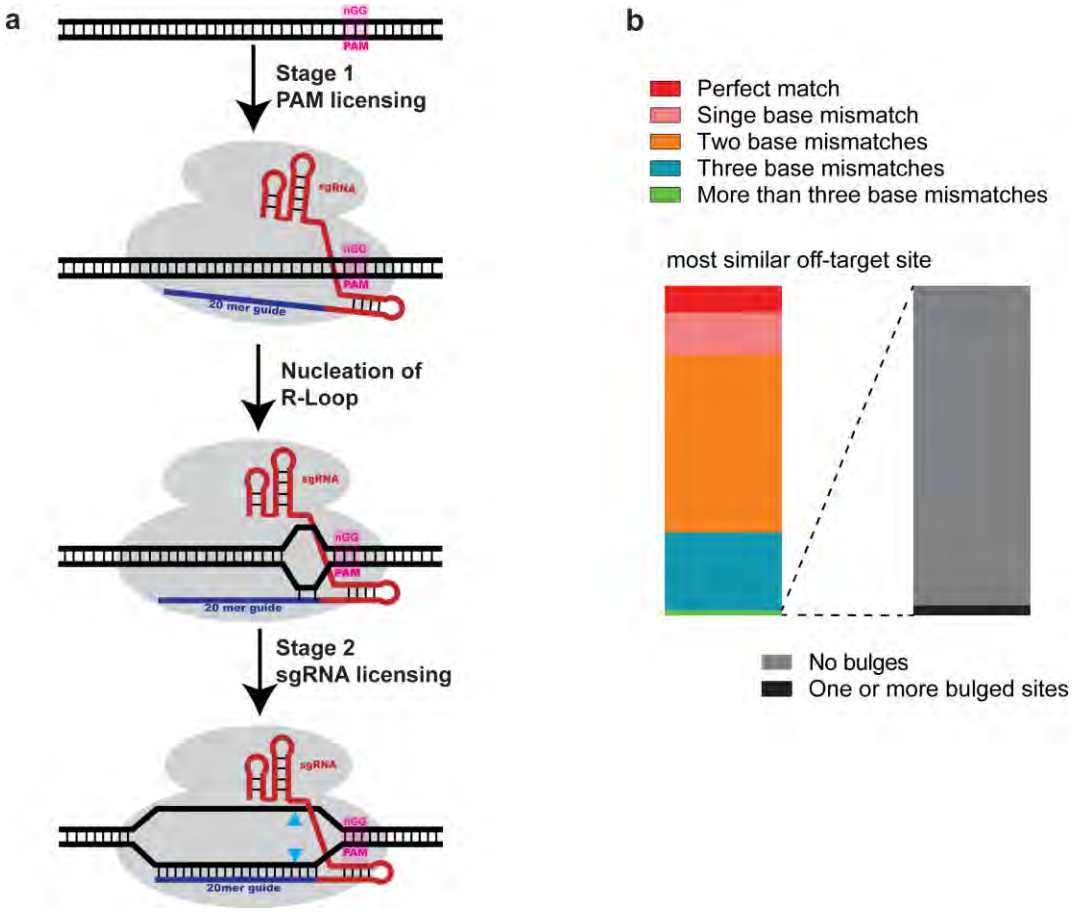


Figure 1. 3 Stages of target-site recognition by Cas9
a) The SpCas9-sgRNA complex proceeds through two stages of licensing prior to DNA cleavage. In the first stage target acquisition proceeds through recognition of the neighboring PAM sequence (nGG preferred for SpCas9). If 3' end of the programmed guide is complementary with the DNA sequence neighboring the PAM element an R-loop is initiated. Guide complementarity throughout the DNA sequence is assessed through extension of the R-loop, which once confirmed results in the allosteric activation of the RuvC and HNH nuclease domains resulting in a DSB (blue triangles). **b)** Only a small fraction of guide sequences have no near cognate sequences with the genome. [left stacked bar] Genome-wide CRISPRseek[278] assessment of the distribution of guide sequences for SpCas9 that have high quality off target sites. A representative set of 124,793 guide RNAs targeting human exons sequences are binned based on the off-target site with the fewest mismatches to the guide sequence. A "perfect match" indicates the presence of an off-target site perfectly complementary to the guide sequence (red segment), which would be anticipated to be an active site for cleavage. 98.4% of these guide sequences have at least one off-target site with 3 or fewer mismatches to the guide sequence. The remaining 1.6% of guides (light green segment) would be the best candidates for precise genome editing. [Right stacked bar] Guides within the light green segment were binned based on those with or without a single base bulged off-target site[270] but with no other mismatches (gray- absence of bulged off-targets, black- one or more bulged off-targets).

1.9.4 Lessons from unbiased methods for assessing nuclease fidelity

Before the development of unbiased methods for the characterization of nuclease off-target activity, computational algorithms based on guide-sequence similarity were the primary method for identifying potential off-target sites. Whereas these algorithms originally focused on binning off-target sites on the basis of the number of mismatches in the target sequence, more recent algorithms incorporate a scoring function whereby the position and type of mismatch are used to further stratify the ranking of these potential off-target sites[182, 256, 267, 268, 278-285] (**Table 1.1**). This information can be utilized to identify guide sequences with a low probability of off-target activity. However, even simple filtering based on excluding guides with near-cognate matches in the genome leads to a sizable decrease in the density of available target sequences (**Figure 1.3.b**), suggesting that alternative approaches to improve nuclease precision will be important. Currently the limited data on genome-wide Cas9-sgRNA off-target activity preclude the training robust predictive models that can accurately rank the relative cleavage rate of sites with multiple guide-target mismatches[286]. Consequently, although existing computational methods can identify active off-target sites for a guide sequence, the specificity of these predictions is poor.

Improving future computational models requires 'unbiased' genome-wide assessments of nuclease activity for a panel of different sgRNAs. A new suite of off-target detection methods have recently been described that can identify sites with low cleavage activity[244, 273, 286-290]. These methods focus on one of two approaches to identify the genomic positions of nuclease-induced DSBs: 1) direct capture of DSBs or 2) the detection

of surrogates associated with DSB repair activity. These methods have different strengths and weaknesses and vary in their sensitivity, expense and ease of implementation (sections **1.9.7** and **1.9.8**).

Digenome-seq[273] (*in vitro* Cas9-digested whole-genome sequencing) and BLESS[244, 291] (direct *in situ* break labeling, enrichment on streptavidin and next-generation sequencing) identify off-target sites through the detection of Cas9-induced DSBs within the genome. BLESS directly labels unrepaired DSBs in fixed cells via ligation of a biotin-labeled adapter, which allows the purification and sequencing of these genomic regions. Digenome-seq uses *in vitro* Cas9-sgRNA digestion of purified genomic DNA followed by WGS to identify the locations of DSBs as interruptions in mapped sequencing reads. Digenome-seq seems to be more sensitive than BLESS, as the latter only detected off-target sites with indel rates ≥1 %. Recently, SITE-seq[289] (selective enrichment and identification of tagged genomic DNA ends) and CIRCLE-seq[290] (circularization for *in vitro* reporting of cleavage effects by sequencing) used enrichment strategies to reduce the background and thereby improve the sensitivity of Digenome-seq.

The second class of off-target identification methods (integrase-defective lentiviral vector (IDLV) capture [288, 292], HTGTS[287] (high-throughput, genome-wide translocation sequencing) and GUIDE-seq[286] (genome-wide, unbiased identification of double-strand breaks enabled by sequencing)) rely on erroneous NHEJ-mediated DNA repair (exogenous DNA capture or translocation) to identify Cas9-induced breakpoints within the genome. IDLV-capture[288, 292] tags nuclease-induced DSBs via the insertion of exogenously supplied IDLV DNA, whereas GUIDE-seq[286] utilizes the insertion of modified blunt-ended duplex

oligonucleotides at DSBs. The tagging of DSBs by oligonucleotides is more efficient than that by IDLV, which enabled GUIDE-seq to identify off-target sites that were undiscovered by IDLV-capture with identical sgRNAs. Finally, HTGTS[287] captures off-target sites as genomic rearrangement events with a target locus. For a common sgRNA (VEGFa site 1) HTGTS identified a repertoire of off-target sites concordant with the GUIDE-seq method: all off-target sites identified by GUIDE-seq (21 total) were identified by HTGTS (38 total) and the ranked order of activity for the top 5 off-target sites was identical.

Together the data from these studies demonstrate that different sgRNAs display variability in the number of active off-target sites and the frequency of DSBs at those sites. Some sgRNAs led to promiscuous activity at more than 100 off-target sequences, whereas others were not associated with any detectable off-target events[286]. In general there is an anticorrelation between a genomic site's number of mismatches to a guide sequence and the likelihood that the site will be cleaved by Cas9, which is consistent with activity assays[256, 257]. Nonetheless, the parameters that define active noncognate sequences remain nebulous, as sites with up to six mismatches[273, 286] or a single base bulge[244, 286] between the guide and target sequence show activity. The availability of unbiased techniques for whole-genome surveys of nuclease activity provides an opportunity to construct more highly parameterized computational models of guide-target interactions to achieve the accurate prediction of functional off-target sequences for any given sgRNA.

1.9.5 Methods to enhance Cas9-sgRNA precision

Because of the imperfect precision of the wild-type SpCas9 nuclease, a number of strategies have been developed to increase the fidelity of sequence cleavage within the

genome (**Figure 1.4** and section **1.9.9**). These methods range from careful target-site selection and guide design to variations in Cas9 function that restrict conditions for DSB formation. These are discussed as independent approaches, but in some cases they could potentially be combined.

Perhaps the most straightforward approach is to use the growing number of computational tools to predict guide sequences that will have the fewest off-target sites[182, 256, 267, 268, 278-285] (as described above; **Table 1.1**). Careful choice of the guide can have a dramatic impact on the number of active off-target sites[244, 273, 286-288]. Modification of the guide sequence has also proven advantageous. Truncation of the guide from 20 nt to 17 nt or 18 nt (tru-gRNA) preserves nearly full activity of the SpCas9-tru-gRNA nuclease at the majority of target sites[268, 286, 293] while reducing activity at many off-target sites[286, 293]. Alternately, appending two extra guanine nucleotides on the 5′ end of the guide sequence has been shown to improve nuclease precision[198, 273]. In addition, delivering Cas9 and its sgRNA as a ribonucleoprotein complex (RNP) limits its temporal activity thereby improving its precision[208, 209, 294-297].

Modifications to the Cas9 nuclease have also proven effective in improving precision. Inactivation of one of the two nuclease domains in Cas9 generates a nickase (Cas9n) that cleaves only a single strand of the target sequence[170]. By programming Cas9 nickases with two different sgRNAs targeting neighboring sites on opposite DNA strands, one can generate a DSB[198, 261, 276, 298]. The combinatorial requirement for two active proximal nickases to generate a DSB reduces the likelihood of off-target DSBs[198, 261, 287, 298]. However, at some sites in transformed cell lines individual nickases can yield high rates

of single base mutations (~10 %), although the precise mechanism is undefined[190, 299]. A related strategy mimicking ZFN and TALEN platforms, uses a catalytically inactive Cas9 (dCas9) as a DNA-recognition platform for the site-specific delivery of the FokI nuclease domain[190, 299, 300]. FokI-dCas9 offers improved precision relative to standard Cas9 because of the requirement for FokI dimerization for the efficient generation of a DSB. However, single FokI-dCas9 monomers can generate a low level of mutagenic activity at some sequences[190, 299]. Another strategy relies on the attenuation of non-sequence specific DNA contact residues to reduce the overall binding energy of Cas9 to the DNA. Three variants (enhanced specificity Cas9[301], high-fidelity Cas9[302], and hyperaccurate Cas9[303]) improved the targeting accuracy of SpCas9 substantially.

One can also modulate Cas9 activity by manipulating the Cas9 PAM-interacting domain (PID; **Figure 1.2**). Structural and biochemical studies identified the location of PID and demonstrated that PAM specificity could be switched by substituting PIDs from orthologous Cas9s[205, 253]. Recently, selection-based approaches were used to identify mutations in the PID that generate SpCas9 and variants with distinct PAM preferences[304]. Not only these SpCas9 variants not only expand the targeting range of SpCas9, but some variants also show improved precision on the basis of genome-wide activity analysis. Similar approach has been extended to select SaCas9 variants with different PAM preferences[305]. In an alternate approach, mutations in the PID that attenuate the DNA-binding affinity of Cas9 can render its function dependent on DNA recognition by an attached programmable DNA-binding domain (pDBD). The resulting Cas9^{MT}-pDBD

system, which is the focus of this thesis, provides improved precision by adding an additional specificity determinant to target site recognition[306].

Exerting temporal control over nuclease activity is another appealing method for improving precision. One can regulate nuclease activity by breaking the Cas9 protein into two independent components (split-Cas9)[307-310] or adding an allosteric switch[311] or an interrupted sequence (intein-Cas9)[310, 312] in which assembly or intein excision, respectively, can be exogenously regulated. The switch to an active state can be driven through the delivery of a small molecule[308, 311, 312] or light of a suitable wavelength[309].

Finally, there are a multitude of Type II CRISPR-Cas9 systems[244, 313], only a small number of which have been characterized in any detail[242, 244, 245, 314]. In addition, Type V CRISPR-Cas systems[152] that use Cpf1 (Cas12-a) instead of Cas9 as the nuclease effector also show promise for editing eukaryotic genomes[174, 315-319]. Some of these alternative systems have more favorable characteristics for the precise editing of vertebrate genomes owing to differences in PAM stringencies and sgRNA mismatch sensitivities[320-322]. Looking forward, it seems likely that the preferred editing platform, in particular for therapeutic applications, will be a more compact nuclease (Cas9 or Cpf1 orthologue) with improved targeting range (for example, a simple PAM element) and greater sensitivity to mismatches within its recognition site. This system likely does not exist in a native form, but will be improved through design and selection approaches to achieve the desired goal of robust activity and high precision.

Table 1.1 Computational programs for sgRNA off-target analysis.

Algorithm Name	Off-target search features
CRISPRseek[278]	Allows discovery of non-cognate sites with user-defined maximum number of mismatches; Scoring based on number and position of mismatches; Available as a Bioconductor package; Search and scoring parameters can be modified by the user and program can be included in computational pipelines.
Cas-OFFinder[279]	Allows discovery of non-cognate sites with up to 9 mismatches and up to 2 DNA and RNA bulges; does not score/rank off-target sites; Program is available for download.
CRISPR Design Tool[256]	Allows discovery of non-cognate sites with up to 4 mismatches; Scoring based on number and position of mismatches; No user-defined features.
COSMID[280]	Allows discovery of non-cognate sites with up to 3 mismatches and up to 2 DNA and RNA bulges; Scoring based on number and position of mismatches.
CropIT[281]	Allows discovery of non-cognate sites with up to 9 mismatches; Scoring based on number and position of mismatches; Incorporates DNase hypersensitivity profiles from the ENCODE dataset for ranking off-target sites within the human genome; limited user-defined parameters.
CasFinder[282]	Allows discovery of non-cognate sites with user defined number of mismatches; Scoring based on number and position of mismatches; Available as a perl package; parameters can be modified by the user.
E-CRISP[283]	Allows discovery of non-cognate sites with up to 7 consecutive mismatches at the 5' most end and up to 1 mismatch in the remaining 13 positions or up to 3 mismatches total; Scoring based on number of mismatches, genomic location of the off-target site and its sequence composition; limited user-defined parameters.
ChopChop[267]	Allows discovery of non-cognate sites with up to 2 mismatches; Lists off-target sites by number of mismatches.
ZiFiT[182, 284]	Allows discovery of non-cognate sites with up to 3 mismatches, where the position and type of mismatch are indicated.
CasOT[285]	Allows discovery of non-cognate sites with up to 6 mismatches; Program is available for download.

1.9.6 Practical considerations for the design of Cas9-based editing experiments

There are a multitude of factors (Cas9 platform, sgRNA target site, etc.) to consider during the design phase of Cas9-based projects. Below we suggest some strategies that may be useful in design of nuclease-based experiments.

Importantly, not all guide sequences can program Cas9 for highly efficient genomic cleavage. The use of SpCas9 for genome-wide gene-inactivation screens[199-202, 268] is producing a large amount of activity data that can be used to train computational models to predict the most favorable sequences for Cas9 cleavage[258, 264, 266, 268]. Algorithms that predict sgRNA activity (for example, sgRNA Designer[264], Sequence Scan for CRISPR[266], sgRNA Scorer[258] or CRISPRscan[268]) can be used with their built-in off-target assessments or in combination with algorithms that evaluate additional aspects of off-target sites for promising guide sequences (for example, heteroduplex bulge by COSMID[280]; **Table 1.1**) to identify targets that are highly active and have few potential off-target sites. The activity of the chosen guides can then be verified in the appropriate biological system[323].

When targeting protein-coding genes to generate null alleles, the most robust method to achieve a null is targeting an exon containing a critical domain for protein function[324]. Simply targeting an early coding exon can prove ineffective, as alternate transcriptional start sites or alternative splicing may allow the generation of functional transcripts that bypass the Cas9-induced lesions. For non-coding RNAs or regulatory elements, complete excision of the genomic segment[128, 325-329] may provide the greatest likelihood that a loss-of-function allele is generated.

The targeted insertion of DNA through HDR-based approaches also has important caveats. Optimal lengths of homology arms for single-stranded oligonucleotides[116, 209] (ssODN) and duplex DNA donors[206, 218] have been defined in certain systems, but these parameters may vary in different biological settings. Importantly, these insertions do not always occur via precise HDR; depending on the type of donor, they can also occur through NHEJ-mediated pathways at one end (or both ends) of the donor insertion site[109, 206, 214, 330, 331]. The ratio of NHEJ-generated indels relative to HDR-mediated insertions at the target site in a population of treated cells or embryos can be quantified directly via SMRT sequencing[218] or Illumina sequencing when an ssODN is used as a HDR donor. In addition, for large donor DNAs, Southern blotting or linear amplification-mediated (LAM)-PCR[332]—whichever is more appropriate for the biological system—should be standard practice to verify the absence of random donor integration in the genome.

The tolerance for the level of nuclease promiscuity will obviously depend on the application. Researchers using Cas9 nucleases for the genomic alteration of model organisms or cell lines for basic science investigations will be more comfortable with low-frequency, in contrast to nucleases destined for therapeutic applications involving the modification of millions to billions of stem or progenitor cells. In basic science investigations, concerns about observed phenotypes resulting from collateral genomic damage can be alleviated by the use of two or more independent cell lines or animal models, ideally generated using different sgRNAs that target the same locus. When possible, genetic complementation can be used to validate the phenotype-genotype association. For therapeutic applications, in-depth unbiased off-target analysis of nuclease

activity should be carried out on the relevant cell type(s). Once a baseline of nuclease precision is established, off-target analysis can be used to optimize the nuclease dose and the delivery method to identify the conditions that produce the most favorable activity profiles with high precision.

The nuclease field is advancing at a breathtaking rate. Translating this newfound technical ability into the creation of breakthrough therapeutics that profoundly alter disease outcomes for patients is no longer science fiction; in fact, it is likely that this will be accomplished in the foreseeable future. Achieving this goal will require continued improvements in nuclease precision, the ability to detect unintended insults to the genome, and our understanding of the complexities of DNA repair. The result of these efforts, if advanced prudently, will enable the deployment of exciting gene- and cell-based therapeutics that will radically improve treatment options for a number of diseases.

1.9.7 Unbiased methods for off-target analysis

1.9.7.1 Whole Genome Sequencing (WGS)

The most exhaustive method to assess changes to the genome. It is only feasible to analyze a small number of nuclease-treated clones (or animals), which limits the ability to extrapolate this data to off-target lesion rates in a population of nuclease-treated cells.

1.9.7.2 Direct DSB capture

These methods directly identify DSBs generated by the Cas9.

Digenome-Seq[273]: This method allows the dose dependent *in vitro* assessment of Cas9-sgRNA activity, which permits identification of weakly active off-target sites. Genomic loci with mapped sequencing reads that have aligned 5′ ends are characteristic of

potential off-targets. Digenome-seq cannot capture the influence of local chromatin structure or folding on Cas9 activity[258], since it is performed *in vitro*.

BLESS[244, 291]: This method labels unrepaired DSBs in fixed nuclei of treated cells via ligation of a biotinylated adaptor. This tag allows solid support based enrichment of these loci for deep sequencing analysis. Overall there is a good correlation between measured lesion rates and the DSB score assigned based on BLESS analysis. The persistence of DSBs prior to repair, if short, may limit the sensitivity of this approach.

SITE-seq[289]: This method also uses *in vitro* digestion of genomic DNA with Cas9-sgRNA RNP complexes. High molecular weight genomic DNA is treated with an RNP to generate blunt end DSBs. The ends are 3' adenylated and ligated with biotinylated P5 adaptors. Then, the DNA is fragmented, end repaired, and ligated to a P7 adaptor. Next, biotin selection and PCR amplification is applied to enrich the with Cas9 activity related DNA pieces. Similar Digenome-seq, Site-seq also cannot capture the influence of local chromatin structure or folding on Cas9 activity.

CIRCLE-seq[290]: This method also uses *in vitro* digestion of genomic DNA with Cas9-sgRNA RNP complexes. Fragmented genomic DNA is circularized via intra-molecular ligation. The circular DNA is treated with an RNP and then adaptors are ligated to the broken ends for deep sequencing. Although CIRCLE-seq is very sensitive method, the influence of local chromatin structure or folding on Cas9 activity also cannot be assessed.

1.9.7.3 Repair based methods

These methods exploit erroneous NHEJ-mediated repair to tag DSBs through either exogenous DNA capture or genomic rearrangements.

IDLV-capture[288, 292]: This method uses NHEJ-mediated integration of IDLV DNA to tag DSBs. IDLV integration sites are detected by LAM-PCR[332138] followed by deep-sequencing. This method has been successfully used for ZFNs, TALENs, and Cas9 but due to low sensitivity it cannot capture inefficient off-target sites (<1% indel frequency).

GUIDE-seq[286]: This method uses NHEJ-mediated integration of phosphorothioate-protected double stranded oligonucleotides to tag genomic DSBs that are then enriched by PCR and deep-sequenced. The modified oligonucleotides can achieve high insertion frequencies in some cell lines (20-50 % of total indel frequency), which enhances its sensitivity relative to IDLV capture. Off-target sites with indel rates >0.1 % can be detected. There is a good correlation between the indel rate and GUIDE-seq oligonucleotide insertion rate. Across analyzed sgRNA datasets, the majority of off-target sites with two base mismatches and a subset of three base mismatches were active. This method works preferentially with blunt DSBs, so it is less sensitive when used with Cas9 nickases or FokI-dCas9 fusions[286, 299].

HTGTS[287]: This method combines LAM-PCR and genome-wide translocation sequencing[333] to capture genomic rearrangements caused by improper ligation of DSBs between the nuclease target and a spatially proximal off-target site. The target site is used as the 'bait' to capture off-target 'prey' sites, although highly active off-target sites can also be used as the 'bait' affording a similar distribution of genomic DSBs. HTGTS is both

sensitive and applicable to other Cas9 variants that improve precision (Cas9 nickases or FokI-dCas9).

1.9.8 Discussion on unbiased methods for off-target analysis

With a number of methods available for performing unbiased off-target analysis, end-users must choose the most appropriate method for their experimental application. The choice of off-target detection method depends on many factors including: (i) its sensitivity and specificity, (ii) its ease of implementation and data analysis, (iii) the Cas9 variant employed, and (iv) the cell type.

Among these methods, GUIDE-seq[286] and HTGTS[287] perform similarly with both giving the most comprehensive list of off-target sites and the highest levels of sensitivity. Although GUIDE-seq was shown to be highly accurate, sites obtained in HTGTS were not validated in detail. Digenome-seq[273] can also be highly sensitive but can also miss many off-target sites (for VEGFa TS1, 7 of 21 GUIDE-seq sites and 24 of 38 HTGTS off-target sites were missed) and requires a much higher depth of sequencing. Reduction of the background during the library preparations of SITE-seq[289] and CIRCLE-seq[290] make them very sensitive platforms to assess off-target cleavages. However, the presence of *in cellulo* off-target activities may require confirmation.

There is some background that is associated with both Digenome-seq and BLESS[244], as less than 50% of the identified off-target sites withstand validation. The data analysis for BLESS and GUIDE-seq requires filtering to remove DSB hotspots within the genome. GUIDE-seq had a higher rate of validated sites, but the data processing identifying

peaks also involved pre-filtering sequences based on the similarity to the target sequence (up to 8 base mismatch was allowed).

Although most methods should be relatively easy to perform, the LAM-PCR step used in HTGTS[287] and IDLV-capture[288] methods may require more technical expertise. A JOVE video is available describing the LAM-PCR technique[292]. Importantly computational analysis of the Illumina sequencing data can be challenging for some of these approaches. In particular for HTGTS[287] where translocation sequences will contain elements that map to two different sites within the genome. Users will likely require bioinformatics support for the data processing from these approaches.

Although all of the unbiased off-target analysis methods work with wild-type Cas9, some methods may be insensitive to Cas9 variants that do not leave blunt DSBs. For example GUIDE-seq has low sensitivity to Cas9-nickases or FokI-dCas9 dimers as they produce DSBs with overhangs upon cleavage[299]. However, WGS, Digenome-seq, SITE-seq, CIRCLE-seq and HTGTS should remain sensitive to Cas9 variants.

Since Cas9 binding[250, 259] and hence activity can potentially be influenced by chromatin architecture, the repertoire of off-target sites can vary between different cell lines[286]. Consequently, off-target analysis should be performed in the relevant cell type in which the nuclease will be applied. Although methods like WGS, BLESS[244], Digenome-seq[273], SITE-seq[289], and CIRCLE-seq[290] in principle, can be performed for any cell line, the efficiency of other methods may be cell line dependent. For example, the cell-type sensitivity of GUIDE-seq may depend on both the efficiency of the insertion of GUIDE-seq oligonucleotides and the accommodation of the introduction of foreign DNA. The

latter may induce innate immune responses in some cell types such as T-cells[334, 335]. Likewise, some cell types may be more tolerant and amenable to translocations than others, which could impact the sensitivity of HTGTS output. Thus, choosing the best method for off-target identification will be both user and application dependent.

1.9.9 Additional information on methods to enhance Cas9-sgRNA precision

These methods can be categorized into four primary strategies:

1.9.9.1 Computational selection of the best guides

Computational tools can be used to choose guides that are predicted to have high on-target and low off-target activity (**Table 1.1**). This filtering inherently reduces the density of Cas9 targets within the genome (**Figure 1.3.b**), which may prove problematic for editing applications that have a limited sequence window.

1.9.9.2 Modification of the guide sequence

Truncated guides (tru-gRNA[286, 293]): Truncating guides from 5′ end (17 or 18nt) improves precision by reducing the stability of the Cas9-sgRNA complex[254, 255, 336], possibly although the exact mechanism remains unclear. GUIDE-seq analysis indicates a dramatic reduction in off-target activity for Cas9-tru-gRNAs relative to full-length Cas9-sgRNAs[286]. Nonetheless, residual nuclease activity persists at a small number of off-target sites.

5′ Guanine[198, 273]: In addition to standard 20nt guide for SpCas9, two guanine nucleotides are added on to the 5′ end of the guide sequence. This modification appears to function in part through nuclease attenuation, which can impact activity at both the target and off-target sites[268, 273, 277].

1.9.9.3 Delivery of the Cas9-sgRNA ribonucleoprotein complex

The delivery of Cas9 nuclease and the sgRNA via plasmid transfection or viral transduction typically results in expression of the Cas9-sgRNA complex beyond that needed to achieve target site DSB formation[294], which increases the potential for off-target activity[289]. In contrast, delivery of Cas9 and sgRNA as a ribonucleoprotein complex[191, 208, 209, 294-297] or within a lipid nanoparticle[239] can more readily control the cellular concentration of Cas9-sgRNA and restrict the time window of its activity.

1.9.9.4 Modification of the Cas9 nuclease

Dual Cas9 nickases (Cas9n[198, 261, 276, 298]): The strand that is nicked is determined by the nuclease domain (RuvC or HNH) that is inactivated[170, 253]. When used with a pair of guides to generate nicks on both strands that leave 5′ overhangs, DSBs are generated that stimulate NHEJ or HDR[261, 298]. Although single nicks are much less mutagenic than DSBs, they can lead to mutagenesis in the context of some sequences[190, 198, 276, 300]. The requirement for two neighboring Cas9 sites reduces the density of targetable sequences by dual nickases with the genome.

FokI-dCas9[190, 300]: Both the RuvC and HNH nuclease domains of Cas9 are mutated yielding a catalytically inactive Cas9 (dCas9) to deliver the nuclease domain of FokI to specific sites of the genome. FokI-dCas9 can be used in combination with tru-gRNAs to increase nuclease precision[299]. Like Cas9n, FokI-dCas9 has reduced targeting density because a pair of neighboring compatible binding sites is required for DSB formation. However, unlike the nickases, the relative spacing of FokI-dCas9 dimers is much more restricted.

SpCas9 PAM variants[304]: New PAM variants that target NGAG and NGCG were identified through mutagenesis and selection. Notably, the NGCG SpCas9 variant displays few off-target sites based on GUIDE-seq analysis. In addition a D1135E mutant was identified that increases the stringency of SpCas9 for the standard NGG PAM, which improves its precision.

eSpCas9_v1.1[301], SpCas9-HF1[302], HypaSpCas9[303]: In these variants, non-sequence specific DNA recognition residues of SpCas9 is mutated to reduce its excess binding energy to the DNA. Mechanistically these mutations extend the period of SpCas9 to stay on the inactive state. Therefore, nuclease activities at some sites are reduced in comparison to wild type SpCas9.

SpCas9^{MT}-pDBD[306]: In this system, Cys2-His2 zinc finger arrays or transcription activator-like effector (TALE) domains are engineered to deliver the mutant Cas9 to a desired target site in the genome. Target site recognition by the pDBD is necessary for nuclease activity, which reduces the number of active off-target sites.

Base editors[337-339]: In this system, the Cas9 is either nuclease dead or a D10A nickase, and fused to a cytosine deaminase effector. Base editors convert a PAM distal cytosine into uracil, thereby a C:G base pair to T:A[337, 338]. This utility is particularly useful for gene correction based therapies where the underlying cause is a point mutation. Recently, directed evolution based study described the development of adenine deaminase, which converts a PAM distal adenine into inosine, thereby a A:T base pair to G:C[339].

Split[307-310]- or Intein-Cas9[310, 312]: Due to the bilobed architecture of Cas9, it can be divided (or interrupted) into two inactive segments, where reassembly (or intein

excision) restores activity to the complex. The association of the two domains split-Cas9 can be regulated through the introduction of drug-dependent[308] or light-dependent[309] dimerization domains. Likewise, intein excision can be regulated through drug-dependent control[312]. Besides temporal control, dividing SpCas9 into two fragments allows its packaging in two separate AAV vectors[310].

Allosterically-regulated Cas9[311]: Another way to inducible nuclease activity is to add an allosteric switch to Cas9. This is achieved by insertion of estrogen receptor-α ligand-binding domain (ER-LBD) at a hotspot (residue 231) of the SpCas9 protein. This insertion blocks the nuclease activity of SpCas9. Upon addition of a small molecule ligand, the ER-LBD undergoes conformational change and reactivates the SpCas9.

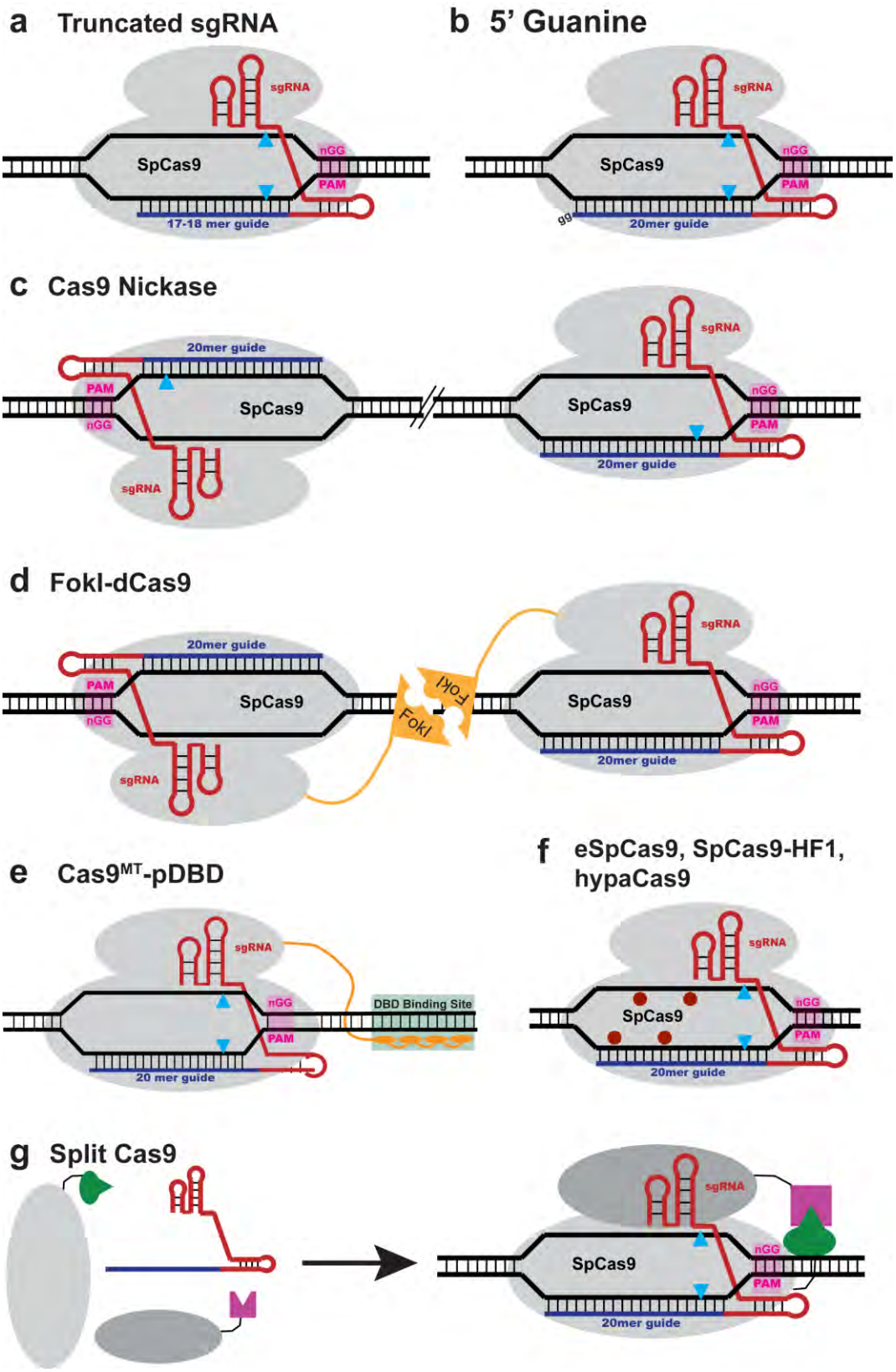

a Truncated sgRNA
sgRNA
SpCas9
nGG
PAM
17-18 mer guide
b 5' Guanine
sgRNA
SpCas9
nGG
PAM
gg
20mer guide
c Cas9 Nickase
20mer guide
PAM
nGG
SpCas9
sgRNA
sgRNA
SpCas9
nGG
PAM
20mer guide
d FokI-dCas9
20mer guide
PAM
nGG
SpCas9
sgRNA
FokI
FokI
sgRNA
SpCas9
nGG
PAM
20mer guide
e Cas9MT-pDBD
sgRNA
nGG
PAM
DBD Binding Site
20 mer guide
f eSpCas9, SpCas9-HF1, hypaCas9
sgRNA
SpCas9
nGG
PAM
20mer guide
g Split Cas9
sgRNA
SpCas9
nGG
PAM
20mer guide

Figure 1. 4 Schematic of SpCas9 variants with improved precision
a) Truncated sgRNA where the guide is reduced to 17-18nt in length. **b)** Addition of two Guanines to the 5' end of the guide. **c)** Dual Cas9 nickases (Cas9n) where one of the catalytic domains has loss-of-function mutation. When used in pairs with the proper orientation of sites the nickases generating a DSB with 5' overhangs. **d)** FokI-dCas9 fusions where the catalytically inactive SpCas9 acts as programmable DNA-binding unit. Two FokI-dCas9 complexes must bind with the proper orientation and spacing to allow dimerization of FokI and DSB formation. **e)** $Cas9^{MT}$-pDBD where a programmable DNA-binding domain (pDBD) is fused to a Cas9 that has attenuated DNA-binding affinity through a PID mutation. Consequently, nuclease activity requires DNA recognition by the pDBD as well as a compatible PAM and sgRNA recognition sequence. **f)** SpCas9-HF1, eSpCas9_v1.1, HypaSpCas9 variants are generated through mutations in non-sequence specific DNA interaction residues of SpCas9 (red circles). These mutations reduce the DNA-binding energy of SpCas9 and extend its inactive confirmation time to improve. **g)** Split-Cas9 where the SpCas9 is divided into two catalytically inactive segments fused to a (inducible) dimerization domain. Upon stimulus, dimerization occurs to reunite the components of SpCas9 and restore nuclease activity.

CHAPTER II: DNA-binding-domain fusions enhance the targeting range and precision of Cas9

The CRISPR-Cas9 system is commonly used in biomedical research; however, the precision of Cas9 is suboptimal for applications that involve editing a large population of cells (for example, gene therapy). Variations on the standard Cas9 system have yielded improvements in the precision of targeted DNA cleavage, but they often restrict the range of targetable sequences. It remains unclear whether these variants can limit lesions to a single site in the human genome over a large cohort of treated cells. Here we show that by fusing a programmable DNA-binding domain (pDBD) to Cas9 and attenuating Cas9's inherent DNA-binding affinity, we were able to produce a Cas9-pDBD chimera with dramatically improved precision and an increased targeting range. Because the specificity and affinity of this framework can be easily tuned, Cas9-pDBDs provide a flexible system that can be tailored to achieve extremely precise genome editing at nearly any genomic locus.

2.1 Introduction

The CRISPR-Cas9 genome engineering system is revolutionizing biological sciences owing to its simplicity and efficacy [185, 340, 341]. The most commonly studied Cas9 nuclease (SpCas9) originates from *Streptococcus pyogenes* [170]. SpCas9 and its associated guide RNA license a DNA sequence for cleavage on the basis of two stages of sequence interrogation [170, 205, 248, 249, 254] (**Figure 2.1a**): (1) compatibility of the protospacer-adjacent motif (PAM) element with the specificity of the PAM-interacting domain, and (2)

complementarity of the guide RNA sequence to the target site. As it is straightforward to program Cas9 to cleave a desired target site through the incorporation of a complementary single guide RNA (sgRNA) [170], the primary constraint on Cas9 targeting is the presence of a compatible PAM element [170, 190, 256]. The PAM-interacting domain of wild-type SpCas9 preferentially recognizes an NGG element [170], although it can inefficiently utilize other PAM sequences (for example, NAG and NGA) [256, 260]. The simplicity of the SpCas9-sgRNA system allows facile editing of genomes in a variety of organisms and cell lines [185, 340, 341].

The precision of SpCas9 is suboptimal for most gene therapy applications involving editing of a large population of cells [184, 342]. Numerous studies have demonstrated that SpCas9 can cleave the genome at unintended sites [256, 257, 269, 270, 273, 286-288], with some guides acting at more than 100 off-target sites [286]. Recent genome-wide analyses of SpCas9 precision indicate that the majority of genomic loci that differ from the guide RNA sequence at two nucleotides and a subset of genomic loci that differ at three nucleotides are cleaved with moderate activity [273, 286-288]. For some guides, off-target sites that differ by up to six nucleotides can be inefficiently cleaved [273, 286-288], and bulges can be accommodated in the sgRNA:DNA heteroduplex [270]. In this light, we assessed the general frequency of potential off-target sites with three or fewer mismatches for SpCas9 guide RNAs in exons or promoter regions using CRISPRseek [278, 343]. We found that the vast majority of guides (~98% in exons and ~99% in promoters) had one or more off-target sites with three or fewer mismatches (**Figure 2.1b-e**) and thus were likely to have some level of off-target activity. Because off-target breaks have the potential to cause both local

mutagenesis and genomic rearrangements [286, 287, 344, 345] (for example, segmental deletions, inversions and translocations), the resulting collateral damage from SpCas9 treatment could have adverse consequences in therapeutic applications.

Decreased off-target cleavage rates have been associated with several modifications to the structure or delivery of the CRISPR-Cas9 system, such as changes to the guide sequence length and composition [198, 293], the use of a pair of Cas9 nickases [198, 261, 298] or FokI-dCas9 nucleases [190, 300], inducible assembly of split Cas9 [307-309, 312], Cas9 PAM variants with enhanced specificity [304] and the delivery of Cas9:sgRNA ribonucleoprotein complexes [294, 295, 297]. However, it remains unknown whether these variations can restrict cleavage to a single site in the human genome over a large cohort of treated cells [274, 342]. In addition, some of the most promising approaches (for example, paired nickases and dimeric FokI-dCas9) restrict the targetable sequence space by requiring the proximity of two sequences compatible with Cas9 recognition.

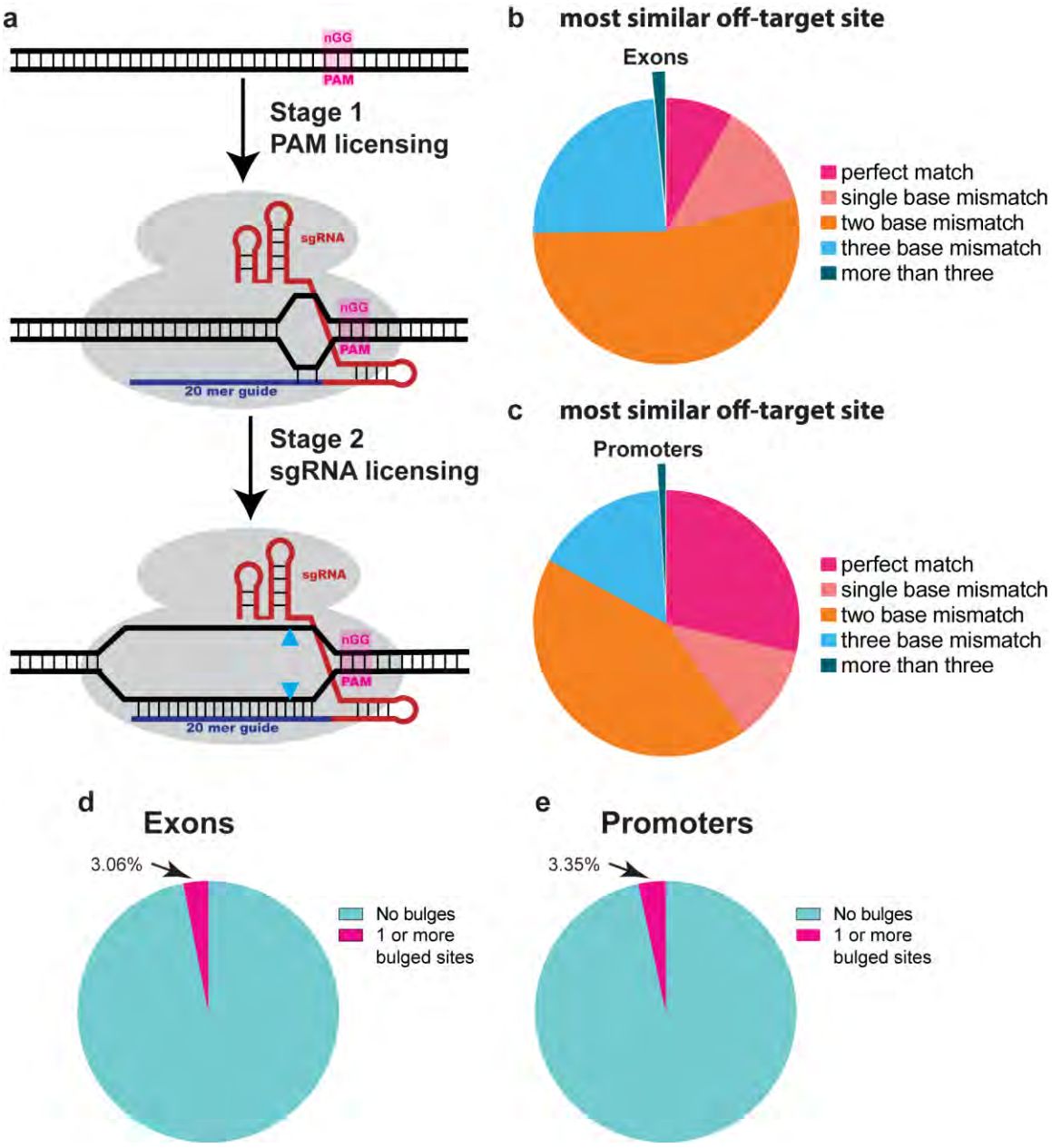

a
nGG
PAM
Stage 1
PAM licensing
sgRNA
nGG
PAM
20 mer guide
Stage 2
sgRNA licensing
sgRNA
nGG
PAM
20 mer guide
b
most similar off-target site
Exons
perfect match
single base mismatch
two base mismatch
three base mismatch
more than three
c
most similar off-target site
Promoters
perfect match
single base mismatch
two base mismatch
three base mismatch
more than three
d
Exons
3.06%
No bulges
1 or more
bulged sites
e
Promoters
3.35%
No bulges
1 or more
bulged sites

Figure 2. 1 Overview of potential SpCas9 off-target sites in the human genome
(**a**) Schematic of the SpCas9:sgRNA system and the two sequential stages of licensing required for cleavage: Stage 1 – PAM recognition (nGG is highly preferred) and Stage 2 – complementary R-loop formation between the 20 nucleotide guide RNA and the interrogated DNA sequence. (**b)** Genome-wide analysis using CRISPRseek[278] of the potential off target sites for a representative set of 124,793 guide RNAs targeting human exons sequences. Guides were binned based on the predicted off-target site with the smallest number of mismatches to the guide sequence. A perfect match indicates the presence of an off-target site with a perfect guide match (magenta wedge). Only 1.6% of these guide sequences do not have an off-target site with 3 or fewer mismatches to the guide sequence (teal wedge). This subset would be the best candidates for precise genome editing. The vast majority of guides typically have many potential off-target sequences with 3 or fewer mismatches. (**c)** Genome-wide analysis of the minimum number of mismatches in off-target sites for a representative set of 55,687 guide RNAs targeting human promoter regions (binned as describe above). Only 1% of these guide sequences do not have an off-target site with 3 or fewer mismatches to the guide sequence (teal wedge). (**d,e**) Guide RNAs targeting gene exons (**d**) or promoters € with no predicted off-targets with <= 3 mismatches (teal wedge from **b,c**) are analyzed for off-target sites with potential bulges in the sgRNA:DNA heteroduplex[270]. Magenta wedges indicate the fraction of guides that have one or more off-target sites that have perfect complementarity with the exception of a single bulge.

We envisioned an improved Cas9 platform in which the precision of target recognition was augmented by the incorporation of pDBDs such as Cys_2-His_2 zinc-finger proteins [63] (ZFPs) or transcription activator–like effectors [43] (TALEs) (**Figure 2.2**). Both of these pDBD platforms can be programmed to recognize nearly any sequence in the genome [37, 43, 63, 346]. Indeed, pDBDs have been used with great success as targeting domains for programmable nucleases through the incorporation of nonspecific FokI nuclease domains (zinc-finger nucleases [63] and TALE nucleases [43]) or sequence-specific nuclease domains (e.g., megaTAL [347]). One favorable characteristic of pDBDs is their inherent modularity whereby specificity and affinity can be rationally tuned by adjustments to the number and composition of incorporated modules and the linkage between modules [348, 349]. Here we demonstrate that the fusion of a pDBD to a mutant SpCas9 with attenuated DNA-binding affinity generated a chimeric nuclease with a broader sequence-targeting range and dramatically improved precision. This SpCas9-pDBD platform has favorable properties for genome engineering applications. In addition, our analysis of these SpCas9-pDBD chimeras provides new insights into the barriers involved in licensing target-site cleavage by a SpCas9:sgRNA complex.

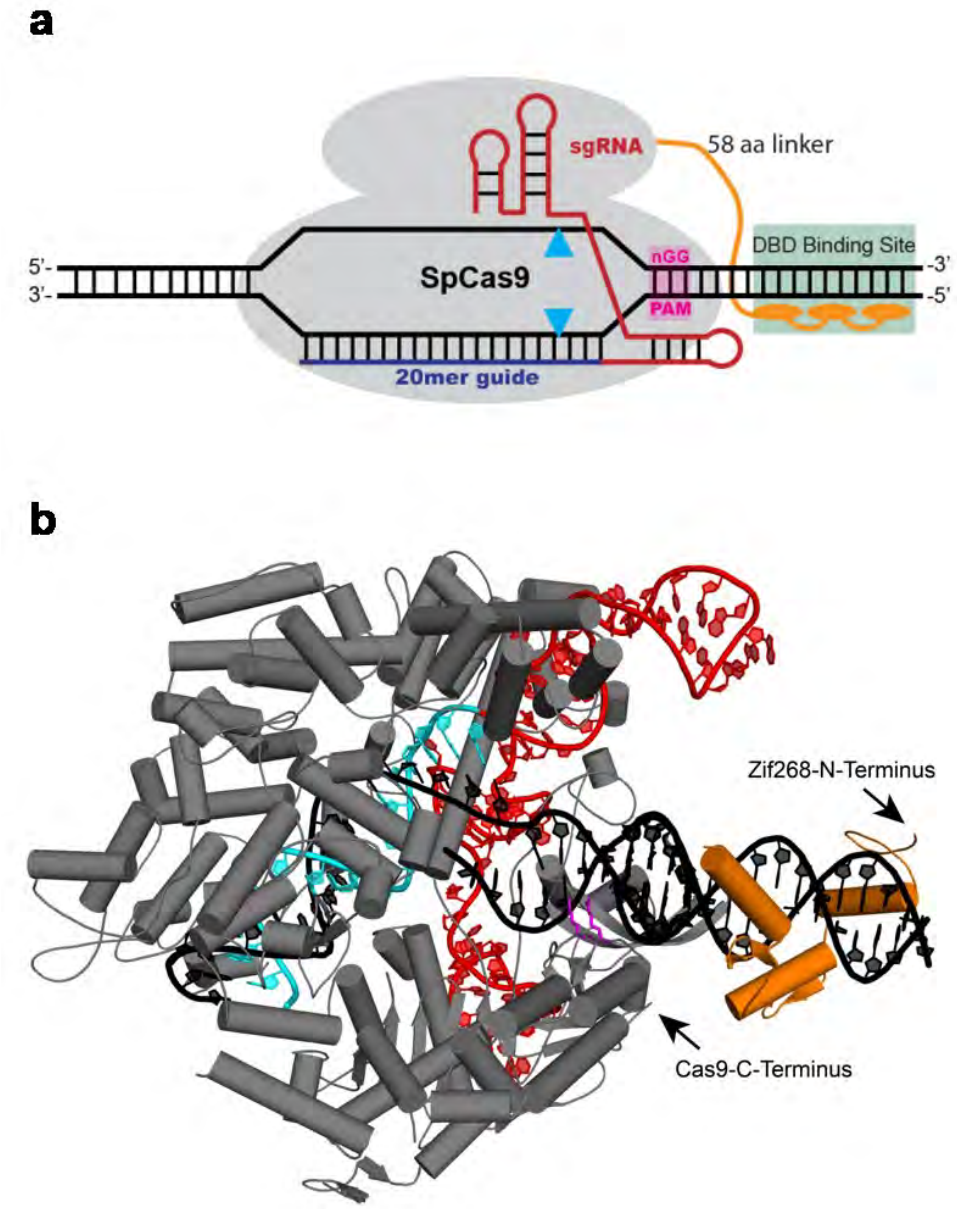


Figure 2. 2 Overview of the SpCas9-pDBD designs
a) Schematic of the SpCas9:sgRNA system fused to a pDBD (orange) that recognizes a binding site 3' to the PAM**. b)** A hybrid model containing the structure of SpCas9[205] (grey, PAM recognition residues magenta) with an sgRNA (20 nucleotide guide region cyan, remaining nucleotides red) and complementary target DNA (black) with the structure of Zif268[22] (orange) placed with a binding site 11 bp from the PAM recognition sequence (Watson strand), where the two structures were superimposed on a B-DNA model constructed using 3DNA[350]. In parallel with the spacing parameter analysis in **Figure 2.3**, the structural model suggests that there is ample room for a ZFP to dock proximally to SpCas9 downstream of the PAM element.

2.2 Materials and Methods

2.2.1 Plasmid constructs:

Our SpCas9-pDBD experiments used the following plasmids: All sgRNAs were expressed via a U6 promoter from pLKO1-puro[351]. All SpCas9 and SpCas9-DBD fusions were expressed via pCS2-Dest gateway plasmid under chicken β-globin promoter[352]. ZFPs were assembled as gBlocks (Integrated DNA Technologies) from finger modules on the basis of previously described recognition preferences[35, 36]. ZFPs were cloned into a pCS2-Dest-SpCas9 plasmid backbone cloned thorough BspEI and XhoI sites. TALEs were assembled via golden gate assembly[111] into our JDS TALE plasmids[353]. Assembled TALEs were cloned into BbsI digested pCS2-Dest-SpCas9-TALEntry backbone through Acc65I and BamHI sites. Key plasmids of this study will be deposited at Addgene for distribution to the community. Plasmid reporter assays of nuclease activity used the restoration of GFP activity through single-strand annealing (SSA)-mediated repair of an inactive GFP construct using the M427 plasmid developed by the Porteus laboratory[354]. SpCas9 target sites were cloned into plasmid M427 via ligation-independent methods after SbfI digestion. Mutations in the PAM-interacting domain of SpCas9 were generated by cassette mutagenesis.

2.2.2 Cell culture assay

HEK293T cells obtained from our collaborator M. Green (UMass Medical School, Worcester, Massachusetts, USA) were cultured in high-glucose DMEM with 10% FBS and 1% penicillin-streptomycin (Gibco) in a 37 °C incubator with 5% CO_2. These cells were

not verified or tested for mycoplasma contamination. For transient transfection, we used early to mid-passage cells (passage numbers 5–25). Approximately 1.6×10^5 cells were transfected with 50 ng SpCas9-pDBD–expressing plasmid, 50 ng sgRNA-expressing plasmid and 100 ng mCherry plasmid with Polyfect transfection reagent (Qiagen) in a 24-well format according to the manufacturer's suggested protocol. For the SSA-reporter assay, 150 ng M427 SSA-reporter plasmid was also included in the cotransfection mix.

2.2.3 Western blotting

HEK293T cells were transfected with 500 ng Cas9 and 500 ng sgRNA-expressing plasmid in a six-well plate with Lipofectamine 3000 transfection reagent (Invitrogen) according to the manufacturer's suggested protocol. 48 h after transfection, cells were harvested and lysed with 100 μl of RIPA buffer. 8 μl of cell lysate was used for electrophoresis and blotting. The blots were probed with anti-hemagglutinin (Sigma, H9658) and anti–α-tubulin (Sigma, T6074) as primary antibodies, and then with horseradish peroxidase–conjugated anti-mouse IgG (Abcam, ab6808) and anti-rabbit IgG as secondary antibodies, respectively. For visualization we used Immobilon Western Chemiluminescent HRP substrate (EMD Millipore, WBKLS0100).

2.2.4 Flow cytometry reporter assay

48 h after transfection, cells were trypsinized and collected in a microcentrifuge tube. Cells were centrifuged at 500*g* for 2 min, washed once with 1× PBS, recentrifuged at 500*g* for 2 min and resuspended in 1× PBS for flow cytometry (Becton Dickinson FACScan). For FACS analysis, 10,000 events were counted for each sample. To minimize

the effect of differences in the efficiency of transfection among samples, cells were initially gated for mCherry expression, and the percentage of EGFP-expressing cells (nuclease-positive events) was quantified in mCherry-positive cells. All of the experimental replicates were performed in triplicate on different days; data are reported as mean ± s.e.m.

2.2.5 Genomic targeting analysis with T7EI

72 h after transfection, cells were harvested and genomic DNA was extracted using a DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's suggested protocol. 50 ng of input DNA was PCR-amplified using T7EI primers that were specific for each genomic region (**Appendix 1**) with Phusion High Fidelity DNA Polymerase (New England BioLabs): (98 °C, 15 s; 67 °C, 25 s; 72 °C, 18 s) for 30 cycles. 10 μl of PCR product was hybridized and treated with 0.5 μl of T7EI (New England BioLabs) in 1× NEB Buffer 2 for 45 min[12]. The samples were run on a 2.5% agarose gel and quantified with ImageJ software[355]. Insertion-deletion percentages were calculated as previously described[12]. Experiments for T7EI analysis were performed in triplicate on different days; data are reported as mean ± s.e.m.

2.2.6 Targeted deep-sequencing-based off-target analysis for SpCas9-pDBDs

For the generation of each amplicon, we used two-step PCR amplification to first amplify the genomic segments and then install the barcodes and indexes. In the first step, we used 'locus-specific primers' bearing common overhangs with tails complementary to the TruSeq adaptor sequences (**Appendix 1**). 50 ng of input DNA was PCR amplified with Phusion High Fidelity DNA Polymerase (New England BioLabs): (98 °C, 15 s; 67 °C 25

s; 72 °C 18 s) for 30 cycles. 5 µl of each PCR reaction was gel-quantified by ImageJ[355] against a reference ladder, and equal amounts from each genomic-locus PCR were pooled for each treatment group (15 different treatment groups). The pooled PCR products from each group were run on a 2% agarose gel, and the DNA from the expected product size (between 100 and 200 bp) was extracted and purified with a QIAquick gel extraction kit (Qiagen). In the second step, we amplified the purified pool from each treatment group with a 'universal forward primer' and an 'indexed reverse primer' to reconstitute the TruSeq adaptors. 2 ng of input DNA was PCR amplified with Phusion High Fidelity DNA Polymerase (New England BioLabs): (98 °C, 15 s; 61 °C, 25 s; 72 °C, 18 s) for nine cycles. 5 µl of each PCR reaction was gel-quantified by ImageJ[355], and then equal amounts of the products from each treatment group were mixed and run on a 2% agarose gel. Full-size products (~250 bp in length) were gel-extracted and purified with a QIAquick gel extraction kit (Qiagen). The purified library was deep-sequenced using a paired-end 150-bp MiSeq run.

Sequences from each genomic locus in a specific index were identified on the basis of a perfect match to the final 11 bp of the proximal genomic primer used for locus amplification. Insertions or deletions in the SpCas9 target region were defined on the basis of the distance between a 'prefix' sequence at the 5' end of each off-target site (typically 10 bp) and a 'suffix' sequence at the 3' end of each off-target site (typically 10 bp)[356], where there were typically 33 bp between these elements in the unmodified locus. Distances that were greater than expected were binned as insertions, and distances that were shorter than expected were binned as deletions. Reads that did not contain the suffix

sequence were marked as undefined. For some loci, the background sequencing error rate was high. For example, for OT2-1 a homopolymer sequence in the guide region led to a high error rate.

All statistical analyses were performed using R, a system for statistical computation and graphics[357]. Log-odds ratios of lesions were calculated for the on-target and off-target sites of each individual Cas9 treatment group versus the untreated control for each of the three independent experiments. A *t*-test was applied to assess whether the log-odds ratio was significantly different from 0, that is, whether there was a significant difference in lesion odds between each individual Cas9 treatment group and the untreated control for the on-target and off-target sites. We obtained odds ratios and their 99% confidence intervals by taking the exponents of the estimated log-odds ratios and their 99% confidence intervals. These analyses were also applied to the sum of the lesion rates across all three replicates (combined). To adjust for multiple comparisons, we adjusted *P* values using the Benjamini-Hochberg method[358]. Only loci that had significant Benjamini-Hochberg–adjusted *P* values in the combined data for the treatment group relative to the control were considered significant.

2.2.7 GUIDE-seq off-target analysis for SpCas9-pDBDs

We performed GUIDE-seq with some modifications to the original protocol[286]. Importantly, there is an error in the original publication with regard to the GSP1 and GSP2 primer sets, which list incompatible combinations. It was necessary to properly sort the primer sets for the positive (+) and negative (–) strands to achieve successful library amplification.

Nuclease_off_+_GSP1

GGATCTCGACGCTCTCCCTGTTT**AATTGAGTTGTCATATGTTAATAAC** +

Nuclease_off_–_GSP1

GGATCTCGACGCTCTCCCTATACCGTTATTAACATATGACA –

Nuclease_off_+_GSP2

CCTCTCTATGGGCAGTCGGTG**ATTTGAGTTGTCATATGTTAATAAC**GGTA +

Nuclease_off_–_GSP2

CCTCTCTATGGGCAGTCGGTGATACATATGACAACTCAATTAAAC –

In addition, our protocol differed from the published protocol[286] in the following manner: in a 24-well format, HEK293T cells were transfected with 250 ng of Cas9, 150 ng of sgRNA, 50 ng of GFP and 10 pmol of annealed GUIDE-seq oligonucleotide using Lipofectamine 3000 transfection reagent (Invitrogen) according to the manufacturer's suggested protocol. 48 h after transfection, genomic DNA was extracted with a DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's suggested protocol. Library preparations were done with original adaptors according to protocols described by Joung and colleagues[286], with each library barcoded for pooled sequencing. The barcoded, purified libraries were deep-sequenced as a pool using two paired-end 150-bp MiSeq runs.

Reads containing identical molecular indices and identical starting 8-bp elements on read 1 were pooled into one unique read. The initial 30 bp and the final 50 bp of the unique read 2 sequences were clipped for removal of the adaptor sequence and low-quality sequences and then mapped to the human genome (hg19) using Bowtie2. Peaks containing mapped unique reads were identified using the pile-up program ESAT

(http://garberlab.umassmed.edu/software/esat/) with a window of 25 bp with a 15-bp overlap. Neighboring windows that were on different strands of the genome and less than 50 bp apart were merged using Bioconductor package ChIPpeakAnno[359, 360]. Peaks that were present with multiple different guides (hot spots[286]) or that did not contain unique reads for both sense and antisense libraries[286] were discarded. The remaining peaks were searched for sequence elements that were complementary to the nuclease target site using CRISPRseek[278]. Only peaks that harbored a sequence with fewer than seven mismatches to the target site were considered potential off-target sites. These regions are reported in **Appendix 2**, and the numbers of reads from the sense and the antisense libraries were combined into the final read number.

2.2.8 CRISPRseek analysis of potential off-target site for SpCas9 sgRNAs

Human hg19 exon and promoter sequences were fetched using Bioconductor packages ChIPpeakAnno[359, 360] and TxDb.Hsapiens.UCSC.hg19.knownGene. A subset of 16,500 exons and 192 promoter sequences of 2 kb each were selected for sgRNA searching and genome-wide off-target analysis using Bioconductor package CRISPRseek[278, 343] with the default settings (both NGG and NAG PAMs were allowed), except BSgenomeName = BSgenome.Hsapiens.UCSC.hg19, annotateExon = FALSE, outputUniqueREs = FALSE, exportAllgRNAs = "fasta" and fetchSequence = FALSE. After sgRNAs with on-targets or/and off-targets in the haplotype blocks had been excluded, there were 124,793 unique sgRNAs from exon sequences and 55,687 unique guide RNAs from promoter sequences included in the analysis. Each guide was binned on the basis of either the off-target site with the fewest mismatches to the guide sequence or the sum of the off-target scores for

the top ten off-target sites. The fraction of guides in each bin for exons or promoters is displayed as a pie chart (**Figure 2.1**).

2.3 Results

2.3.1 Defining the properties of the SpCas9-pDBD framework

To define the parameters necessary for the function of a SpCas9-pDBD chimera, we assayed the cleavage of a Cas9:sgRNA target site with a suboptimal NAG PAM using a plasmid reporter assay[354] (**Figure 2.3**). We examined the ability of a pDBD fused (ZFP or TALE) to SpCas9 to enhance nuclease activity when the pDBD binding sites were located at different positions and orientations relative to the Cas9 target site (**Figure 2.4a**). In pilot experiments, we observed the most robust activity when we used a C-terminal fusion of a ZFP or a TALE to SpCas9 when the pDBD binding sites were positioned 3' to the PAM element (M.F.B. and S.A.W., unpublished results). Both SpCas9-ZFP and SpCas9-TALE proteins increased nuclease activity on an NAG PAM target to a level similar to the activity of wild-type SpCas9 on an NGG PAM (**Figure 2.4a**) while being expressed at similar levels (**Figure 2.4b**). SpCas9-pDBD nuclease activity remained dependent on the length of the guide sequence (**Figure 2.4c**), confirming that the chimera retained the guide-dependent licensing stage for sequence cleavage. To define the functional PAM elements for SpCas9-pDBD, we examined activity at each of the 16 possible sequence combinations. In contrast to wild-type SpCas9, SpCas9-pDBD demonstrated high activity for NAG, NGA and NGC as well as for the standard NGG PAM (**Figure 2.4d-e**). Accounting for reverse complements of the functional PAM elements, the

SpCas9-pDBD chimeras recognized 7 of the 16 possible dinucleotide sequence combinations. The increased targeting range for SpCas9-pDBDs was also observed at genomic target sites (**Figure 2.5**). Because of the smaller size of SpCas9-ZFPs relative to SpCas9-TALEs, which confers advantages for certain viral-delivery systems[361], we focused primarily on SpCas9-ZFP chimeras for the immediate development of this platform.

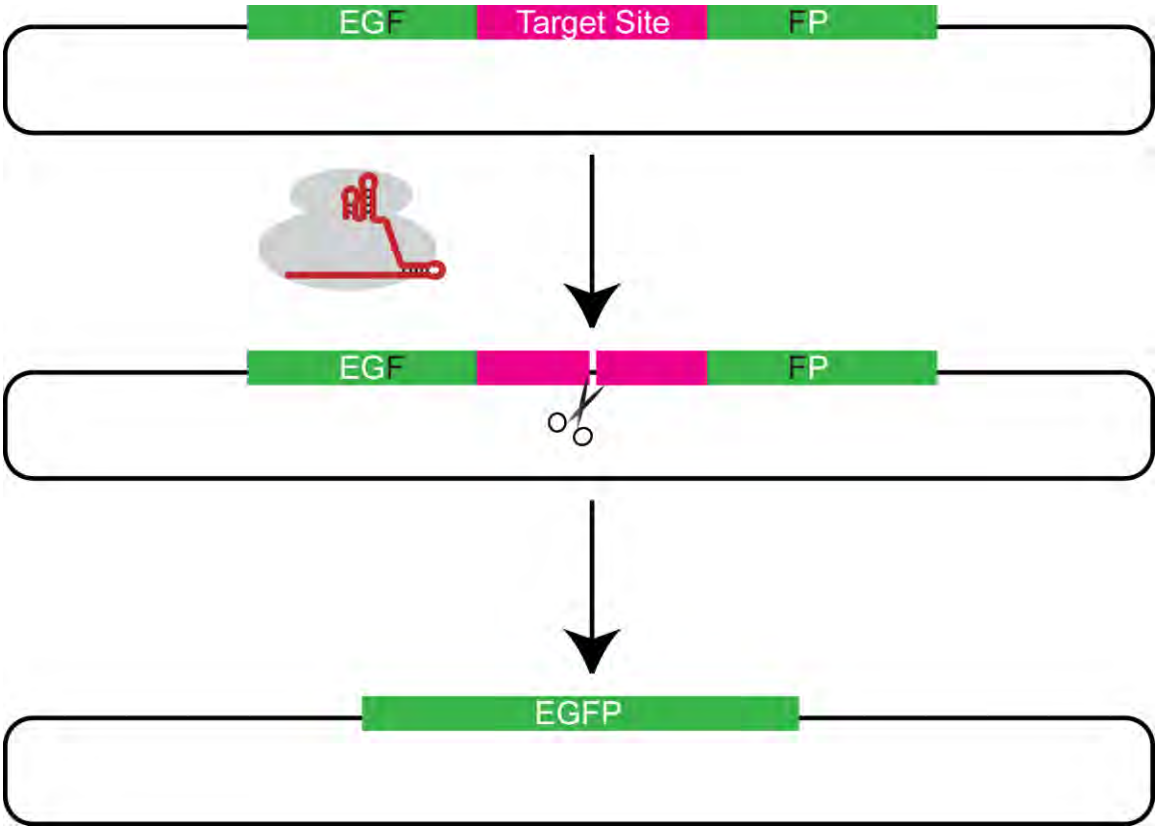


Figure 2. 3 Schematic overview of plasmid-based nuclease activity reporter assay
The reporter plasmid carriers broken EGFP expression construct. A programmable nuclease target site is cloned in between two repeated sequences (indicated by "F"). Once the nuclease creates a double-strand break at the target site on the plasmid DNA, the DSB is repaired via cellular single-strand annealing pathway. This results in functional restoration of the EGFP reporter gene.

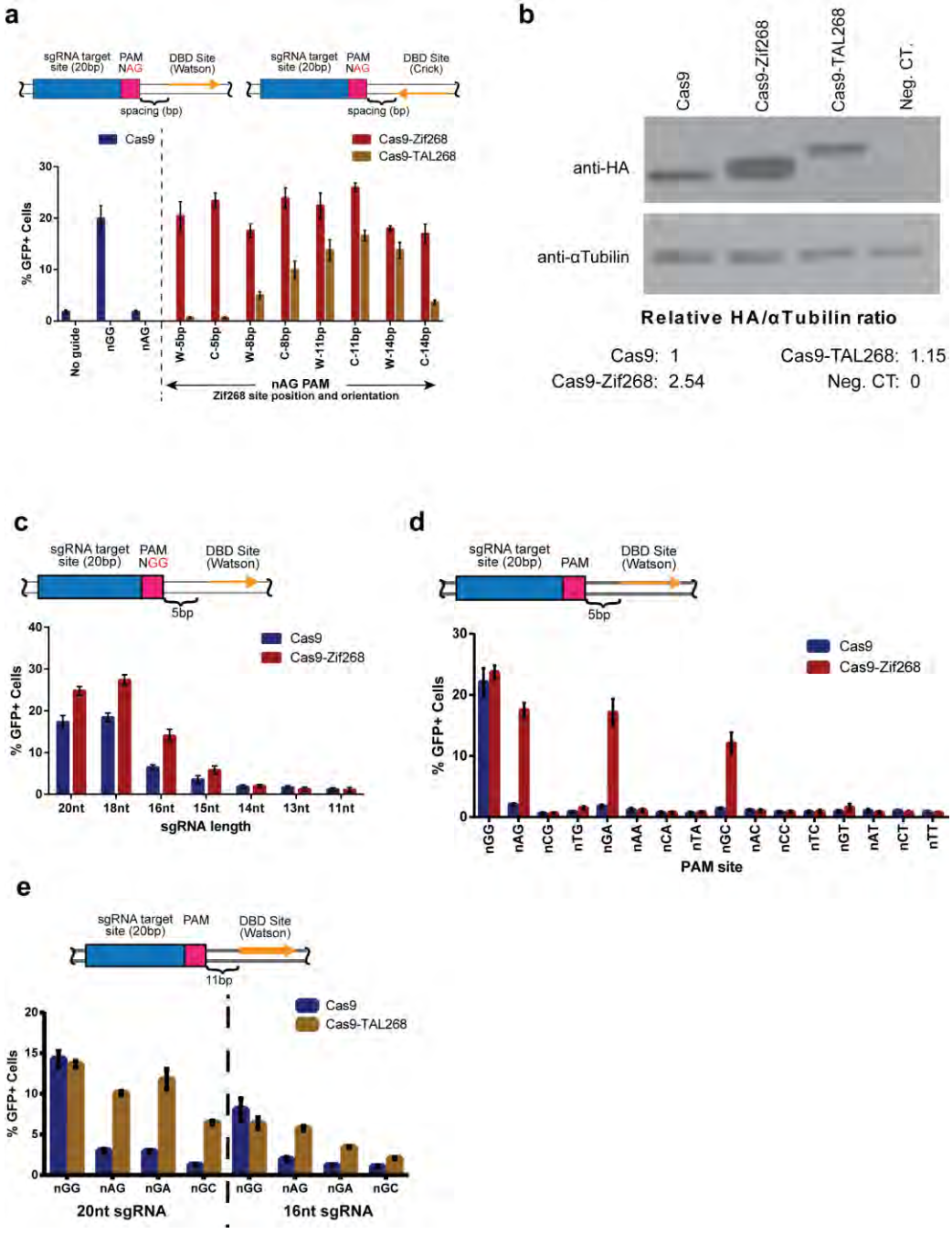

a
sgRNA target site (20bp)
PAM NAG
DBD Site (Watson)
DBD Site (Crick)
spacing (bp)
Cas9
Cas9-Zif268
Cas9-TAL268
% GFP+ Cells
No guide
nGG
nAG
W-5bp
C-5bp
W-8bp
C-8bp
W-11bp
C-11bp
W-14bp
C-14bp
nAG PAM
Zif268 site position and orientation
b
Cas9
Cas9-Zif268
Cas9-TAL268
Neg. CT.
anti-HA
anti-αTubilin
Relative HA/αTubilin ratio
Cas9: 1
Cas9-TAL268: 1.15
Cas9-Zif268: 2.54
Neg. CT: 0
c
sgRNA target site (20bp)
PAM NGG
DBD Site (Watson)
5bp
Cas9
Cas9-Zif268
% GFP+ Cells
20nt
18nt
16nt
15nt
14nt
13nt
11nt
sgRNA length
d
sgRNA target site (20bp)
PAM
DBD Site (Watson)
5bp
Cas9
Cas9-Zif268
% GFP+ Cells
nGG
nAG
nCG
nTG
nGA
nAA
nCA
nTA
nGC
nAC
nCC
nTC
nGT
nAT
nCT
nTT
PAM site
e
sgRNA target site (20bp)
PAM
DBD Site (Watson)
11bp
Cas9
Cas9-TAL268
% GFP+ Cells
nGG
nAG
nGA
nGC
nGG
nAG
nGA
nGC
20nt sgRNA
16nt sgRNA

Figure 2. 4 Development of an SpCas9-pDBD framework
a) Top, schematic of the pDBD binding-site orientation and spacing parameters examined. The position and 5'-to-3' orientation of the pDBD binding site relative to the PAM element of the SpCas9 binding site are represented by orange arrows (Watson (W) and Crick (C)). Bottom, activity profiles of SpCas9 (blue, on an NGG or NAG PAM), SpCas9-Zif268 (red, NAG PAM) and SpCas9-TAL268 (brown, NAG PAM) in the GFP reporter assay on a common sgRNA target site. The pDBD-site orientation was either W or C, and spacing was 5, 8, 11 or 14 bp from the PAM. **b)** HEK293T cells are transfected with the indicated Cas9 plasmid (see methods for details), which has triple HA-tag. (Top) Full length protein is probed with anti-HA antibody. (Bottom) Alpha-tubulin is used as loading control (Neg. CT; negative control). **C**) Activity profile of SpCas9 (blue) and SpCas9-Zif268 (red) in the GFP reporter assay[354] with sgRNAs of various lengths truncated from the 5' end of the guide and an nGG PAM target site. **d)** Activity profile of SpCas9 (blue) and SpCas9-Zif268 (red) in the GFP reporter assay on a common target site with different PAM sequences and a neighboring Zif268 site. **e)** Activity profile of SpCas9 (blue) and SpCas9-TAL268 (brown) in the GFP reporter assay with sgRNAs of 20nt vs 16nt lengths on nGG, nAG, nGA, nGC PAM target sites. Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m (standard error of the mean).

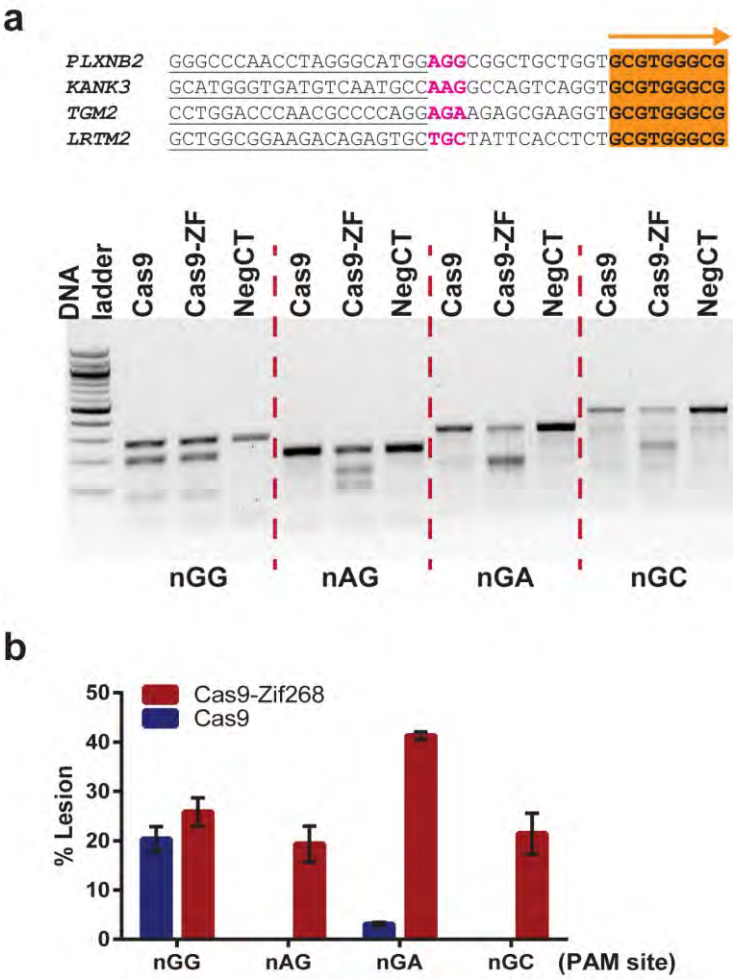


Figure 2. 5 Activity profile of SpCas9-pDBDs at target sites with different PAMs
a) Top, SpCas9 or SpCas9-Zif268 programmed independently with four different sgRNAs targeting four different genomic sites with neighboring Zif268 binding sites (highlighted in orange). Bottom, SpCas9 cut efficiently only at the target site with an NGG PAM, but SpCas9-Zif268 cut efficiently at additional target sites with NAG, NGA and NGC PAMs. **b)** Genomic regions were PCR amplified, and lesions (indicating cleavage and mutagenic non-homologous end joining) were detected by T7EI assay. Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

2.3.2 Attenuating the DNA-binding activity of SpCas9

The fusion of a pDBD to SpCas9 should increase nuclease precision if target cleavage is dependent on DNA recognition by the pDBD. To test this, we attenuated the DNA-binding affinity of SpCas9 by independently mutating the key PAM recognition residues[205] (Arg1333 and Arg1335) to either lysine or serine (**Figure 2.6a**). In the plasmid reporter assay, all four mutations decreased the nuclease activity of SpCas9 to background levels. A ZFP fusion in the presence of a complementary binding site restored nuclease activity in all mutants except R1335S (SpCas9^{MT4}) (**Figure 2.6b**). We found that R1335K (SpCas9^{MT3}) lacked activity with the NAG PAM even as a SpCas9-ZFP fusion. This prompted a broader assessment of PAM specificity for the three active SpCas9-ZFP mutants, which revealed a preference for alternate PAMs that preserved the remaining arginine-guanine interaction[205] (i.e., R1333 mutants preferred NNG PAMs, whereas the R1335K mutant preferred NGN PAMs; **Figure 2.6c**). The activity of each SpCas9 mutant was also characterized on compatible genomic target sites with an NGG PAM. R1333K (SpCas9^{MT1}) retained independent activity on a subset of target sequences, whereas R1333S (SpCas9^{MT2}) and R1335K (SpCas9^{MT3}) showed only background activity, which could be restored to wild-type levels in the presence of a ZFP fusion (**Figure 2.7** and **Figure 2.8**). To confirm that the ZFP-dependent restoration of activity was general, we assessed the nuclease activity of three additional SpCas9^{MT3}-ZFP fusions, two of which restored nuclease function (**Figure 2.9** and **Table 2.1**). Thus, altering the affinity of PAM recognition through mutation generated SpCas9 variants that were dependent on the attached pDBD for efficient function. This pDBD dependence established a third stage of

target-site licensing for our SpCas9^{MT3}-pDBDs, which we believed would increase their precision.

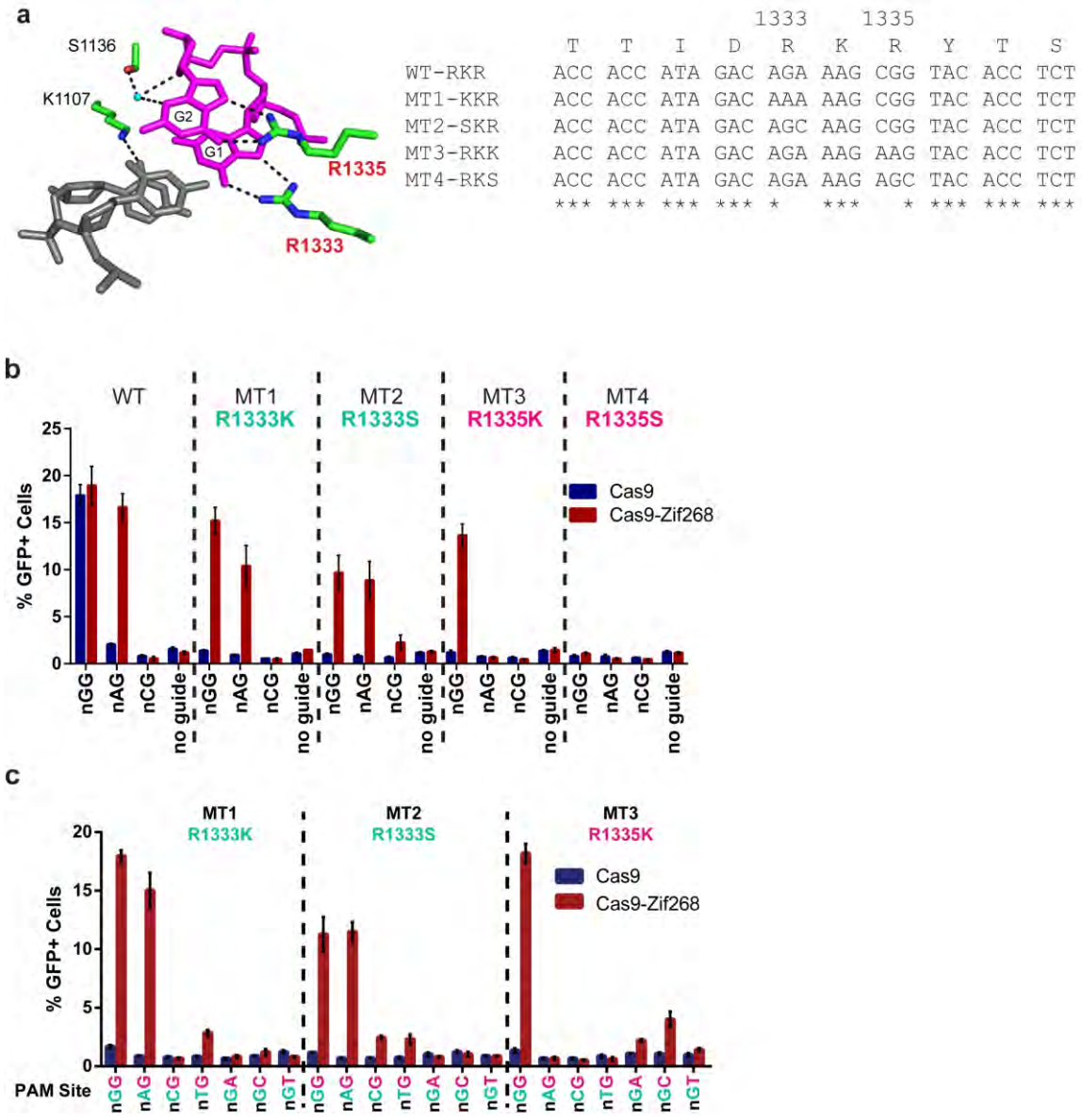


Figure 2. 6 Attenuating the nuclease activity of SpCas9
a) Left, four PAM-interacting amino acids neighboring the NGG PAM (magenta) in the structure of SpCas9[205]. Arginines at positions 1,333 and 1,335 were mutated to attenuate the DNA-binding affinity of SpCas9. Right, local sequences of the PAM interacting domain mutants at positions 1333 or 1335 of SpCas9 examined in this study. **b)** Activity profiles of SpCas9 (blue) and SpCas9-Zif268 (red) bearing lysine and serine substitutions at positions 1,333 and 1,335 in the PAM-interaction domain in comparison to wild-type (WT) SpCas9. **b)** Analysis of SpCas9 mutant activity on different nGn or nnG PAM-containing target sites in the GFP reporter assay. Mutations that alter the interaction of R1333 with its guanine contact (nGn, teal) reveal modest activity at nnG PAMs. Correspondingly, mutations that alter the interaction of R1335 with its guanine contact (nnG, magenta) reveal modest activity at nGn PAMs. Reporter assays were performed in HEK293T cells. Bars represent means from three independent biological replicates performed on different days. Error bars indicate ±s.e.m.

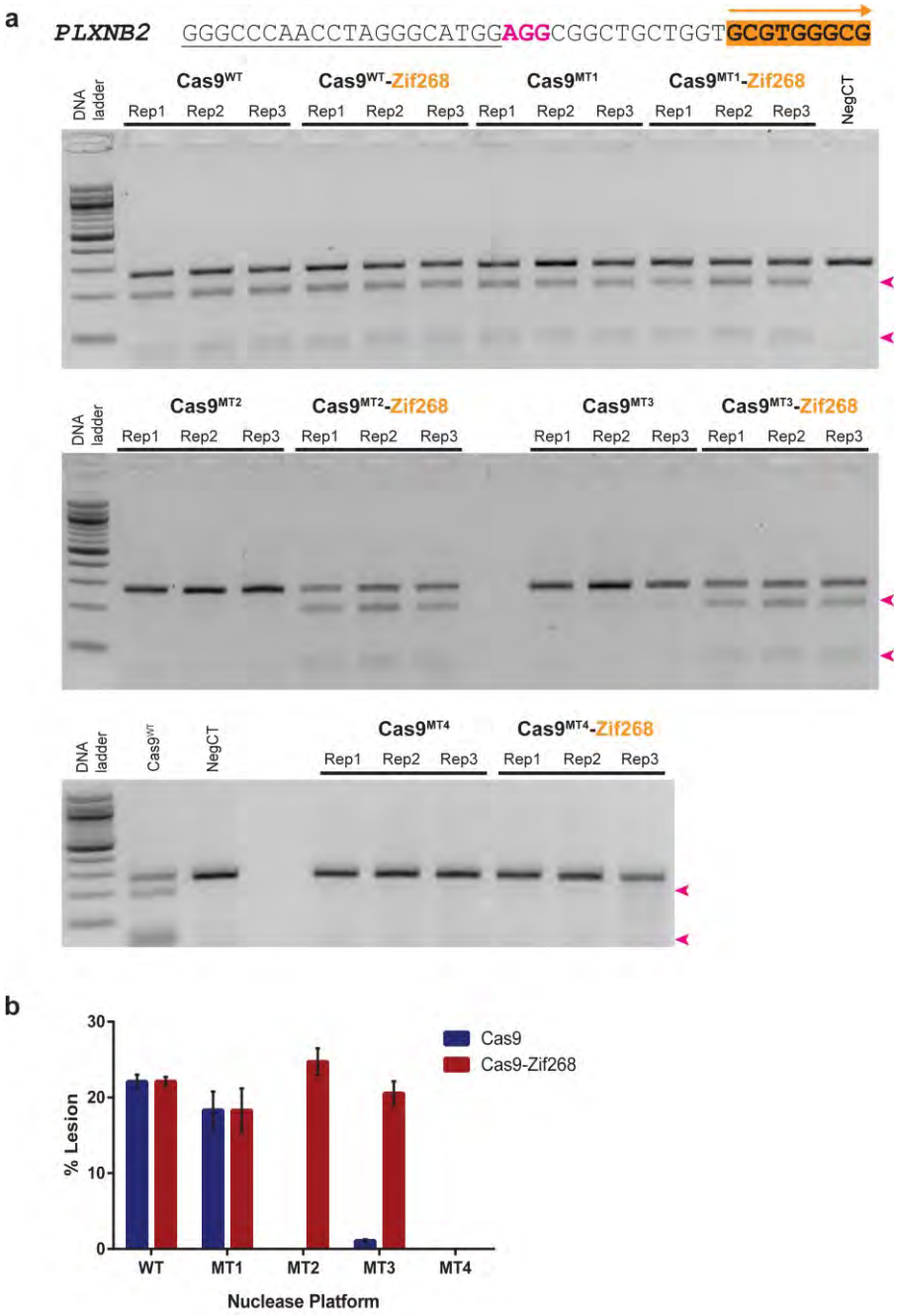


Figure 2. 7 Genomic activity profile of SpCas9 mutants

a) Analysis of the genomic activity profile of SpCas9 mutants (MT1, MT2, MT3 & MT4) independently and as SpCas9-Zif268 fusions at the *PLXNB2* locus at a target site with an nGG PAM and a Zif268 binding site 11 bp away on the Watson strand. T7EI assay data from PCR products spanning the target site in three independent biological replicates (Rep1, Rep2, Rep3) performed on different days in HEK293T cells. Cleaved products are indicated by magenta arrowheads. **b)** Quantification of average T7EI-based lesion rates. Bars represent means from three independent biological replicates. Error bars indicate ±s.e.m.

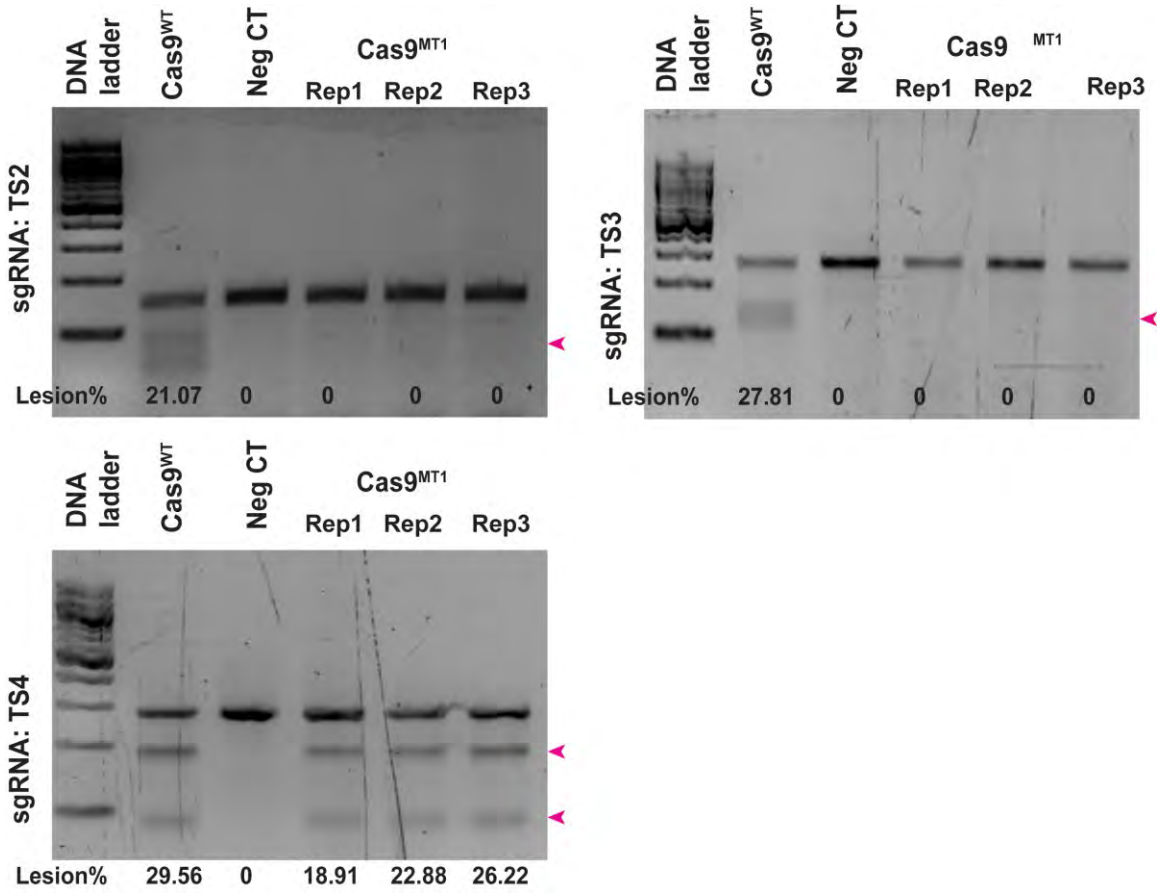


Figure 2. 8 Analysis of the genomic activity profile of SpCas9^{MT1}
T7EI assay data from PCR products spanning the target site in three independent biological replicates (Rep1, Rep2, Rep3) performed on different days in HEK293T cells. Cleaved products are indicated by magenta arrowheads.

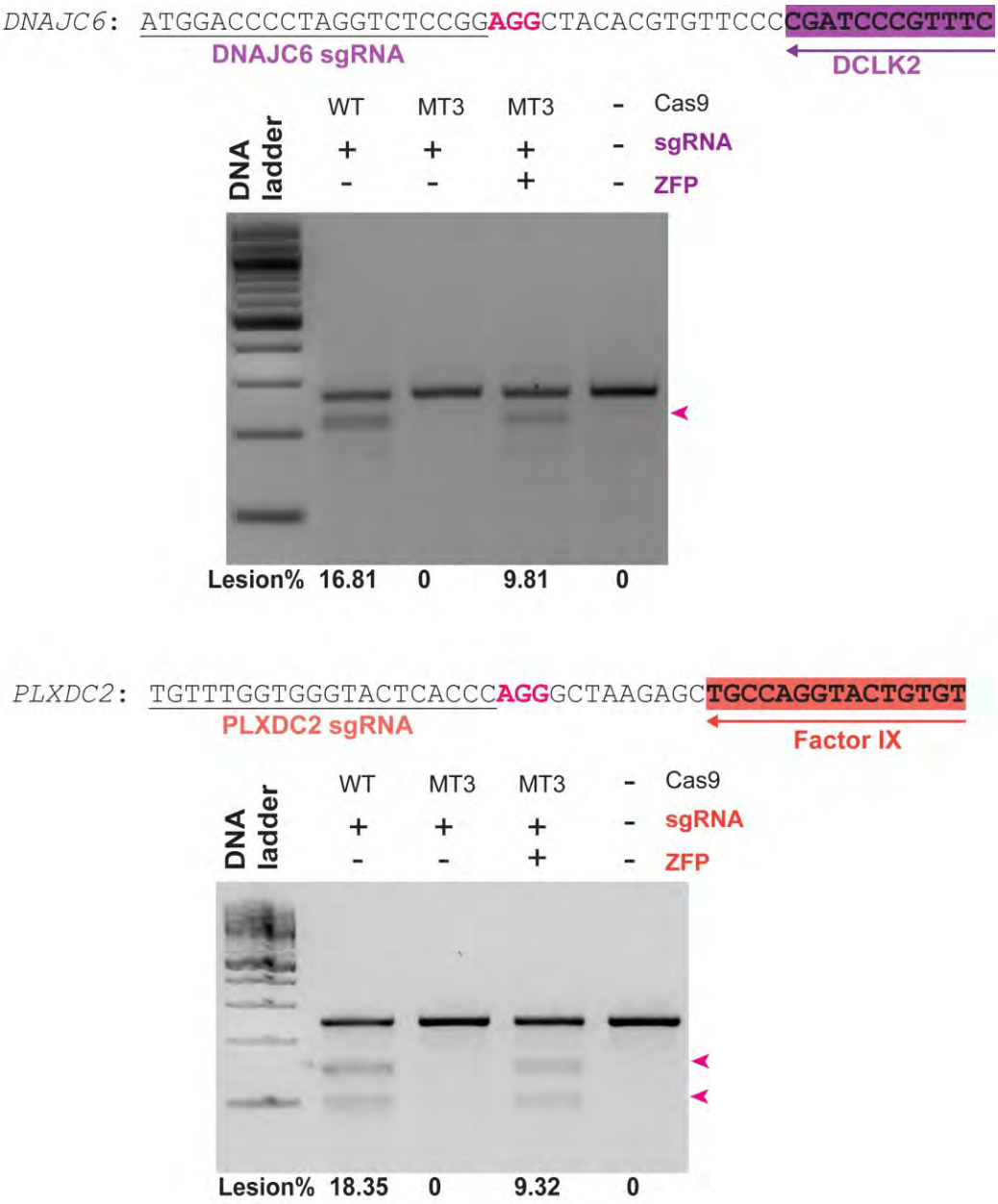


Figure 2. 9 Analysis of the genomic activity of other SpCas9^{MT3}-ZFPs
Activity of SpCas9^{MT3}-ZFP^{DCLK2} and SpCas9^{MT3}-ZFP^{F9} at DNAJC6 and PLXDC2 sites respectively. These sequences have compatible binding sites for the DCLK2[35] and Factor IX[107] ZFPs. T7EI assay data from PCR products spanning the target site from single experiment done in HEK293T cells. Cleaved products are indicated by magenta arrowheads. Similar analysis of SpCas9^{MT3}-ZFP^{HEBP2} (targeting a compatible binding site for the HEBP2 ZFP[36]) at GPRC5B did not detect any lesions for this SpCas9MT3-ZFP fusion (data not shown).

Table 2.1: Summary of SpCas9^{MT3}-pDBD nuclease activities (T7EI)

pDBD Name	Type	Target Sequence	sgRNA	Activity (%Lesion)
ZFP^{TS2}	4 Finger ZFP	GCGGGCAGGGGC	TS2	36.64
ZFP^{TS2*}	4 Finger ZFP	GCAGGGGCCGGA	TS2	23.04
ZFP^{TS3}	4 Finger ZFP	GGCGTTGGAGCG	TS3	26.75
ZFP^{TS4}	4 Finger ZFP	CCGGTTGATGTG	TS4	12.86
Zif268	3 Finger ZFP	GCGTGGGCG	PLXNB2	25.81
ZFP^{DCLK2}	4 Finger ZFP	GAAACGGGATCG	DNAJC6	9.32
ZFP^{FactorIX}	5 Finger ZFP	ACACAGTACCTGGCA	PLXDC2	9.90
ZFP^{HEBP2}	4 Finger ZFP	GAAAAGTATCAA	GPRC5B	N.D
TAL268	8.5 Module TALE	TGCGTGGGCG	PLXNB2	N.D
TALE^{TS3-S}	9.5 Module TALE	TTGGAGCGGGG	TS3	8.00
TALE^{TS3-L}	15.5 Module TALE	TTGGAGCGGGGAGAAGG	TS3	16.26
TALE^{TS4-S}	9.5 Module TALE	TCAACCGGTGG	TS4	2.01**
TALE^{TS4-L}	15.5 Module TALE	TCAACCGGTGGCGCATT	TS4	1.82**

N.D: Not Detected

**: Not above background independent activity for SpCas9^{MT3}.

Table 2.1: Summary of the SpCas9^{MT3}-pDBDs tested in this study.
7 out of 8 ZFPs and 1 out of 3 TALEs are able to restore nuclease activity to the SpCas9^{MT3} mutant when fused to the C-terminus. Genomic nuclease activities are measured by T7EI assay.

2.3.3 Assessing the precision of SpCas9^{MT3}-ZFP fusions

Next we compared the precision of SpCas9^{MT3}-ZFPs to that of SpCas9 using sgRNAs with previously defined off-target sites[257, 293]. We programmed three different four-finger ZFPs to recognize 12-bp sequences neighboring sgRNA target sites 2, 3 and 4 (TS2, TS3 and TS4, respectively) for use as SpCas9^{MT3}-ZFP fusions (**Figure 2.10a**). The activity of SpCas9, SpCas9^{MT3} and SpCas9^{MT3}-ZFP with the corresponding sgRNA was compared at each target site. In all cases SpCas9^{MT3} dramatically decreased the cleavage efficiency, which was restored by the cognate ZFP fusion (**Figure 2.10b**). The activity of SpCas9^{MT3}-ZFP was dependent on the presence of both a cognate sgRNA and a cognate ZFP (**Figure 2.10c**). Consistent with the dependence on ZFP binding, truncation of one zinc finger from either end of ZFP^{TS3} decreased the activity of SpCas9^{MT3}-ZFP^{TS3} at TS3, and the removal of two zinc fingers abrogated activity (**Figure 2.11**). The additional stage of target-site licensing supplied by the pDBD dramatically increased the precision of SpCas9^{MT3}-ZFP^{TS3} relative to that of wild-type SpCas9; lesion rates for sgRNA^{TS3} at off-target site 2 (OT3-2), the most active off-target site, were 22% by T7 endonuclease I (T7EI) assay with wild-type Cas9 but were undetectable with SpCas9^{MT3}-ZFP^{TS3} (**Figure 2.10c**). We also programmed two TALE arrays to target SpCas9^{MT3} to TS3 and TS4 (TALE^{TS3} and TALE^{TS4}). Nuclease activity at TS3, but not at TS4, was restored by the related SpCas9^{MT3}-TALE fusion (**Figure 2.12**).

To examine the catalytic tolerance of the SpCas9^{MT3}-ZFP^{TS3}:sgRNA complex for mismatches between the guide and a target sequence, we used a set of guides that progressively shifted blocks of two base mismatches from the 5’ to the 3’ end of the guide

sequence. SpCas9^{MT3}-ZFP^{TS3} had a lower tolerance for mismatches between the guide and target site than did wild-type SpCas9 (SpCas9^{WT}), whereas SpCas9^{WT}-ZFP^{TS3} seemed to have a modestly increased tolerance for mismatches (**Figure 2.10d**). SpCas9^{MT3}-ZFPs also exhibited decreased activity with truncated sgRNAs^{293} (**Figure 2.13**), consistent with the requirement for a higher degree of guide–target site complementarity to achieve efficient cleavage.

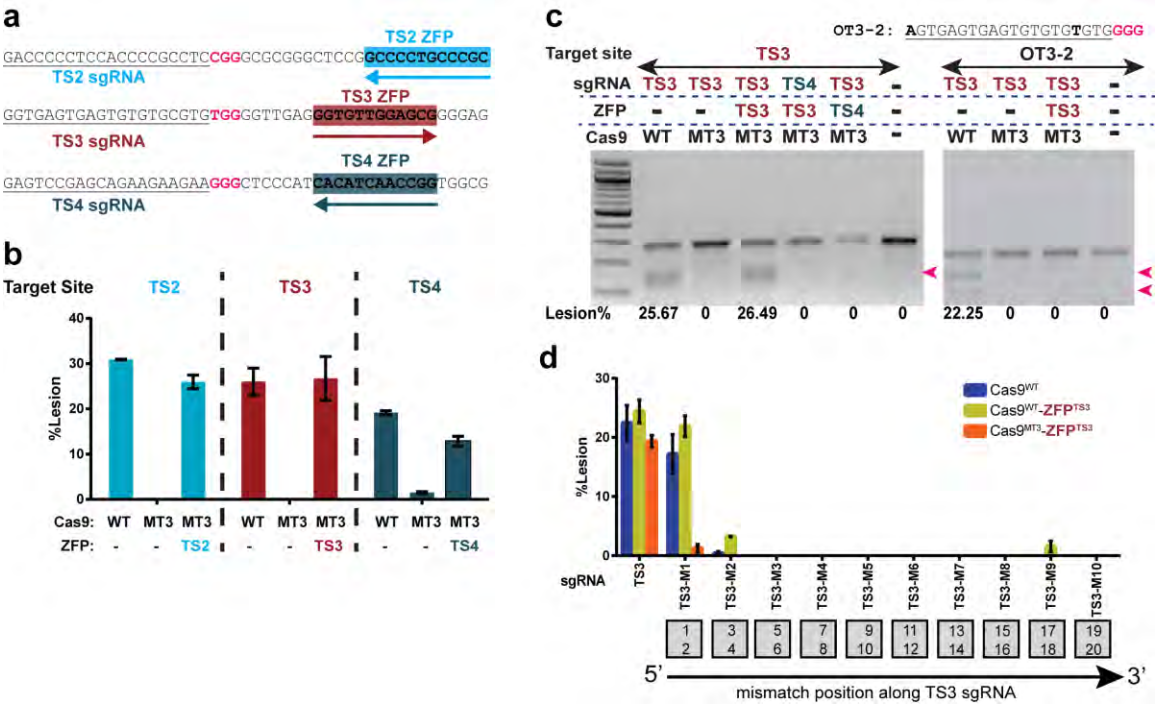


Figure 2. 10 SpCas9^{MT}-ZFP chimeras have improved precision

a) Sequences of TS2, TS3 and TS4 for the SpCas9:sgRNAs described by Joung and colleagues[257, 293]. The 12-bp ZFP binding sites for TS2, TS3 and TS4 are highlighted in cyan, red and teal, respectively, with arrows indicating the bound strands. **b**) Lesion rates determined by T7EI assay for SpCas9, SpCas9^{MT3} and SpCas9^{MT3}-ZFP at TS2, TS3 and TS4. Data are from three independent biological replicates performed on different days in HEK293T cells. **c**) Representative T7EI assay comparing lesion rates at TS3 and OT3-2[293] for various SpCas9-chimera:sgRNA combinations. The activity at the target site for SpCas9^{MT3}-ZFP was dependent on the cognate sgRNA and ZFP, where SpCas9^{MT3}-ZFP^{TS3} was able to discriminate between TS3 and OT3-2. Cleaved products are indicated by magenta arrowheads. **d**) Genomic target-site cleavage activity by SpCas9, SpCas9^{WT}-ZFP^{TS3} and SpCas9^{MT3}-ZFP^{TS3} in response to dinucleotide mismatches placed at different positions in the guide sequence targeting the TS3 site. Top, T7EI assay data from PCR products spanning the TS3 site in three independent biological replicates performed on different days in HEK293T cells. Bottom, schematic indicating the position of the dinucleotide mismatches across the guide sequence. SpCas9^{MT3}-ZFP^{TS3} demonstrated superior discrimination relative to SpCas9 for dinucleotide mismatches in the sgRNA recognition sequence. Error bars indicate ±s.e.m.

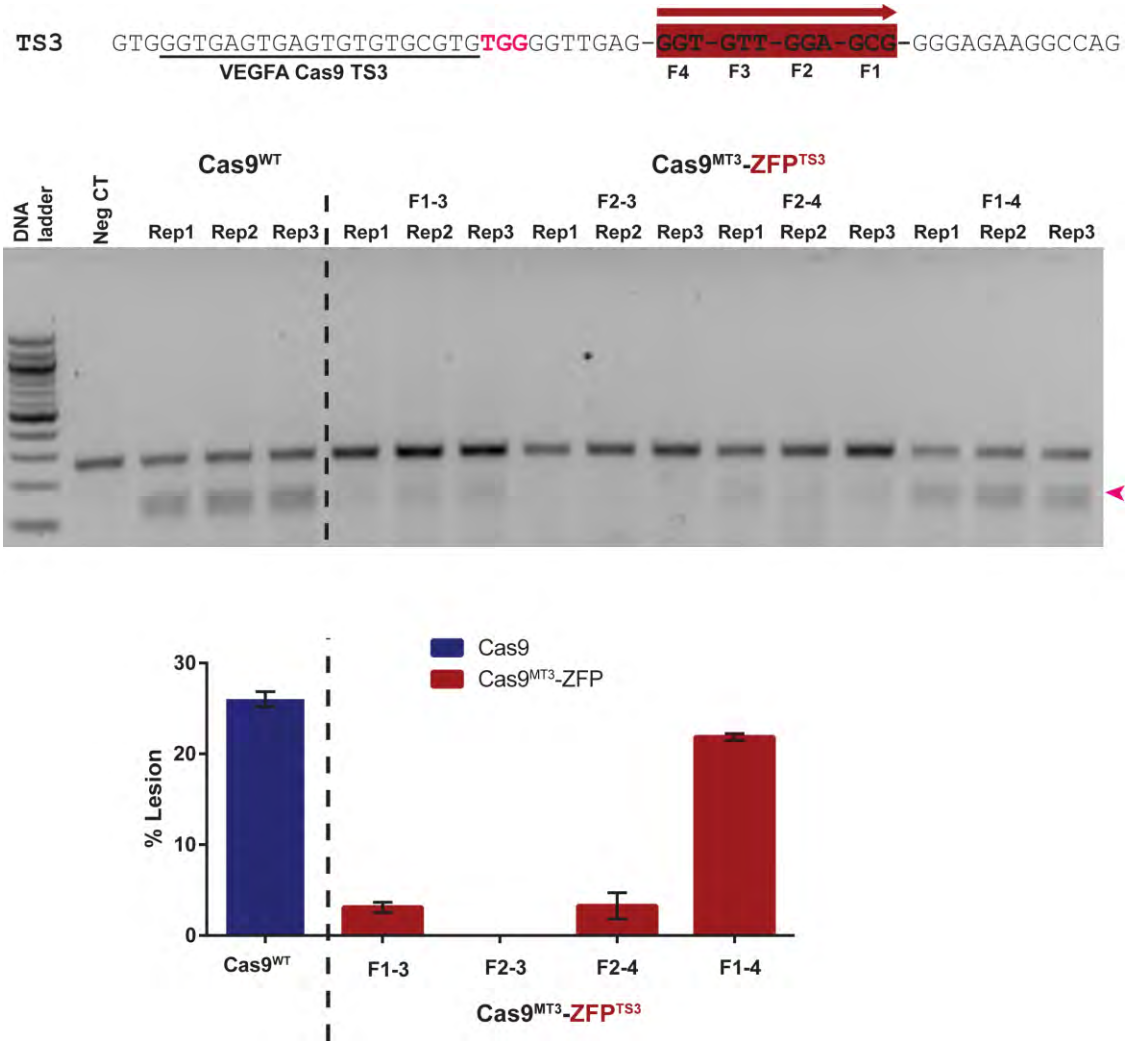


Figure 2. 11 Effect of the number of zinc fingers on SpCas9^{MT3}-ZFP^{TS3} activity
Both Cas9^{WT} and SpCas9^{MT3}-ZFP^{TS3} with four fingers (**F1-4**) achieve efficient target cleavage. Removing a single finger from either end of the zinc finger array (**F1-3** or **F2-4**) dramatically reduces the activity of the SpCas9^{MT3}-ZFP chimera. Cleaved products are indicated by magenta arrowheads. The bar graph displays the mean lesion rate in three independent biological replicates (Rep1, Rep2, Rep3) performed on different days in HEK293T cells. Error bars indicate standard error of the mean.

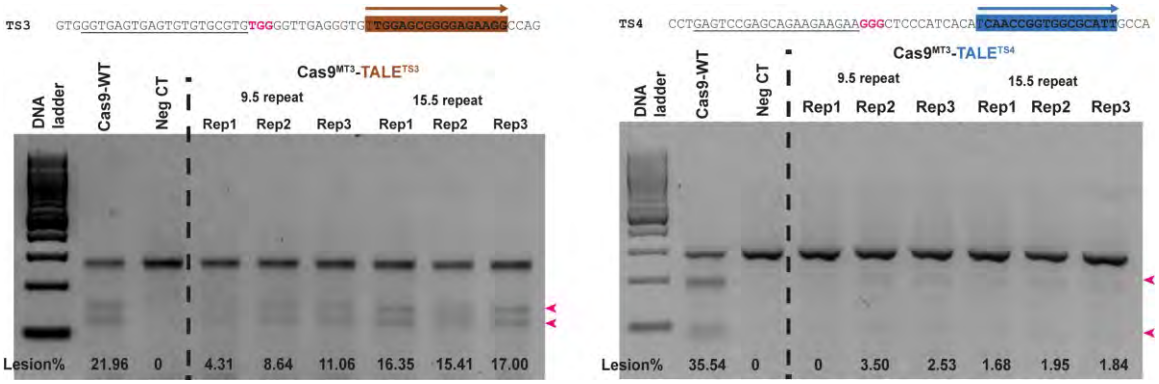


Figure 2. 12 Activity profiles of SpCas9^{MT3}-TALE^{TS3} and SpCas9^{MT3}-TALE^{TS4}
An arrow indicates the strand (Watson) of the highlighted sequence that is bound by the TALE. Two different TALE repeat lengths (9.5 and 15.5) were examined at each target site. T7EI assay data from PCR products spanning the target site in three independent biological replicates (Rep1, Rep2, Rep3) performed on different days in HEK293T cells. Cleaved products are indicated by magenta arrowheads.

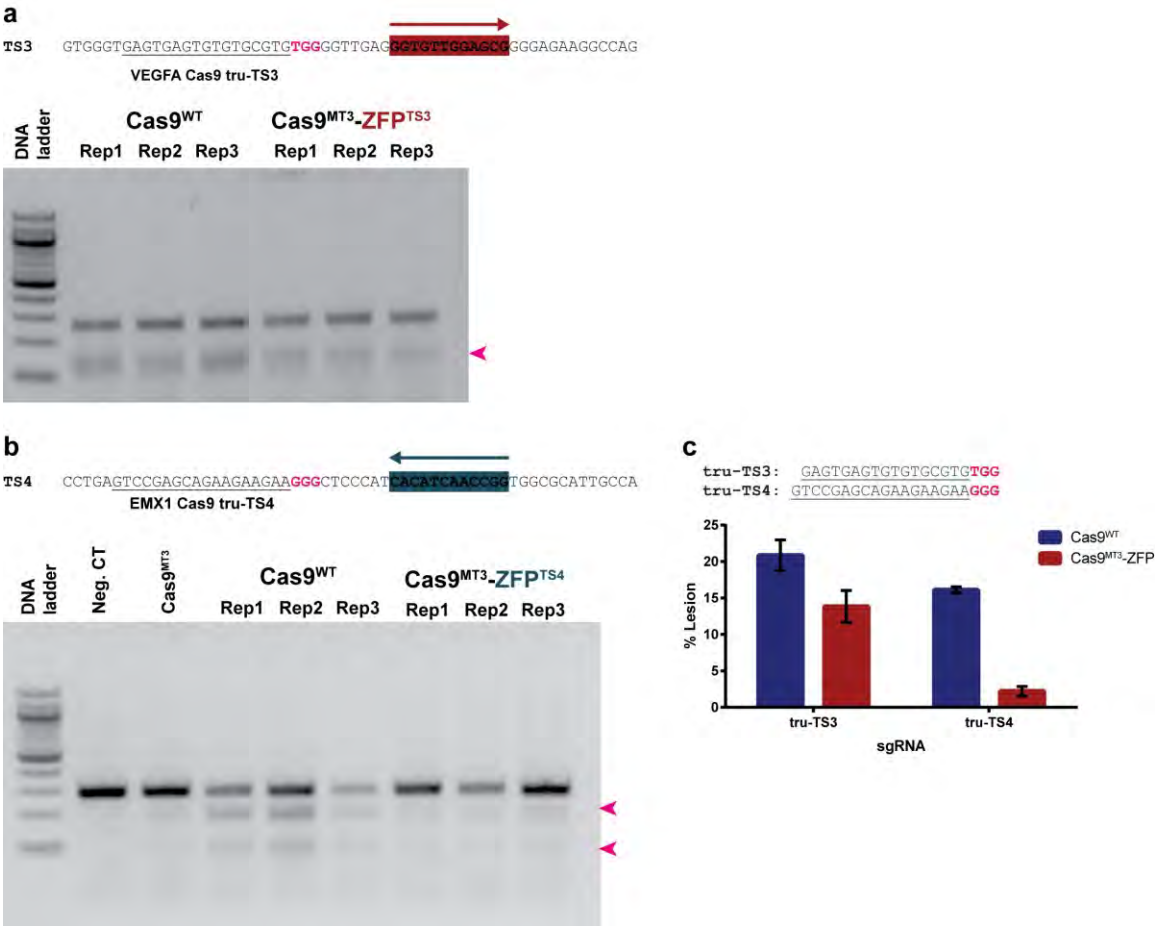


Figure 2. 13 Activity profile of SpCas9^{MT3}-ZFP$^{TS3/TS4}$ with tru-sgRNAs

a) Nuclease activity based on T7EI assay for SpCas9WT and SpCas9^{MT3}-ZFP^{TS3} with a 17 nucleotide truncated guide at the TS3 target site. **b)** Nuclease activity based on T7EI assay for SpCas9WT and SpCas9^{MT3}-ZFP^{TS4} with an 18 nucleotide truncated guide at the TS4 target site. Cleaved products are indicated by magenta arrowheads. **c)** Target sites for the TS3 and TS4 tru-sgRNAs and graph showing the average activity at each target site in three independent biological replicates performed on different days in HEK293T cells. Error bars indicate standard error of the mean. For both TS3 and TS4, the SpCas9^{MT3}-ZFP chimera is more sensitive to the truncation of the guide sequence, which is consistent with the greater sensitivity of this system to guide mismatches.

2.3.4 Deep-sequencing analysis of off-target activity

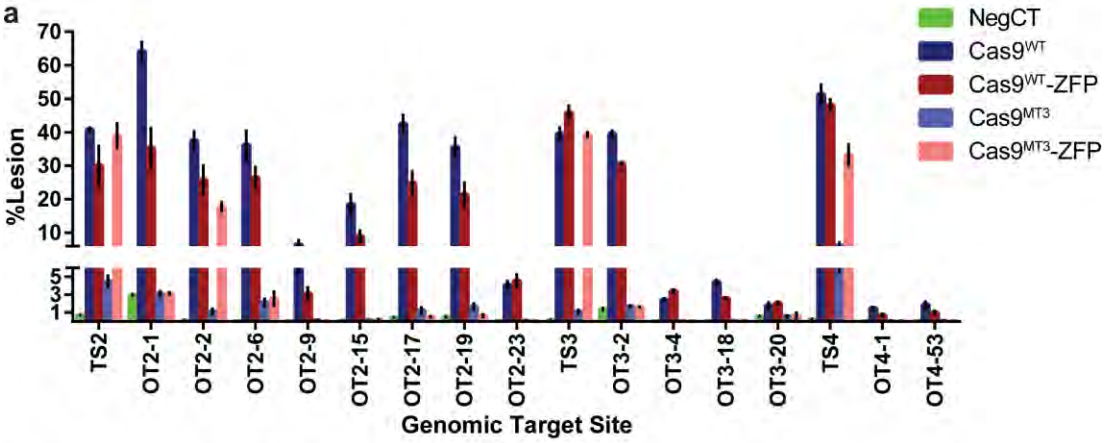
To more broadly assess improvements in precision, we deep-sequenced PCR products spanning previously defined off-target sites for sgRNA$^{TS2:TS3:TS4}$ [257, 293], as well as several additional genomic loci with favorable ZFP$^{TS2:TS3:TS4}$ recognition sites and some complementarity to the TS2, TS3 or TS4 guide sequences. We compared the nuclease activity of SpCas9, SpCas9^{MT3}, SpCas9^{WT}-ZFP$^{TS2:TS3:TS4}$ and SpCas9^{MT3}-ZFP$^{TS2:TS3:TS4}$ at these off-target sites and found that SpCas9^{MT3}-ZFP$^{TS2:TS3:TS4}$ dramatically increased the precision of target-site cleavage (**Figure 2.14a**). In most cases, using SpCas9^{MT3}-ZFP$^{TS2:TS3:TS4}$ decreased lesion rates at off-target sites to background levels, resulting in up to 150-fold improvements in the specificity ratio (**Figure 2.14b**). Only one off-target site (OT2-2), which had a neighboring sequence similar to the expected ZFP^{TS2} recognition sequence (**Figure 2.15a**), still had high lesion rates. One other site (OT2-6) showed some residual activity for both SpCas9^{MT3} and SpCas9^{MT3}-ZFP^{TS2} that was above the background error rate in our sequencing data. Overall, these data demonstrate a dramatic enhancement in precision for SpCas9^{MT3}-ZFPs relative to that of standard SpCas9 at previously defined active off-target sites.

One potential advantage of the SpCas9-pDBD system over other Cas9 platforms is the ability to rapidly tune the affinity and specificity of the attached pDBD to further optimize its precision. Consequently, we sought to improve the precision of SpCas9^{MT3}-ZFP^{TS2} by truncating the ZFP to reduce its affinity for target site OT2-2. Constructs with a truncation of either of the terminal zinc fingers showed high activity at the target site ((**Figure 2.14c**). However, these truncations decreased or eliminated off-target activity at

OT2-2, reflecting a profound improvement in the precision of SpCas9^{MT3}-ZFP^{TS2} (**Figure 2.14c** and **Figure 2.15b-c**). Similarly, utilization of a ZFP (TS2*) that recognized an alternate sequence neighboring the TS2 guide target site also abolished off-target activity at OT2-2, confirming that cleavage by SpCas9^{MT3} at this off-target site was ZFP dependent (**Figure 2.14c**). Given the improvements in precision realized through these simple adjustments in the composition of the ZFP, it should be possible to achieve even greater enhancements in precision via more focused modification of the ZFP composition and the linker connecting it to SpCas9.

Finally, we used genome-wide, unbiased identification of double-stranded breaks enabled by sequencing (GUIDE-seq)[286] to provide an unbiased assessment of the propensity for SpCas9^{MT3}-ZFPs to cleave at alternate off-target sites in the genome (**Figure 2.16**). Using modified versions of the original protocol and bioinformatics pipeline, we assessed genome-wide double-stranded break induction by SpCas9 and SpCas9^{MT3}-ZFP$^{TS2:TS3:TS4}$ (Online Methods). This analysis revealed dramatically enhanced precision of the SpCas9^{MT3}-ZFPs compared with that of SpCas9 for all three target sites (**Figure 2.17a-d** and **Appendix 2**). For SpCas9^{MT3}-ZFP^{TS3} and SpCas9^{MT3}-ZFP^{TS4}, we did not detect nuclease-dependent oligonucleotide capture at any site other than the target site. For SpCas9^{MT3}-ZFP^{TS2}, which retained two active off-target sites that overlapped with SpCas9, cleavage activity at all of the alternate sequences was dramatically decreased compared with that of SpCas9. In addition, there was one new weak off-target site (OTG2-42) for SpCas9^{MT3}-ZFP^{TS2}. These data demonstrate that use of the SpCas9^{MT3}-ZFP fusion

decreased cleavage at wild-type SpCas9 off-target sites without generating a new class of highly active ZFP-mediated off-target sites.



b

Specificty Ratio

Site		Cas9-WT	Cas9-MT3-ZFP	Improvement (Fold)
TS2	OT2-1	0.6	12.3	19
	OT2-2#	1.1	2.2	2
	OT2-6#	1.1	15.4	14
	OT2-9	6.2	973.0	156
	OT2-15	2.2	185.3	84
	OT2-17	1.0	73.4	77
	OT2-19	1.1	58.1	51
	OT2-23	9.9	778.4	79
TS3	OT3-2	1.0	24.5	25
	OT3-4	16.3	1321.7	81
	OT3-18	9.0	660.8	74
	OT3-20	22.2	59.2	3
TS4	OT4-1	34.7	665.2	19
	OT4-53	27.3	475.1	17

#Sites where the Cas9MT3-ZFP lesion rates are above background level

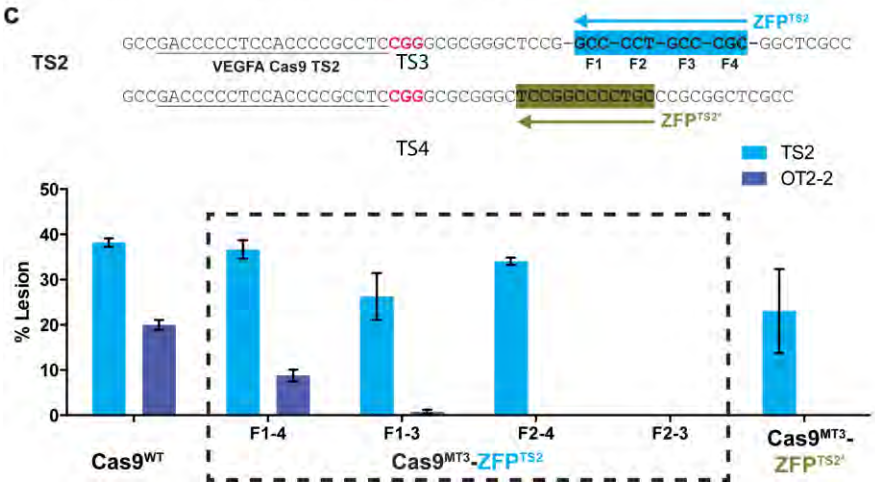


Figure 2. 14 Deep-sequencing analysis of SpCas9MT3-ZFP chimera precision
a) Lesion rates for target sites and off-target sites with statistically significant activity ($P <= 0.01$) assayed by deep-sequencing of PCR products spanning each genomic locus for SpCas9 (blue), $SpCas9^{MT3}$ (light blue), $SpCas9^{WT}$-ZFP (red) and $SpCas9^{MT3}$-ZFP (pink). NegCT indicates untransfected negative control. **b**) Specificity ratio of $SpCas9^{MT3}$-ZFP relative to that of $SpCas9^{WT}$ for the target-site lesion rate relative to each off-target lesion rate (Specificity ratio = (Target site lesion rate)/(Off-target lesion rate)). **c**) Comparison of average lesion rates at TS2 and OT2-2 determined by T7EI assay for $SpCas9^{WT}$ and $SpCas9^{MT3}$-ZFP^{TS2} variants that altered the number of zinc fingers or changed them completely (TS2*). The binding site for ZFP^{TS2*} is indicated in green. Removing finger 1 (F2-4) or finger 4 (F1-3) from the four-finger TS2 ZFP array (F1-4) at most modestly affected the target-site activity, but it dramatically improved precision (dashed box). Data are from three independent biological replicates performed on different days in HEK293T cells (**Figure 2.14**). Error bars indicate ±s.e.m.

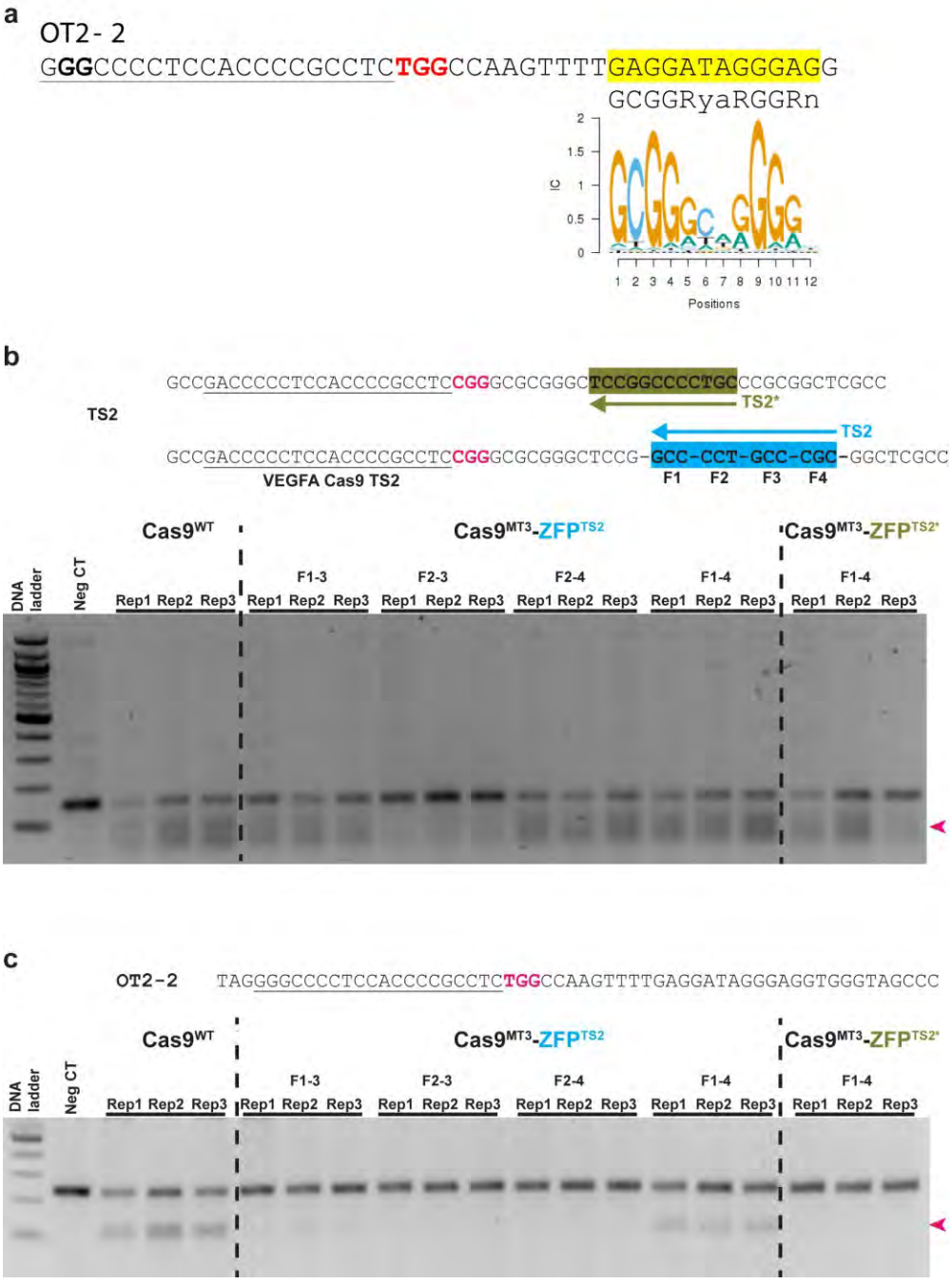

a
OT2- 2
GGGCCCCTCCACCCCGCCTCTGGCCAAGTTTTGAGGATAGGGAGG
GCGGRyaRGGRn
Positions
b
GCCGACCCCCTCCACCCCGCCTCCGGGCGCGGGCTCCGGCCCCTGCCCGCGGCTCGCC
TS2
TS2*
GCCGACCCCCTCCACCCCGCCTCCGGGCGCGGGCTCCG-GCC-CCT-GCC-CGC-GGCTCGCC
VEGFA Cas9 TS2
F1 F2 F3 F4
Cas9WT
Cas9MT3-ZFPTS2
Cas9MT3-ZFPTS2*
DNA ladder
Neg CT
F1-3
F2-3
F2-4
F1-4
Rep1 Rep2 Rep3
c
OT2-2 TAGGGGCCCCTCCACCCCGCCTCTGGCCAAGTTTTGAGGATAGGGAGGTGGGTAGCCC
Cas9WT
Cas9MT3-ZFPTS2
Cas9MT3-ZFPTS2*
DNA ladder
Neg CT
F1-3
F2-3
F2-4
F1-4
Rep1 Rep2 Rep3

Figure 2. 15 SpCas9^{MT3}-pDBD activity can be tuned to further improve precision
a) The sequence complementary to the guide is underlined with the two mismatched positions in bold. The nGG PAM is red and the potential ZFP^{TS2} binding site highlighted in yellow. Below the genomic sequence is predicted consensus recognition motif and sequence logo for ZFP^{TS2} based on a Random Forest model of ZFP recognition[362]. The predicted recognition motif only differs substantially at one position in the finger 4 binding site (C versus A). (**b-c)** T7EI activity profile of SpCas9^{MT3}-ZFP^{TS2} at the TS2 genomic locus and OT2-2 as a function of the number of fingers. **b)** Both Cas9^{WT} and SpCas9^{MT3}-ZFP^{TS2} with four fingers (**F1-4**) result in efficient cleavage at the TS2 target site (magenta arrowheads indicate cleaved products). Removing a single finger from either end of the zinc finger array (**F1-3** or **F2-4**) at most modestly reduces activity of the SpCas9^{MT3}-ZFP chimera. Removing a both terminal fingers from the zinc finger array (**F2-3**) dramatically reduces activity of the SpCas9^{MT3}-ZFP chimera. Construction of an alternate ZFP (TS2*) that recognizes an overlapping target site can also promote target cleavage. **c)** Both Cas9^{WT} and SpCas9^{MT3}-ZFP^{TS2} with four fingers (**F1-4**) result in efficient cleavage at the OT2-2 off-target site (magenta arrowheads indicate cleaved products). Removing a single finger from either end of the zinc finger array (**F1-3** or **F2-4**) dramatically reduces activity of the SpCas9^{MT3}-ZFP chimera. As does the utilization of an alternate ZFP (TS2*) that recognizes a different target site. Data from three independent biological replicates (Rep1, Rep2, Rep3) performed on different days in HEK293T cells.

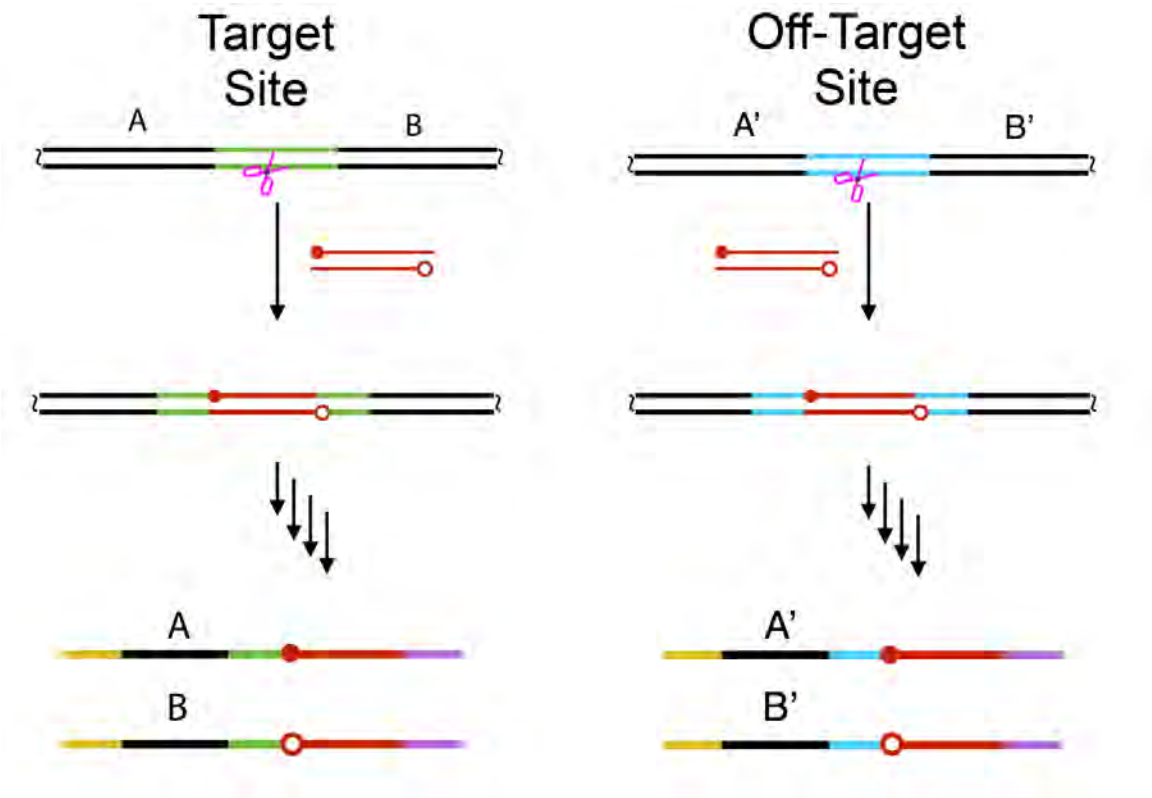


Figure 2. 16 Schematic overview of GUIDE-seq library preparation
In addition to sgRNA and Cas9 expression constructs, cells are provided with end-protected double-stranded, short GUIDE-seq oligonucleotides (red lines with circles). Once Cas9 (magenta scissors) generates DSB(s) at the target site (left, green) or an off-target site (right, cyan), the GUIDE-seq oligo is incorporated at the cleavage sites. To identify genome-wide Cas9 active sites, genomic DNA of the treated cells are isolated, mechanically sheared, end repaired, and ligated to deep sequencing adaptors (yellow and purple). The adaptor-ligated DNA pieces are amplified and enriched within the library prior to deep sequencing.

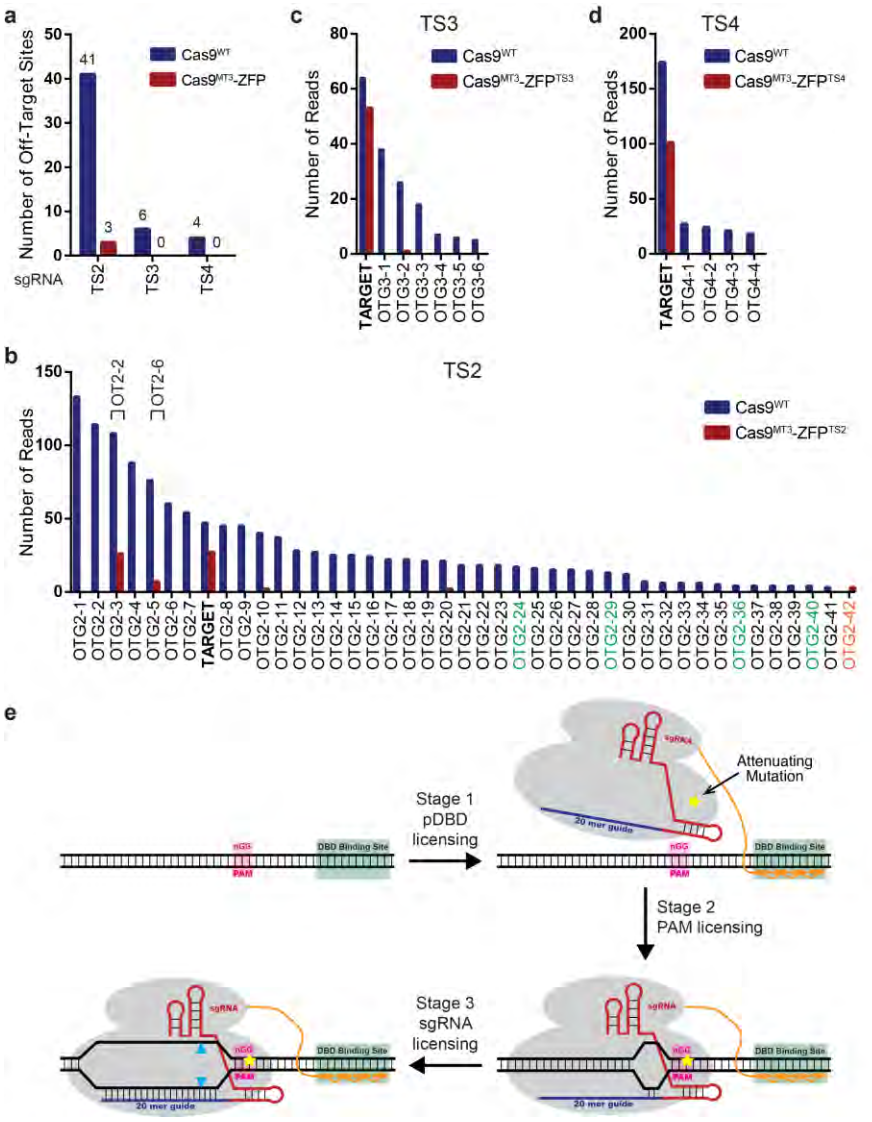


Figure 2. 17 Genome-wide off-target analysis of SpCas9^{MT3}-ZFPs by GUIDE-seq
a) Number of off-target sites with nuclease activity detected for SpCas9^{WT} (blue) and SpCas9^{MT3}-ZFP (red) with TS2, TS3 and TS4 guides. (**b**–**d**) Number of unique reads captured by GUIDE-seq for nuclease active sites in the genome (TS2 **b**), TS3 **c**) and TS4 **d**) target sites (bold) and off-target sites). Previously defined off-target sites are in black font[257, 286], and potential new off-target sites that were identified in this analysis are in green font for SpCas9^{WT} and orange font for SpCas9^{MT3}-ZFP. Some sites (e.g., OGT2-10 and OGT2-20) contained only reads from a single library for SpCas9^{MT3}-ZFP and so are not binned as off-target sites in **a**). Detailed information on the sites and counts is presented in **Appendix 2**. **e**) Model of the three stages of target-site licensing that are necessary for SpCas9^{MT3}-pDBD to cleave DNA. Because of the modification of SpCas9 (mutation, indicated by yellow stars), the efficient engagement of a sequence for PAM recognition or guide RNA complementarity requires the presence of a neighboring DNA sequence that can be bound by the attached pDBD. This requirement for pDBD binding adds a third stage of target-site licensing for efficient cleavage.

2.4 Discussion

Our analysis of the activity of SpCas9-pDBD chimeras provides new insights into the mechanism of target-site licensing by SpCas9 and methods to exploit this mechanism to improve precision. Fusion of a pDBD to SpCas9 allows for efficient use of a broader repertoire of PAM sequences by SpCas9. However, even for SpCas9-pDBDs, there remains a dichotomy between functional and inactive PAMs. The broader targeting range of SpCas9-pDBDs probably reflects the bypass of a kinetic barrier to R-loop formation that follows PAM recognition, as proposed by Seidel and colleagues[249]. We believe that the pDBD tethering of SpCas9 leads to activity at a target site containing a suboptimal PAM by increasing the effective concentration of SpCas9 around the target site and hence stabilizing the SpCas9-PAM interaction[363]. For wild-type SpCas9, only high-affinity (NGG) PAM sites consistently have sufficient residence time to facilitate efficient progression to R-loop formation, but pDBD tethering increases the likelihood that SpCas9:sgRNA will be able to overcome this barrier at suboptimal PAMs. Our data also support an allosteric licensing mechanism, as described by Doudna and colleagues[248], which is likely to restrict Cas9 nuclease activity for the majority of sequence combinations in the PAM element even with the increased local concentration afforded by pDBD tethering. The enhanced sensitivity to guide–target site heteroduplex stability observed for our $SpCas9^{MT3}$-ZFP^{TS3} chimera (**Figure 2.10d** and **Figure 2.13**) further supports the interplay between PAM recognition and guide complementarity in the licensing of nuclease activity.

We found that mutations to the SpCas9 PAM-interacting domain introduced a third stage of licensing (pDBD site recognition) for efficient target-site cleavage in the SpCas9^{MT}-pDBD system (**Figure 2.16e**). The weakened interaction between mutant Cas9 and the PAM sequence necessitated an increased effective concentration for nuclease function that was achieved by the high-affinity interaction of the tethered pDBD with its target site. This combination dramatically improved precision as assessed by targeted deep-sequencing and GUIDE-seq analysis. Compared with previous GUIDE-seq analysis of TS2, TS3 and TS4 for SpCas9, we detected five, three and three of the top five off-target sites that were previously described[286]. The discrepancy between the studies could be due to our lower sequencing depth, the use of an alternate cell line or different delivery methods. Nonetheless, on the basis of our analysis we can exclude the presence of a new class of highly active off-target sites generated by the fusion of the ZFP to Cas9.

This system has important advantages over previously described Cas9 variant systems that improve precision[190, 198, 261, 293, 298, 300]. The SpCas9^{MT}-pDBD system increased the targeting range of the nuclease by expanding the repertoire of highly active PAM sequences. This is in contrast to dimeric systems (e.g., dual nickases and FokI-dCas9 nucleases) that have a more restricted targeting range because of the requirement for a pair of compatible target sequences. Moreover, our system should be compatible with either of these dimeric nuclease variants, providing a further potential increase in precision while also expanding the number of compatible target sites for these platforms. In addition, the affinity and the specificity of the pDBD component can be easily tuned to achieve the desired level of nuclease activity and precision for demanding gene therapy applications.

We programmed our SpCas9-ZFPs targeting TS2, TS3 or TS4 with four-finger ZFPs, as we believed that these would provide the optimal balance of specificity and affinity. In the case of SpCas9^{MT3}-ZFP^{TS3}, this proved prudent (**Figure 2.11**). However, for SpCas9^{MT3}-ZFP^{TS2}, we achieved improved precision with a three-finger ZFP, which demonstrates the flexibility provided by modular pDBDs. (More details on ZFP design for Cas9-ZFPs are given in **section 2.4.1**. Readers can refer to our website (http://mccb.umassmed.edu/Cas9-pDBD_search.html) for assistance with the identification of target sites and compatible ZFP sequences.) In addition to pDBD tuning, further optimization of the linker length and its composition can provide improvements in precision (and potentially activity) by further restricting the relative orientation and spacing of the SpCas9 and pDBD. Finally, it should be possible to generate Cas9-pDBD fusions for other Cas9 orthologs that have superior characteristics for gene therapy applications (for example, more compact Cas9 nucleases[242, 244] for viral delivery). Ultimately, for gene therapy applications where precision, activity and target-site location are of paramount importance, the expanded targeting range and precision achieved by the Cas9-pDBD framework provide a potent platform for the optimization of nuclease-based reagents that cleave a single target site in the human genome.

2.4.1 Discussion on Designing and Construction of Cas9-pDBDs

(1) Cas9-ZFP fusions

In principle Zinc Finger Arrays (ZFAs) containing from three to six fingers can be designed for the construction of Cas9-ZFPs, which will bind 9bp to 18bp target sites respectively (3 bp per finger). Based on the results of our studies with the TS2, TS3 and

TS4 Cas9-ZFPs, we would recommend the construction of four-finger ZFAs when possible for the initial testing of Cas9-ZFPs at a particular target site. For Cas9-ZFPs containing a 58 aa linker the target site can be 5 to 14 bp downstream of the last base pair of the PAM triplet and can be on either the Watson or the Crick strand. If longer ZFAs are desired (5 or 6 fingers), we recommend using one or more TGSQKP linkers to break the array into 2 or 3 finger module sets[107]. Other modified linkers can be utilized to skip a base between pairs of zinc finger modules[364, 365] to achieve more favorable recognition by neighboring arrays if desired.

For the commercial design of zinc fingers, Sangamo Biosciences' proprietary zinc finger module archive has a design density likely less than every 10 bp[366], combined with the flexibility of the spacing and orientation, multiple ZFAs can be designed and tested around almost any Cas9 target site. These ZFAs can be purchased from Sigma Aldrich.

In addition, a number of open-source systems have been described for selecting or assembling ZFAs. Highly specific ZFAs can be selected from randomized finger libraries using phage or bacterial selections [28-37, 367], but this process is labor intensive and may be accessible to only few laboratories. By contrast, modular assembly[35, 36, 368-372] wherein pre-characterized single zinc finger modules that recognize 3-base-pair (bp) subsites are joined into arrays, rapidly yields ZFAs that bind desired target sites, and has proven to be an effective method for the creation of active Cas9-ZFPs in this manuscript. For modular assembly, a number of zinc finger archives have been described focusing on single-finger (1F)[37, 38, 369, 371] and two-finger (2F) modules[35, 36, 368, 370, 373]. Using phage-based selections, Barbas lab identified 1F-modules that target 49 of the 64 triplets[29-32, 369]. The Kim

laboratory has reported 1F-modules recognizing 38 of the 64 triplets[371]. We have also published a curated archive of 1F-modules that bind 27 of 64 triplets[38]. Recently, using bacterial-one-hybrid based selections Noyes lab defined zinc finger modules that can recognize each of the 64 DNA triplets allowing targeting virtually any DNA sequence[37]. In addition, two-finger archives have been published that take into account finger-finger interface and therefore can yield ZFAs with higher specificity but the targeting range of these 2F archives is more limited[35, 36, 368, 370]. The 1F and 2F archives that our laboratory has described can be used to design a ZFP roughly every 10 bp, whereas some of the other finger archives can achieve even higher design densities.

With the number of finger archives now available, it is possible to design a ZFA targeting almost every DNA sequence. Moreover, there are a number of tools available to help users to identify the best target site and design the ZFA. We have designed a web-based tool for the identification of Cas9-ZFP target sites for which ZFAs can be designed from our zinc finger archive. This site (http://mccb.umassmed.edu/Cas9-pDBD_search.html) provides a simple scoring function for the evaluation of ZFAs with higher activity based on the number of arginine-guanine contacts that are present. Tools from other laboratories are available for the construction of ZFAs. The "Zinc Finger Tools" published by Barbas lab can identify target sites for single ZFAs and design ZFAs using their archive of 49 1F-modules[374].
(http://www.scripps.edu/barbas/zfdesign/zfdesignhome.php). The Joung laboratory has developed a suite of tools "ZiFiT" that allows the design of ZFAs for a particular target sequence[284] (http://zifit.partners.org/ZiFiT/). In addition, a zinc finger tool developed by

Noyes laboratory can be used to design zinc finger arrays one finger at a time for a desired target sites[37] (http://zf.princeton.edu/b1h/dna.html). This tool provides multiple zinc finger(s) for every DNA triplet but does not identify the best zinc finger site in a given target sequence.

(2) Cas9-TALE fusions

When designing TALE-arrays for Cas9-TALE fusion, we recommend a minimum of a 10 bp target site (excluding the 5' T) approximately 10-14 bp downstream and on the Watson strand relative to the NGG PAM site. The principal requirement for finding a target site is a 5' T but other rules of target recognition can be important for site selection[121]. Multiple programs are available that allow design of single TAL-arrays including TALE-NT[375] (https://tale-nt.cac.cornell.edu/) and SAPTA TAL Targeter Tool[376]. (http://bao.rice.edu/Research/BioinformaticTools/TAL_targeter.html)

CHAPTER III: Inducible Cas9-DNA-binding-domain fusions for precise genome editing

SpCas9-programmable DNA-binding domain (SpCas9-pDBD) chimeras introduce an extra stage of target site licensing that substantially improves Cas9 specificity. In this study, we integrate drug-dependent dimerization domains into the SpCas9-pDBD framework to provide temporal control over nuclease activity while retaining the high specificity and extended targeting range of the parent covalent system. In addition, we show that the SpCas9-pDBD system has higher precision at challenging target sites than other previously described precision variants of SpCas9. These features make the SpCas9-pDBD nuclease system an attractive nuclease platform for therapeutic genome editing applications where precision is of paramount importance.

3.1 Introduction

CRISPR-Cas9 nucleases (and variants thereof) are being developed for a variety of gene therapy applications[5, 7], but their off-target activity can create undesired collateral damage within the genome of treated cells[377, 378]. We have previously described a PAM-interacting domain (PID)-attenuated SpCas9-DNA-binding domain chimera (SpCas9-pDBD) that reduces the off-target activity of Cas9 and expands the range of PAMs that can be effectively utilized within a genomic region of interest[306]. This nuclease platform has superior precision to many other types of SpCas9 variants engineered for improved precision[377-379].

The in vivo delivery of nuclease-based gene therapy reagents presents challenges for many cell types or organ systems[7]. When delivered using viral vectors, Cas9 expression can potentially persist for an extended time period, which can reduce nuclease precision[289]. One potential solution to this problem is to utilize nuclease systems whose activity can be regulated through the administration of a small molecule or light stimulus. A number of different regulatable nuclease systems have been described[308, 309, 311, 312, 380-383], and many display improved precision compared to constitutively expressed nucleases in small surveys of previously defined off-target sites[308, 312]. However, to our knowledge no unbiased, genome-wide analysis of precision has been reported on one of these regulatable nuclease platforms. Herein we report the development of drug-inducible SpCas9-pDBDs and the genome-wide analysis of their precision. These data demonstrate that SpCas9-pDBD systems are more precise than other existing nuclease platforms, and thus provide an appealing technology platform for therapeutic genome editing applications.

3.2 Materials and Methods

3.2.1 Plasmid constructs

Our dimeric SpCas9-pDBD experiments employed the following plasmids: All sgRNAs are expressed via a U6 promoter from a pLKO1-puro-based vector, as previously described[306]. All SpCas9 and SpCas9-DBD constructs are expressed via a CMV IE94 promoter from a pCS2-Dest gateway plasmid[306]. The ZFP sequences (Zif268, TS2 ZFP, TS3 ZFP) and TALE sequences (TS3 TALE) used to target the SpCas9-ZFP, SpCas9-DD-ZFP and Split-SpCas9-ZFP constructs were previously described[306]. Representative

SpCas9-DD-ZFP, Split-SpCas9-ZFP, SpCas9-DD-TALE and Split-SpCas9-TALE plasmid constructs used for the activity, precision and off-target analysis will be deposited with Addgene. The mutations associated with the eSpCas9_v1.1[301] and SpCas9-HF1[302] sequences were introduced into our pCS2-Dest SpCas9WT plasmid by cassette mutagenesis, where eSpCas9_v1.1 was amplified from the deposited Addgene plasmid (#71814). Plasmid reporter assays of nuclease activity detected the restoration of GFP activity through SSA-mediated repair of an inactive GFP construct based on the M427 plasmid developed by the Porteus laboratory[354]. SpCas9 target sites were cloned into plasmid M427 via ligation independent methods following SbfI digestion.

3.2.2 Cell culture assay

Human Embryonic Kidney (HEK293T) obtained from our collaborator M. Green (UMass Medical School, Worcester, Massachusetts, USA) cells were cultured in high glucose DMEM with 10% FBS and 1% Penicillin/Streptomycin (Gibco) in a 37°C incubator with 5% CO2. These cells were authenticated by University of Arizona Genetics Core and tested for mycoplasma contamination at regular intervals. For transient transfection to assay nuclease activity, we used early to mid-passage cells (passage number 5-25). Approximately $1.6X10^5$ cells were transfected using Polyfect transfection reagent (Qiagen) in 24-well format according to the manufacturer's suggested protocol. Each transfection contained 50ng sgRNA expressing plasmid and 25 ng mCherry plasmid. In addition the following nuclease expression plasmid amounts were included in the transfections: 50 ng for monomeric nuclease systems ($Cas9^{WT}$, $Cas9^{MT3}$ or $Cas9^{MT3}$-ZFP, SpCas9-HF1 or eSpCas9 v1.1), 100 ng of Cas9-DD expression plasmid and 50 ng pDBD-

DD expression plasmid for DD dimeric systems; 100 ng of each Split-Cas9 component for Split-Cas9 systems. A pBlueScript II plasmid was added to bring the total DNA mass of plasmid to 300ng per transfection. For the SSA-reporter assay, 150ng M427 SSA-reporter plasmid was also included in the co-transfection mix and 100 ng of cherry plasmid was used.

3.2.3 Flow cytometry reporter assay

The flow cytometry experiments were done as we described previously[306]. Briefly, 48 hours post-transfection cells were harvested and resuspended in 1XPBS for flow cytometry (Becton Dickonson FACScan). For FACS analysis, 10000 events were counted from each sample. To minimize the effect of differences in the efficiency of transfection among samples, cells were initially gated for mCherry-expression, and the percentage of EGFP expressing cells (nuclease positive events) were quantified within mCherry positive cells. All of the experimental replicates were performed in triplicate on different days with data reported as mean values with error bars indicating the standard error of the mean.

3.2.4 Targeted deep-sequencing-based off-target analysis

Regions flanking target and off-target sites were PCR amplified using locus-specific primers bearing tails complementary to the Truseq adapters as described previously[306] (**Appendix 1**). 25-50ng input DNA was PCR amplified with Q5 High-Fidelity DNA Polymerase (New England Biolabs) with the addition of Q5 High GC Enhancer: (98°C, 15s; 59°C 25s; 72°C 20s) x30 cycles. 0.1 µl of each PCR reaction was amplified with barcoded primers to reconstitute the TruSeq adaptors using the Q5 High-

Fidelity DNA Polymerase (New England Biolabs) with the addition of Q5 High GC Enhancer: (98°C, 15s; 59°C, 25s; 72°C, 20s) x9 cycles. Equal amounts of the products were pooled and gel purified. The purified library was deep sequenced using a paired-end 150bp Illumina MiSeq run.

MiSeq data analysis was performed using a suite of Unix-based software tools. First, the quality of paired-end sequencing reads (R1 and R2 fastq files) was assessed using FastQC. Raw paired-end reads were combined using paired end read merger (PEAR)[384] to generate single merged high-quality full-length reads. Reads were then filtered by quality (using Filter FASTQ[385]) to remove those with a mean PHRED quality score under 30 and a minimum per base score under 24. Each group of reads was then aligned to a corresponding reference sequence using BWA (version 0.7.5) and SAMtools (version 0.1.19). To determine indel frequency, size and distribution, all edited reads from each experimental replicate were combined and aligned, as described above. Indel types and frequencies were then cataloged in a text output format at each base using bam-readcount (https://github.com/genome/bam-readcount). For each treatment group, the average background indel frequencies (based on indel type, position and frequency) of the triplicate negative control group were subtracted to obtain the nuclease-dependent indel frequencies.

3.2.5 GUIDE-seq off-target analysis

We performed GUIDE-seq[286] with some modifications as described previously[306]. Briefly, HEK293T cells were transfected using Lipofectamine 3000 transfection reagent (Invitrogen) according to the manufacturer's suggested protocol with 50 ng of $Cas9^{WT}$, $Cas9^{MT3}$ or $Cas9^{MT3}$-ZFP expression plasmid for monomeric systems, 200 ng of Cas9

expression plasmid and 100 ng pDBD expression plasmid for the 'DD' system, or 150 ng of each expression plasmid for the Split-Cas9 components. Each transfection also included 100 ng sgRNA expressing plasmid, 50 ng of a GFP expression plasmid and 10 pmol of annealed GUIDE-seq oligonucleotide. 48 h after transfection, genomic DNA was extracted with a DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's suggested protocol. Library preparations were done with original adaptors according to protocols described by Joung and colleagues[286], with each library barcoded within the P5 and P7 adaptors for pooled sequencing. The barcoded, purified libraries were deep-sequenced as a pool using two paired-end 150-bp Illumina MiSeq runs.

Deep sequencing data from the GUIDE-seq experiment was analyzed using the GUIDEseq v1.1.17 Bioconductor Package[386] The upstream and downstream windows for peak aggregation we set to 50 bp. Off-target site identification parameters were set as follows: min.reads = 1, max.mismatch = 6, PAM.pattern = "(NAG|NGG|NGA)$", allowed.mismatch.PAM = 2. The potential off-target sites identified for each nuclease are listed in **Appendix 2**. The Specificity Ratio is calculated as the sum of the unique GUIDE-seq reads at the target site divided by all of the unique reads at all of the off-target sites.

3.3 Results

Rapamycin-induced heterodimerization of FKRP and FRB domains is one of the most commonly employed technologies for associating two proteins[383]. It has been used successfully in a variety of regulatable Split-Cas9 architectures for genome editing and gene regulation[308, 309, 381, 387]. We examined the ability of the FKRP-FRB system to regulate the nuclease activity of a SpCas9-zinc finger protein (SpCas9-ZFP) in two different

formats (**Figure 3.1a**): a configuration where rapamycin-dependent FKRP and FRB dimerization domains (DD) replace the covalent linker between wild-type or PID-attenuated[306] SpCas9 and the zinc finger protein (SpCas9-DD-ZFP), or a configuration where a previously described rapamycin-inducible Split-Cas9 architecture[308] (wild-type or PID-attenuated[306]) is fused covalently to the ZFP (Split-SpCas9-ZFP). A hallmark of the covalent SpCas9-ZFP system is its ability to utilize non-canonical PAM sequences[306]. Consequently, to confirm that these different dimeric nuclease configurations retain an extended targeting range, they were evaluated at a series of target sequences with a sub-optimal NAG PAM using a GFP reporter assay[354]. In these target sequences the distance between the protospacer and the ZFP binding site was varied, as was their relative orientation. We found that the SpCas9-DD-ZFP nuclease displayed distance-dependent activity in the presence of rapamycin, whereas the Split-SpCas9-ZFP nuclease displayed a more uniform activity profile (**Figure 3.1b**). The distance-dependent activity of the SpCas9-DD-ZFP system likely originates from steric restrictions associated with the interaction of the FKBP and FRB domains at separation distances of 8 bp or less.

Next, we evaluated the activity of the dimeric SpCas9-ZFP nucleases at a genomic target site. To demonstrate the functionality of the associated ZFP, we chose a target site within the PMPCA locus contained a suboptimal NAG PAM with a Zif268 binding site 14 bp downstream of the PAM sequence. Pilot experiments revealed good nuclease activity in the presence of rapamycin when a nuclear localization sequence (NLS) was present within both dimeric components (**Figure 3.2**). However we also observed modest nuclease

activity in the absence of rapamycin for these constructs. Similar background activity has been observed with other inducible Cas9 architectures[308].

These results prompted an analysis of the impact of localization sequence composition within these constructs on their nuclease activity. For the SpCas9-DD-ZFP dimer, we achieved promising induction of nuclease activity when FKBP is fused to SpCas9 in the absence of either a nuclear export sequence (NES) or NLS, and when FRB is fused C-terminally to the ZFP with both NESs and NLSs appended (**Figure 3.2a**). For the Split-SpCas9-ZFP dimer, a similar strategy of withholding a localization sequence from one Cas9 fragment and including both NLSs and an NES within the other Cas9 fragment maximized the degree of nuclease activity and inducibility (**Figure 3.2b**). As described by Zetsche and colleagues for Split-SpCas9[310], we found that the incorporation of a single NES on one nuclease component and the NLS on the other nuclease component with the dimeric SpCas9-ZFP systems decreased the background nuclease activity in the absence of rapamycin. However, this combination of localization sequences also restricted the maximum level of inducible nuclease activity that could be achieved. Thus, both SpCas9-ZFP dimerization systems achieved robust, drug-dependent nuclease activity at a genomic target site with a suboptimal PAM, which confirms their extended targeting range. In addition, we found that using a PID-attenuated SpCas9 variant (R1333S; SpCas9MT2)[306] in the Cas9-DD-ZFP framework achieved robust activity at the NAG PAM target site with lower background in the absence of rapamycin (**Figure 3.2a**), demonstrating that the PID-attenuated SpCas9s are compatible with this dimerization system.

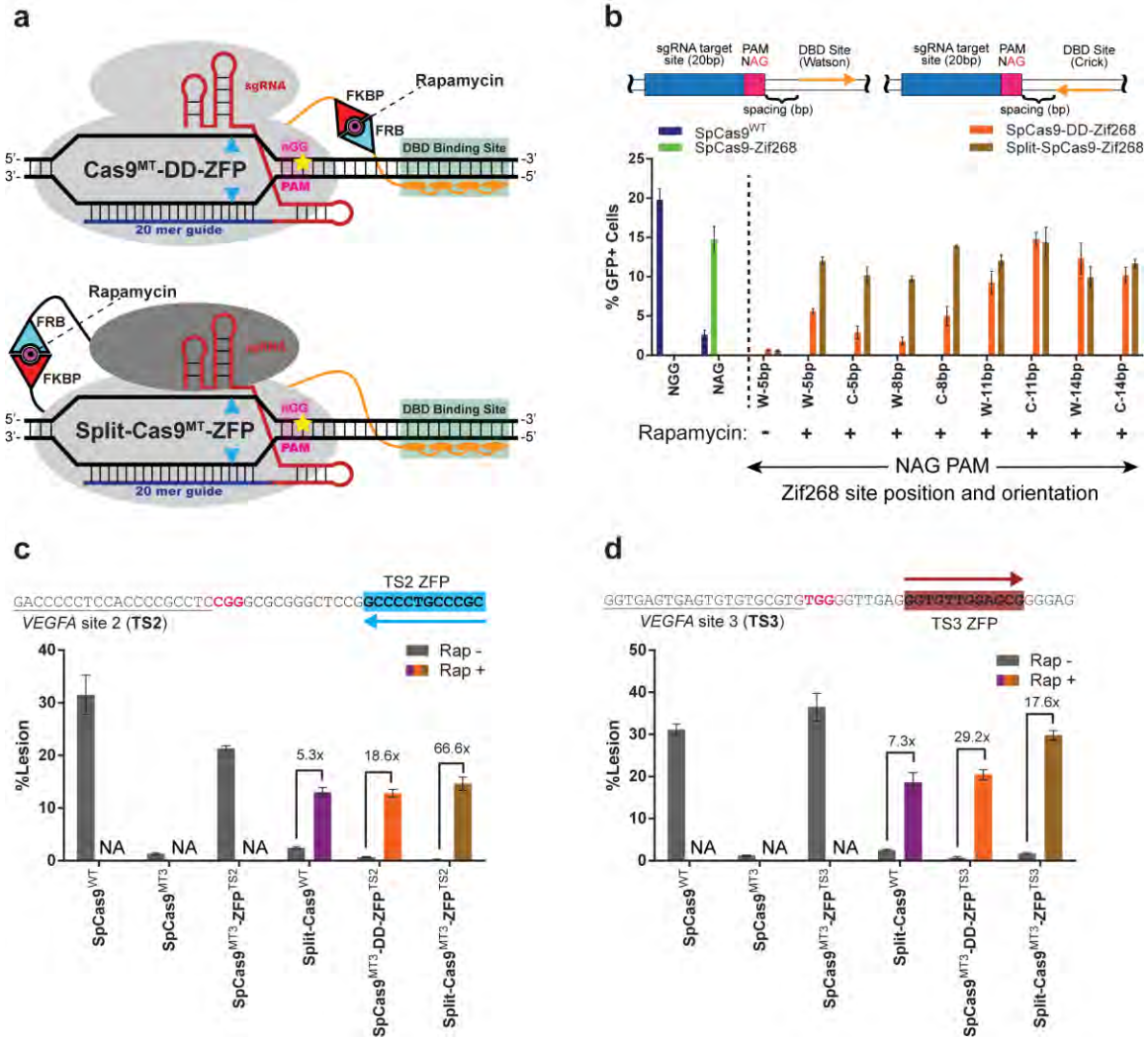

a
sgRNA
FKBP
Rapamycin
FRB
Cas9MT-DD-ZFP
nGG
PAM
DBD Binding Site
20 mer guide
5'-
3'-
-3'
-5'
Split-Cas9MT-ZFP
b
sgRNA target site (20bp)
PAM NAG
DBD Site (Watson)
DBD Site (Crick)
spacing (bp)
SpCas9WT
SpCas9-Zif268
SpCas9-DD-Zif268
Split-SpCas9-Zif268
% GFP+ Cells
NGG
NAG
W-5bp
C-5bp
W-8bp
C-8bp
W-11bp
C-11bp
W-14bp
C-14bp
Rapamycin:
NAG PAM
Zif268 site position and orientation
c
TS2 ZFP
GACCCCCTCCACCCCGCCTCCGGGCGCGGGGCTCCGGCCCCCTGCCCGC
VEGFA site 2 (TS2)
Rap -
Rap +
%Lesion
NA
5.3x
18.6x
66.6x
SpCas9WT
SpCas9MT3
SpCas9MT3-ZFPTS2
Split-Cas9WT
SpCas9MT3-DD-ZFPTS2
Split-Cas9MT3-ZFPTS2
d
GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGGTGTTGGACCGGGGAG
VEGFA site 3 (TS3)
TS3 ZFP
7.3x
29.2x
17.6x
SpCas9MT3-ZFPTS3
Split-Cas9WT
SpCas9MT3-DD-ZFPTS3
Split-Cas9MT3-ZFPTS3

Figure 3. 1 Development of drug-inducible dimeric SpCas9-pDBD frameworks
(**a**) Schematic of two different dimeric SpCas9-pDBD:sgRNA systems. In Cas9^{MT}-DD-ZFP (top) the rapamycin-induced dimerization of FKBP (red) and FRB (blue) domains associates the Cas9^{MT}:sgRNA complex with the ZFP (orange) that recognizes a binding site 3' to the PAM. In Split-Cas9^{MT}-ZFP (bottom) the rapamycin induced dimerization of FKBP and FRB associates the N- (1-573) and C-terminal (574-1368) fragments of the Split-Cas9^{308}:sgRNA complex where the ZFP is covalently fused to the C-terminal fragment. The yellow star indicates attenuation of PAM recognition through mutagenesis[306] (**b**) Top, schematic of the examined pDBD binding-site orientation and spacing parameters. The position and 5'-to-3' orientation of the pDBD binding site relative to the PAM element of the SpCas9 binding site are represented by orange arrows (Watson (W) and Crick (C)). Bottom, activity profiles of SpCas9^{WT} (blue, on an NGG or NAG PAM), SpCas9-Zif268 (green, NAG PAM), SpCas9-DD-Zif268 (orange, NAG PAM) and Split-SpCas9-Zif268 (brown, NAG PAM) in the GFP reporter assay on a common sgRNA target site. The pDBD-site spacing is 5, 8, 11 or 14 bp from the PAM on either strand. The presence or absence of rapamycin in each experiment is indicated below the x-axis (**c, d**) Lesion rates determined by deep sequencing for SpCas9^{WT}, SpCas9^{MT3}, SpCas9^{MT3}-ZFP, Split-SpCas9^{WT}, SpCas9^{MT3}-DD-ZFP and Split-SpCas9^{MT3}-ZFP at the *VEGFA* TS2 (**c**) and TS3 (**d**) target sites, respectively. The activity of the dimeric constructs was assayed +/- rapamycin (NA: not available). Sequences of TS2 and TS3 regions are shown above the bar chart, where the 12-bp ZFP binding site is highlighted and arrows indicate the orientation of ZFP binding site. Fold induction in presence of rapamycin is shown. Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

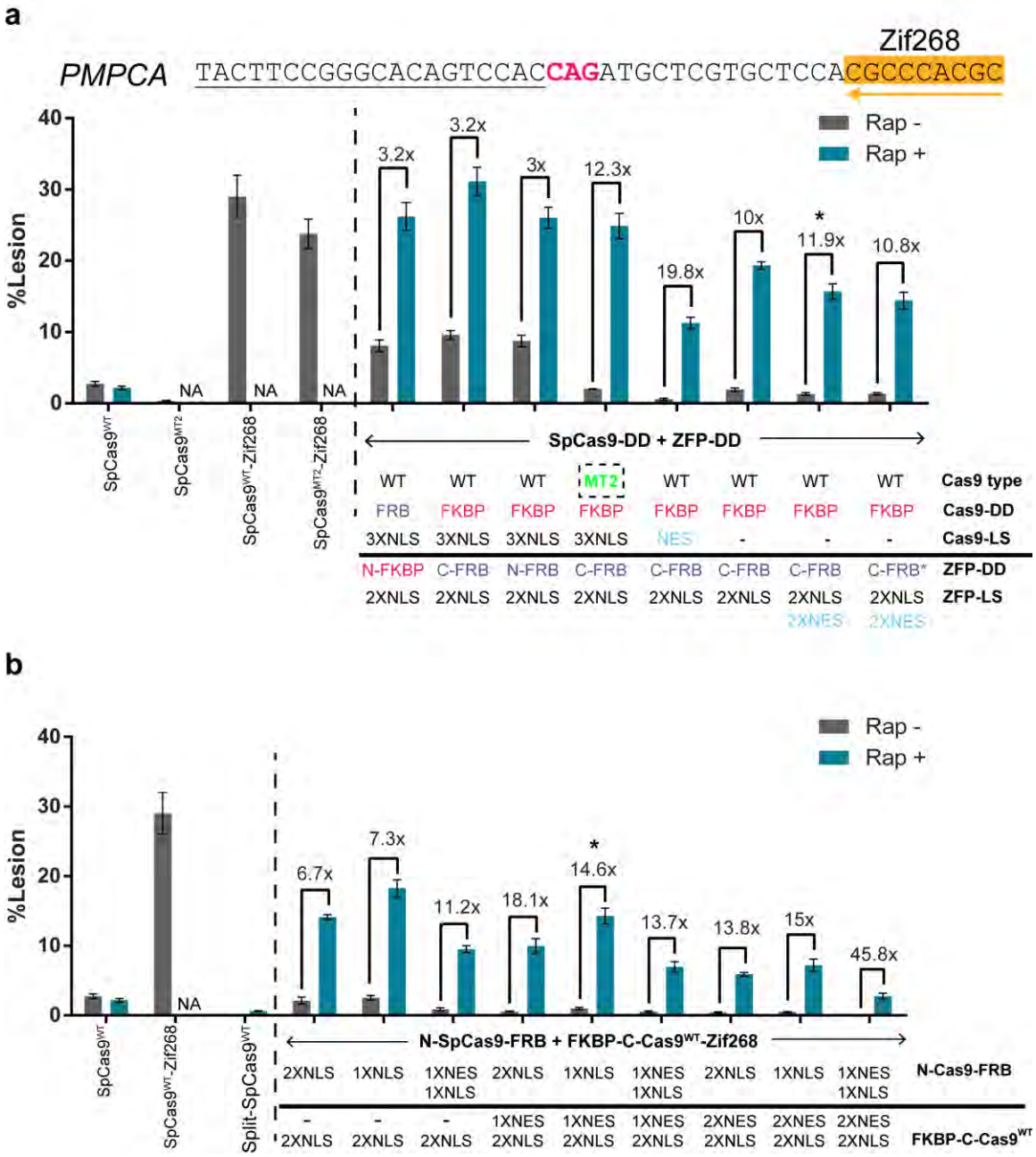

a
PMPCA
TACTTCCGGGCACAGTCCACCAGATGCTCGTGCTCCACGCCCACGC
Zif268
Rap -
Rap +
%Lesion
3.2x
3.2x
3x
12.3x
19.8x
10x
*
11.9x
10.8x
NA
NA
NA
SpCas9-DD + ZFP-DD
SpCas9WT
SpCas9MT2
SpCas9WT-Zif268
SpCas9MT2-Zif268
WT WT WT MT2 WT WT WT WT Cas9 type
FRB FKBP FKBP FKBP FKBP FKBP FKBP FKBP Cas9-DD
3XNLS 3XNLS 3XNLS 3XNLS NES - - - Cas9-LS
N-FKBP C-FRB N-FRB C-FRB C-FRB C-FRB C-FRB C-FRB* ZFP-DD
2XNLS 2XNLS 2XNLS 2XNLS 2XNLS 2XNLS 2XNLS 2XNLS ZFP-LS
2XNES 2XNES
b
Rap -
Rap +
%Lesion
6.7x
7.3x
11.2x
18.1x
*
14.6x
13.7x
13.8x
15x
45.8x
NA
N-SpCas9-FRB + FKBP-C-Cas9WT-Zif268
SpCas9WT
SpCas9WT-Zif268
Split-SpCas9WT
2XNLS 1XNLS 1XNES 1XNLS 2XNLS 1XNLS 1XNES 1XNLS 2XNLS 1XNLS 1XNES 1XNLS N-Cas9-FRB
- - - 1XNES 1XNES 1XNES 2XNES 2XNES 2XNES
2XNLS 2XNLS 2XNLS 2XNLS 2XNLS 2XNLS 2XNLS 2XNLS 2XNLS FKBP-C-Cas9WT

Figure 3. 2 Activity profiles dimeric SpCas9-Zif268 nuclease compositions
The sequence of *PMPCA* target region is shown at the top, where the 9-bp Zif268 binding site 14 bases from the PAM is highlighted with an arrow indicating recognition of the Crick strand. Lesion rates determined by deep sequencing for SpCas9^{WT}, SpCas9^{MT2}, SpCas9^{WT}-Zif268, SpCas9^{MT2}-Zif268, Split-SpCas9^{WT}, and different compositions of the SpCas9-DD-Zif268 and Split-SpCas9-Zif268 dimers. (**a**) Impact of the organization of the FRB and FKBP domains and the types of localization sequences on the activity of SpCas9-DD-Zif268. The SpCas9^{WT} and SpCas9^{MT2} (R1333S) mutant[306] respectively display little to no activity on this target site in the absence of a fusion to Zif268. Robust nuclease activity is observed when Zif268 is covalently fused to either of these constructs. For the SpCas9-DD-Zif268 architecture, the impact of three different parameters on nuclease activity in the absence and presence of rapamycin was examined: Cas9 nuclease variant (Cas9 type), the fusion partner for the FKRP and FRB dimerization domains (DD; Cas9-DD and ZFP-DD), and the type and number of localization sequences (LS) on each component (Cas9-LS and ZFP-LS). A favorable combination (marked with an asterisk) of overall activity and drug-dependent behavior was achieved with Cas9^{WT} when the Cas9 domain lacked an NLS and Zif268 fused to a combination of NLS and NES elements. Favorable activity profiles were also observed when SpCas9^{MT2} was substituted for SpCas9^{WT} when an NLS was present on both components (Cas9 type = MT2). For the ZFP-DD fusion the position of the DD relative to Zif268 is indicated (C-FRB = FRB on the C-terminus of Zif268) where FRB* indicates the PLF mutant of FRB that destabilizes this domain[388]. (**b**) Impact of the organization of the types and number of localization sequences on the activity of Split-SpCas9^{WT}-Zif268. The Split-SpCas9^{WT} (NLS on N-terminal component & NES on C-terminal component) described by Zetsche and colleagues[308] that is the basis for this dimeric architecture displays minimal activity on this target site. However, by fusing Zif268 to the C-terminal component FKBP-C-Cas9^{WT}, robust activity can be achieved in the presence of rapamycin. As observed for the SpCas9-DD-Zif268 architecture, a promising combination of fold induction and overall activity in the presence of rapamycin was achieved when one fragment (N-Cas9-FRB) lacks a localization sequence and the other fragment (FKBP-C-Cas9^{WT}-Zif268) contains a combination of NLS and NES elements (marked with an asterisk). Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

In preparation for the analysis of the genome-wide specificity of these dimeric SpCas9-ZFP systems, we evaluated their activity at two target sites within the *VEGFA* locus (target site 2 & target site 3; TS2 & TS3) previously used for assessing precision of different SpCas9 variants[257, 286, 293, 302, 306]. For these NGG PAM target sites, our dimeric constructs utilized a PID-attenuated SpCas9 variant (R1335K; SpCas9^{MT3}) that requires recruitment by a pDBD to the genomic locus for robust activity[306]. The resulting extra stage of target site licensing increases the precision of the chimeric nuclease. In the context of the attenuated dimeric SpCas9^{MT3}-ZFP systems, we assessed the impact of the incorporated localization sequences on their drug-dependent nuclease activity (**Figures 3.3 & 3.4**). As with the wild-type SpCas9-ZFP dimers at the *PMPCA* locus, dimeric SpCas9^{MT3}-ZFP nuclease activity at the TS2 and TS3 target sites was sensitive to the type and number of localization sequences that were fused to each component. For particular combinations of localization sequences, both dimeric SpCas9^{MT3}-ZFP systems displayed a greater degree of rapamycin-inducible nuclease activity than the Split-SpCas9 system described by Zetsche and colleagues[308] (**Figure 3.1c-d**). All of these inducible nuclease systems display somewhat reduced activity relative to monomeric SpCas9, which can be attributed in part to the ~16 hour delay prior to adding rapamycin following transient transfection. The dimeric SpCas9^{MT3} system also functioned at the TS3 target site when a TALE DBD programmed to bind a neighboring sequence replaced the ZFP (**Figure 3.5**). However, the maximal nuclease activity of the TALE-fused dimeric systems was lower than the ZFP-fused systems. Overall the nuclease activity and fold induction of both dimeric SpCas9^{MT3}-

ZFP systems compare favorably with the previously described inducible Split-SpCas9 architecture[308].

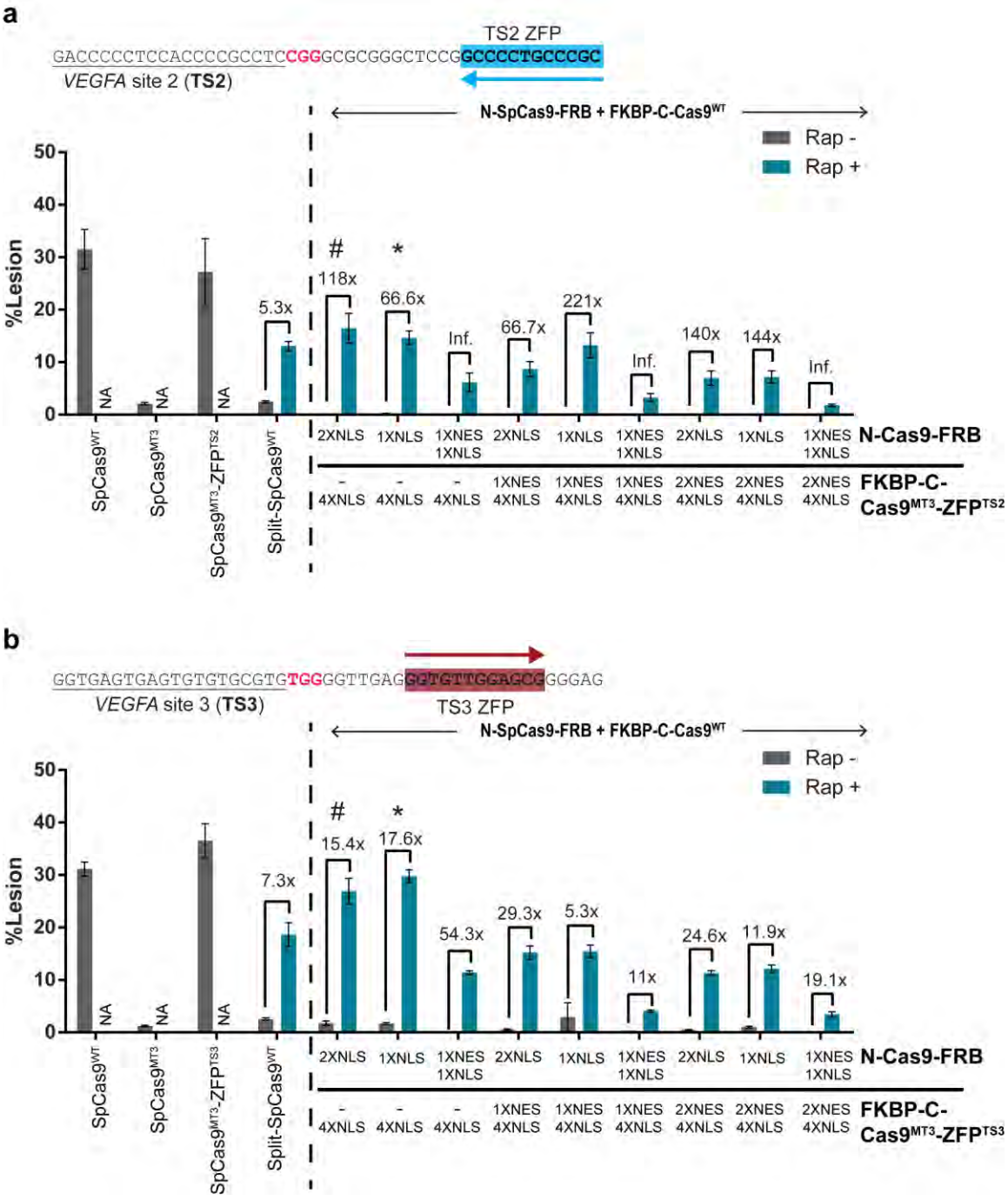


Figure 3. 3 Activity profiles of Split-SpCas9^{MT3}-ZFP nuclease compositions
The sequences of the TS2 and TS3 regions are shown above each bar chart of nuclease activity, where the 12-bp ZFP binding site is highlighted with an arrow indicating the bound strand (Watson = top pointing right; Crick = bottom pointing left). (**a**) Impact of the types of localization sequences incorporated within each fragment of Split-SpCas9^{MT3}-ZFP^{TS2} on nuclease activity at the *VEGFA* TS2 target site. SpCas9^{WT} displays robust activity on the target site. The SpCas9^{MT3} (R1335K) mutant[306] displays little activity on this target site, however covalent fusion of the ZFP to the C-terminal fragment restores nuclease activity. Split-SpCas9^{WT} (NLS on N-terminal component & NES on C-terminal component described by Zetsche and colleagues[308]), which is the framework for this dimeric architecture, displays inducible activity on this target site with modest background activity in the absence of rapamycin. For the Split-SpCas9^{MT3}-ZFP^{TS2} architecture, the most favorable combination of overall activity and fold induction was achieved when the N-terminal Cas9 fragment (N-Cas9-FRB) lacks a localization sequence and the C-terminal Cas9 fragment (FKBP-C-Cas9^{MT3}-ZFP^{TS2}) contained a 2xNLS. (**b**) Impact of the types of localization sequences incorporated within each fragment of Split-SpCas9^{MT3}-ZFP^{TS3} on nuclease activity at the *VEGFA* TS3 target site. For the SpCas9^{MT3}-DD-ZFP^{TS3} the most favorable combination of high activity and fold induction was achieved when the N-terminal Cas9 fragment (N-Cas9-FRB) lacks a localization sequence and the C-terminal Cas9 fragment (FKBP-C-Cas9^{MT3}-ZFP^{TS3}) contains a 2xNLS. Lesion rates were determined by deep sequencing. The constructs utilized for GUIDE-seq analysis are indicated by '#' and for targeted amplicon deep sequencing of off-target sites to assess nuclease activity are indicated by an asterisk. Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

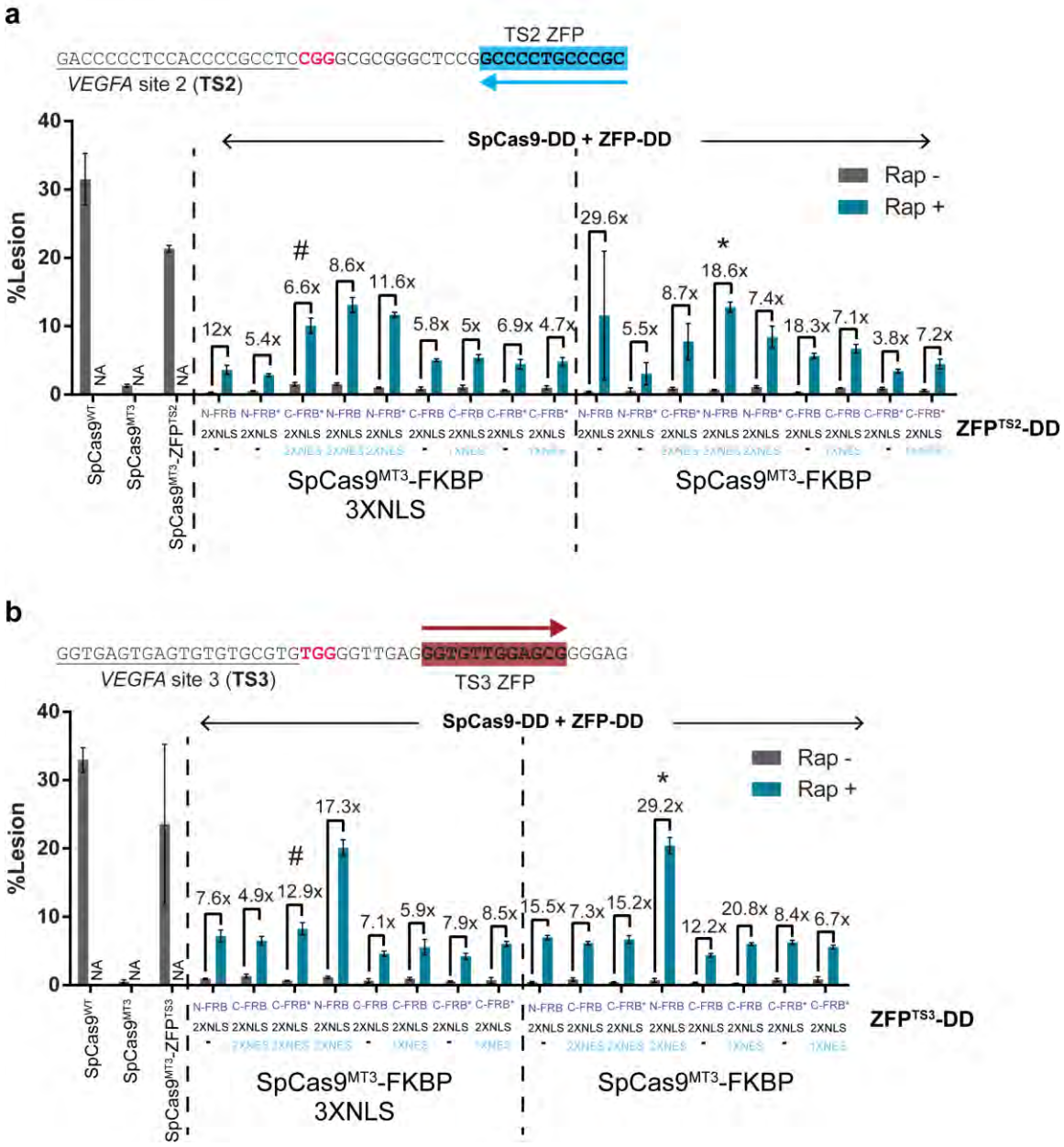

a
TS2 ZFP
GACCCCCTCCACCCCGCCTCCGGGCGCGGGCTCCGGCCCCTGCCCGC
VEGFA site 2 (TS2)
%Lesion
SpCas9-DD + ZFP-DD
Rap -
Rap +
29.6x
18.6x
12x
5.4x
6.6x
8.6x
11.6x
5.8x
5x
6.9x
4.7x
5.5x
8.7x
7.4x
18.3x
7.1x
3.8x
7.2x
SpCas9WT
SpCas9MT3
SpCas9MT3-ZFPTS2
SpCas9MT3-FKBP 3XNLS
SpCas9MT3-FKBP
ZFPTS2-DD
b
GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGGTGTTGGAGCGGGGAG
VEGFA site 3 (TS3)
TS3 ZFP
SpCas9-DD + ZFP-DD
Rap -
Rap +
29.2x
17.3x
7.6x
4.9x
12.9x
7.1x
5.9x
7.9x
8.5x
15.5x
7.3x
15.2x
12.2x
20.8x
8.4x
6.7x
SpCas9WT
SpCas9MT3
SpCas9MT3-ZFPTS3
SpCas9MT3-FKBP 3XNLS
SpCas9MT3-FKBP
ZFPTS3-DD

Figure 3. 4 Activity profiles of SpCas9^{MT3}-DD-ZFP nuclease compositions

The sequences of the TS2 and TS3 regions are shown above each bar chart of nuclease activity, where the 12-bp ZFP binding site is highlighted with an arrow indicating the bound strand (Watson = top pointing right; Crick = bottom pointing left). (**a**) Impact of nuclease composition on the activity of SpCas9^{MT3}-DD-ZFP^{TS2} at the *VEGFA* TS2 target site. SpCas9^{WT} displays robust activity on the target site. The SpCas9^{MT3} (R1335K) mutant[306] displays little activity on this target site, however covalent fusion of the ZFP restores nuclease activity. For the SpCas9^{MT3}-DD-ZFP^{TS2} architecture, the impact of two different parameters on rapamycin-inducible nuclease activity were evaluated: the type and number of localization sequences (LS) on each component (Cas9-LS and ZFP-LS), and the FRB type and fusion position relative to the ZFP. (FRB* indicates the PLF mutant of FRB that destabilizes this domain[388].) Fold-activation of drug-dependent nuclease activity was maximized when the SpCas9^{MT3}-FKBP domain lacked an NLS. The most favorable combination of overall activity was achieved with SpCas9^{MT3}-FKBP when FRB is fused to the N-terminus of the ZFP with a combination of NLS and NES elements. (**b**) Impact of the position of fusion and the type of the FRB domain on the ZFP and the type of localization sequences on each component on the activity of SpCas9^{MT3}-DD-ZFP^{TS3} at the *VEGFA* TS3 target site. For the SpCas9^{MT3}-DD-ZFP^{TS3} the activity was more sensitive to the organization of the ZFP-DD component. As at the TS2 target site, the most favorable combination of overall activity was achieved with Cas9^{MT3} when FRB is fused to the N-terminus of the ZFP with a combination of NLS and NES elements. Lesion rates were determined by deep sequencing. The constructs utilized for GUIDE-seq analysis are indicated by '#' and for targeted amplicon deep sequencing of off-target sites to assess nuclease activity are indicated by an asterisk. Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

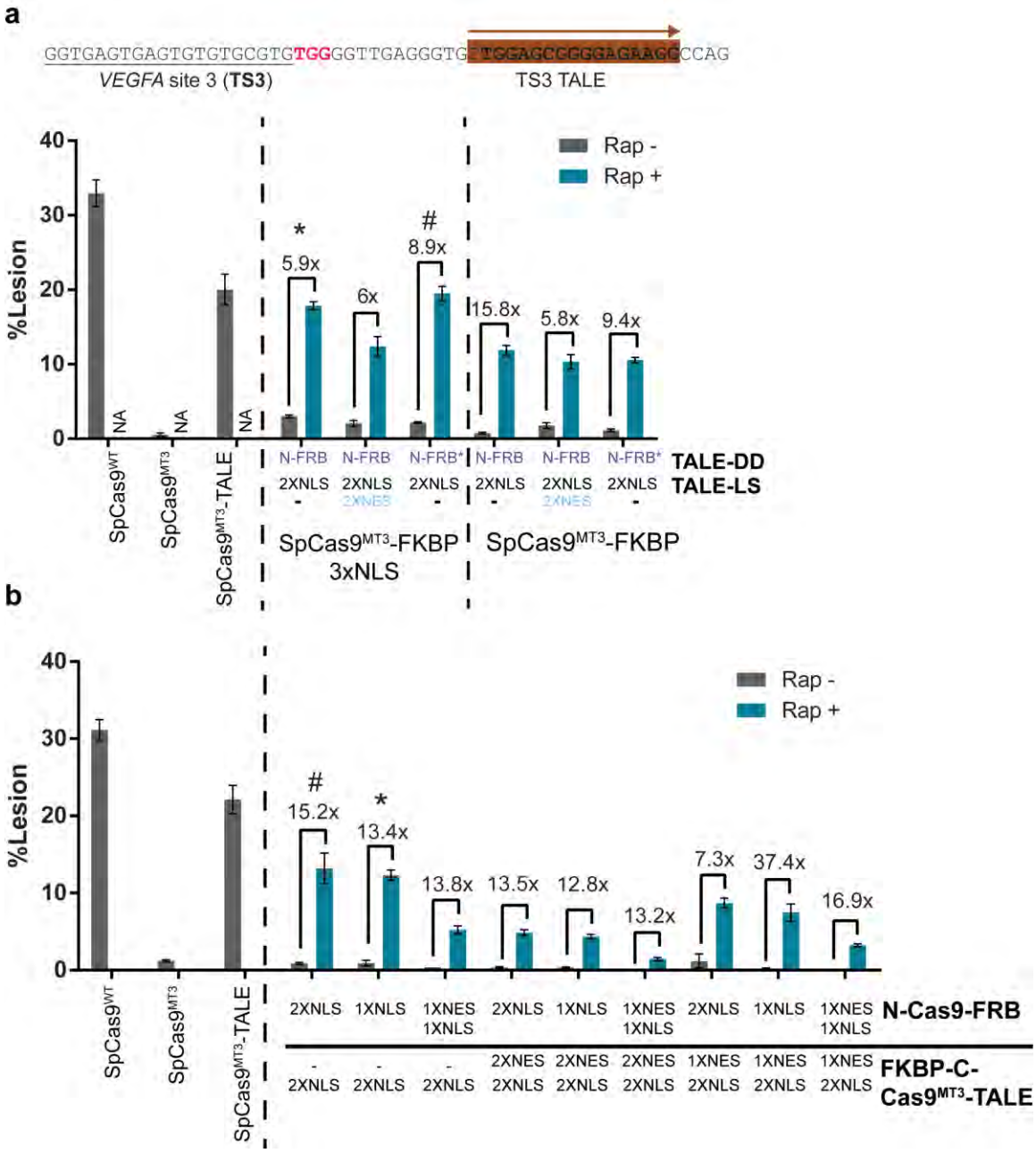


Figure 3. 5 Activity profiles of dimeric SpCas9^{MT3}-TALE^{TS3} nuclease compositions
The sequence of the TS3 region is shown above the bar chart of nuclease activity, where the 17-bp TALE binding site[306] is highlighted with an arrow indicating the bound strand. SpCas9^{WT} displays robust activity on the target site. The SpCas9^{MT3} (R1335K) mutant[306] displays little activity on this target site, however covalent fusion of the TALE restores nuclease activity. A number of different SpCas9^{MT3}-DD-TALE^{TS3} (**a**) and the Split-SpCas9^{MT3}-TALE^{TS3} (**b**) systems display inducible nuclease activity as a function of the appended localization sequences. Lesion rates were determined by deep sequencing. The constructs utilized for GUIDE-seq analysis are indicated by

‘#’ and for targeted amplicon deep sequencing of off-target sites to assess nuclease activity are indicated by an asterisk. Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

One of the valuable attributes of the SpCas9MT3-pDBD system is its enhanced precision relative to wild-type SpCas9[306]. To assess if this characteristic is retained in the dimeric SpCas9^{MT3}-pDBD systems, we utilized GUIDE-seq to perform an unbiased evaluation of their off-target activities[286]. To provide a benchmark for comparison, we also evaluated the off-target activity of the canonical Split-SpCas9 system, which has improved precision relative to wild-type SpCas9[308]. All of these nuclease variants were programmed to target the VEGFA TS2 or TS3 sites, as these guide sequences provide a rigorous test of precision due to the number of similar sequences found within the human genome[257, 286].

Across our GUIDE-seq datasets, the number of unique GUIDE-seq oligonucleotide incorporations at each target site was comparable for all of the nuclease constructs except SpCas9^{WT} at the TS3 target site (**Appendix 2**), which implies that the cellular nuclease activity for the majority of evaluated constructs was similar. For the TS2 sgRNA, there was a reduction in the number of GUIDE-seq peaks within the genome for our SpCas9^{MT3}-ZFP dimeric nucleases relative to both of the previously described Split-SpCas9^{WT} systems[308] (**Figure 3.6a**, *left*). In addition, the Specificity Ratio (the number of the unique GUIDE-seq reads at the TS2 target site relative to the sum of reads at the off-target peaks) was improved for the SpCas9^{MT3}-ZFP dimeric nucleases relative to the canonical Split-SpCas9^{WT} systems (**Figure 3.6a**, *right*). A reduction in the number of GUIDE-seq peaks and an improvement in Specificity Ratio was also observed for the dimeric SpCas9^{MT3}-ZFP nucleases programmed with the TS3 sgRNA (**Figure 3.6b**). The difference in genome-wide nuclease activity profiles was also evident from the number of unique reads associated with individual identified off-target GUIDE-seq peaks (**Appendix 2**). The off-

target peaks for SpCas9^{WT} or the canonical Split-SpCas9^{WT} nucleases displayed a broad distribution of read counts (where a greater number of GUIDE-seq reads implies greater nuclease activity[286]). In contrast, most off-target peaks for the dimeric SpCas9^{MT3}-pDBD nucleases displayed a low read count (**Figure 3.7**), which would suggest that many of these peaks are likely to be either weakly active sites or false positive peaks. In addition, strong new off-target sites are not observed for the dimeric SpCas9^{MT3}-pDBD nucleases, which imply that the fused ZFP is not creating a new class of off-target sequences. Similarly, the SpCas9^{MT3}-TALE nucleases programmed with the TS3 sgRNA displayed improved Specificity Ratios for both the dimeric and covalent systems relative to the other nuclease systems (**Figure 3.8**). Overall these GUIDE-seq datasets reveal a dramatic improvement in precision for the various dimeric SpCas9^{MT3}-pDBD nucleases both with regards to the total number of potentially active off-target sites and the activity levels at the small number of remaining off-target sites.

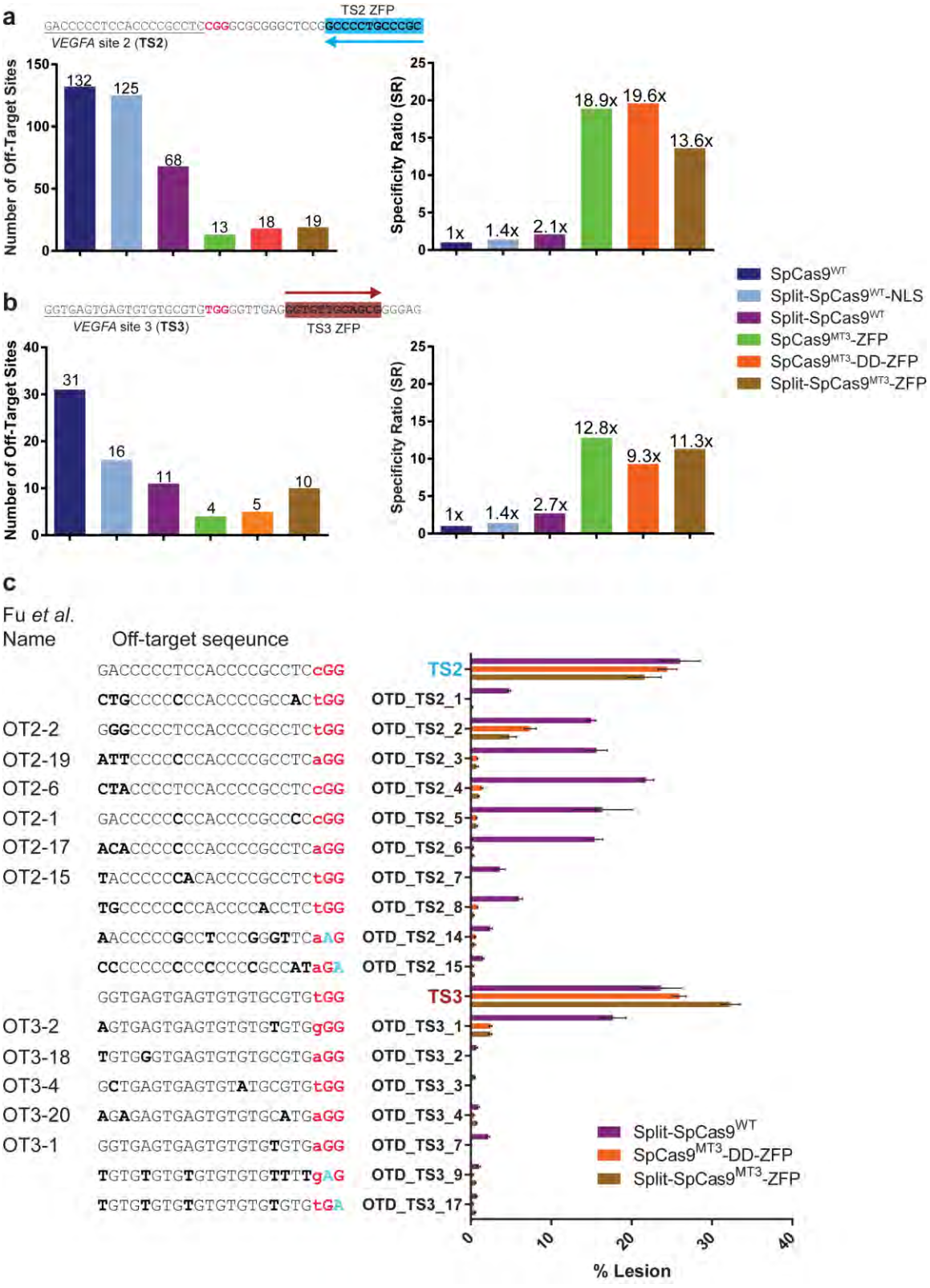

a
TS2 ZFP
GACCCCCTCCACCCCGCCTCCGGGCGCGGGCTCCGGCCCCTGCCCGC
VEGFA site 2 (TS2)
Number of Off-Target Sites
132
125
68
13
18
19
Specificity Ratio (SR)
1x
1.4x
2.1x
18.9x
19.6x
13.6x
b
GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGCTGGTGCAGCCGGGGAG
VEGFA site 3 (TS3)
TS3 ZFP
Number of Off-Target Sites
31
16
11
4
5
10
Specificity Ratio (SR)
1x
1.4x
2.7x
12.8x
9.3x
11.3x
SpCas9WT
Split-SpCas9WT-NLS
Split-SpCas9WT
SpCas9MT3-ZFP
SpCas9MT3-DD-ZFP
Split-SpCas9MT3-ZFP
c
Fu et al.
Name
Off-target seqeunce
GACCCCCTCCACCCCGCCTCcGG
TS2
CTGCCCCCCCACCCCGCCACtGG
OTD_TS2_1
OT2-2
GGGCCCCTCCACCCCGCCTCtGG
OTD_TS2_2
OT2-19
ATTCCCCCCCACCCCGCCTCaGG
OTD_TS2_3
OT2-6
CTACCCCTCCACCCCGCCTCcGG
OTD_TS2_4
OT2-1
GACCCCCCCCACCCCGCCCCcGG
OTD_TS2_5
OT2-17
ACACCCCCCCACCCCGCCTCaGG
OTD_TS2_6
OT2-15
TACCCCCCACACCCCGCCTCtGG
OTD_TS2_7
TGCCCCCCCCACCCCACCTCtGG
OTD_TS2_8
AACCCCCGCCTCCCGGGGTTCaAG
OTD_TS2_14
CCCCCCCCCCCCCCCGCCATaGA
OTD_TS2_15
GGTGAGTGAGTGTGTGCGTGtGG
TS3
AGTGAGTGAGTGTGTGTGTGgGG
OTD_TS3_1
OT3-2
TGTGGGTGAGTGTGTGCGTGaGG
OTD_TS3_2
OT3-18
GCTGAGTGAGTGTATGCGTGtGG
OTD_TS3_3
OT3-4
AGAGAGTGAGTGTGTGCATGaGG
OTD_TS3_4
OT3-20
GGTGAGTGAGTGTGTGTGTGaGG
OTD_TS3_7
OT3-1
TGTGTGTGTGTGTGTGTTTTgAG
OTD_TS3_9
TGTGTGTGTGTGTGTGTGTGtGA
OTD_TS3_17
Split-SpCas9WT
SpCas9MT3-DD-ZFP
Split-SpCas9MT3-ZFP
0
10
20
30
40
% Lesion

Figure 3. 6 Specificity profiles of SpCas9^{MT3}-ZFP nuclease variants
(**a, b**) GUIDE-seq[286] genome-wide off-target analysis of nuclease activity for SpCas9^{WT}, Split-SpCas9^{WT}-dual-NLS[308], Split-SpCas9^{WT}-NLS/NES[308], SpCas9^{MT3}-ZFP, SpCas9^{MT3}-DD-ZFP and Split-SpCas9^{MT3}-ZFP programmed with the *VEGFA* TS2 (**a**) or TS3 (**b**) sgRNAs. The charts indicate (Left) the number of off-target peaks detected in the GUIDE-seq data and (Right) the fold improvement in Specificity Ratio (SR) [number of unique reads at the target site divided by the sum of the unique GUIDE-seq reads associated with the off-target peaks] for each nuclease compared with SpCas9^{WT} (**Appendix 2**). (**c**) Lesion rates for target sites and a subset of off-target sites identified by GUIDE-seq analysis assayed by deep-sequencing of PCR products spanning each genomic locus for Split-SpCas9^{WT}-NLS/NES[308], SpCas9^{MT3}-DD-ZFP and Split-SpCas9^{MT3}-ZFP programmed with the *VEGFA* TS2 or TS3 sgRNAs. Sequences of each of the target or putative off-target sites are shown to the left of the bar chart with the non-cognate positions shown in bold (protospacer) or cyan (PAM). Site IDs from Fu *et al.*[257] are included for comparison. Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

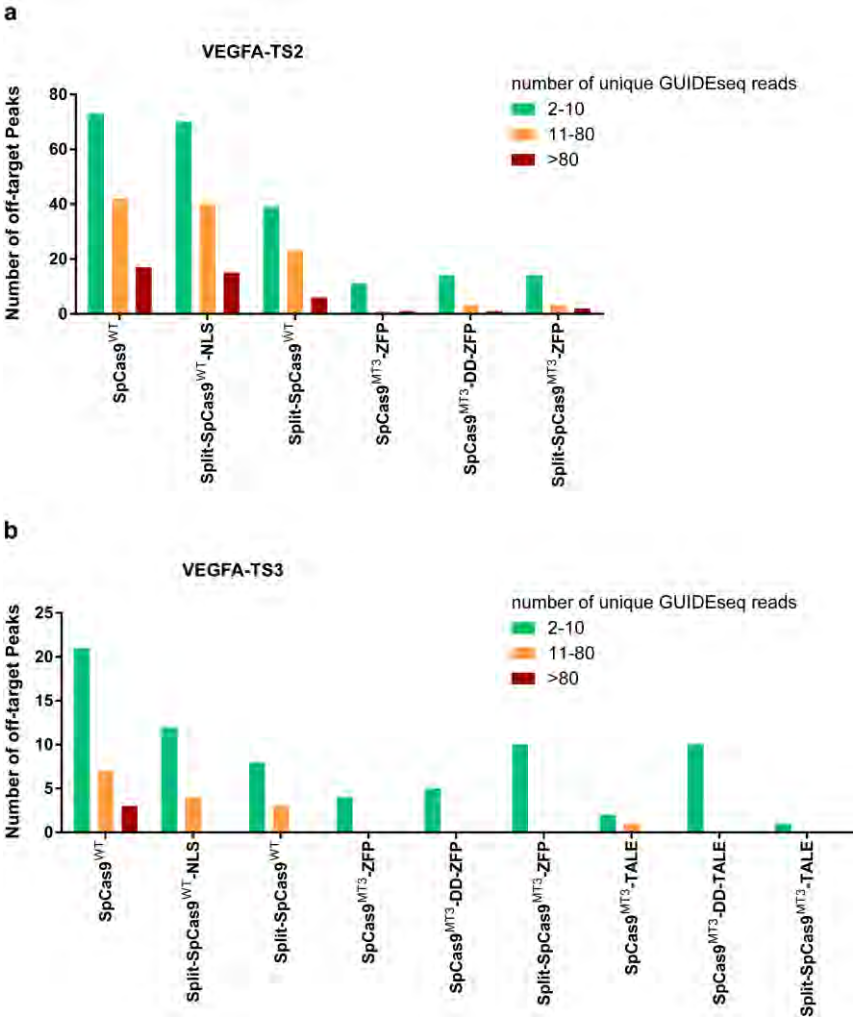


Figure 3. 7 GUIDE-seq analysis of the SpCas9^{MT3}-pDBD nuclease variants

Each graph displays the distribution of GUIDE-seq off-target peaks that fall into different bins based on the number of associated unique reads within each peak for nucleases programmed to target the *VEGFA* TS2 (**a**) and TS3 (**b**) sites. The target site is excluded from this graph. Both the SpCas9^{WT} and two Split-Cas9 architectures described by Zetsche and colleagues[10] (Split-SpCas9^{WT}-NLS [NLS on both components] & Split-SpCas9^{WT} [NLS on N-terminal component & NES on C-terminal component]) have a number of off-target peaks that contain high read counts when programmed with the TS2 sgRNA, which is consistent with high nuclease activity at these sites. For the covalent and dimeric SpCas9^{MT3}-ZFP nucleases programmed with the TS2 sgRNA the small number of identified off-target peaks primarily have a low number of reads. Since the number of GUIDE-seq reads is correlated with nuclease activity[20], this suggests that the off-target sites for the SpCas9^{MT3}-ZFP nucleases have low activity or are perhaps false positives. Although there are not as many GUIDE-seq peaks associated with nucleases programmed with the TS3 sgRNA, the general trend for these nucleases is similar with the majority of off-target peaks associated with the SpCas9^{MT3}-ZFP based nucleases having fewer than 10 reads (**Appendix 2**).

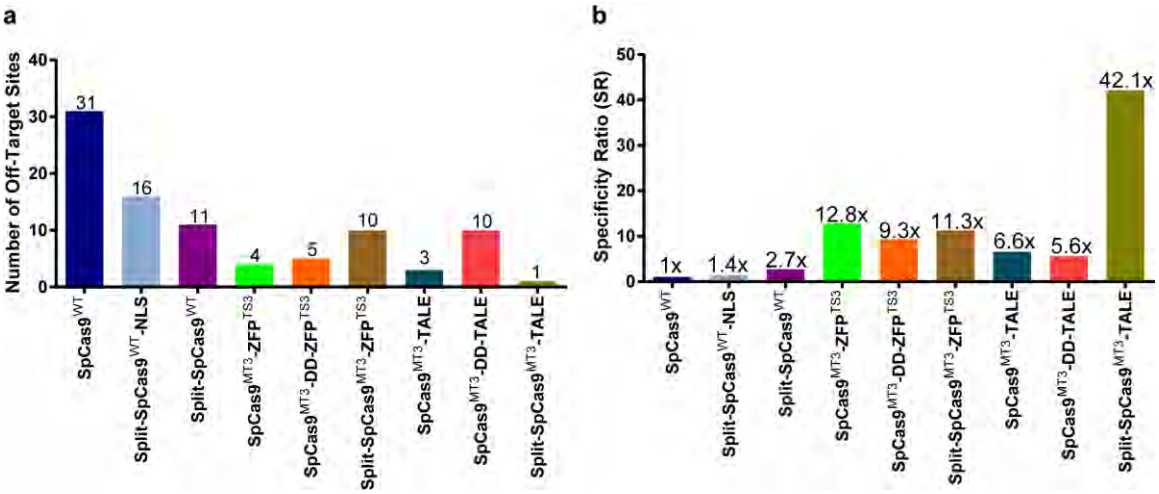


Figure 3. 8 GUIDE-seq analysis of the SpCas9^{MT3}-TALE^{TS3} variants.
(**a**) Number of GUIDE-seq peaks that are associated with each nuclease. This chart contains the nucleases listed in **Figure 3.6b** as well as the covalent SpCas9^{MT3}-TALE^{TS3} nuclease, and the SpCas9^{MT3}-DD-TALE^{TS3} and Split-SpCas9^{MT3}-TALE^{TS3} dimeric nucleases. The SpCas9^{MT3}-TALE-based nucleases all display a reduction in the number of peaks, although when SpCas9^{MT3}-DD-TALE^{TS3} is compared to the Split-Cas9^{WT} constructs the reduction is modest. The Split-SpCas9^{MT3}-TALE^{TS3} had the smallest number of off-target sites (1), but it also had the smallest number of reads at the target site – implying a lower cellular level of nuclease activity, which makes direct comparisons with the other constructs less direct. (**b**) Specificity Ratio (SR) for each nuclease, [number of unique reads at the target site divided by the sum of the unique GUIDE-seq reads associated with the off-target peaks]. This chart contains the nucleases listed in **Figure 3.6b** as well as the covalent SpCas9^{MT3}-TALE^{TS3} nuclease, and SpCas9^{MT3}-DD-TALE^{TS3} and Split-SpCas9^{MT3}-TALE^{TS3} dimeric nucleases. The SpCas9^{MT3}-TALE-based nucleases all display improved SRs relative to the SpCas^{WT}-based constructs, where they are similar to the SpCas9^{MT3}-ZFP-based nucleases. The one outlier (Split-SpCas9^{MT3}-TALE^{TS3}) had the lowest nuclease activity at the target site, which may reduce its activity at off-target sites (**Appendix 2**).

To confirm the improved precision of dimeric SpCas9^{MT3}-pDBD nucleases, we performed deep sequencing of PCR amplicons spanning a representative set of off-target sites within the genomes of nuclease treated cells. In this analysis we utilized dimeric SpCas9^{MT3}-ZFP systems with asymmetric localization sequences that have the highest on-target activity (**Figures 3.3 & 3.4**). These dimeric SpCas9^{MT3}-ZFP systems display a marked reduction in nuclease activity at off-target sites associated with the TS2 and TS3 sgRNAs in comparison with the NES-containing Split-Cas9^{WT} system (**Figure 3.6c**). Similar improvements in off-target activity were observed for the dimeric SpCas9^{MT3}-TALE nuclease programmed with the TS3 sgRNA (**Figure 3.9**).

These data demonstrate that the outstanding precision and greater PAM targeting range of the covalent Cas9-pDBD systems can be extended to inducible dimeric systems derived from this framework. The results mirror the improvements in specificity that were observed with the covalent SpCas9^{MT3}-ZFP system when compared to wild-type SpCas9 at the TS2 target site, where only a single prominent off-target site (OT2-2) displayed substantial activity[306]. We previously showed that remaining off-target activity within the covalent SpCas9^{MT3}-ZFP can be further attenuated by tuning the binding affinity of the associated ZFP[306]. We anticipate that this will also be the case for these dimeric systems. Although our dimeric systems were constructed using the FKRP/FRB domains to generate a rapamycin inducible system, they should also be amenable to the incorporation of other types of switchable domains (e.g. drug-dependent[309, 381, 389], light-dependent[309, 390]) for the temporal regulation of nuclease activity.

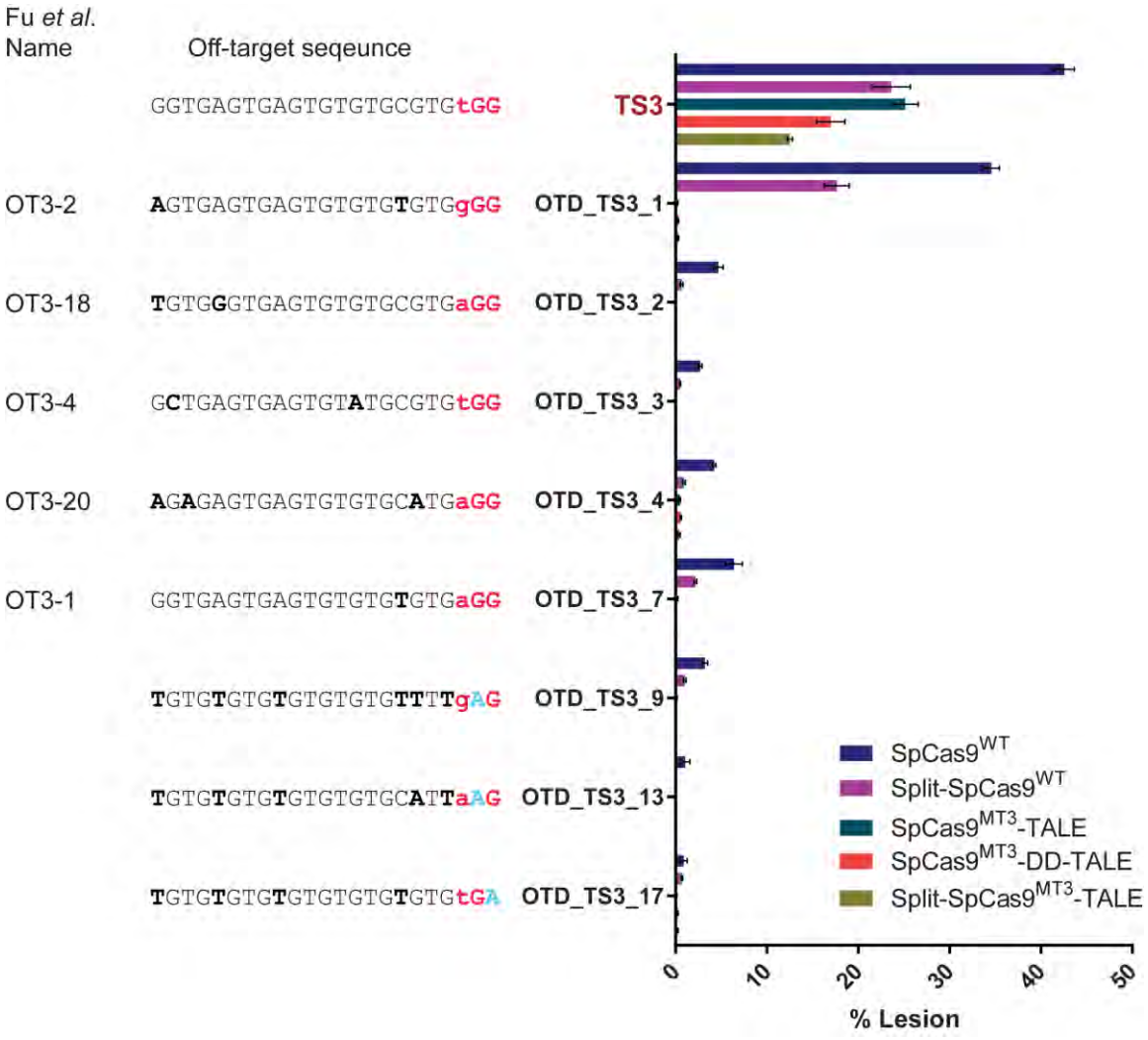


Figure 3. 9 Specificity analysis of SpCas9^{MT3}-TALE^{TS3} nuclease variants
Lesion rates for nucleases programmed to target *VEGFA* TS3 site. All of the non-standard nuclease variants (Split-SpCas9^{WT} [NLS on N-terminal component & NES on C-terminal component], SpCas9^{MT3}-TALE^{TS3}, SpCas9^{MT3}-DD-TALE^{TS3} and Split-SpCas9^{MT3}-TALE^{TS3}) have lower activity at the TS3 target site than SpCas9^{WT}. All of the SpCas9^{MT3}-TALE-based nucleases have lower activity at the off-target sites, where the difference in activity at OT3-2 are particularly dramatic. Site IDs from Fu *et al.*[257] are included for comparison. Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

Recently, two non-specific DNA contact attenuated SpCas9 systems were described that have improved nuclease precision (SpCas9-HF1[302] and eSpCas9_v1.1[301]). We performed a side-by-side comparison of these Cas9 variants and our covalent SpCas9^{MT3}-pDBD system to assess their relative levels of precision. Based on the deep-sequencing analysis of previously defined off-target sites, the precision of the covalent SpCas9^{MT3}-pDBD system compares favorably against SpCas9-HF1[302] and eSpCas9_v1.1[301] variants when loaded with challenging sgRNAs (TS2 and TS3; **Figure 3.10**). In addition, the covalent SpCas9^{MT3}-pDBD system can tolerate the presence of an extra G incorporated into the 5' end of the guide sequence (G+20nt; **Figure 3.11**), which is necessary for guide expression from U6 and T7 promoters and is frequently incorporated as an extra based in the guide when the genomic target site does not begin with a G. However, the addition of an extra G at the 5' end of the guide (G+20nt) diminishes the activities of both SpCas9-HF1 and eSpCas9_v1.1 similar to the attenuation previously described for SpCas9-HF1 when a G mismatch is present at the last position of the guide sequence (G+19nt[302]).

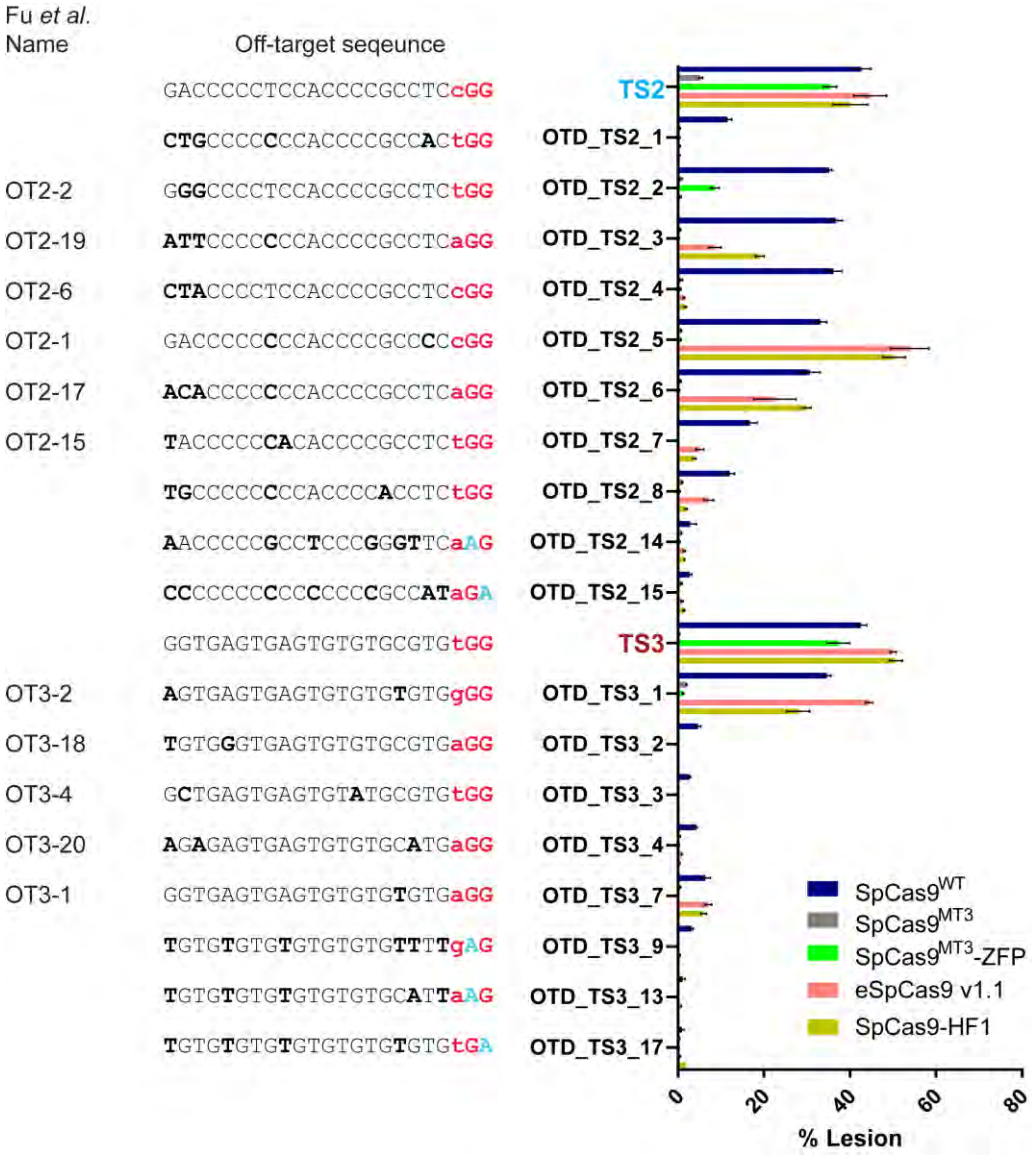


Figure 3. 10 Specificity comparison of SpCas9 variants

Covalent SpCas9^{MT3}-ZFP nucleases have improved precision relative to other high fidelity Cas9 variants. Lesion rates for nucleases programmed with the *VEGFA* TS2 or TS3 sgRNA at the target site and a subset of off-target sites identified by GUIDE-seq analysis for SpCas9^{WT}. Lesion rates were assayed by deep-sequencing of PCR products spanning each genomic locus for cells treated with SpCas9^{WT}, SpCas9^{MT3}, SpCas9^{MT3}-ZFP, SpCas9-HF1 or eSpCas9_v1.1 and the appropriate sgRNA. Although SpCas9-HF1[302] and eSpCas9_v1.1[301] have greatly improved precision relative to SpCas9^{WT}, there are still a few off-target sites that have near wild-type activity. SpCas9^{MT3}-ZFP nucleases achieve a much more dramatic reduction in activity at these off-target sites. Site IDs from Fu *et al.*[257] are included for comparison. Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

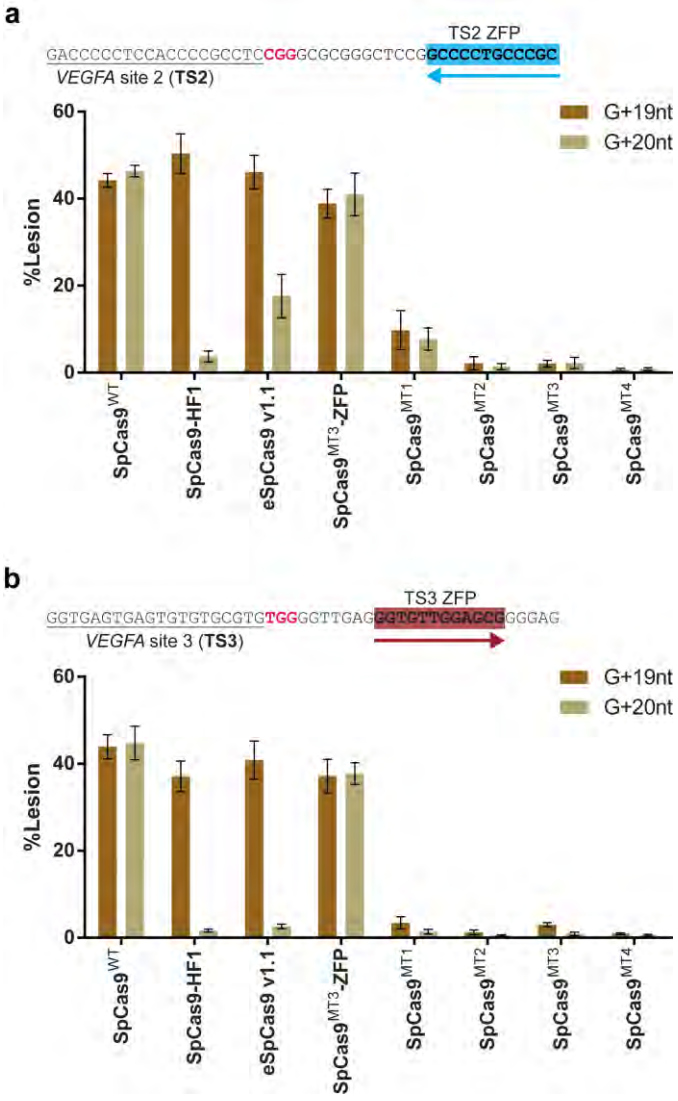


Figure 3. 11 The effects of an extra 5' G at the guide sequence on nuclease activity
Comparison of nuclease activity at the *VEGFA* TS2 or TS3 target sites for a variety of nuclease constructs. Both of these target sites contain a G at the 5' end of the protospacer, so a fully complementary guide (G+19) can be produced from a U6 promoter, which requires a G at the transcription start site for efficient initiation. SpCas9^{WT}, SpCas9-HF1, eSpCas9_v1.1 and SpCas9^{MT3}-ZFP display robust activity at the TS2 and TS3 target sites when programed with the G+19 guides. However, when an extra G is added onto the guide sequence (G+20) there is a dramatic reduction in activity for SpCas9-HF1 and eSpCas9_v1.1 at both target sites. The activities of SpCas9^{WT} and SpCas9^{MT3}-ZFP are unaffected. For reference, the attenuated SpCas9 PAM recognition mutants (MT1=R1333K, MT2=R1333S, MT3=R1335K and MT4=R1335S,) are included without a ZFP fusion. These have only limited activity in the absence of a ZFP fusion. Data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

3.4 Discussion

In this study, we developed two working prototypes of drug inducible Cas9-pDBD chimeras to achieve highly specific genome editing with temporal control. Drug inducible Cas9-pDBD chimeras retain the features—enhanced targeting range and improved specificity—of the original Cas9-pDBD[306]. We also show that the SpCas9^{MT3}-pDBD system has higher precision at challenging target sites than other previously described specific SpCas9 variants, SpCas9-HF1[302] and eSpCas9_v1.1[301]. Unlike the SpCas9-HF1 and eSpCas9_v1.1 systems, the SpCas9^{MT3}-pDBD platforms require the design of DNA-binding domains for the target site of interest. This extra effort makes the SpCas9^{MT3}-pDBD systems most amenable to therapeutic applications, where the improved nuclease precision and targeting range are worth the added investment needed to construct these systems. Recent advances in the characterization of zinc fingers recognizing all possible triplets and the selection of ZFPs with outstanding specificity from these libraries[37, 64] should facilitate the construction of ZFPs for a broad variety of sequences. Finally, we anticipate that the highly precise genome editing systems constructed on the SpCas9-pDBD framework will be adaptable to other Cas9 isoforms (e.g. NmCas9[242] or SaCas9[244]) and perhaps Cpf1-based systems[174].

CHAPTER IV: Orthogonal Cas9-Cas9 chimeras provide a versatile platform for genome editing

The development of robust, versatile and accurate toolsets are critical for the development of therapeutic genome editing applications. In this study, we establish RNA-programmable Cas9-Cas9 chimeras, in single- and dual-nuclease formats, as versatile genome engineering systems. In both of these formats, Cas9-Cas9 fusions display an expanded targeting repertoire and achieve highly specific genome editing. Dual-nuclease Cas9-Cas9 chimeras have distinct advantages over monomeric Cas9s including higher target site activity and the generation of predictable precise deletion products between their target sites. At a therapeutically relevant site within the *BCL11a* erythroid enhancer, Cas9-Cas9 nucleases produced precise deletions that comprised up to 97% of all sequence alterations. Thus Cas9-Cas9 chimeras represent an important new tool that could be particularly valuable for therapeutic genome editing applications where a precise cleavage position and defined end products are desirable.

4.1 Introduction

The Class 2 CRISPR-Cas bacterial adaptive immune system has been used for a wide variety of applications since being repurposed for programmable genome editing and gene regulation[172, 391]. The development of these tools for therapeutic genome editing applications is well underway with numerous investigations examining *ex vivo* and *in vivo* therapeutic approaches[5, 392]. The Type II effector protein Cas9 from *S. pyogenes* (SpCas9) is one of the most widely used of these nucleases due to its robust activity and broad targeting range[170, 173]. Target site recognition involves Cas9 binding its PAM recognition element and the complementarity of the complexed guide RNA[170, 205, 248, 393]. Once fully engaged, SpCas9 typically generates a blunt double-strand break (DSBs) at the target site[170]. In some instances the accuracy of the wild type SpCas9 ($SpCas9^{WT}$) nuclease is imperfect, leading to cleavage of "off-target" sites within the genome[256, 257, 286]. The resulting collateral damage to the genome is suboptimal for many therapeutic applications.

We and others have used protein and RNA engineering strategies to improve the specificity of SpCas9 for therapeutic genome editing applications[377, 378]. We previously reported that a chimera between PAM-interaction attenuated SpCas9 ($SpCas9^{MT}$) and a programmable DNA-binding domain (pDBD; ZFP or TALE) enhances the targeting range and specificity of SpCas9[306]. In the $SpCas9^{MT}$-pDBD chimeras the pDBD provides an additional stage of target site licensing prior to cleavage. The first stage is mediated by pDBD recognition of a sequence downstream of the PAM. The increased effective

concentration of SpCas9^{MT} upon pDBD binding facilitates recognition of the PAM element and initiation of R-loop formation[248, 306]. If sufficient complementarity exists between the sgRNA and the target site, cleavage of the DNA strands occurs. When programmed to target challenging repetitive sites in human genome, SpCas9^{MT}-pDBD chimeras achieve higher specificity than other high-fidelity SpCas9 variants (**Figure 3.10**). However, building functional pDBDs requires some level of expertise in pDBD assembly, which creates a barrier to this platform's adoption.

In this study, we assessed the feasibility of substituting the pDBD within the Cas9^{MT}-pDBD platform with an orthogonal Cas9 from either *N. meningiditis* (NmCas9)[242, 314] or *S. aureus* (SaCas9)[244] to develop an entirely RNA-programmable nuclease platform spanning two linked Cas9 domains. We constructed these Cas9-Cas9 chimeras in both single- and dual-nuclease formats. In the single-nuclease format, the attenuated SpCas9^{MT} domain is fused to a nuclease-dead NmCas9 or SaCas9, whose orthogonal guide is programmed to target a neighboring DNA sequence. Here the nuclease-dead Cas9 (dNmCas9 or dSaCas9) should act like a pDBD to deliver SpCas9^{MT} to the target site, thereby permitting target site recognition through the increased effective concentration of the SpCas9^{MT} nuclease (**Figure 4.1a**). Similar to SpCas9^{MT}-pDBDs[306], SpCas9^{MT}-dNm/SaCas9 chimeras achieve a high-level of specificity as assessed via GUIDE-seq[286] and targeted amplicon sequencing. In the dual-nuclease format, both nucleases within the orthogonal Cas9-Cas9 fusions are active. We hypothesized that synchronous cleavage of the genome at two neighboring positions will primarily produce segmental deletions with defined junctions (referred as precise deletions), as is observed to varying extents when

independent nucleases are targeted to neighboring sites within a genome[173, 189, 193]. When programmed to target composite sites within the human genome, SpCas9^{WT}-Nm/SaCas9^{WT} nuclease fusions produce a larger fraction of precise deletions, as high as 97% of all lesions, than a pair of independent Cas9 monomers used simultaneously. Similar to SpCas9-pDBD chimeras[306], Cas9-Cas9 fusions in both the single- and dual-nuclease format expand the targeting range of SpCas9 by allowing the recognition of suboptimal PAMs.

These dual nucleases should particularly be useful for the disruption of therapeutically relevant regulatory elements within a genome. In sickle cell disease, one proposed therapeutic approach is to induce the expression of the fetal γ-globin gene by deleting the GATA1 binding motif within the erythroid-lineage specific regulatory element (enhancer +58kb) of the *BCL11A* gene[96-99, 394]. Here, we show that Cas9-Cas9 fusions programmed to target sites spanning the GATA1 element can delete this regulatory element with greater efficiency and accuracy than separate Cas9/sgRNA complexes.

4.2 Materials and Methods

4.2.1 Plasmid Constructs

Our SpCas9-Sa/NmCas9 experiments employed the following plasmids: All sgRNAs are individually expressed under a U6 promoter from a pBluescript II SK(+)-based vector. All single- or dual-Cas9 nuclease constructs are expressed via a CMV IE94 promoter from a pCS2-Dest gateway plasmid[306]. NmCas9 and SaCas9 open reading frames for nuclease construction were obtained from Addgene (#48670 & #61591). The nuclease-

dead versions of these constructs (dNmCas9: D16A, D587A, H588A and N611A; dSaCas9: D10A and N580A) are generated via site-directed mutagenesis. These plasmids will be deposited to Addgene for community distribution. We used the single-strand annealing-based plasmid reporter assay developed by Porteus laboratory[354] to monitor nuclease activity. Nuclease target sequences are cloned into the M427 plasmid in between EcoRI and SbfI sites.

4.2.2 Cell culture and transfection

We received the Human Embryonic Kidney (HEK293T) cells from our collaborator M. Green (UMass Medical School, Worcester, Massachusetts, USA). The cell line was authenticated by University of Arizona Genetics Core and tested for mycoplasma contamination at regular intervals. HEK293T were cultured in high glucose DMEM with 10% FBS and 1% Penicillin/Streptomycin (Gibco) in a 37°C incubator with 5% CO_2. We used cells at a passage number from 5 until 25 for transient transfection to assay nuclease activity. In 24-well format, about $1.6x10^5$ cells were transfected by Polyfect transfection reagent (Qiagen) according to the manufacturer's suggested protocol. For both single and dual nucleases we used 50 ng of each sgRNA-expressing plasmid and 50 ng mCherry-expressing plasmid, 50 ng of single nuclease (SpCas9 or NmCas9 or SaCas9) or 100 ng Cas9-Cas9 fusion expressing plasmid. In addition, pBluescript II SK(+) was also added to the co-transfection mix to bring the total DNA mass to 300 ng per transfection. For the SSA-reporter assay, an additional 150 ng M427 reporter plasmid was added.

4.2.3 GFP Reporter Assay

Flow cytometry experiments were performed as described previously[306]. 48 h after transfection, cells were harvested and resuspended in 1xPBS for flow cytometry (Becton Dickonson FACScan). 10000 events were counted from each sample for FACS analysis. To adjust the transfection efficiency differences in between samples, cells were initially gated for mCherry-expression, and the percentage of EGFP expressing cells (nuclease positive events) were quantified within mCherry positive cells. Experiments were performed in three replicates on different days. The data are reported as mean values with error bars indicating the standard error of the mean.

4.2.4 Immunofluorescence

HEK293T cells are transfected in 6-well format via Polyfect transfection reagent (Qiagen) using the manufacturer's suggested protocol with 300 ng Cas9-Cas9 fusion expression plasmid and 150 ng of each sgRNA expression plasmid on a cover slip. 48 h following transfection, transfection media was removed, cells were washed with 1x PBS and fixed with 4% formaldehyde in 1x PBS for 15 min at room temperature. Following blocking (blocking solution: 2% BSA, 0.3% Triton X-100 within 1x PBS), samples were stained with mouse anti-hemagglutinin (Sigma, H9658, 1:500), and Alexa 488 donkey anti-mouse IgG (H+L; Invitrogen, A-21202, 1:2000), sequentially. VECTASHIELD Mounting Medium with DAPI (Vector Laboratories, H-1200) was used to stain the nuclei and to mount the samples on slide. Images were taken with Zeiss AxioPlan 2 IE Motorized Microscope System.

4.2.5 Target and off-targetsite lesion type and frequency analysis by deep sequencing

Library construction for deep sequencing is modified from our previous report[306]. Briefly, 72 h after transfection, cells are harvested and genomic DNA extracted with GenElute Mammalian Genomic DNA Miniprep Kit (Sigma). Genomic loci spanning the target and off-target sites were PCR amplified with locus-specific primers carrying tails complementary to the Truseq adapters (**Appendix 1**). 50 ng input genomic DNA was PCR amplified with Q5 High-Fidelity DNA Polymerase (New England Biolabs): (98°C, 15s; 67°C 25s; 72°C 20s) x30 cycles. For the construction of the UMI-based library, 50 ng input genomic DNA was first linearly pre-amplified with 10 nM final concentration 5p-BCL11A_enh58_UMI primer using the Q5 High-Fidelity DNA Polymerase (New England Biolabs): (98°C, 60s; 67°C, 25s; 72°C, 20s) x10 cycles. To the same reaction mix, 500 nM final concentration 5p-DS_constant and 3p-BCL11A_enh58_DS primers were added for another round of amplification (98°C, 60s; 67°C, 25s; 72°C, 20s) for 30 cycles. Next, 0.1 µl of each PCR reaction was amplified with barcoded primers to reconstitute the TruSeq adaptors using the Q5 High-Fidelity DNA Polymerase (New England Biolabs): (98°C, 15s; 67°C, 25s; 72°C, 20s) x10 cycles. Equal amounts of the products were pooled and gel purified. The purified library was deep sequenced using a paired-end 150bp Illumina MiSeq run.

MiSeq data analysis was done with the help of Unix-based software tools. First, we employed FastQC[395] to determine the quality of paired-end sequencing reads (R1 and R2 fastq files). Next, we used paired end read merger (PEAR)[384] to pool raw paired-end reads

and generate single merged high-quality full-length reads. Reads were then filtered according to quality via FASTQ[385] for a mean PHRED quality score above 30 and a minimum per base score above 24. After that, we used BWA (version 0.7.5) and SAMtools (version 0.1.19) for aligning each group of filtered reads to a corresponding reference sequence. To determine lesion type, frequency, size and distribution, all edited reads from each experimental replicate were combined and aligned, as described above. Alignments were categorized into seven classes: unedited, SpCas9 indels, Nm/SaCas9 indels, precise deletions, imprecise deletions, SpCas9 + Nm/SaCas9 indels, and inversions. Lesion types and frequencies were then cataloged in a text output format at each base using bam-readcount. For each treatment group, the average background lesion frequencies (based on lesion type, position and frequency) of the triplicate negative control group were subtracted to obtain the nuclease-dependent lesion frequencies. Next, using R, a system for statistical computation and graphics[357], we assessed whether the Cas9-Cas9 fusions resulted in different lesion rates from two independent Cas9s. Percent of lesion rates were transformed using logit function followed by one-way analysis of variance (ANOVA) with the Randomized Complete Block Design (**Figures 4.3, 4.5, 4.6, and 4.9**) or Completely Randomized Design (**Figure 4.12**). BH-adjusted p-values were calculated to counteract the problem of multiple comparisons of the data shown in **Figure 4.12** (Ref. 358).

For UMI analysis, we first used BWA (version 0.7.5) and SAMtools (version 0.1.19) for aligning each group of filtered reads to a corresponding reference sequence. Next, we used a custom Python and PySAM script to process mapped reads into counts of UMI-labeled reads for each target. Alignments were categorized into seven classes:

unedited, SpCas9 indels, Nm/SaCas9 indels, precise deletions, imprecise deletions, SpCas9 + Nm/SaCas9 indels, and inversions. Next, we identified UMI duplicates and the minimal set of amplicons that can account for the full set of reads with unique UMIs. Then, the reads with unique UMI were counted. The same pipeline was used separately for each sample, and the resulting UMIs number tables were concatenated and loaded into GraphPad Prism 7 for data visualization.

4.2.6 GUIDE-seq

GUIDE-seq[286] was performed with the adjustments to the original protocol as described previously[306]. Briefly, HEK293T cells were transfected using Polyfect transfection reagent (Qiagen) according to the manufacturer's suggested protocol with 50 ng of single nuclease ($SpCas9^{WT}$, $NmCas9^{WT}$, $SaCas9^{WT}$) or 100 ng Cas9-Cas9 fusion expression plasmid, 50 ng of each sgRNA expressing plasmid, 50 ng of a mCherry expression plasmid and 5 pmol of annealed GUIDE-seq oligonucleotide. 72 h after transfection, genomic DNA was extracted with GenElute Mammalian Genomic DNA Miniprep Kit (Sigma) according to the manufacturer's suggested protocol. We prepared the GUIDE-seq library with the original adaptors according to protocols described by Joung and colleagues[286]. Each library was indexed within the P5 and P7 adaptors for multiplex sequencing. The libraries were deep-sequenced as a pool using two paired-end 150-bp Illumina MiSeq runs.

Deep sequencing data from the GUIDE-seq experiment was analyzed using the Bioconductor Package GUIDEseq (v1.4.1) [386]. The window size for peak aggregation was

set to 50 bp. Off-target site identification parameters were set for SpCas9 as follows: min.reads = 2, min.reads.per.lib = 1, distance.threshold = 70, min.peak.score.1strandOnly = 2, upstream = 20, downstream = 20, max.mismatch = 6, PAM.pattern = "NNN$", allowed.mismatch.PAM = 2. For NmCas9, same parameters were used except the following: PAM.size = 8, PAM = "NNNNGHTT", PAM.pattern = "NNNNGNNN$", allowed.mismatch.PAM = 3, max.mismatch = 9. For SaCas9, same parameters were used as for SpCas9 except the following: PAM.size = 6, PAM = "NNGRRT", PAM.pattern = "NNGNNN$", allowed.mismatch.PAM = 3, max.mismatch = 6. The potential off-target sites identified for each nuclease are listed in **Appendix 2**. The Specificity Ratio is calculated as the sum of the unique GUIDE-seq reads at the target site divided by all of the unique reads at all of the computationally identified off-target sites.

4.3 Results

4.3.1 Single nuclease Cas9MT-dCas9 fusions facilitate highly accurate editing

To identify binding site parameters necessary for functional Cas9-Cas9 fusion activity, we used a plasmid reporter assay[354] to detect DSB formation. In this assay, two protospacers with optimal cognate PAM elements (for SpCas9[170, 304] and NmCas9[242, 322] or SpCas9 and SaCas9[244, 305]) are arrayed in tandem in four configurations with various intervening spacings (**Figure 4.1b** and **Figure 4.2b**). In this assay, SpCas9^{WT} displays robust nuclease activity while the R1335K mutant (SpCas9^{MT3})[306] has substantially reduced activity. Fusion of nuclease-dead NmCas9 or SaCas9 to SpCas9^{MT3} (SpCas9^{MT3}-dNmCas9 or SpCas9^{MT3}-dSaCas9) restores nuclease activity in configurations where the

SpCas9 protospacer is upstream of the orthogonal Cas9 target site (**Figure 4.1b**). Notably, the C-terminal fusions of dNm/SaCas9 are more successful at restoring the loss of activity of SpCas9^{MT3} than N-terminal fusions with the single linker configuration that we tested (**Figure 4.2b**). In our pilot experiments, Cas9-dCas9 fusions displayed a low level of activity at genomic sites. We reasoned that this could be due to poor nuclear localization of the fusion protein. Immunofluorescent imaging of the Cas9-Cas9 fusions indicated that the position and number of nuclear localization signals (NLSs) impact nuclear import efficiency of the Cas9-Cas9 fusions (**Figure 4.2c**). NLSs within the linker between the two Cas9 domains appeared to have little function. Addition of NLSs at N and C termini was necessary for efficient nuclear localization of the Cas9-Cas9 fusions. We used this architecture for the remainder of this study.

One of the salient features of the SpCas9^{MT}-pDBD chimera is that this platform has improved specificity relative to SpCas9^{WT} (Ref. 306). To evaluate the specificity of Cas9-Cas9 fusions, we programmed SpCas9^{MT3}-dNmCas9 and SpCas9^{MT3}-dSaCas9 fusions to recognize a SpCas9 site (*VEGFA-TS3*)[257, 286], which has numerous highly active off-target sites (**Figure 4.1c**). At the *VEGFA-TS3* site, both SpCas9^{MT3}-dNmCas9 and SpCas9^{MT3}-dSaCas9 fusions display similar levels of on-target activity (**Figure 4.1d**). We used GUIDE-seq[286] to assess the genome-wide specificity of these nucleases. In comparison to SpCas9^{WT}, both SpCas9^{MT3}-dNmCas9 and SpCas9^{MT3}-dSaCas9 fusions have a substantially reduced number of GUIDE-seq peaks (**Figure 4.1e** and **Appendix 2**). In addition to the reduction in the numbers of GUIDE-seq peaks, the Specificity Ratios (number of unique capture events at the target site divided by sum of unique capture events

at all off-target sites) for the SpCas9^{MT3}-dNmCas9 and SpCas9^{MT3}-dSaCas9 fusions are higher than for SpCas9^{WT} (**Figure 4.1f**). To validate the off-target editing rates, we used targeted amplicon deep sequencing for a representative set of off-target sites within the genomes of the treated cells. Despite being comparably active to SpCas9^{WT} at the target site, SpCas9^{MT3}-dNmCas9 and SpCas9^{MT3}-dSaCas9 fusions present dramatically better discrimination against near-cognate off-target sites (**Figure 4.1g**). Overall, GUIDE-seq and targeted amplicon deep sequencing datasets indicate that SpCas9^{MT}-dNm/SaCas9 fusions substantially improve the specificity of SpCas9^{WT} similar to the levels of achieved by SpCas9-pDBDs[306].

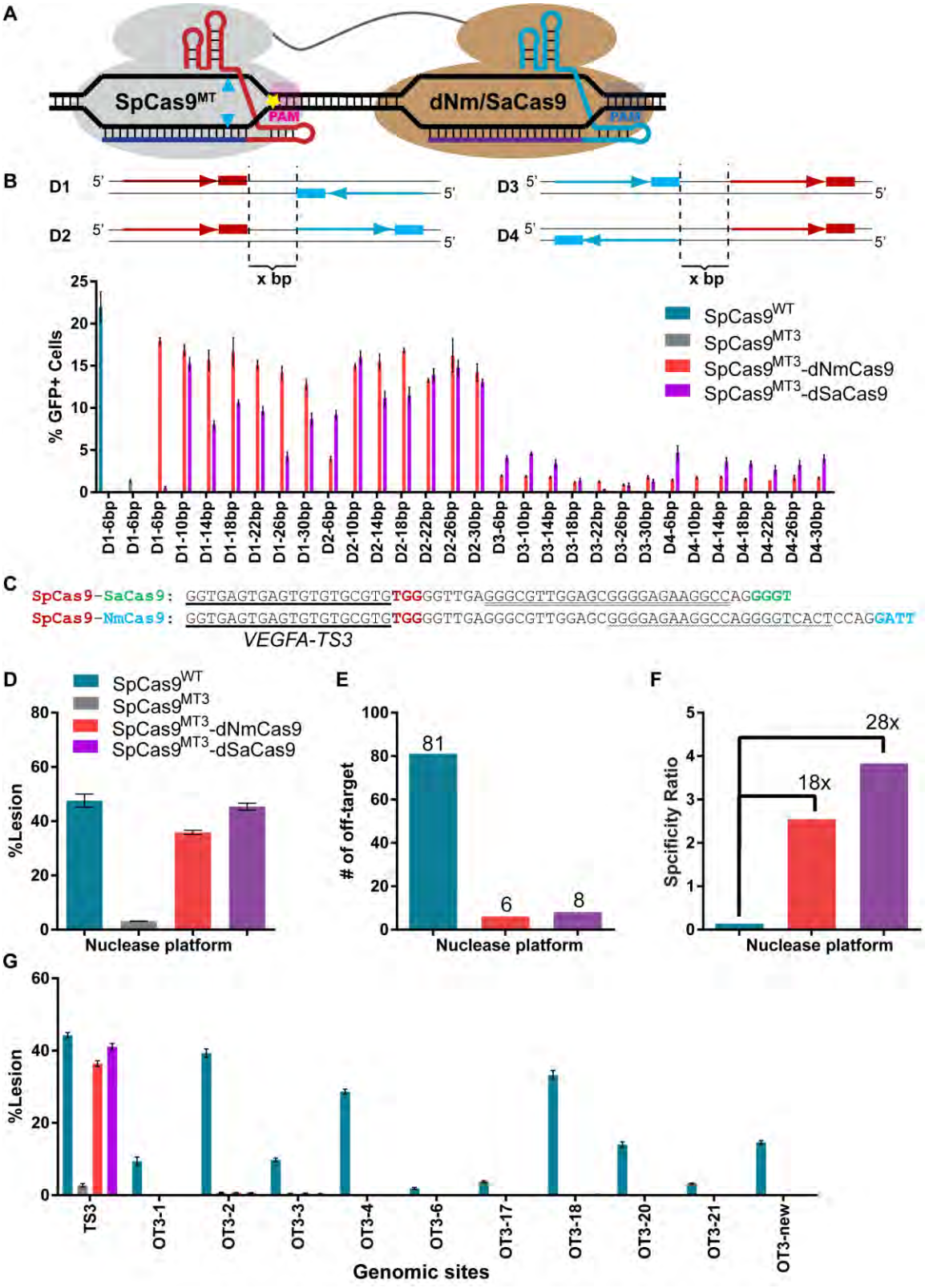

A
SpCas9MT
PAM
dNm/SaCas9
PAM
B
D1
D2
D3
D4
5'
x bp
% GFP+ Cells
SpCas9WT
SpCas9MT3
SpCas9MT3-dNmCas9
SpCas9MT3-dSaCas9
D1-6bp
D1-6bp
D1-6bp
D1-10bp
D1-14bp
D1-18bp
D1-22bp
D1-26bp
D1-30bp
D2-6bp
D2-10bp
D2-14bp
D2-18bp
D2-22bp
D2-26bp
D2-30bp
D3-6bp
D3-10bp
D3-14bp
D3-18bp
D3-22bp
D3-26bp
D3-30bp
D4-6bp
D4-10bp
D4-14bp
D4-18bp
D4-22bp
D4-26bp
D4-30bp
C
SpCas9-SaCas9: GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGGCGTTGGAGCGGGGAGAAGGCCAGGGGT
SpCas9-NmCas9: GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGGCGTTGGAGCGGGGAGAAGGCCAGGGGTCACTCCAGGATT
VEGFA-TS3
D
SpCas9WT
SpCas9MT3
SpCas9MT3-dNmCas9
SpCas9MT3-dSaCas9
%Lesion
Nuclease platform
E
# of off-target
81
6
8
Nuclease platform
F
Spcificity Ratio
28x
18x
Nuclease platform
G
%Lesion
TS3
OT3-1
OT3-2
OT3-3
OT3-4
OT3-6
OT3-17
OT3-18
OT3-20
OT3-21
OT3-new
Genomic sites

Figure 4. 1 Development of a functional Cas9-Cas9 nuclease framework
(**a**) Schematic of SpCas9^{MT}-dNm/SaCas9 fusions. PAM-interaction attenuated SpCas9^{306} (yellow star) is C-terminally fused to a nuclease dead Cas9 from *N. meningiditis* or *S. aureus*. Each Cas9 is loaded with its cognate sgRNA. (**b**) Top, schematic of parameters tested for target site organization. Four composite target site configurations are tested (D1:D4). Red arrow line and the rectangle represents SpCas9 protospacer in 5' to 3' orientation and PAM respectively. Blue arrow line and the rectangle represents Nm/SaCas9 protospacer in 5' to 3' orientation and PAM respectively. Two dashed lines indicate the closest nucleotide in between the binding site of two orthogonal Cas9, and x represents the number of nucleotides in between. Bottom, activity profiles of SpCas9 (blue), SpCas9^{MT3} (R1335K; gray), SpCas9^{MT3}-dNmCas9 (pink), and SpCas9^{MT3}-dSaCas9 (purple) in GFP reporter assay. (**c**) Sequences of dual Cas9 genomic target sites at *VEGFA* locus. SpCas9 protospacer is bold underlined and its PAM is in red, SaCas9 protospacer is double underlined and its PAM is green, NmCas9 protospacer is wavy underlined and its PAM is in blue. (**d**) Lesion rates of the nuclease platforms are determined by deep sequencing. (**e, f**) Genome-wide off-target analysis of the nuclease platforms determined via GUIDE-seq[286] (**Appendix 2**). (**e**) The number of off-target peaks detected for the given nuclease. (**f**) Fold improvement of specificity ratio of the Cas9-dCas9 framework relative to SpCas9^{WT}. (**g**) Deep sequencing determined lesion rates of the nucleases at small set of off-target sites discovered within the GUIDE-seq data. GUIDE-seq result is from single experiment, whereas GFP reporter assay and deep sequencing data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

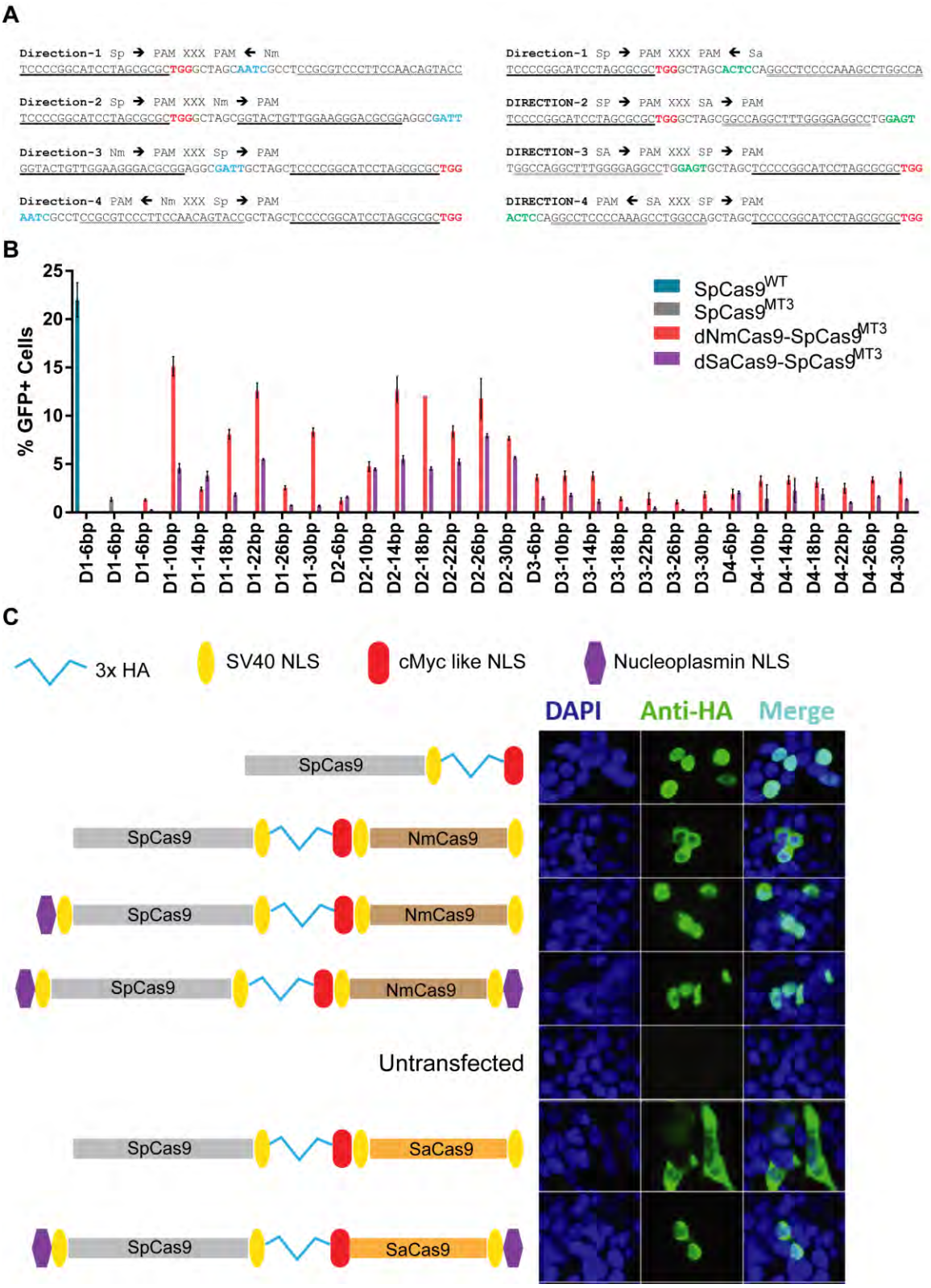

A
Direction-1 Sp → PAM XXX PAM ← Nm
TCCCCGGCATCCTAGCGCGCTGGGCTAGCAATCGCCTCCGCGTCCCTTCCAACAGTACC
Direction-2 Sp → PAM XXX Nm → PAM
TCCCCGGCATCCTAGCGCGCTGGGCTAGCGGTACTGTTGGAAGGGACGCGGAGGCGATT
Direction-3 Nm → PAM XXX Sp → PAM
GGTACTGTTGGAAGGGACGCGGAGGCGATTGCTAGCTCCCCGGCATCCTAGCGCGCTGG
Direction-4 PAM ← Nm XXX Sp → PAM
AATCGCCTCCGCGTCCCTTCCAACAGTACCGCTAGCTCCCCGGCATCCTAGCGCGCTGG
Direction-1 Sp → PAM XXX PAM ← Sa
TCCCCGGCATCCTAGCGCGCTGGGCTAGCACTCCAGGCCTCCCCAAAGCCTGGCCA
DIRECTION-2 SP → PAM XXX SA → PAM
TCCCCGGCATCCTAGCGCGCTGGGCTAGCGGCCAGGCTTTGGGGAGGCCTGGAGT
DIRECTION-3 SA → PAM XXX SP → PAM
TGGCCAGGCTTTGGGGAGGCCTGGAGTGCTAGCTCCCCGGCATCCTAGCGCGCTGG
DIRECTION-4 PAM ← SA XXX SP → PAM
ACTCCAGGCCTCCCCAAAGCCTGGCCAGCTAGCTCCCCGGCATCCTAGCGCGCTGG
B
% GFP+ Cells
25
20
15
10
5
0
SpCas9WT
SpCas9MT3
dNmCas9-SpCas9MT3
dSaCas9-SpCas9MT3
D1-6bp
D1-6bp
D1-6bp
D1-10bp
D1-14bp
D1-18bp
D1-22bp
D1-26bp
D1-30bp
D2-6bp
D2-10bp
D2-14bp
D2-18bp
D2-22bp
D2-26bp
D2-30bp
D3-6bp
D3-10bp
D3-14bp
D3-18bp
D3-22bp
D3-26bp
D3-30bp
D4-6bp
D4-10bp
D4-14bp
D4-18bp
D4-22bp
D4-26bp
D4-30bp
C
3x HA
SV40 NLS
cMyc like NLS
Nucleoplasmin NLS
DAPI
Anti-HA
Merge
SpCas9
SpCas9
NmCas9
SpCas9
NmCas9
SpCas9
NmCas9
Untransfected
SpCas9
SaCas9
SpCas9
SaCas9

Figure 4. 2 Identification of the functional parameters for Cas9-Ca9 fusion activity
(**a**) Sequence information of the target sites tested for SpCas9-SaCas9 and SpCas9-NmCas9 fusions. SpCas9 protospacer is bold underlined and its PAM is in red, SaCas9 protospacer is double underlined and its PAM is green, NmCas9 protospacer is wavy underlined and its PAM is in blue. (**b**) Activity profiles of SpCas9 (blue), SpCas9^{MT3} (R1335K; gray), dNmCas9-SpCas9^{MT3} (pink), and dSaCas9-SpCas9^{MT3} (purple) in GFP reporter assay. The data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m. (**c**) Cellular localization profile of Cas9-Cas9 fusion proteins in cell. On left is the schematic of various compositions of the Cas9-Cas9 fusion protein. On right, immunofluorescence imaging of nucleus (blue) and Cas9-Cas9 proteins (green).

4.3.2 Cas9-Cas9 dual nucleases primarily generate precise segmental deletions

Unlike SpCas9-pDBDs, the Cas9-Cas9 fusions can also be used as dual nucleases. We hypothesized that when a pair of Cas9 nucleases dock together on a target site, they will generate two DSBs synchronously. Since rejoining of the broken ends without resection via canonical NHEJ is the predominant DSB repair response in mammalian cells[10], the primary editing outcome of Cas9-Cas9 dual nucleases should be the precise deletion of the intervening segment between the cleavage sites. Previous studies have demonstrated that wild-type Cas9 nucleases that are targeted to a pair of neighboring genomic sequences can produce precise deletions but with variable efficiencies[173, 189, 193]. Such variability in editing outcomes is possibly due to asynchronous cleavage due to differences in the efficiency of recognition or cleavage by the nuclease at two different sequences. Cas9-Cas9 dual nucleases should improve the level of synchrony by delivering both nucleases simultaneously to a pair of target sequences.

To test this hypothesis, we programmed SpCas9^{WT}-NmCas9^{WT} dual nucleases to target the *VEGFA* locus in HEK293T cells. PCR amplification of the *VEGFA* locus from nuclease-treated cells indicates that dual-Cas9 nucleases, either independently or as fusions, generate precise segmental deletions, as anticipated (**Figure 4.3a**). In principle, two DSBs at a nearby locus in the genome can produce at least six possible repair outcomes: random indels at the first nuclease cut site (SpCas9 indel), random indels at the second nuclease cut site (NmCas9 indel), random indels at both nuclease cut sites (Sp&NmCas9 indel), precise deletions, imprecise deletions, and inversions. To monitor the presence and distribution of each set of repair events quantitatively, we applied targeted

amplicon deep sequencing of the genomic DNAs from cells treated with the different nuclease platforms. Analysis of the *VEGFA* amplicon sequencing data indicate two superior features of Cas9-Cas9 dual nuclease fusions ($SpCas9^{WT}$-$NmCas9^{WT}$, $SpCas9^{MT}$-$NmCas9^{WT}$) over two independent Cas9s ($SpCas9^{WT}$ + $NmCas9^{WT}$): Cas9-Cas9 fusions display higher levels of genome modification and the fraction of these events that are precise deletions than two independent Cas9 nucleases (**Figure 4.3b**). We observed similar editing outcomes for SpCas9 and SaCas9 dual nucleases (**Figure 4.4**).

To test the generality of this phenomenon, we initially screened 41 genomic sites for the activity profiles of single and dual nucleases, and for the types of lesions that are produced within the genome. Overall, SpCas9 has a higher median nuclease activity than SaCas9 and NmCas9. The lesion rates of the dual independent nucleases are similar to the levels of achieved by SpCas9 alone, but the composition of the lesions changes due to the production of precise deletions (**Figure 4.5a-b**). In comparison to dual independent nucleases, Cas9-Cas9 dual nucleases not only have higher overall activity, but also produce precise segmental deletions more efficiently (**Figure 4.5.c**).

Next, we selected a representative set of 12 sites from these 41 sites that span different activity profiles for the individual nucleases to assess in greater depth the nuclease activities of single and dual nucleases for wild-type and PAM-interaction-deficient[306] forms. $SpCas9^{WT}$-Sa/$NmCas9^{WT}$ dual-nuclease fusions have higher overall activity than SpCas9 and two independent Cas9s (**Figure 4.3c**). This enhancement is likely due to cooperativity between the fused nucleases, where strong binding of one of the Cas9s

increases the effective concentration and consequently the activity of the other nuclease. The total activity levels of SpCas9^{MT}-Sa/NmCas9^{WT} fusions are associated with the activity levels of the orthogonal Cas9:sgRNA complexes (**Figure 4.6**). SpCas9^{WT}-Sa/NmCas9^{WT} dual-nuclease fusions double the production of precise deletions relative to two independent Cas9s (**Figure 4.3d**). More importantly, precise deletions are the predominant products of Cas9-Cas9 dual nucleases, encompassing on average >90% of all lesion types (**Figure 4.3e** and **Figure 4.6**). These data suggest that generation of two synchronous nearby breaks by Cas9-Cas9 dual nucleases are preferentially repaired via canonical NHEJ.

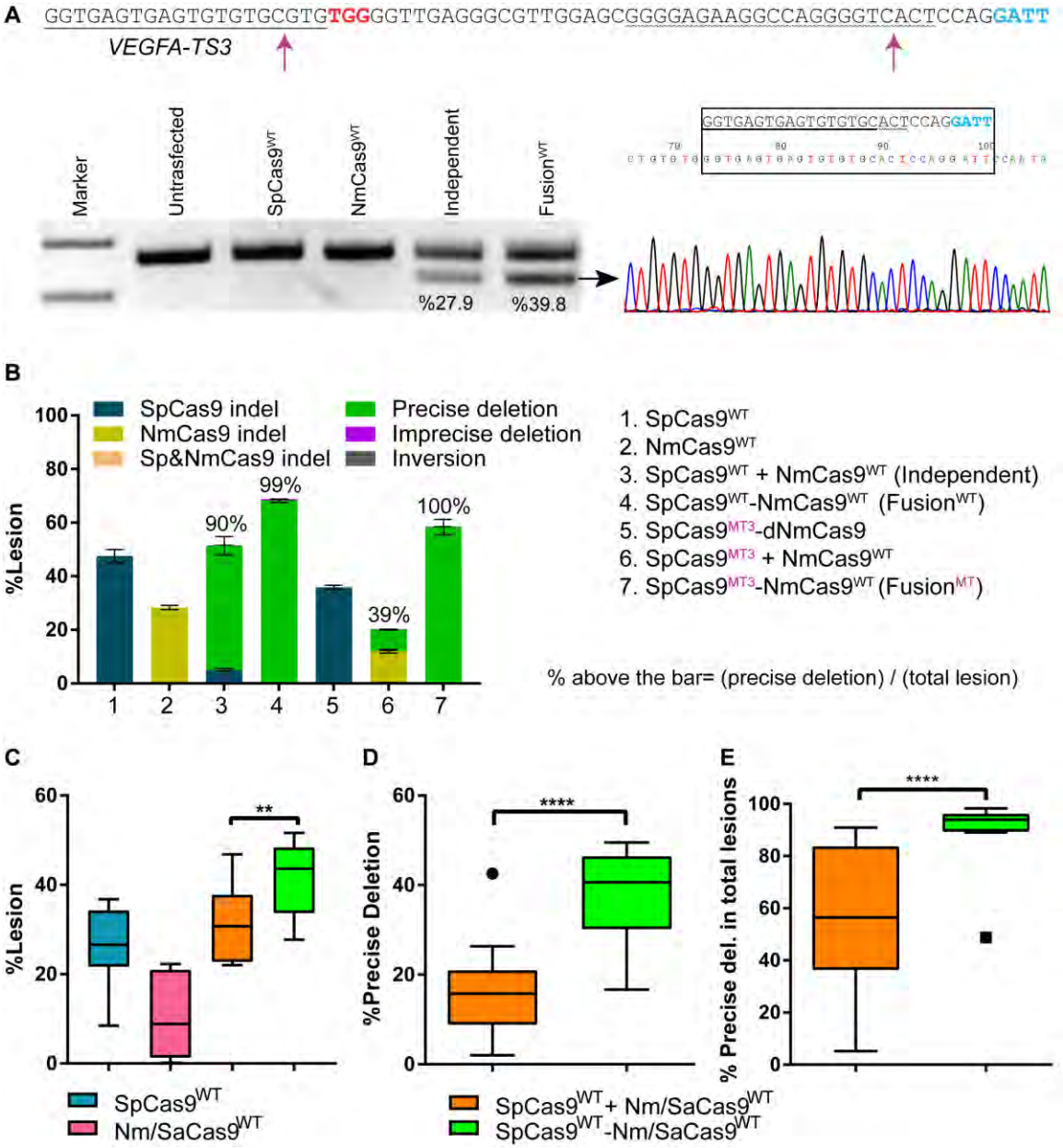

A
GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGGCGTTGGAGCGGGGAGAAGGCCAGGGGTCACTCCAGGATT
VEGFA-TS3
Marker
Untrasfected
SpCas9WT
NmCas9WT
Independent
FusionWT
%27.9
%39.8
B
SpCas9 indel
NmCas9 indel
Sp&NmCas9 indel
Precise deletion
Imprecise deletion
Inversion
%Lesion
99%
90%
100%
39%
1. SpCas9WT
2. NmCas9WT
3. SpCas9WT + NmCas9WT (Independent)
4. SpCas9WT-NmCas9WT (FusionWT)
5. SpCas9MT3-dNmCas9
6. SpCas9MT3 + NmCas9WT
7. SpCas9MT3-NmCas9WT (FusionMT)
% above the bar= (precise deletion) / (total lesion)
C
%Lesion
**
SpCas9WT
Nm/SaCas9WT
D
%Precise Deletion
****
SpCas9WT+ Nm/SaCas9WT
SpCas9WT-Nm/SaCas9WT
E
% Precise del. in total lesions
****

Figure 4. 3 Cas9-Cas9 dual nucleases generate uniform deletion products
(**a**) Top, sequence of SpCas9-NmCas9 target site: SpCas9 protospacer is bold underlined and its PAM is in red, NmCas9 protospacer is wavy underlined and its PAM is in blue. Purple arrows indicate the positions of the double-strand break formation. Bottom left, genomic region containing the target site is PCR amplified; higher band is the wild-type sequence or sequences with small indels, lower band is the segmental deletion product generated by dual nucleases. Bottom right, chromatogram from Sanger sequencing of the lower band extracted from the gel (black arrow). The main product is the perfect junction of two double-strand break sites yielding a precise deletion (black rectangle). (**b**) Lesion rates and types are determined by the deep sequencing. Single nucleases generate small indels at their corresponding cleavage sites, whereas dual nucleases (independent or fusion) may generate six types of lesion products. The majority of the lesions produced by dual nuclease fusions is a precise deletion. $SpCas9^{MT}$-dNmCas9 fusions behave like a monomeric SpCas9. (**c**, **d**, **e**) Activity profiles of $SpCas9^{WT}$ (blue), $Nm/SaCas9^{WT}$ (pink), $SpCas9^{WT}$ + $Nm/SaCas9^{WT}$ (orange), and $SpCas9^{WT}$- $Nm/SaCas9^{WT}$ (green) nucleases at 12 genomic sites (6 D1 and 6 D2 configuration) are determined by deep sequencing. (**c**) Total lesion rates for the given nucleases. Typically, SpCas9 is more active than Nm/SaCas9. Total lesion rates of the Cas9-Cas9 dual nucleases are higher than the monomeric Cas9s used in combination. (**d**) Cas9-Cas9 fusions generate higher rates of precise deletions in the target genome than two independent Cas9 monomers. (**e**) Cas9-Cas9 dual nucleases primarily generate exact deletion products whereas lesion types of the two independent monomeric Cas9s are site-dependent. Each Box plot is drawn by GraphPad Prism, where the box represent 25^{th} and 75^{th} percentile and the middle line is the median. Whiskers and outliers are defined by Tukey method. Statistical significance is determined by one-way analysis of variance (ANOVA), "**" and "****" denote P <0.01 and <0.0001 respectively. Deep sequencing data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

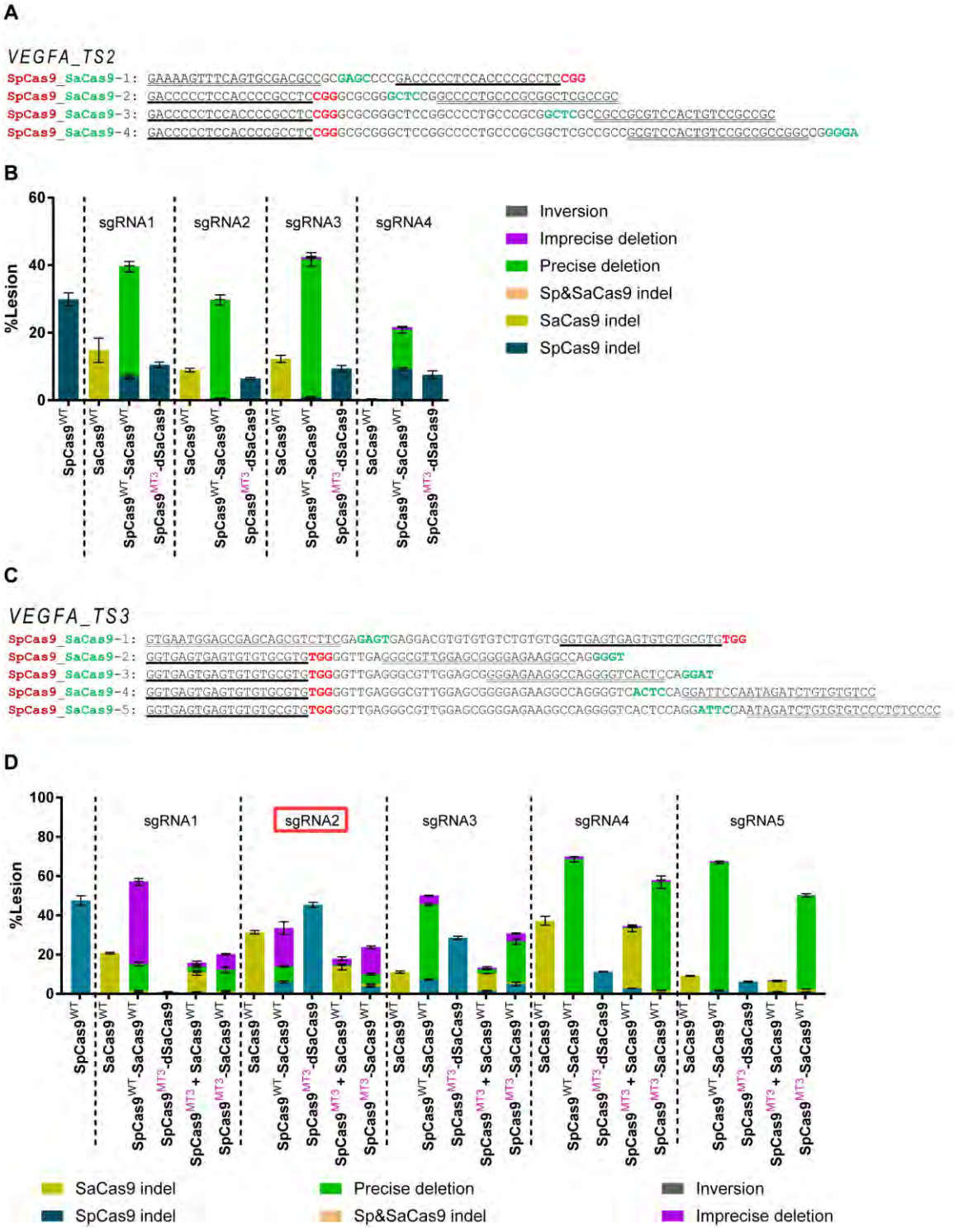

A
VEGFA_TS2
SpCas9_SaCas9-1: GAAAAGTTTCAGTGCGACGCCGCGAGCCCCGACCCCCTCCACCCCGCCTCCGG
SpCas9_SaCas9-2: GACCCCCTCCACCCCGCCTCCGGGCGCGGGCTCCGGCCCCTGCCCGCGGCTCGCCGC
SpCas9_SaCas9-3: GACCCCCTCCACCCCGCCTCCGGGCGCGGGCTCCGGCCCCTGCCCGCGGCTCGCCGCGCGTCCACTGTCCGCCGC
SpCas9_SaCas9-4: GACCCCCTCCACCCCGCCTCCGGGCGCGGGCTCCGGCCCCTGCCCGCGGCTCGCCGCCGCGTCCACTGTCCGCCGCCGGCCGGGGA
B
%Lesion
sgRNA1
sgRNA2
sgRNA3
sgRNA4
Inversion
Imprecise deletion
Precise deletion
Sp&SaCas9 indel
SaCas9 indel
SpCas9 indel
C
VEGFA_TS3
SpCas9_SaCas9-1: GTGAATGGAGCGAGCAGCGTCTTCGAGAGTGAGGACGTGTGTGTCTGTGTGGGTGAGTGAGTGTGTGCGTGTGG
SpCas9_SaCas9-2: GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGGCGTTGGAGCGGGGAGAAGGCCAGGGGT
SpCas9_SaCas9-3: GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGGCGTTGGAGCGGGGAGAAGGCCAGGGGTCACTCCAGGAT
SpCas9_SaCas9-4: GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGGCGTTGGAGCGGGGAGAAGGCCAGGGGTCACTCCAGGATTCCAATAGATCTGTGTGTCC
SpCas9_SaCas9-5: GGTGAGTGAGTGTGTGCGTGTGGGGTTGAGGGCGTTGGAGCGGGGAGAAGGCCAGGGGTCACTCCAGGATTCCAATAGATCTGTGTGTCCCTCTCCCC
D
%Lesion
sgRNA1
sgRNA2
sgRNA3
sgRNA4
sgRNA5
SaCas9 indel
SpCas9 indel
Precise deletion
Sp&SaCas9 indel
Inversion
Imprecise deletion

Figure 4. 4 Analysis of lesion type and frequency profile of SpCas9-SaCas9 fusions
(**a**) Sequence information of the *VEGFA-TS2* target sites for SpCas9-SaCas9 and SpCas9-NmCas9 fusions. SpCas9 protospacer is bold underlined and its PAM is in red, SaCas9 protospacer is double underlined and its PAM is green. (**b**) Lesion rates and types are determined by the deep sequencing. Single nucleases generate small indels at their cognate cleavage sites, whereas dual nucleases (independent or fusion) may generate six types of lesion products. (**c**) Sequence information of the *VEGFA-TS3* target sites for SpCas9-SaCas9 and SpCas9-NmCas9 fusions. SpCas9 protospacer is bold underlined and its PAM is in red, SaCas9 protospacer is double underlined and its PAM is green. (**d**) Lesion rates and types are determined by the deep sequencing. Single nucleases generate small indels at their cognate cleavage sites, whereas dual nucleases (independent or fusion) may generate six types of lesion products. Red rectangle indicate the SaCas9 sgRNA that is used for specificity analysis of SpCas9^{MT3}-dSaCas9 combination. Deep sequencing data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

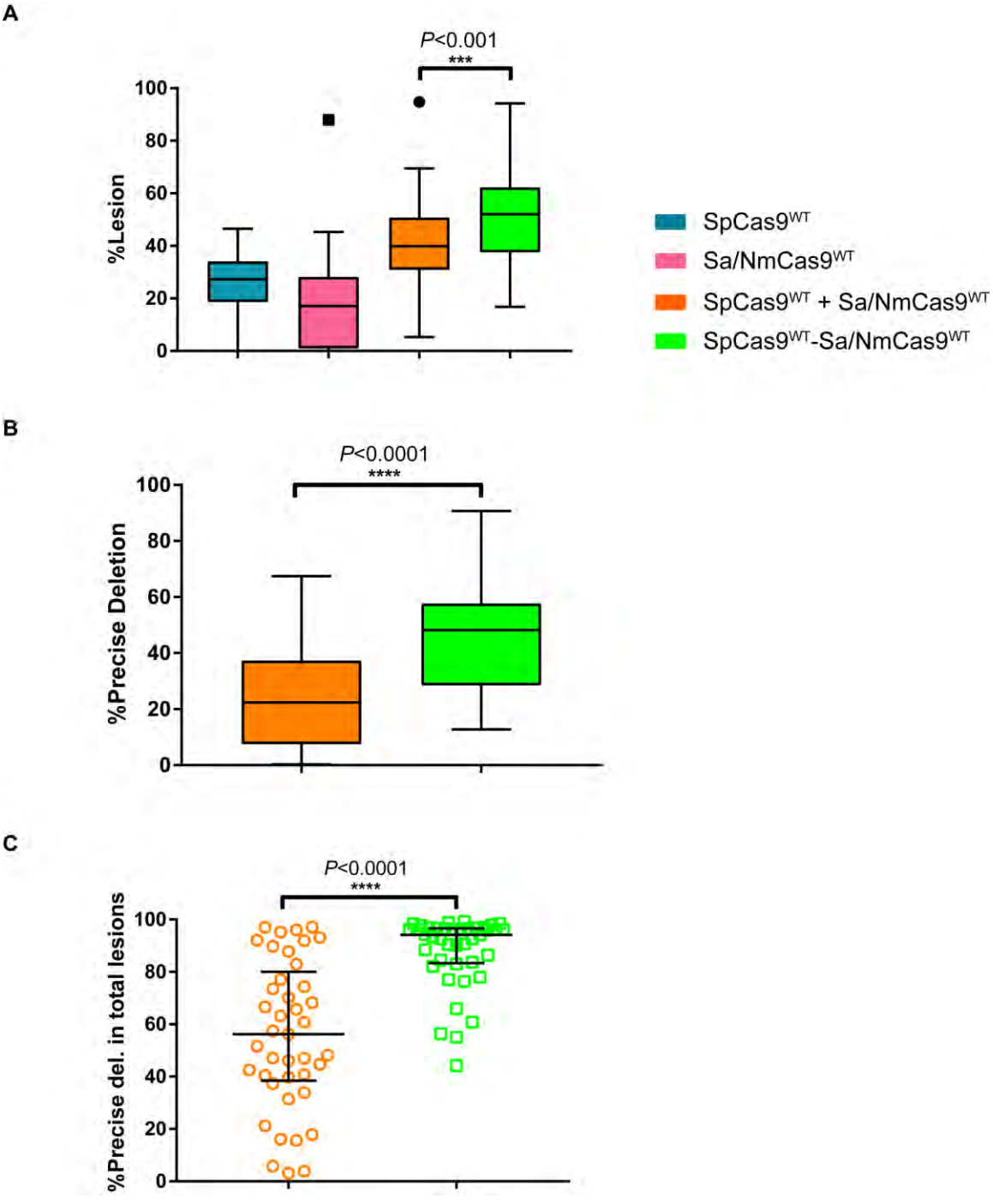

A
P<0.001
***
100
80
60
40
20
0
%Lesion
SpCas9WT
Sa/NmCas9WT
SpCas9WT + Sa/NmCas9WT
SpCas9WT-Sa/NmCas9WT
B
P<0.0001
****
100
80
60
40
20
0
%Precise Deletion
C
P<0.0001
****
100
80
60
40
20
0
%Precise del. in total lesions

Figure 4. 5 Activity profiles of singe and dual nucleases at 41 genomic sites
$SpCas9^{WT}$ (blue), $Nm/SaCas9^{WT}$ (pink), $SpCas9^{WT}$ + $Nm/SaCas9^{WT}$ (orange), and $SpCas9^{WT}$- $Nm/SaCas9^{WT}$ (green) nucleases activities are determined by deep sequencing. (**a**) Total lesion rates for the given nucleases. Typically, SpCas9 is more active than Nm/SaCas9. Total lesion rates of the Cas9-Cas9 dual nucleases are higher than the monomeric Cas9s used in combination. (**b**) Cas9-Cas9 fusions generate higher rates of precise deletions in the target genome than two independent Cas9 monomers. (**c**) Cas9-Cas9 dual nucleases primarily generate exact deletion products whereas lesion types of the two independent monomeric Cas9s are site-dependent. Each Box plot is drawn by GraphPad Prism, where the box represent 25^{th} and 75^{th} percentile and the middle line is the median. Whiskers and outliers are defined by Tukey method. The scatter dot plot is drawn by GraphPad Prism, where middle line indicates the median and top and bottom lines borders the interquartile ranges. Statistical significance is determined by one-way analysis of variance (ANOVA), "**" and "****" denote P <0.01 and <0.0001 respectively. Deep sequencing data are from single replicate in HEK293T cells.

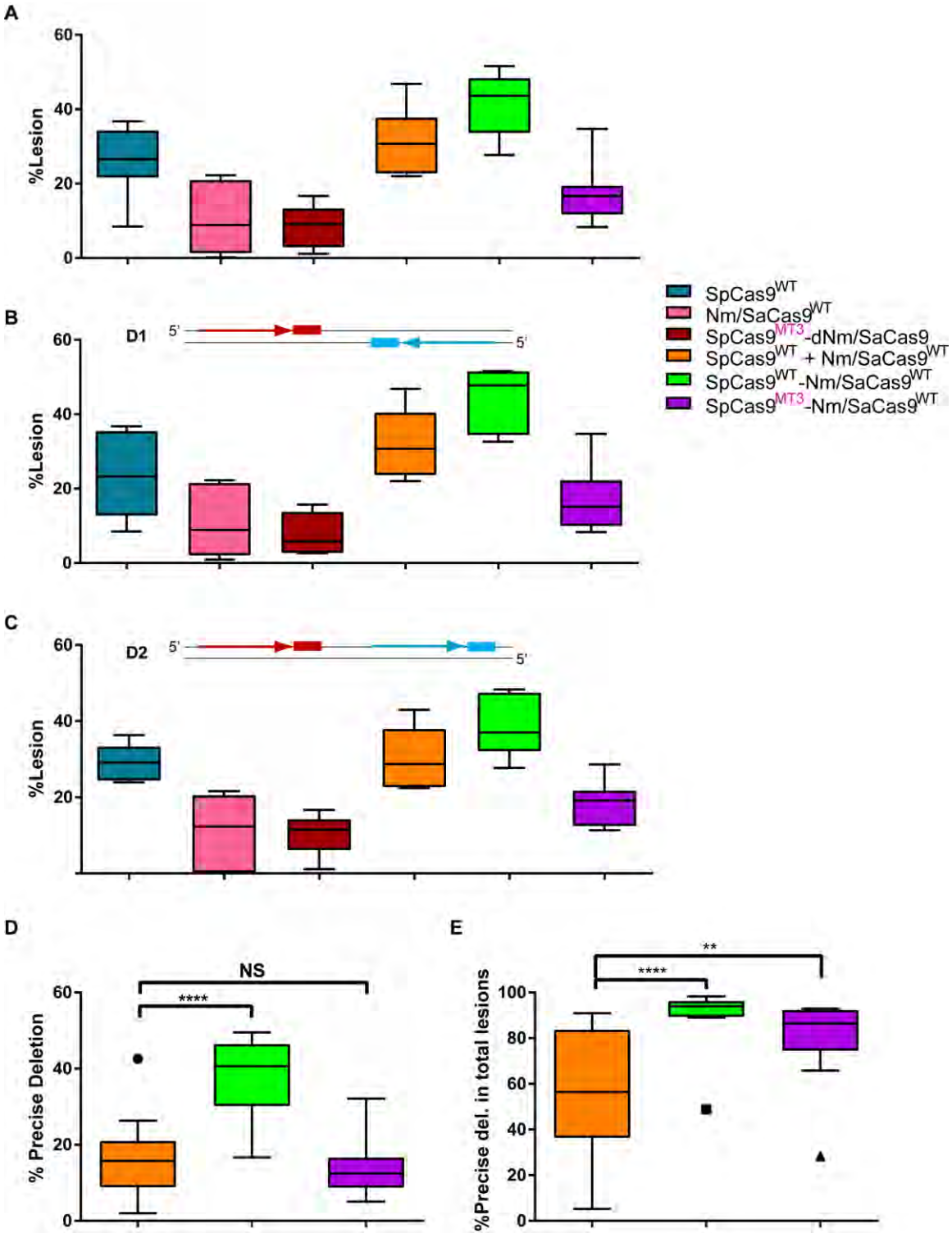

A
B
C
D
E
%Lesion
D1
D2
5'
5'
SpCas9WT
Nm/SaCas9WT
SpCas9MT3-dNm/SaCas9
SpCas9WT + Nm/SaCas9WT
SpCas9WT-Nm/SaCas9WT
SpCas9MT3-Nm/SaCas9WT
% Precise Deletion
NS
****
%Precise del. in total lesions
**
****

Figure 4. 6 Activity profiles of singe and dual nucleases at 12 genomic sites

Activity profiles of SpCas9^{WT} (blue), Nm/SaCas9^{WT} (pink), SpCas9^{MT3}-dNm/SaCas9 (brown), SpCas9^{WT} + Nm/SaCas9^{WT} (orange), SpCas9^{WT}- Nm/SaCas9^{WT} (green), and SpCas9^{MT3}-Nm/SaCas9^{WT} (purple) nucleases at 12 genomic sites (6 D1 and 6 D2 configuration) are determined by deep sequencing. (**a**) Total lesion rates for the given nucleases at 12 genomic sites. (**b**) Total lesion rates for the given nucleases at 6 genomic sites with D1 configuration. (**c**). Total lesion rates for the given nucleases at 6 genomic sites with D2 configuration. (**d**) SpCas9^{WT}-Nm/SaCas9^{WT} fusions generate higher rates of precise deletions in the target genome than two independent Cas9 monomers. (**e**) Cas9-Cas9 dual nucleases primarily generate exact deletion products whereas lesion types of the two independent monomeric Cas9s are site-dependent. Each Box plot is drawn by GraphPad Prism, where the box represents 25th and 75th percentile and the middle line is the median. Whiskers and outliers are defined by Tukey method. Statistical significance is determined by one-way analysis of variance (ANOVA), "**" and "****" denote, P <0.01 and P <0.0001 respectively. Deep sequencing data are from three independent biological replicates performed on different days in HEK293T cells.

4.3.3 Defining the targeting range of the Cas9-Cas9 fusions

Another salient feature of SpCas9-pDBD fusions is their increased targeting range achieved through their functionality at suboptimal PAM sequences[306]. To examine the targeting range of Cas9-Cas9 fusions, we designed several SpCas9 guides that target protospacers with suboptimal PAMs in tandem with an SaCas9 target site with an optimal N_2GRRT PAM sequence[244] (**Figure 4.7a**). Wild-type SpCas9 has very low or no activity on these sites, as expected (**Figure 4.7b**). However, in single- and dual-nuclease formats, SpCas9-SaCas9 fusions display nuclease activity at the SpCas9 sites with NAG, NTG, NCG, NGA, NGT, and NGC suboptimal PAMs. These data reflect the ability of the SpCas9-SaCas9 fusions to utilize the presence of a single guanine in the SpCas9 PAM element as a functional cleavage site (**Figure 4.7b**). Considering the flexibility in the target site spacing for Cas9-Cas9 fusions (**Figure 4.1b**), the target site density of the Cas9-Cas9 fusions is likely the sum of the targeting repertoire of the SaCas9 and NmCas9 nucleases.

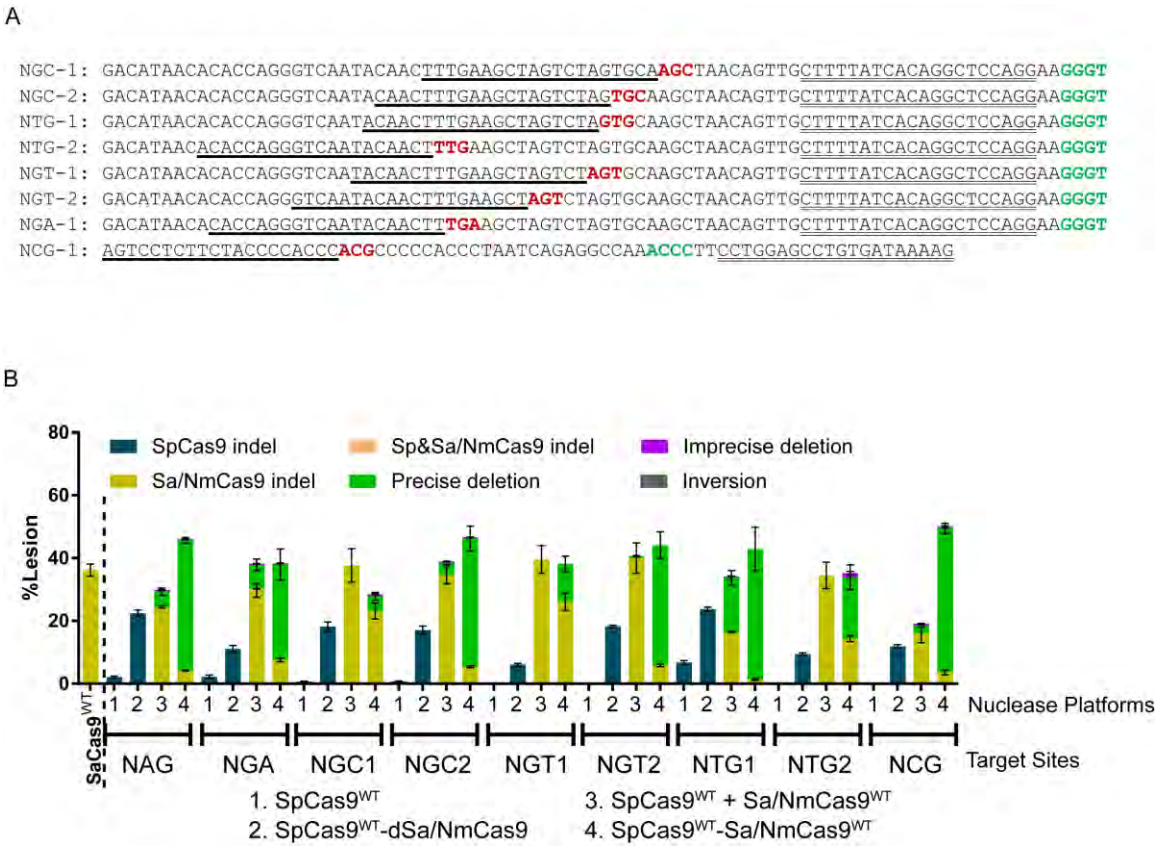


Figure 4. 7 Cas9-Cas9 fusions expand the PAM usage of SpCas9

(**a**) Target site information; SpCas9 protospacer is bold underlined and a suboptimal PAM element is in red, SaCas9 protospacer is double underlined and the cognate PAM element is green. (**b**) Lesion rates and types at the tandem target sites are determined by the deep sequencing with bulk analysis. SaCas9 generates robust editing whereas SpCas9 displays low or no activity. In Cas9-Cas9 fusion format, SpCas9 cuts effectively at these protospacers as observed from the SpCas9-dSaCas9 fusions or the fused wild-type nucleases. Deep sequencing data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

4.3.4 Cas9-Cas9 dual nucleases accurately delete functional elements in genome

Increased nuclease activity, enhanced targeting range, and the generation of uniform editing products favor the Cas9-Cas9 dual nucleases for the disruption of a gene or a regulatory element within genome. To apply this platform at a therapeutically relevant genomic locus, we tested the ability of Cas9-Cas9 dual nucleases to delete the GATA1 binding motif within the *BCL11A* erythroid-lineage specific enhancer (+58kb) element[97, 98]. Disruption of this regulatory element in the genomes of CD34+ hematopoietic stem and progenitor cells (HSPCs) silences *BCL11A* expression in the erythroid lineage, and thereby increases production of fetal γ-globin protein in differentiated red blood cells[99]. *Ex vivo* genome editing of this locus in HSPCs in conjunction with autologous bone marrow transplantation is a potential therapeutic approach for the treatment for sickle cell disease[96, 394].

To efficiently delete the GATA1 binding motif, we programmed Cas9-Cas9 fusions to target twelve different sites spanning this regulatory element in HEK293T cells (**Figure 4.8**). Similar to our analysis at other genomic loci, Cas9-Cas9 fusions effectively generate precise segmental deletions at most but not all of the examined sites (**Figure 4.9**). Notably, in SpCas9^{WT}-Sa/NmCas9^{WT} dual-nuclease fusion format, all three Cas9s (Sp/Nm/SaCas9s) effectively cut protospacers with suboptimal PAM sequences: SpCas9^{304} at NAG PAMs (GATA1-TS7, TS8, TS9, TS10, and TS11) and GATA1-TS5 and TS9), NmCas9$^{242, 322}$ at N_4GCTT (GATA1-TS4), N_4GTTT (GATA1-TS5), and N_4GACT (GATA1-TS6) PAMs, SaCas9$^{244, 305}$ at N_2GGGC (GATA1-TS10) and N_2GAGG (GATA1-TS12) PAMs. Among the twelve sites, we focused on four with the most promising activity, GATA1-TS7,

GATA1-TS9, GATA1-TS10, and GATA1-TS11, for further characterization and specificity analysis (**Figure 4.10a**).

Figure 4. 8 Cas9-Cas9 fusion target sequences in the *BCL11A* enhancer +58kb
Target site information of Cas9-Cas9 fusions for deletion of GATA1 binding element (highlighted in gray) in *BCL11A* enhancer +58kb. SpCas9 protospacer is bold underlined and PAM element is in red, SaCas9 protospacer is double underlined and the PAM element is green, and NmCas9 protospacer is wavy underlined and the PAM element is blue.

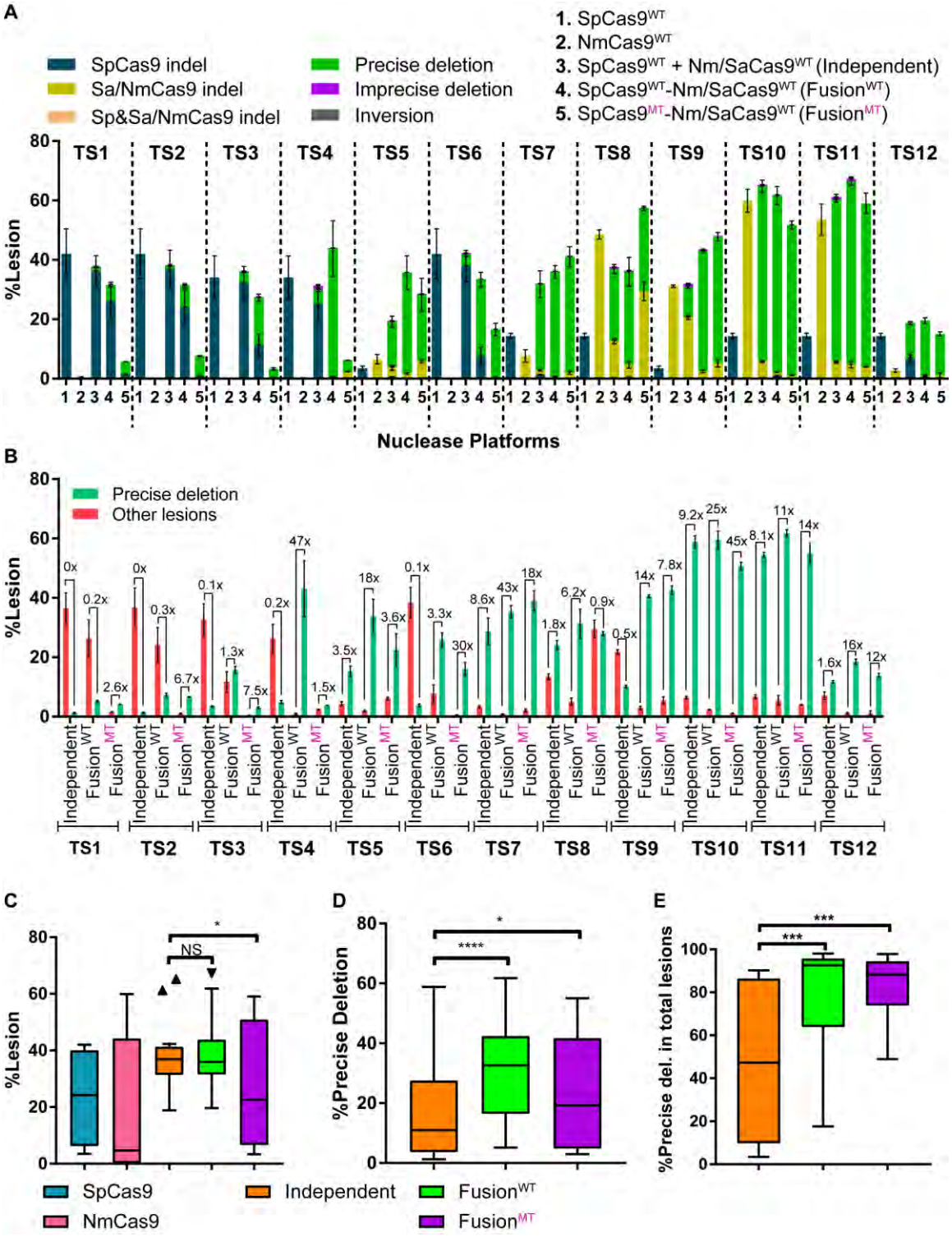

A
1. SpCas9WT
2. NmCas9WT
3. SpCas9WT + Nm/SaCas9WT (Independent)
4. SpCas9WT-Nm/SaCas9WT (FusionWT)
5. SpCas9MT-Nm/SaCas9WT (FusionMT)
SpCas9 indel
Sa/NmCas9 indel
Sp&Sa/NmCas9 indel
Precise deletion
Imprecise deletion
Inversion
%Lesion
TS1 TS2 TS3 TS4 TS5 TS6 TS7 TS8 TS9 TS10 TS11 TS12
Nuclease Platforms
B
Precise deletion
Other lesions
%Lesion
Independent
FusionWT
FusionMT
C
%Lesion
NS
*
SpCas9
NmCas9
Independent
FusionWT
FusionMT
D
%Precise Deletion
*
****
E
%Precise del. in total lesions
***
***

Figure 4. 9 Activity profiles of nucleases at *BCL11A* enhancer +58kb GATA1 element
(**a**) Lesion rates and types are determined by the deep sequencing. Single nucleases generate small indels at their corresponding cleavage sites, whereas dual nucleases (independent or fusion^{WT} or fusion^{MT}) may generate six types of lesion products. (**b**) Ratio distributions of precise deletions relative to other types of lesions. (**c**, **d**, **e**) Overview of the nuclease activities at 12 GATA1 target sites: (**c**) Total lesion rates, (**d**) precise deletion rates, and (**e**) fraction of precise deletion among all lesions. Each Box plot is drawn by GraphPad Prism, where the box represent 25th and 75th percentile and the middle line is the median. Whiskers and outliers are defined by Tukey method. Statistical significance is determined by one-way analysis of variance (ANOVA), "*", "***" and "****" denote P <0.05, P <0.001 and P <0.0001 respectively. Deep sequencing data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

One potential caveat on the precise quantification of the lesion levels is the PCR amplification bias of a precise deletion product due to its shorter length[396]. To address this possibility, we developed an assay based on unique-molecular identifiers (UMI)[397] and linear amplification mediated (LAM) PCR[332] (**Figure 4.11a-c**). In this approach, the genomic DNA is pre-amplified linearly with a single primer that is locus-specific, and that contains UMI and non-cognate adaptor sequences. Next, the UMI-containing single stranded DNA is selectively amplified and barcoded for deep sequencing. We used the UMI-correction method to measure the lesion profiles of four GATA1 target sites within the *BCL11A* enhancer +58kb locus. For single nucleases, there is no significant difference in the total lesion levels between bulk and UMI-corrected analyses. For dual nucleases, there is slight overestimation of the precise deletion levels in bulk sequencing data relative to the UMI-corrected analysis (**Figure 4.11d-e**). Since the SpCas9 target sites have suboptimal NAG PAMs, the activity levels of monomeric SpCas9 are modest. Cas9-Cas9 dual nucleases display increased activity at all of these target sites relative to the dual

independent nucleases. Notably, Cas9-Cas9 fusions containing the attenuated R1333S SpCas9 mutant (SpCas9^{MT2}) (Ref. 306) also retain high activity at all four target sites with NAG PAMs, concordant with our previous findings (**Figure 4.10c**). Similar to previous observations at other sites, even after UMI-correction, up to 97% of all lesions generated by Cas9-Cas9 dual nucleases are precise deletions (**Figure 4.10c-d**). These data suggest that Cas9-Cas9 dual nucleases are a promising platform for the generation of uniform editing products at therapeutically important sites.

.

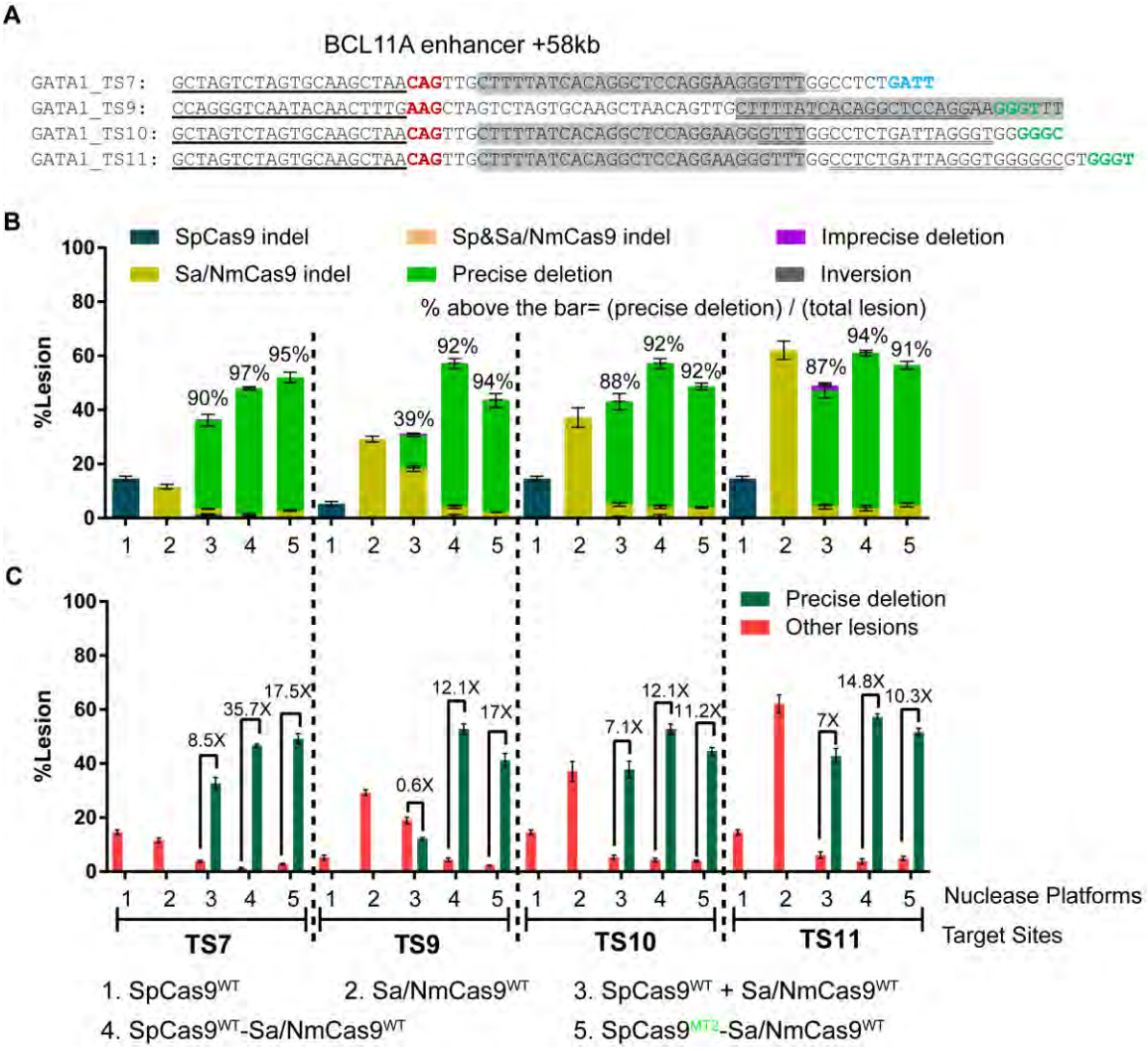


Figure 4. 10 Cas9-Cas9 fusions effectively delete *BCL11A* enhancer +58kb GATA1 element
(**a**) Sequence information of the four target sites chosen for more detailed assessment of the application of Cas9-Cas9 fusions to the deletion of the GATA1 binding element (highlighted in gray) at *BCL11A* enhancer +58kb. (**b, c**) Lesion rates and types at four target sites spanning the GATA1 element are determined by the deep sequencing after UMI-correction. (**c**) Ratio distributions of precise deletions relative to other types of lesions. Deep sequencing data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

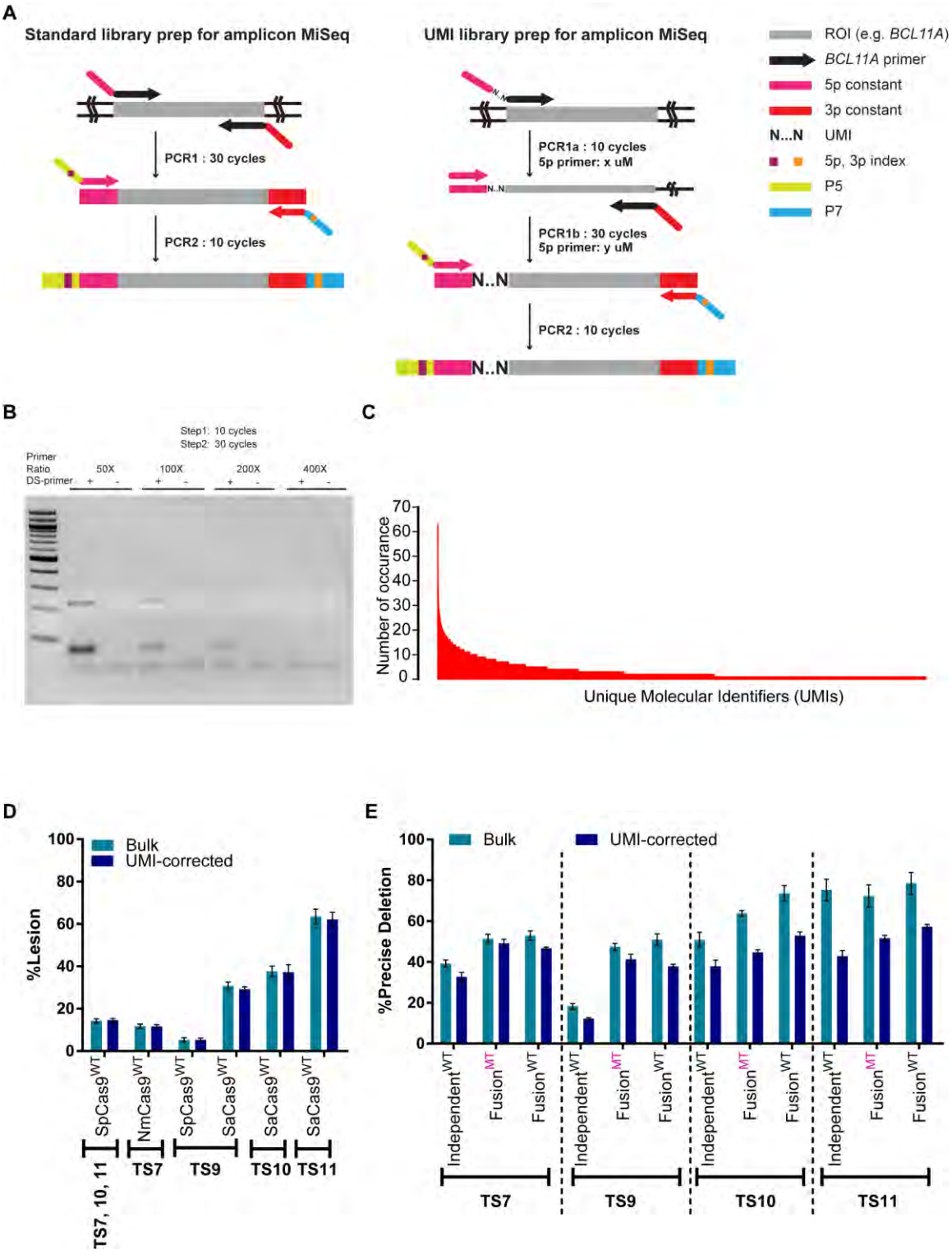

A
Standard library prep for amplicon MiSeq
PCR1 : 30 cycles
PCR2 : 10 cycles
UMI library prep for amplicon MiSeq
PCR1a : 10 cycles
5p primer: x uM
PCR1b : 30 cycles
5p primer: y uM
N..N
PCR2 : 10 cycles
ROI (e.g. BCL11A)
BCL11A primer
5p constant
3p constant
N...N UMI
5p, 3p index
P5
P7
B
Step1: 10 cycles
Step2: 30 cycles
Primer Ratio
DS-primer
50X
100X
200X
400X
C
Number of occurance
Unique Molecular Identifiers (UMIs)
D
Bulk
UMI-corrected
%Lesion
SpCas9WT
NmCas9WT
SpCas9WT
SaCas9WT
SaCas9WT
SaCas9WT
TS7, 10, 11
TS7
TS9
TS10
TS11
E
Bulk
UMI-corrected
%Precise Deletion
IndependentWT
FusionMT
FusionWT
TS7
TS9
TS10
TS11

Figure 4. 11 Accurate quantification of the segmental deletions
Accurate quantification of the segmental deletions via unique molecular identifiers (UMI) and linear amplification mediated (LAM) PCR. (**a**) Schematic of the library construction process for deep sequencing. Left, standard library prep where the region of interest (gray box) is first amplified with locus specific oligos (black arrows) bearing constant sequences (magenta and red) for TruSeq adaptor recognition. Next, the amplicon is further amplified from the constant sequences with oligos bearing indexes (purple and orange squares) and overhangs for P7 (light green) and P5 (blue). Right, library construction with for UMI-correction, where the genomic DNA is pre-amplified with oligos bearing UMI (N…N) and 5' constant sequence (magenta) via LAM PCR. Next, the single stand DNA is amplified with 5' constant sequence and 3' locus specific primer (black arrow) bearing 3' constant sequence (red). Then, the amplicon is further amplified from the constant sequences with oligos bearing indexes (purple and orange squares) and overhangs for P7 (light green) and P5 (blue). (**b**) An example agarose gel electrophoresis picture showing the outcome of PCR 1b of the UMI-correction library. The primer ratio indicates the value of 5' constant primer concentration (y μM) divided by 5' BCL11A_enh58_UMI primer concentration (x μM). In the absence of the 5' constant primer, the amplification product is not visible. (**c**) A representative histogram plotting the distribution of particular UMI abundance of TS11_SpCas9^{WT}-SaCas9^{WT}_Replicate-3 sample in the deep sequencing library. (**d**) For single nucleases the analysis of total lesions indicate that indels do not generate amplification bias. (**e**) For dual nucleases, the rate of precise deletions are a bit over-estimated via bulk analysis. Deep sequencing data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m.

To evaluate the targeting specificity of these nucleases at the four GATA1 target sites, we performed GUIDE-seq genome-wide specificity analysis[286]. We screened for active off-target sites for each individual wild-type Cas9 and the wild type Cas9-Cas9 dual nucleases. We observed robust GUIDE-seq oligo incorporation at the target sites for all of the wild-type Cas9-Cas9 dual nucleases (**Figure 4.12a**). A small number of potential off-target sites were identified for the TS7 and TS9 based on the GUIDE-seq analysis. A larger number of off-target sites were associated with the SaCas9 guide for the TS11 target site. The number of off-target sites identified by GUIDE-seq and the unique counts that are associated with each site are similar between the Cas9-Cas9 fusions and the individual

Cas9 nucleases (**Appendix 2**), suggesting that there is not a dramatic difference in the off-target activity between the individual nucleases and the Cas9-Cas9 fusions. The three off-target sites identified for NmCas9 have nine-nucleotide mismatches to the guide sequence, and are so divergent that they potentially represent false positive sequences.

To assess the editing rate at potential off-target sites, we performed amplicon deep sequencing at these regions within the genome. We evaluated ten potential off-target sites for each nuclease, which were identified by either the GUIDE-seq analysis[286] or computationally predicted via Cas-OFFinder[198] when insufficient high-quality sites were identified by GUIDE-seq. We evaluated the off-target activities of two different versions of the Cas9-Cas9 dual-nuclease fusions, SpCas9^{WT}-Sa/NmCas9^{WT} or SpCas9^{MT2}-Sa/NmCas9^{WT}, in comparison with the independent SpCas9^{WT} and Sa/NmCas9^{WT} nucleases delivered simultaneously. The deep sequencing analysis indicates that the Cas9-Cas9 dual nuclease fusions display similar levels of off-target activity when compared to independent dual-Cas9 nucleases. At two of the SpCas9 off-target sites (OT9-4 and OT9-8), the use of the attenuated SpCas9^{MT2}-SaCas9^{WT} nuclease dramatically reduces the off-target activity compared to the independent wild-type SpCas9 or the SpCas9^{WT}-SaCas9^{WT} nuclease (**Figure 4.12b**). These data demonstrate that Cas9-Cas9 dual nucleases achieve robust editing at the target site without generating a new class of off-target sites. The generation of uniform, accurate editing products within the *BCL11A* enhancer +58kb locus highlights the utility of Cas9-Cas9 dual nucleases as a promising genome editing platform for the deletion of therapeutically relevant regulatory elements[96, 394].

A

Target Site (TS)	Nuclease platform	Number of off-target sites	Number of unique reads at TS	Sum of unique reads at off-target sites
GATA1_TS7	SpCas9^{WT}	0	79	0
	NmCas9^{WT}	3	256	47
	SpCas9^{WT}-NmCas9^{WT}	3	177	14
GATA1_TS9	SpCas9^{WT}	0	0	0
	SaCas9^{WT}	0	1704	0
	SpCas9^{WT}-SaCas9^{WT}	5	1809	47
GATA1_TS10	SpCas9^{WT}	0	79	0
	SaCas9^{WT}	0	662	0
	SpCas9^{WT}-SaCas9^{WT}	0	116	0
GATA1_TS11	SpCas9^{WT}	0	79	0
	SaCas9^{WT}	43	1248	1938
	SpCas9^{WT}-SaCas9^{WT}	17	210	388

B

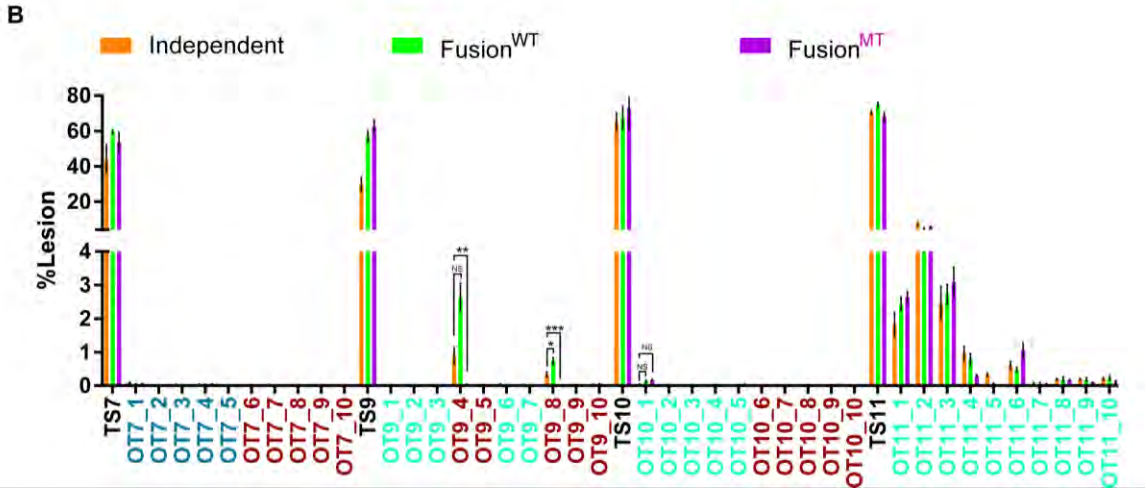


Figure 4. 12 Cas9-Cas9 fusions achieve robust and specific genome editing

(**a**) Summary of the GUIDE-seq genome-wide off-target analysis of SpCas9^{WT}, Sa/NmCas9^{WT}, and SpCas9^{WT}-Sa/NmCas9^{WT} at four GATA1 target sites (**Appendix 2**). (**b**) Deep sequencing determined lesion rates for these nucleases at small set of off-target sites discovered by the GUIDE-seq data or predicted by CasOFFinder. The names of SpCas9, NmCas9 and SaCas9 off-target sites are colored in dark red, blue, and green. The GUIDE-seq result is from single experiment, deep sequencing data are from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate ±s.e.m. Statistical significance is determined by one-way analysis of variance (ANOVA), "*", "**", "***", and "NS" denote BH-adjusted *P*-values of <0.05, <0.01, <0.001, and "not significant" respectively.

4.4 Discussion

In this study, we have expanded the CRISPR/Cas9 toolset by developing orthogonal Cas9-Cas9 chimeras in single- and dual-nuclease formats. Unlike the original Cas9-pDBD platform, which required some expertise in protein engineering, the entirely RNA-programmable Cas9-Cas9 fusions should be completely accessible to the broader scientific community. In both single- and dual-nuclease formats, Cas9-Cas9 fusions act similarly to SpCas9-pDBDs with regards to enhanced targeting range and improved specificity[306]. The presence of a single guanine within the PAM is sufficient for SpCas9 to cleave its targets in the context of Cas9-Cas9 fusions. In principle, similar suboptimal PAM usage should be applicable for SaCas9 and NmCas9 when the SpCas9 targets an optimal NGG PAM element. For homology-directed repair applications, the ability to target sites with suboptimal PAM elements may allow Cas9-Cas9 fusions to generate a DSB closer to the site of the desired sequence conversion[398]. This feature can also be useful for allele-specific targeting by Cas9-Cas9 nucleases by placing the PAM recognition at a polymorphic site. In the dual-nuclease format, $SpCas9^{WT}$-Sa/$NmCas9^{WT}$ fusions display superior nuclease activity and primarily produce uniform, predictable lesions within the targeted genome. These fusions are also compatible with PAM-interaction-attenuated SpCas9 to provide an additional level of target site specificity. Overall, RNA-programmable Cas9-Cas9 fusions offer superior features to independent nucleases: expanded targeting range, improved specificity, and efficient generation of uniform editing products.

Our analysis of the DNA repair products produced by Cas9-Cas9 fusions provides insights into methods to increase the efficiency of DSB-mediated genome sequence alterations by programmable nucleases. Canonical NHEJ (cNHEJ) is typically the default choice for DSB repair in most stages of the cell cycle[10]. Precise ligation by cNHEJ of the ends of a DSB generated by a single nuclease restores the target site sequence, and in the context of an active nuclease leads to another round of cleavage. The predominant production of precise deletions after the generation of two synchronous blunt DSBs at a composite target site supports the cellular preference for cNHEJ for DSB repair. Consistently, SpCas9^{MT}-Sa/NmCas9^{WT} dual nuclease fusions display higher total lesion rates than SpCas9^{MT}-dSa/NmCas9 single nuclease fusions, where the latter only rely on less frequent non-canonical NHEJ mediated DSB repair to produce a stable end product (**Figure 4.3b** and **Figure 4.4**). This is particularly striking for dual nucleases at target sites where one of the Cas9 monomers is weakly active, such as at some of the Nm/SaCas9 target sites (TS5, TS7 & TS12; **Figure 4.9**) or at SpCas9 protospacers with suboptimal PAMs (**Figure 4.7**). DNA cleavage at these sites by the independent nuclease is inefficient, and since cNHEJ repair restores the native DNA sequence, they are marked by low lesion rates. Cas9 nucleases employed in combination reveal the activity of the weak nuclease in the form of precise deletions between the cleavage sites. This increase in lesion rates could also be due in part to the relaxation of the local chromatin architecture as observed for proximal CRISPR targeting[399].

One of the hurdles for therapeutic genome editing applications is the uncertainty of the functional activity of the lesions that are produced by individual nucleases when relying

on imprecise DNA repair. At a therapeutically relevant site, Cas9-Cas9 dual nuclease fusions produce defined precise deletions comprising up to 97% of the modified genomes, which should produce specific alleles that have activities that can be defined in model systems prior to advancement in therapeutic applications. Cas9-Cas9 fusion mediated production of precise deletions should be applicable to the development of therapeutic genome editing strategies for a number of genetic disorders. As the immediate extension of the results described in this study, efficient excision of the core regulatory elements within the *BCL11A* erythroid enhancer in CD34+ HSPCs via Cas9-Cas9 ribonucleoproteins is likely to achieve higher rates of inactivation than the production of local indels by a single nuclease[97-99]. This increased efficiency should reduce the number of modified cells that need to be administered to a patient for the treatment of β-hemoglobinopathies, such as sickle cell disease[96].

CHAPTER V: Discussion and future outlook

5.1 Summary of the chimeric Cas9 platforms

The most commonly employed CRISPR-Cas9 system uses the Cas9 effector from *S. pyogenes* due to its efficacy and relatively broad targeting range; the PAM requirement of "GG" occurs every 8 base pairs in a random genome. However, the wild type SpCas9 has been reported to generate double strand breaks at unintended sites in human genome[378]. Chapter 2 describes the development of a chimeric SpCas9-pDBD (programmable DNA-binding domain, ZFP or TALE) platform to expand the targeting range and improve the specificity of SpCas9. Upon fusion to a pDBD that is targeting a downstream DNA sequence, SpCas9 is able to cleave target sites with suboptimal PAM elements: NAG, NGA, and NGC. Attenuation mutations of the PAM interaction residues of SpCas9 ($SpCas9^{MT}$) make the nuclease activity dependent on the fused pDBD targeting a downstream sequence. This additional stage of target site recognition provides substantial improvement to the nuclease specificity. Furthermore, the SpCas9-pDBD framework can be tuned to achieve extreme genome editing specificity.

Continuous expression of programmable nucleases increases the likelihood of off-target nuclease activity[289]. Additional level of targeting specificity can be achieved by adding a mechanism for temporal control over the nuclease activity. Chapter 1 provides an overview of some of the platforms that offer temporal control in nuclease activity. Chapter 3 describes the development of two SpCas9-pDBD variants that are inducible by a small molecule. The first variant involves the splitting the covalent linkage between SpCas9 and the pDBD through the addition of FKBP:FRB heterodimerization domains. The second

variant is developed from previously described Split-Cas9 architecture[308] with a PAM-interaction attenuation mutation and fusion to a pDBD. Both of these variants achieve rapamycin inducible genome editing with similar levels of specificity to the SpCas9-DBD platform.

Although SpCas9-pDBD platforms offer accurate genome editing, the challenges inherent in the construction of pDBDs—particularly ZFPs—limit their usage by the broader scientific community. Chapter 4 describes the conversion of the chimeric Cas9 platforms into an entirely RNA-programmable system. Substitution of the pDBD with a nuclease dead Cas9 orthologue—from *N. meningiditis*[242] or *S. aureus*[244]—simplifies the construction of chimeric Cas9 platforms with three-stages of target site recognition. Cas9-Cas9 fusions can be utilized as a single or a dual nuclease. Both nucleases achieve accurate genome editing with the expansion of PAM usage of SpCas9 to a single guanine. Furthermore, dual nuclease Cas9-Cas9 fusions predominantly produce a defined editing product (precise deletion) as opposed to random indels at the target site. This feature is particularly useful to disrupt therapeutically relevant genes and regulatory elements.

5.2 The future of the chimeric Cas9 platforms

Similar to other programmable nucleases, the chimeric Cas9 platforms described in this thesis can be used in various basic genome engineering applications in cell lines and model organisms. There can also be cases where the usage of a chimeric Cas9 platform becomes even more advantageous over standard Cas9 nucleases. For example, in gene correction applications, keeping the DSB close to the conversion site increases the HDR rate[398]. Being able to target sites with suboptimal PAM elements may allow chimeric Cas9s

to generate a DSB closer to the site of the desired sequence conversion. Suboptimal PAM licensing can also be useful to for allele-specific targeting of these nucleases by placing the PAM recognition at a polymorphic site. In dividing cells, generation of a DSB in one of alleles can trigger HDR templated from the other non-cut allele in similar mechanism to the gene drive phenomenon[400] described below. In autosomal dominant diseases with heterozygous genotype, nuclease targeting of the disease causing allele should be converted to the other allele. In addition to PAM expansion, the improved specificity of these platforms makes them appealing for therapeutic genome editing application for genetic disorders. Such applications can be either *ex vivo* where the nuclease can be delivered as mRNA or ribonucleoprotein complex or *in vivo* where the nuclease can be delivered within viral or non-viral carriers. Particularly, AAV-mediated delivery of nucleases may be the most therapeutically relevant but challenging due the limited cargo capacity. Therefore, the development of Cas9-pDBD platforms employing smaller Cas9 orthologues[242, 244, 245] should provide a solution. For example, an NmCas9 fused with 4 finger ZFP is around 3.7kb. Including the additional regulatory sequences, the expression cassette should be within the packaging limit (~4.7 kb) of single AAV[401]. For viral deliveries, the constitutive expression of nucleases may increase the likelihood of off-target cleavage no matter how specific the nuclease platform is. Therefore, incorporation of a temporal control at the expression or the activity stages of any nucleases would be useful.

Cas9-Cas9 fusions can provide additional utility in therapeutic genome editing applications. The data in Chapter 4 suggest that dual nuclease fusions can effectively and accurately delete the GATA1 binding motif within the erythroid-lineage specific enhancer

element of *BCL11A* gene[97, 98]. The immediate extension of this application is its recapitulation in CD34+ hematopoietic stem and progenitor cells as a potential *ex vivo* therapeutic approach for sickle cell disease[95, 96]. Since the GATA1 deletion is a canonical NHEJ mediated event, the rate of editing in long term progenitor cells should be comparable to the bulk population[99]. The Cas9-Cas9 fusions can be delivered into CD34+ HSPCs as electroporation of an mRNA or an RNP. The latter may be associated with additional challenges such as the difficulty in its bacterial expression as a single polypeptide. Yet, by employing a split intein system[402, 403], the two Cas9 component can first be expressed and purified independently, and then ligated to form the fusion protein. Another application area of the Cas9-Cas9 dual nucleases is functional genomics screening. Because Cas9-Cas9 dual nuclease fusions typically have higher activity and produce more uniform lesions than single Cas9s, a Cas9-Cas9 fusion based loss-of-function screening platform should produce cleaner outcomes. In addition to single and dual nucleases, Cas9-Cas9 should be able used as a single nickase, a dual nickase or a nuclease nickase combination. Such fusions will generate synchronous nicks and (or) double strand breaks at the target site to form various end products which may trigger different cellular repair machineries[9-11]. The HDR potentials of such cleavage products are worth investigation. Overall, the chimeric Cas9 platforms described in this thesis offer highly specific, versatile genome editing applications.

5.3 The future of genome engineering

Novel applications of programmable nucleases are happening at a breathtaking pace. In addition to what has been projected in Chapter 1, there are other avenues of

genome editing applications that will impact the future of medicine and the ecosystem. One such application is their use as gene drives, a term to describe inheritance bias of a particular gene against the odds of natural genetics[404, 405]. In order to initiate a gene drive, a Cas9 and sgRNA expression cassette with homology arms against the flanking sequences of the sgRNA target site is inserted into a genome. Germline expression of the Cas9 and sgRNA from the incorporated allele generates a DSB on the other allele, which is repaired via HDR templated from the Cas9 sgRNA inserted allele, which results in gene conversion. Once transferred into germline, mating of the gene drive carrying organism with a wild type mate predominantly gives rise to progeny carrying homozygous gene drive[400]. Although being controversial with regards to their safety and the environmental risks associated, gene drives have the potential to control vector based diseases (e.g. malaria and dengue fever), and invasive species[405].

Another interesting application of programmable nucleases is in xenotransplantation. The term corresponds to the transplantation of living organs (or tissues) from one species into another species to treat organ failure of the recipient. Due to the persistent shortage of organs from compatible human donors, xenotransplantation is one potential solution to this problem[406]. Pigs are considered as suitable organisms for the development of human transplantable organs. However, there are critical barriers associated with the xenotransplantation of pig organs into humans: immunological rejection, compatibility of the organs, and potential transmission of porcine viruses[406]. Programmable nuclease-directed genome engineering is being used to engineer porcine

cell lines and generate transgenic animals to investigate and improve their feasibility for xenotransplantation[407, 408].

Ever since the birth of the genome editing concept, the possibility of human germline editing has been debated. In earlier 2017, three research groups reported the use of CRISPR-Cas9 nucleases in human embryos for studying the gene function in early human embryogenesis[409] or the correction of disease causing mutations in monogenic diseases[410, 411]. These proof-of-principle studies provide insight with regards to feasibility of the technique and potential challenges associated. Particularly, Ma *et al*. investigated the HDR efficacy and embryo mosaicism upon the co-injection of CRISPR-Cas9 nucleases and ssODN donor template either at the time of or after fertilization. The former approach reported to be more effective with regards to reducing mosaicism. Interestingly, the researchers argue that the HDR in human embryos is mainly templated from the non-targeted allele rather than the exogenous donor supplied[410]. Although being controversial[412, 413], if verified, inter-homologue templated HDR will have utility in therapeutic gene correction applications. Based on these studies, it is likely that more of therapeutic genome editing experiments in human embryos are underway. There are ethical concerns associated with human embryo genome editing which is discussed further in a later section **5.3.2**.

5.3.1 Limitations

Despite the success of programmable nucleases, there are still hurdles that need to be overcome for therapeutic genome editing applications: (i) safety and toxicity, (ii) efficacy and control of editing outcome, (iii) and delivery.

Nuclease-directed genome editing depends on creation of a site-specific double strand break in the genome and its repair via cellular machineries. The presence of DSBs at unintended genomic positions may result in confounding point mutations, chromosomal breaks and translocations. Therefore, the specificity of any nuclease platform must be thoroughly assessed prior to clinical application. The core work of this thesis is mainly focused on the specificity of CRISPR-Cas9 nucleases. As reviewed in Chapter 1, there are many methods to identify and reduce the off-target mutations produced by CRISPR-Cas9 nucleases. At present, the prudent method for assessing off-target activities of Cas9 nucleases involves a two-step verification: a genome-wide unbiased screening for identification of potential off-target sites in the cell type of interest, and their verification with targeted amplicon deep sequencing. The chimeric Cas9 platforms described in Chapters 2, 3 and 4 are of the most specific genome editing platforms available. Although their specificity needs to be verified in a therapeutically relevant system, the observed level of specificity of these nucleases should be sufficient for *ex vivo* therapeutic applications. Since the sensitivity level of the current deep sequencing methodologies is about 0.01%[8], an advance in mutation detection levels is needed to fully assess the specificity of these nucleases for treating larger populations of cells for therapeutic use.

Another remaining challenge is the efficient editing of the genomes of quiescent and post-mitotic cells. Previous research indicates that the activity of CRISPR-Cas nucleases is impacted by the chromatin accessibility of the target site[414, 415]. The effect of chromatin accessibility can be neglected for many dividing cells type due the changes in chromatin accessibility during the cell cycle. However, when targeting slow or non-

dividing cells in a therapeutic setting, such as the long term progenitors within the bulk CD34+ HSPC population or differentiated neurons, the impact of chromatin accessibility on nuclease activity may be greater. In addition to the potential chromatin accessibility problems, driving HDR in these cells still stands as a barrier.

Efficient delivery of the nucleases into target tissue and organs is as important as engineering functional nucleases. For *ex vivo* therapeutics, electroporation of the Cas9 mRNA or RNP has proven to be effective[213, 217]. Additional optimization of parameters such as nuclear localization signal composition or additional RNA modifications may provide useful improvements. On the other hand, for *in vivo* therapeutics, nuclease delivery is still a problem. Adeno-associated viruses have being used in clinical gene therapy applications and offer advantageous features such as low rates of integration, cargo expression, and tropism to different tissues[401]. Yet, the relatively small cargo capacity of AAVs may limit their use for co-delivery of the components (including the HDR template) of genome editing. Nanoparticle-based non-viral delivery vectors[239] or local delivery of the RNP[416] can be promising but further assessment on their ability to transport the editing components to different tissues and organs is needed.

Finally, another point related to the safety and the toxicity of these nucleases is immunogenicity. Either the viral delivery agents[417] or the bacterial Cas9 protein itself may run into immune rejection by the host in an *in viv*o therapeutic setting. Instead of the continuous expression of the Cas9 protein and the sgRNA from viral expression cassette, the non-viral delivery (nanoparticles) of the Cas9 expressing mRNA and the sgRNA may

have advantageous. In such setting, the modifications to the Cas9 encoding mRNA and the sgRNA are needed to circumvent the host's immune responses[191, 239].

5.3.2 Ethical concerns

As the genome editing field advances, the debate over human genome editing applications is heating up. Two main aspects of human genome editing are particularly debated: medical use and personal enhancement (e.g. improved athletic ability). Recently, two surveys reported similar trends of public opinion on human genome editing[418, 419]. Although there is support for adult medical genome editing applications, there is opposition for adult enhancement genome editing applications. For human germline editing, there is weaker support—perhaps a neutral stand—for medical applications, yet strong rejection of enhancement applications[418, 419]. One of the surveys also points out that *in vitro* basic research on human embryo editing that does not culminate pregnancy is supported for continuation and public funding[419]. Manipulation of human embryos, even for medically relevant reasons is a delicate subject. There are two types of germline therapies in limited use: pre-implantation genetic diagnosis (PGD) and mitochondrial replacement therapy (MRT). The first report of PGD was almost three decades ago[420]. Following *in vitro* fertilization, PGD has been applied to screen and select embryos against monogenic disorders and cancer or for increased pregnancy changes and human leukocyte antigen matching savior siblings[421]. MRT is a relatively newer subject than PGD. If the mother has mitochondrial genetic disorder and the PGD is not a suitable option, MRT is appended to *in vitro* fertilization to use the mitochondrial DNA of a healthy donor[422]. In 2015, the United Kingdom became the first country where MRT is legal. Despite the ongoing debate,

embryo genome editing—once its safety is assured—is likely be allowed as a complementary therapeutic approach only to cases where pre-implantation genetic diagnosis and mitochondrial replacement therapies are not applicable and there is no conventional treatment option.

5.3.3 Summary

The future of genome editing is wide open; yet, there are challenges and concerns on the table. Up until this point, researchers have developed and tested the potential of programmable nucleases on eukaryotic genomes for various applications. There is growing interest in the biotechnology and pharmacology industry to direct these tools towards economically viable medical products and applications. Conversion of a working prototype into a complete reliable product requires effort, patience, and professionalism. Although the broad utility and wide potentials of genome editing is appealing to us (researchers), an untimed mistake at clinical level can lead to major setback if not a catastrophe for the field. Having regulations to control the pace of this conversion and is critical to build a reputation and gain public's trust.

BIBLIOGRAPHY

1. Capecchi, M.R. Altering the genome by homologous recombination. *Science* **244**, 1288-1292 (1989).
2. Capecchi, M.R. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. *Nat Rev Genet* **6**, 507-512 (2005).
3. Choulika, A., Perrin, A., Dujon, B. & Nicolas, J.F. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. *Mol Cell Biol* **15**, 1968-1973 (1995).
4. Smih, F., Rouet, P., Romanienko, P.J. & Jasin, M. Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. *Nucleic Acids Res* **23**, 5012-5019 (1995).
5. Maeder, M.L. & Gersbach, C.A. Genome-editing Technologies for Gene and Cell Therapy. *Mol Ther* **24**, 430-446 (2016).
6. Carroll, D. Genome engineering with targetable nucleases. *Annu Rev Biochem* **83**, 409-439 (2014).
7. Cox, D.B., Platt, R.J. & Zhang, F. Therapeutic genome editing: prospects and challenges. *Nat Med* **21**, 121-131 (2015).
8. Porteus, M.H. Towards a new era in medicine: therapeutic genome editing. *Genome Biol* **16**, 286 (2015).
9. Jasin, M. & Haber, J.E. The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. *DNA Repair (Amst)* **44**, 6-16 (2016).
10. Chang, H.H.Y., Pannunzio, N.R., Adachi, N. & Lieber, M.R. Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nat Rev Mol Cell Biol* **18**, 495-506 (2017).
11. Chang, H.H. et al. Different DNA End Configurations Dictate Which NHEJ Components Are Most Important for Joining Efficiency. *J Biol Chem* **291**, 24377-24389 (2016).
12. Gupta, A. et al. Targeted chromosomal deletions and inversions in zebrafish. *Genome Res* **23**, 1008-1017 (2013).
13. Piganeau, M. et al. Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome Res* **23**, 1182-1193 (2013).
14. Chen, F. et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. *Nat Methods* **8**, 753-755 (2011).
15. Bothmer, A. et al. Characterization of the interplay between DNA repair and CRISPR/Cas9-induced DNA lesions at an endogenous locus. *Nat Commun* **8**, 13905 (2017).
16. Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A. & Luscombe, N.M. A census of human transcription factors: function, expression and evolution. *Nat Rev Genet* **10**, 252-263 (2009).
17. Miller, J., McLachlan, A.D. & Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. *EMBO J* **4**, 1609-1614 (1985).

18. Wolfe, S.A., Nekludova, L. & Pabo, C.O. DNA recognition by Cys2His2 zinc finger proteins. *Annu Rev Biophys Biomol Struct* **29**, 183-212 (2000).
19. Parraga, G. et al. Zinc-dependent structure of a single-finger domain of yeast ADR1. *Science* **241**, 1489-1492 (1988).
20. Lee, M.S., Gippert, G.P., Soman, K.V., Case, D.A. & Wright, P.E. Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* **245**, 635-637 (1989).
21. Pavletich, N.P. & Pabo, C.O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. *Science* **252**, 809-817 (1991).
22. Elrod-Erickson, M., Rould, M.A., Nekludova, L. & Pabo, C.O. Zif268 protein-DNA complex refined at 1.6 A: a model system for understanding zinc finger-DNA interactions. *Structure* **4**, 1171-1180 (1996).
23. Elrod-Erickson, M., Benson, T.E. & Pabo, C.O. High-resolution structures of variant Zif268-DNA complexes: implications for understanding zinc finger-DNA recognition. *Structure* **6**, 451-464 (1998).
24. Rebar, E.J. & Pabo, C.O. Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671-673 (1994).
25. Jamieson, A.C., Kim, S.H. & Wells, J.A. In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry* **33**, 5689-5695 (1994).
26. Choo, Y. & Klug, A. Selection of DNA binding sites for zinc fingers using rationally randomized DNA reveals coded interactions. *Proc Natl Acad Sci U S A* **91**, 11168-11172 (1994).
27. Wu, H., Yang, W.P. & Barbas, C.F., 3rd Building zinc fingers by selection: toward a therapeutic application. *Proc Natl Acad Sci U S A* **92**, 344-348 (1995).
28. Greisman, H.A. & Pabo, C.O. A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science* **275**, 657-661 (1997).
29. Segal, D.J., Dreier, B., Beerli, R.R. & Barbas, C.F., 3rd Toward controlling gene expression at will: selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc Natl Acad Sci U S A* **96**, 2758-2763 (1999).
30. Dreier, B., Segal, D.J. & Barbas, C.F., 3rd Insights into the molecular recognition of the 5'-GNN-3' family of DNA sequences by zinc finger domains. *J Mol Biol* **303**, 489-502 (2000).
31. Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. & Barbas, C.F., 3rd Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **276**, 29466-29478 (2001).
32. Dreier, B. et al. Development of zinc finger domains for recognition of the 5'-CNN-3' family DNA sequences and their use in the construction of artificial transcription factors. *J Biol Chem* **280**, 35588-35597 (2005).
33. Maeder, M.L. et al. Rapid "open-source" engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol Cell* **31**, 294-301 (2008).

34. Meng, X., Noyes, M.B., Zhu, L.J., Lawson, N.D. & Wolfe, S.A. Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat Biotechnol* **26**, 695-701 (2008).
35. Gupta, A. et al. An optimized two-finger archive for ZFN-mediated gene targeting. *Nat Methods* **9**, 588-590 (2012).
36. Zhu, C. et al. Using defined finger-finger interfaces as units of assembly for constructing zinc-finger nucleases. *Nucleic Acids Res* **41**, 2455-2465 (2013).
37. Persikov, A.V. et al. A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. *Nucleic Acids Res* **43**, 1965-1984 (2015).
38. Zhu, C. et al. Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. *Development* **138**, 4555-4564 (2011).
39. Kay, S., Hahn, S., Marois, E., Hause, G. & Bonas, U. A bacterial effector acts as a plant transcription factor and induces a cell size regulator. *Science* **318**, 648-651 (2007).
40. Romer, P. et al. Plant pathogen recognition mediated by promoter activation of the pepper Bs3 resistance gene. *Science* **318**, 645-648 (2007).
41. Boch, J. & Bonas, U. Xanthomonas AvrBs3 family-type III effectors: discovery and function. *Annu Rev Phytopathol* **48**, 419-436 (2010).
42. Herbers, K.C.-S., J; Bonas, U Race-specificity of plant resistance to bacterial spot disease determined by repetitive motifs in a bacterial avirulence protein. *Nature* **356**, 172-174 (1992).
43. Joung, J.K. & Sander, J.D. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* **14**, 49-55 (2013).
44. Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. *Science* **326**, 1509-1512 (2009).
45. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. *Science* **326**, 1501 (2009).
46. Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J. & Stoddard, B.L. The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science* **335**, 716-719 (2012).
47. Deng, D. et al. Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science* **335**, 720-723 (2012).
48. Lamb, B.M., Mercer, A.C. & Barbas, C.F., 3rd Directed evolution of the TALE N-terminal domain for recognition of all 5' bases. *Nucleic Acids Res* **41**, 9779-9785 (2013).
49. Sugisaki, H. & Kanazawa, S. New restriction endonucleases from Flavobacterium okeanokoites (FokI) and Micrococcus luteus (MluI). *Gene* **16**, 73-78 (1981).
50. Li, L., Wu, L.P. & Chandrasegaran, S. Functional domains in Fok I restriction endonuclease. *Proc Natl Acad Sci U S A* **89**, 4275-4279 (1992).
51. Wah, D.A., Hirsch, J.A., Dorner, L.F., Schildkraut, I. & Aggarwal, A.K. Structure of the multimodular endonuclease FokI bound to DNA. *Nature* **388**, 97-100 (1997).
52. Kim, Y.G. & Chandrasegaran, S. Chimeric restriction endonuclease. *Proc Natl Acad Sci U S A* **91**, 883-887 (1994).

53. Kim, Y.G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc Natl Acad Sci U S A* **93**, 1156-1160 (1996).
54. Kim, Y.G., Smith, J., Durgesha, M. & Chandrasegaran, S. Chimeric restriction enzyme: Gal4 fusion to FokI cleavage domain. *Biol Chem* **379**, 489-495 (1998).
55. Christian, M. et al. Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics* **186**, 757-761 (2010).
56. Li, T. et al. TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. *Nucleic Acids Res* **39**, 359-372 (2011).
57. Waugh, D.S. & Sauer, R.T. Single amino acid substitutions uncouple the DNA binding and strand scission activities of Fok I endonuclease. *Proc Natl Acad Sci U S A* **90**, 9596-9600 (1993).
58. Kaczorowski, T., Skowron, P. & Podhajska, A.J. Purification and characterization of the FokI restriction endonuclease. *Gene* **80**, 209-216 (1989).
59. Skowron, P., Kaczorowski, T., Tucholski, J. & Podhajska, A.J. Atypical DNA-binding properties of class-IIS restriction endonucleases: evidence for recognition of the cognate sequence by a FokI monomer. *Gene* **125**, 1-10 (1993).
60. Bitinaite, J., Wah, D.A., Aggarwal, A.K. & Schildkraut, I. FokI dimerization is required for DNA cleavage. *Proc Natl Acad Sci U S A* **95**, 10570-10575 (1998).
61. Wah, D.A., Bitinaite, J., Schildkraut, I. & Aggarwal, A.K. Structure of FokI has implications for DNA cleavage. *Proc Natl Acad Sci U S A* **95**, 10564-10569 (1998).
62. Catto, L.E., Ganguly, S., Milsom, S.E., Welsh, A.J. & Halford, S.E. Protein assembly and DNA looping by the FokI restriction endonuclease. *Nucleic Acids Res* **34**, 1711-1720 (2006).
63. Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. Genome editing with engineered zinc finger nucleases. *Nat Rev Genet* **11**, 636-646 (2010).
64. Oakes, B.L. et al. Multi-reporter selection for the design of active and more specific zinc-finger nucleases for genome editing. *Nat Commun* **7**, 10194 (2016).
65. Klug, A. The discovery of zinc fingers and their applications in gene regulation and genome manipulation. *Annu Rev Biochem* **79**, 213-231 (2010).
66. Emerson, R.O. & Thomas, J.H. Adaptive evolution in zinc finger transcription factors. *PLoS Genet* **5**, e1000325 (2009).
67. Choo, Y. & Klug, A. A role in DNA binding for the linker sequences of the first three zinc fingers of TFIIIA. *Nucleic Acids Res* **21**, 3341-3346 (1993).
68. Cook, W.J. et al. Mutations in the zinc-finger region of the yeast regulatory protein ADR1 affect both DNA binding and transcriptional activation. *J Biol Chem* **269**, 9374-9379 (1994).
69. Kim, J.S. & Pabo, C.O. Getting a handhold on DNA: design of poly-zinc finger proteins with femtomolar dissociation constants. *Proc Natl Acad Sci U S A* **95**, 2812-2817 (1998).
70. Moore, M., Choo, Y. & Klug, A. Design of polyzinc finger peptides with structured linkers. *Proc Natl Acad Sci U S A* **98**, 1432-1436 (2001).

71. Moore, M., Klug, A. & Choo, Y. Improved DNA binding specificity from polyzinc finger peptides by using strings of two-finger units. *Proc Natl Acad Sci U S A* **98**, 1437-1441 (2001).
72. Szczepek, M. et al. Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nat Biotechnol* **25**, 786-793 (2007).
73. Miller, J.C. et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nat Biotechnol* **25**, 778-785 (2007).
74. Doyon, Y. et al. Enhancing zinc-finger-nuclease activity with improved obligate heterodimeric architectures. *Nat Methods* **8**, 74-79 (2011).
75. Ramirez, C.L. et al. Engineered zinc finger nickases induce homology-directed repair with reduced mutagenic effects. *Nucleic Acids Res* **40**, 5560-5568 (2012).
76. Kim, E. et al. Precision genome engineering with programmable DNA-nicking enzymes. *Genome Res* **22**, 1327-1333 (2012).
77. Bibikova, M. et al. Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol Cell Biol* **21**, 289-297 (2001).
78. Porteus, M.H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. *Science* **300**, 763 (2003).
79. Urnov, F.D. et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646-651 (2005).
80. Porteus, M.H. Mammalian gene targeting with designed zinc finger nucleases. *Mol Ther* **13**, 438-446 (2006).
81. Wright, D.A. et al. High-frequency homologous recombination in plants mediated by zinc-finger nucleases. *Plant J* **44**, 693-705 (2005).
82. Lloyd, A., Plaisier, C.L., Carroll, D. & Drews, G.N. Targeted mutagenesis using zinc-finger nucleases in Arabidopsis. *Proc Natl Acad Sci U S A* **102**, 2232-2237 (2005).
83. Bibikova, M., Golic, M., Golic, K.G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. *Genetics* **161**, 1169-1175 (2002).
84. Bibikova, M., Beumer, K., Trautman, J.K. & Carroll, D. Enhancing gene targeting with designed zinc finger nucleases. *Science* **300**, 764 (2003).
85. Carbery, I.D. et al. Targeted genome modification in mice using zinc-finger nucleases. *Genetics* **186**, 451-459 (2010).
86. Shukla, V.K. et al. Precise genome modification in the crop species Zea mays using zinc-finger nucleases. *Nature* **459**, 437-441 (2009).
87. Hauschild, J. et al. Efficient generation of a biallelic knockout in pigs using zinc-finger nucleases. *Proc Natl Acad Sci U S A* **108**, 12013-12017 (2011).
88. Xiong, K. et al. Targeted editing of goat genome with modular-assembly zinc finger nucleases based on activity prediction by computational molecular modeling. *Mol Biol Rep* **40**, 4251-4256 (2013).
89. Carlson, D.F. et al. Efficient TALEN-mediated gene knockout in livestock. *Proc Natl Acad Sci U S A* **109**, 17382-17387 (2012).
90. Liu, R. et al. Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367-377 (1996).

91. Samson, M. et al. Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene. *Nature* **382**, 722-725 (1996).
92. Perez, E.E. et al. Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat Biotechnol* **26**, 808-816 (2008).
93. Holt, N. et al. Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat Biotechnol* **28**, 839-847 (2010).
94. Tebas, P. et al. Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV. *N Engl J Med* **370**, 901-910 (2014).
95. Lettre, G. & Bauer, D.E. Fetal haemoglobin in sickle-cell disease: from genetic epidemiology to new therapeutic strategies. *Lancet* **387**, 2554-2564 (2016).
96. Canver, M.C. & Orkin, S.H. Customizing the genome as therapy for the beta-hemoglobinopathies. *Blood* **127**, 2536-2545 (2016).
97. Canver, M.C. et al. BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis. *Nature* **527**, 192-197 (2015).
98. Vierstra, J. et al. Functional footprinting of regulatory DNA. *Nat Methods* **12**, 927-930 (2015).
99. Chang, K.H. et al. Long-Term Engraftment and Fetal Globin Induction upon BCL11A Gene Editing in Bone-Marrow-Derived CD34+ Hematopoietic Stem and Progenitor Cells. *Mol Ther Methods Clin Dev* **4**, 137-148 (2017).
100. Naldini, L. Gene therapy returns to centre stage. *Nature* **526**, 351-360 (2015).
101. Hoban, M.D. et al. Correction of the sickle cell disease mutation in human hematopoietic stem/progenitor cells. *Blood* **125**, 2597-2604 (2015).
102. DeKelver, R.C. et al. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res* **20**, 1133-1142 (2010).
103. Lombardo, A. et al. Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Methods* **8**, 861-869 (2011).
104. De Ravin, S.S. et al. Targeted gene addition in human CD34(+) hematopoietic cells for correction of X-linked chronic granulomatous disease. *Nat Biotechnol* **34**, 424-429 (2016).
105. Lombardo, A. et al. Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat Biotechnol* **25**, 1298-1306 (2007).
106. Genovese, P. et al. Targeted genome editing in human repopulating haematopoietic stem cells. *Nature* **510**, 235-240 (2014).
107. Li, H. et al. In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature* **475**, 217-221 (2011).
108. Anguela, X.M. et al. Robust ZFN-mediated genome editing in adult hemophilic mice. *Blood* **122**, 3283-3287 (2013).
109. Sharma, R. et al. In vivo genome editing of the albumin locus as a platform for protein replacement therapy. *Blood* **126**, 1777-1784 (2015).

110. Miller, J.C. et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol* **29**, 143-148 (2011).
111. Cermak, T. et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res* **39**, e82 (2011).
112. Zhang, F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. *Nat Biotechnol* **29**, 149-153 (2011).
113. Schmid-Burgk, J.L., Schmidt, T., Kaiser, V., Honing, K. & Hornung, V. A ligation-independent cloning technique for high-throughput assembly of transcription activator-like effector genes. *Nat Biotechnol* **31**, 76-81 (2013).
114. Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* **30**, 460-465 (2012).
115. Holkers, M. et al. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res* **41**, e63 (2013).
116. Yang, L. et al. Optimization of scarless human stem cell genome editing. *Nucleic Acids Res* **41**, 9049-9061 (2013).
117. Hockemeyer, D. et al. Genetic engineering of human pluripotent cells using TALE nucleases. *Nat Biotechnol* **29**, 731-734 (2011).
118. Osborn, M.J. et al. TALEN-based gene correction for epidermolysis bullosa. *Mol Ther* **21**, 1151-1159 (2013).
119. Mussolino, C. et al. TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity. *Nucleic Acids Res* **42**, 6762-6773 (2014).
120. Guilinger, J.P. et al. Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity. *Nat Methods* **11**, 429-435 (2014).
121. Miller, J.C. et al. Improved specificity of TALE-based genome editing using an expanded RVD repertoire. *Nat Methods* **12**, 465-471 (2015).
122. Kim, Y. et al. A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol* **31**, 251-258 (2013).
123. Wood, A.J. et al. Targeted genome editing across species using ZFNs and TALENs. *Science* **333**, 307 (2011).
124. Sander, J.D. et al. Targeted gene disruption in somatic zebrafish cells using engineered TALENs. *Nat Biotechnol* **29**, 697-698 (2011).
125. Huang, P. et al. Heritable gene targeting in zebrafish using customized TALENs. *Nat Biotechnol* **29**, 699-700 (2011).
126. Tesson, L. et al. Knockout rats generated by embryo microinjection of TALENs. *Nat Biotechnol* **29**, 695-696 (2011).
127. Liu, J. et al. Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy. *J Genet Genomics* **39**, 209-215 (2012).
128. Xiao, A. et al. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res* **41**, e141 (2013).
129. Ousterout, D.G. & Gersbach, C.A. The Development of TALE Nucleases for Biotechnology. *Methods Mol Biol* **1338**, 27-42 (2016).

130. Kim, Y.K. et al. TALEN-based knockout library for human microRNAs. *Nat Struct Mol Biol* **20**, 1458-1464 (2013).
131. Choi, S.M. et al. Efficient drug screening and gene correction for treating liver disease using patient-specific stem cells. *Hepatology* **57**, 2458-2468 (2013).
132. Ding, Q. et al. A TALEN genome-editing system for generating human stem cell-based disease models. *Cell Stem Cell* **12**, 238-251 (2013).
133. Park, C.Y. et al. Targeted inversion and reversion of the blood coagulation factor 8 gene in human iPS cells using TALENs. *Proc Natl Acad Sci U S A* **111**, 9253-9258 (2014).
134. Li, T., Liu, B., Spalding, M.H., Weeks, D.P. & Yang, B. High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nat Biotechnol* **30**, 390-392 (2012).
135. Demorest, Z.L. et al. Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. *BMC Plant Biol* **16**, 225 (2016).
136. Clasen, B.M. et al. Improving cold storage and processing traits in potato through targeted gene knockout. *Plant Biotechnol J* **14**, 169-176 (2016).
137. Cui, C. et al. Gene targeting by TALEN-induced homologous recombination in goats directs production of beta-lactoglobulin-free, high-human lactoferrin milk. *Sci Rep* **5**, 10482 (2015).
138. Mussolino, C. et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res* **39**, 9283-9293 (2011).
139. Bloom, K., Ely, A., Mussolino, C., Cathomen, T. & Arbuthnot, P. Inactivation of hepatitis B virus replication in cultured cells and in vivo with engineered transcription activator-like effector nucleases. *Mol Ther* **21**, 1889-1897 (2013).
140. Chen, J. et al. An efficient antiviral strategy for targeting hepatitis B virus genome using transcription activator-like effector nucleases. *Mol Ther* **22**, 303-311 (2014).
141. Bjurstrom, C.F. et al. Reactivating Fetal Hemoglobin Expression in Human Adult Erythroblasts Through BCL11A Knockdown Using Targeted Endonucleases. *Mol Ther Nucleic Acids* **5**, e351 (2016).
142. Ousterout, D.G. et al. Reading frame correction by targeted genome editing restores dystrophin expression in cells from Duchenne muscular dystrophy patients. *Mol Ther* **21**, 1718-1726 (2013).
143. Voit, R.A., Hendel, A., Pruett-Miller, S.M. & Porteus, M.H. Nuclease-mediated gene editing by homologous recombination of the human globin locus. *Nucleic Acids Res* **42**, 1365-1378 (2014).
144. Ma, N. et al. Transcription activator-like effector nuclease (TALEN)-mediated gene correction in integration-free beta-thalassemia induced pluripotent stem cells. *J Biol Chem* **288**, 34671-34679 (2013).
145. Sweeney, C.L. et al. Targeted Repair of CYBB in X-CGD iPSCs Requires Retention of Intronic Sequences for Expression and Functional Correction. *Mol Ther* **25**, 321-330 (2017).

146. Perez-Pinera, P. et al. Synergistic and tunable human gene activation by combinations of synthetic transcription factors. *Nat Methods* **10**, 239-242 (2013).
147. Konermann, S. et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472-476 (2013).
148. Crocker, J. & Stern, D.L. TALE-mediated modulation of transcriptional enhancers in vivo. *Nat Methods* **10**, 762-767 (2013).
149. Maeder, M.L. et al. Targeted DNA demethylation and activation of endogenous genes using programmable TALE-TET1 fusion proteins. *Nat Biotechnol* **31**, 1137-1142 (2013).
150. Bernstein, D.L., Le Lay, J.E., Ruano, E.G. & Kaestner, K.H. TALE-mediated epigenetic suppression of CDKN2A increases replication in human fibroblasts. *J Clin Invest* **125**, 1998-2006 (2015).
151. Koonin, E.V., Makarova, K.S. & Zhang, F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* **37**, 67-78 (2017).
152. Makarova, K.S. et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**, 722-736 (2015).
153. Bolotin, A., Quinquis, B., Sorokin, A. & Ehrlich, S.D. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**, 2551-2561 (2005).
154. Barrangou, R. et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**, 1709-1712 (2007).
155. Marraffini, L.A. & Sontheimer, E.J. CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* **322**, 1843-1845 (2008).
156. Chen, J.S. & Doudna, J.A. The chemistry of Cas9 and its CRISPR colleagues. *Nature Reviews Chemistry* **1**, 0078 (2017).
157. Shmakov, S. et al. Diversity and evolution of class 2 CRISPR-Cas systems. *Nat Rev Microbiol* **15**, 169-182 (2017).
158. Sontheimer, E.J. & Barrangou, R. The Bacterial Origins of the CRISPR Genome-Editing Revolution. *Hum Gene Ther* **26**, 413-424 (2015).
159. Jackson, S.A. et al. CRISPR-Cas: Adapting to change. *Science* **356** (2017).
160. Nunez, J.K., Harrington, L.B., Kranzusch, P.J., Engelman, A.N. & Doudna, J.A. Foreign DNA capture during CRISPR-Cas adaptive immunity. *Nature* **527**, 535-538 (2015).
161. Nunez, J.K., Lee, A.S., Engelman, A. & Doudna, J.A. Integrase-mediated spacer acquisition during CRISPR-Cas adaptive immunity. *Nature* **519**, 193-198 (2015).
162. Mojica, F.J., Diez-Villasenor, C., Garcia-Martinez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**, 733-740 (2009).
163. Wang, J. et al. Structural and Mechanistic Basis of PAM-Dependent Spacer Acquisition in CRISPR-Cas Systems. *Cell* **163**, 840-853 (2015).
164. Shipman, S.L., Nivala, J., Macklis, J.D. & Church, G.M. Molecular recordings by directed CRISPR spacer acquisition. *Science* **353**, aaf1175 (2016).
165. Brouns, S.J. et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**, 960-964 (2008).

166. Fonfara, I., Richter, H., Bratovic, M., Le Rhun, A. & Charpentier, E. The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature* **532**, 517-521 (2016).
167. East-Seletsky, A. et al. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**, 270-273 (2016).
168. Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**, 602-607 (2011).
169. Garneau, J.E. et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67-71 (2010).
170. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821 (2012).
171. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* **109**, E2579-2586 (2012).
172. Barrangou, R. & Doudna, J.A. Applications of CRISPR technologies in research and beyond. *Nat Biotechnol* **34**, 933-941 (2016).
173. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823 (2013).
174. Zetsche, B. et al. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* **163**, 759-771 (2015).
175. Abudayyeh, O.O. et al. RNA targeting with CRISPR-Cas13. *Nature* **550**, 280-284 (2017).
176. Cox, D.B.T. et al. RNA editing with CRISPR-Cas13. *Science* (2017).
177. Chylinski, K., Makarova, K.S., Charpentier, E. & Koonin, E.V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res* **42**, 6091-6105 (2014).
178. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823-826 (2013).
179. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31**, 230-232 (2013).
180. Jinek, M. et al. RNA-programmed genome editing in human cells. *Elife* **2**, e00471 (2013).
181. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910-918 (2013).
182. Hwang, W.Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat Biotechnol* **31**, 227-229 (2013).
183. Chen, B. et al. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479-1491 (2013).
184. Ledford, H. CRISPR, the disruptor. *Nature* **522**, 20-24 (2015).
185. Doudna, J.A. & Charpentier, E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).

186. Jao, L.E., Wente, S.R. & Chen, W. Efficient multiplex biallelic zebrafish genome editing using a CRISPR nuclease system. *Proc Natl Acad Sci U S A* **110**, 13904-13909 (2013).
187. Swiech, L. et al. In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat Biotechnol* **33**, 102-106 (2015).
188. Kabadi, A.M., Ousterout, D.G., Hilton, I.B. & Gersbach, C.A. Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. *Nucleic Acids Res* **42**, e147 (2014).
189. Mandal, P.K. et al. Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. *Cell Stem Cell* **15**, 643-652 (2014).
190. Tsai, S.Q. et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat Biotechnol* **32**, 569-576 (2014).
191. Hendel, A. et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nat Biotechnol* **33**, 985-989 (2015).
192. Yang, H. et al. One-step generation of mice carrying reporter and conditional alleles by CRISPR/Cas-mediated genome engineering. *Cell* **154**, 1370-1379 (2013).
193. Canver, M.C. et al. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. *J Biol Chem* **289**, 21312-21324 (2014).
194. Lupianez, D.G. et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* **161**, 1012-1025 (2015).
195. Torres, R. et al. Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system. *Nat Commun* **5**, 3964 (2014).
196. Choi, P.S. & Meyerson, M. Targeted genomic rearrangements using CRISPR/Cas technology. *Nat Commun* **5**, 3728 (2014).
197. Ghezraoui, H. et al. Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Mol Cell* **55**, 829-842 (2014).
198. Cho, S.W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24**, 132-141 (2014).
199. Wang, T., Wei, J.J., Sabatini, D.M. & Lander, E.S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80-84 (2014).
200. Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87 (2014).
201. Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera Mdel, C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* **32**, 267-273 (2014).
202. Zhou, Y. et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487-491 (2014).
203. Platt, R.J. et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440-455 (2014).
204. Parnas, O. et al. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell* **162**, 675-686 (2015).

205. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569-573 (2014).
206. Byrne, S.M., Ortiz, L., Mali, P., Aach, J. & Church, G.M. Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells. *Nucleic Acids Res* **43**, e21 (2015).
207. Yin, H. et al. Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol* **32**, 551-553 (2014).
208. Schumann, K. et al. Generation of knock-in primary human T cells using Cas9 ribonucleoproteins. *Proc Natl Acad Sci U S A* **112**, 10437-10442 (2015).
209. 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 (2014).
210. DeWitt, M.A. et al. Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells. *Sci Transl Med* **8**, 360ra134 (2016).
211. Quadros, R.M. et al. Easi-CRISPR: a robust method for one-step generation of mice carrying conditional and insertion alleles using long ssDNA donors and CRISPR ribonucleoproteins. *Genome Biol* **18**, 92 (2017).
212. Yoshimi, K. et al. ssODN-mediated knock-in with CRISPR-Cas for large genomic regions in zygotes. *Nat Commun* **7**, 10431 (2016).
213. De Ravin, S.S. et al. CRISPR-Cas9 gene repair of hematopoietic stem cells from patients with X-linked chronic granulomatous disease. *Sci Transl Med* **9** (2017).
214. Holkers, M. et al. Adenoviral vector DNA for accurate genome editing with engineered nucleases. *Nat Methods* **11**, 1051-1057 (2014).
215. Kaulich, M. et al. Efficient CRISPR-rAAV engineering of endogenous genes to study protein function by allele-specific RNAi. *Nucleic Acids Res* **43**, e45 (2015).
216. Bak, R.O. & Porteus, M.H. CRISPR-Mediated Integration of Large Gene Cassettes Using AAV Donor Vectors. *Cell Rep* **20**, 750-756 (2017).
217. Dever, D.P. et al. CRISPR/Cas9 beta-globin gene targeting in human haematopoietic stem cells. *Nature* **539**, 384-389 (2016).
218. Hendel, A. et al. Quantifying genome-editing outcomes at endogenous loci with SMRT sequencing. *Cell Rep* **7**, 293-305 (2014).
219. Chu, V.T. et al. Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33**, 543-548 (2015).
220. Maruyama, T. et al. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* **33**, 538-542 (2015).
221. Sanchez-Rivera, F.J. et al. Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* **516**, 428-431 (2014).
222. Xue, W. et al. CRISPR-mediated direct mutation of cancer genes in the mouse liver. *Nature* **514**, 380-384 (2014).
223. Chen, Y. et al. Functional disruption of the dystrophin gene in rhesus monkey using CRISPR/Cas9. *Hum Mol Genet* **24**, 3764-3774 (2015).

224. Ablain, J., Durand, E.M., Yang, S., Zhou, Y. & Zon, L.I. A CRISPR/Cas9 vector system for tissue-specific gene disruption in zebrafish. *Dev Cell* **32**, 756-764 (2015).
225. Wang, D. et al. Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses. *Hum Gene Ther* **26**, 432-442 (2015).
226. Zuckermann, M. et al. Somatic CRISPR/Cas9-mediated tumour suppressor disruption enables versatile brain tumour modelling. *Nat Commun* **6**, 7391 (2015).
227. Sanchez-Rivera, F.J. & Jacks, T. Applications of the CRISPR-Cas9 system in cancer biology. *Nat Rev Cancer* **15**, 387-395 (2015).
228. Wang, G. et al. Modeling the mitochondrial cardiomyopathy of Barth syndrome with induced pluripotent stem cell and heart-on-chip technologies. *Nat Med* **20**, 616-623 (2014).
229. Chen, S. et al. Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**, 1246-1260 (2015).
230. Maddalo, D. et al. In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* **516**, 423-427 (2014).
231. Blasco, R.B. et al. Simple and rapid in vivo generation of chromosomal rearrangements using CRISPR/Cas9 technology. *Cell Rep* **9**, 1219-1227 (2014).
232. Sweeney, C.L. et al. CRISPR-Mediated Knockout of Cybb in NSG Mice Establishes a Model of Chronic Granulomatous Disease for Human Stem-Cell Gene Therapy Transplants. *Hum Gene Ther* **28**, 565-575 (2017).
233. Thakore, P.I., Black, J.B., Hilton, I.B. & Gersbach, C.A. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat Methods* **13**, 127-137 (2016).
234. Li, H.L. et al. Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9. *Stem Cell Reports* **4**, 143-154 (2015).
235. Park, C.Y. et al. Functional Correction of Large Factor VIII Gene Chromosomal Inversions in Hemophilia A Patient-Derived iPSCs Using CRISPR-Cas9. *Cell Stem Cell* **17**, 213-220 (2015).
236. Tabebordbar, M. et al. In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* **351**, 407-411 (2016).
237. Nelson, C.E. et al. In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351**, 403-407 (2016).
238. Long, C. et al. Postnatal genome editing partially restores dystrophin expression in a mouse model of muscular dystrophy. *Science* **351**, 400-403 (2016).
239. Yin, H. et al. Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo. *Nat Biotechnol* **34**, 328-333 (2016).
240. Gori, J.L. et al. Delivery and Specificity of CRISPR-Cas9 Genome Editing Technologies for Human Gene Therapy. *Hum Gene Ther* **26**, 443-451 (2015).
241. Wu, Z., Yang, H. & Colosi, P. Effect of genome size on AAV vector packaging. *Mol Ther* **18**, 80-86 (2010).

242. Hou, Z. et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. *Proc Natl Acad Sci U S A* **110**, 15644-15649 (2013).
243. Zhang, Y. et al. Processing-independent CRISPR RNAs limit natural transformation in Neisseria meningitidis. *Mol Cell* **50**, 488-503 (2013).
244. Ran, F.A. et al. In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186-191 (2015).
245. Kim, E. et al. In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat Commun* **8**, 14500 (2017).
246. Hacein-Bey-Abina, S. et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J Clin Invest* **118**, 3132-3142 (2008).
247. Stein, S. et al. Genomic instability and myelodysplasia with monosomy 7 consequent to EVI1 activation after gene therapy for chronic granulomatous disease. *Nat Med* **16**, 198-204 (2010).
248. Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C. & Doudna, J.A. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* **507**, 62-67 (2014).
249. Szczelkun, M.D. et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci U S A* **111**, 9798-9803 (2014).
250. Wu, X. et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat Biotechnol* **32**, 670-676 (2014).
251. Sternberg, S.H., LaFrance, B., Kaplan, M. & Doudna, J.A. Conformational control of DNA target cleavage by CRISPR-Cas9. *Nature* **527**, 110-113 (2015).
252. Jinek, M. et al. Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343**, 1247997 (2014).
253. Nishimasu, H. et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935-949 (2014).
254. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J.A. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477-1481 (2015).
255. Jiang, F. et al. Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. *Science* **351**, 867-871 (2016).
256. Hsu, P.D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**, 827-832 (2013).
257. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31**, 822-826 (2013).
258. Chari, R., Mali, P., Moosburner, M. & Church, G.M. Unraveling CRISPR-Cas9 genome engineering parameters via a library-on-library approach. *Nat Methods* **12**, 823-826 (2015).
259. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32**, 677-683 (2014).

260. Zhang, Y. et al. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. *Sci Rep* **4**, 5405 (2014).
261. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**, 833-838 (2013).
262. Boyle, E.A. et al. High-throughput biochemical profiling reveals sequence determinants of dCas9 off-target binding and unbinding. *Proc Natl Acad Sci U S A* **114**, 5461-5466 (2017).
263. Gagnon, J.A. et al. Efficient mutagenesis by Cas9 protein-mediated oligonucleotide insertion and large-scale assessment of single-guide RNAs. *PLoS One* **9**, e98186 (2014).
264. Doench, J.G. et al. Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* **32**, 1262-1267 (2014).
265. Farboud, B. & Meyer, B.J. Dramatic enhancement of genome editing by CRISPR/Cas9 through improved guide RNA design. *Genetics* **199**, 959-971 (2015).
266. Xu, H. et al. Sequence determinants of improved CRISPR sgRNA design. *Genome Res* **25**, 1147-1157 (2015).
267. Montague, T.G., Cruz, J.M., Gagnon, J.A., Church, G.M. & Valen, E. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. *Nucleic Acids Res* **42**, W401-407 (2014).
268. Moreno-Mateos, M.A. et al. CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat Methods* **12**, 982-988 (2015).
269. Pattanayak, V. et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31**, 839-843 (2013).
270. Lin, Y. et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. *Nucleic Acids Res* **42**, 7473-7485 (2014).
271. Smith, C. et al. Whole-genome sequencing analysis reveals high specificity of CRISPR/Cas9 and TALEN-based genome editing in human iPSCs. *Cell Stem Cell* **15**, 12-13 (2014).
272. Veres, A. et al. 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 (2014).
273. Kim, D. et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods* **12**, 237-243, 231 p following 243 (2015).
274. Tsai, S.Q. & Joung, J.K. What's changed with genome editing? *Cell Stem Cell* **15**, 3-4 (2014).
275. Iyer, V. et al. Off-target mutations are rare in Cas9-modified mice. *Nat Methods* **12**, 479 (2015).
276. Shen, B. et al. Efficient genome modification by CRISPR-Cas9 nickase with minimal off-target effects. *Nat Methods* **11**, 399-402 (2014).

277. Varshney, G.K. et al. High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9. *Genome Res* **25**, 1030-1042 (2015).
278. Zhu, L.J., Holmes, B.R., Aronin, N. & Brodsky, M.H. CRISPRseek: a bioconductor package to identify target-specific guide RNAs for CRISPR-Cas9 genome-editing systems. *PLoS One* **9**, e108424 (2014).
279. Bae, S., Park, J. & Kim, J.S. Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**, 1473-1475 (2014).
280. Cradick, T.J., Qiu, P., Lee, C.M., Fine, E.J. & Bao, G. COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites. *Mol Ther Nucleic Acids* **3**, e214 (2014).
281. Singh, R., Kuscu, C., Quinlan, A., Qi, Y. & Adli, M. Cas9-chromatin binding information enables more accurate CRISPR off-target prediction. *Nucleic Acids Res* **43**, e118 (2015).
282. Aach, J., Mali, P. & Church, G.M. CasFinder: flexible algorithm for identifying specific Cas9 targets in genomes. *bioRxiv* (2014).
283. Heigwer, F., Kerr, G. & Boutros, M. E-CRISP: fast CRISPR target site identification. *Nat Methods* **11**, 122-123 (2014).
284. Sander, J.D. et al. ZiFiT (Zinc Finger Targeter): an updated zinc finger engineering tool. *Nucleic Acids Res* **38**, W462-468 (2010).
285. Xiao, A. et al. CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics* (2014).
286. Tsai, S.Q. et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat Biotechnol* **33**, 187-197 (2015).
287. Frock, R.L. et al. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nat Biotechnol* **33**, 179-186 (2015).
288. Wang, X. et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat Biotechnol* **33**, 175-178 (2015).
289. Cameron, P. et al. Mapping the genomic landscape of CRISPR-Cas9 cleavage. *Nat Methods* **14**, 600-606 (2017).
290. Tsai, S.Q. et al. CIRCLE-seq: a highly sensitive in vitro screen for genome-wide CRISPR-Cas9 nuclease off-targets. *Nat Methods* **14**, 607-614 (2017).
291. Crosetto, N. et al. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods* **10**, 361-365 (2013).
292. Gabriel, R. et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. *Nat Biotechnol* **29**, 816-823 (2011).
293. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat Biotechnol* **32**, 279-284 (2014).
294. Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res* **24**, 1012-1019 (2014).

295. Ramakrishna, S. et al. Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. *Genome Res* **24**, 1020-1027 (2014).
296. D'Astolfo, D.S. et al. Efficient intracellular delivery of native proteins. *Cell* **161**, 674-690 (2015).
297. Zuris, J.A. et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* **33**, 73-80 (2015).
298. Ran, F.A. et al. Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* **154**, 1380-1389 (2013).
299. Wyvekens, N., Topkar, V.V., Khayter, C., Joung, J.K. & Tsai, S.Q. Dimeric CRISPR RNA-Guided FokI-dCas9 Nucleases Directed by Truncated gRNAs for Highly Specific Genome Editing. *Hum Gene Ther* **26**, 425-431 (2015).
300. Guilinger, J.P., Thompson, D.B. & Liu, D.R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat Biotechnol* **32**, 577-582 (2014).
301. Slaymaker, I.M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88 (2016).
302. Kleinstiver, B.P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495 (2016).
303. Chen, J.S. et al. Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* **550**, 407-410 (2017).
304. Kleinstiver, B.P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature* **523**, 481-485 (2015).
305. Kleinstiver, B.P. et al. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* **33**, 1293-1298 (2015).
306. Bolukbasi, M.F. et al. DNA-binding-domain fusions enhance the targeting range and precision of Cas9. *Nat Methods* **12**, 1150-1156 (2015).
307. Wright, A.V. et al. Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci U S A* **112**, 2984-2989 (2015).
308. Zetsche, B., Volz, S.E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **33**, 139-142 (2015).
309. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol* **33**, 755-760 (2015).
310. Truong, D.J. et al. Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res* **43**, 6450-6458 (2015).
311. Oakes, B.L. et al. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat Biotechnol* **34**, 646-651 (2016).
312. Davis, K.M., Pattanayak, V., Thompson, D.B., Zuris, J.A. & Liu, D.R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* **11**, 316-318 (2015).
313. Fonfara, I. et al. Phylogeny of Cas9 determines functional exchangeability of dual-RNA and Cas9 among orthologous type II CRISPR-Cas systems. *Nucleic Acids Res* **42**, 2577-2590 (2014).

314. Esvelt, K.M. et al. Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* **10**, 1116-1121 (2013).
315. Hur, J.K. et al. Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins. *Nat Biotechnol* **34**, 807-808 (2016).
316. Kim, Y. et al. Generation of knockout mice by Cpf1-mediated gene targeting. *Nat Biotechnol* **34**, 808-810 (2016).
317. Begemann, M.B. et al. Precise insertion and guided editing of higher plant genomes using Cpf1 CRISPR nucleases. *Sci Rep* **7**, 11606 (2017).
318. Zetsche, B. et al. Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol* **35**, 31-34 (2017).
319. Kim, H.K. et al. In vivo high-throughput profiling of CRISPR-Cpf1 activity. *Nat Methods* **14**, 153-159 (2017).
320. Kleinstiver, B.P. et al. Genome-wide specificities of CRISPR-Cas Cpf1 nucleases in human cells. *Nat Biotechnol* **34**, 869-874 (2016).
321. Kim, D. et al. Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells. *Nat Biotechnol* **34**, 863-868 (2016).
322. Amrani, N. et al. NmeCas9 is an intrinsically high-fidelity genome editing platform. *bioRxiv* (2017).
323. Hendel, A., Fine, E.J., Bao, G. & Porteus, M.H. Quantifying on- and off-target genome editing. *Trends Biotechnol* **33**, 132-140 (2015).
324. Shi, J. et al. Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* **33**, 661-667 (2015).
325. Kok, F.O. et al. Reverse genetic screening reveals poor correlation between morpholino-induced and mutant phenotypes in zebrafish. *Dev Cell* **32**, 97-108 (2015).
326. Han, J. et al. Efficient in vivo deletion of a large imprinted lncRNA by CRISPR/Cas9. *RNA Biol* **11**, 829-835 (2014).
327. Wang, S., Sengel, C., Emerson, M.M. & Cepko, C.L. A gene regulatory network controls the binary fate decision of rod and bipolar cells in the vertebrate retina. *Dev Cell* **30**, 513-527 (2014).
328. Chen, S. et al. Global microRNA depletion suppresses tumor angiogenesis. *Genes Dev* **28**, 1054-1067 (2014).
329. Hnisz, D. et al. Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell* **58**, 362-370 (2015).
330. Bedell, V.M. et al. In vivo genome editing using a high-efficiency TALEN system. *Nature* **491**, 114-118 (2012).
331. Hwang, W.Y. et al. Heritable and precise zebrafish genome editing using a CRISPR-Cas system. *PLoS One* **8**, e68708 (2013).
332. Schmidt, M. et al. Polyclonal long-term repopulating stem cell clones in a primate model. *Blood* **100**, 2737-2743 (2002).
333. Chiarle, R. et al. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107-119 (2011).
334. Zhao, Y. et al. High-efficiency transfection of primary human and mouse T lymphocytes using RNA electroporation. *Mol Ther* **13**, 151-159 (2006).

335. Ebert, O. et al. Lymphocyte apoptosis: induction by gene transfer techniques. *Gene Ther* **4**, 296-302 (1997).
336. Josephs, E.A. et al. Structure and specificity of the RNA-guided endonuclease Cas9 during DNA interrogation, target binding and cleavage. *Nucleic Acids Res* **43**, 8924-8941 (2015).
337. Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A. & Liu, D.R. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature* **533**, 420-424 (2016).
338. Komor, A.C. et al. Improved base excision repair inhibition and bacteriophage Mu Gam protein yields C:G-to-T:A base editors with higher efficiency and product purity. *Sci Adv* **3**, eaao4774 (2017).
339. Gaudelli, N.M. et al. Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage. *Nature* (2017).
340. Sander, J.D. & Joung, J.K. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotechnol* **32**, 347-355 (2014).
341. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014).
342. Gabriel, R., von Kalle, C. & Schmidt, M. Mapping the precision of genome editing. *Nat Biotechnol* **33**, 150-152 (2015).
343. Zhu, L.J. Overview of guide RNA design tools for CRISPR-Cas9 genome editing technology. *Frontiers in Biology* **10**, 289-296 (2015).
344. Brunet, E. et al. Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci U S A* **106**, 10620-10625 (2009).
345. Lee, H.J., Kim, E. & Kim, J.S. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res* **20**, 81-89 (2010).
346. Hubbard, B.P. et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat Methods* **12**, 939-942 (2015).
347. Boissel, S. et al. megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res* **42**, 2591-2601 (2014).
348. Khalil, A.S. et al. A synthetic biology framework for programming eukaryotic transcription functions. *Cell* **150**, 647-658 (2012).
349. Meckler, J.F. et al. Quantitative analysis of TALE-DNA interactions suggests polarity effects. *Nucleic Acids Res* **41**, 4118-4128 (2013).
350. Lu, X.J. & Olson, W.K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat Protoc* **3**, 1213-1227 (2008).
351. Kearns, N.A. et al. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* **141**, 219-223 (2014).
352. Villefranc, J.A., Amigo, J. & Lawson, N.D. Gateway compatible vectors for analysis of gene function in the zebrafish. *Dev Dyn* **236**, 3077-3087 (2007).
353. Kok, F.O., Gupta, A., Lawson, N.D. & Wolfe, S.A. Construction and application of site-specific artificial nucleases for targeted gene editing. *Methods Mol Biol* **1101**, 267-303 (2014).

354. Wilson, K.A., Chateau, M.L. & Porteus, M.H. Design and Development of Artificial Zinc Finger Transcription Factors and Zinc Finger Nucleases to the hTERT Locus. *Mol Ther Nucleic Acids* **2**, e87 (2013).
355. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nature Methods* **9**, 671-675 (2012).
356. Gupta, A., Meng, X., Zhu, L.J., Lawson, N.D. & Wolfe, S.A. Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Res* **39**, 381-392 (2011).
357. Ihaka, R. & Gentleman, R. R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics* **5**, 299-314 (1996).
358. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).
359. Zhu, L.J. et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
360. Zhu, L.J. Integrative analysis of ChIP-chip and ChIP-seq dataset. *Methods Mol Biol* **1067**, 105-124 (2013).
361. Atkinson, H. & Chalmers, R. Delivering the goods: viral and non-viral gene therapy systems and the inherent limits on cargo DNA and internal sequences. *Genetica* **138**, 485-498 (2010).
362. Gupta, A. et al. An improved predictive recognition model for Cys(2)-His(2) zinc finger proteins. *Nucleic Acids Res* **42**, 4800-4812 (2014).
363. Klemm, J.D. & Pabo, C.O. Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. *Genes Dev* **10**, 27-36 (1996).
364. Yusa, K. et al. Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature* **478**, 391-394 (2011).
365. Mahiny, A.J. et al. In vivo genome editing using nuclease-encoding mRNA corrects SP-B deficiency. *Nat Biotechnol* **33**, 584-586 (2015).
366. Gupta, R.M. & Musunuru, K. Expanding the genetic editing tool kit: ZFNs, TALENs, and CRISPR-Cas9. *J Clin Invest* **124**, 4154-4161 (2014).
367. Maeder, M.L., Thibodeau-Beganny, S., Sander, J.D., Voytas, D.F. & Joung, J.K. Oligomerized pool engineering (OPEN): an 'open-source' protocol for making customized zinc-finger arrays. *Nat Protoc* **4**, 1471-1501 (2009).
368. Sander, J.D. et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat Methods* **8**, 67-69 (2011).
369. Carroll, D., Morton, J.J., Beumer, K.J. & Segal, D.J. Design, construction and in vitro testing of zinc finger nucleases. *Nat Protoc* **1**, 1329-1341 (2006).
370. Kim, S., Lee, M.J., Kim, H., Kang, M. & Kim, J.S. Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nat Methods* **8**, 7 (2011).
371. Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. & Kim, J.S. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res* **19**, 1279-1288 (2009).

372. Bhakta, M.S. et al. Highly active zinc-finger nucleases by extended modular assembly. *Genome Res* **23**, 530-538 (2013).
373. Doyon, Y. et al. Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat Biotechnol* **26**, 702-708 (2008).
374. Mandell, J.G. & Barbas, C.F., 3rd Zinc Finger Tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res* **34**, W516-523 (2006).
375. Doyle, E.L. et al. TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res* **40**, W117-122 (2012).
376. Lin, Y. et al. SAPTA: a new design tool for improving TALE nuclease activity. *Nucleic Acids Res* **42**, e47 (2014).
377. Bolukbasi, M.F., Gupta, A. & Wolfe, S.A. Creating and evaluating accurate CRISPR-Cas9 scalpels for genomic surgery. *Nat Methods* **13**, 41-50 (2016).
378. Tycko, J., Myer, V.E. & Hsu, P.D. Methods for Optimizing CRISPR-Cas9 Genome Editing Specificity. *Mol Cell* **63**, 355-370 (2016).
379. Tsai, S.Q. & Joung, J.K. Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet* **17**, 300-312 (2016).
380. Hu, J.H., Davis, K.M. & Liu, D.R. Chemical Biology Approaches to Genome Editing: Understanding, Controlling, and Delivering Programmable Nucleases. *Cell Chem Biol* **23**, 57-73 (2016).
381. Nguyen, D.P. et al. Ligand-binding domains of nuclear receptors facilitate tight control of split CRISPR activity. *Nat Commun* **7**, 12009 (2016).
382. Richter, F. et al. Switchable Cas9. *Curr Opin Biotechnol* **48**, 119-126 (2017).
383. Rakhit, R., Navarro, R. & Wandless, T.J. Chemical biology strategies for posttranslational control of protein function. *Chem Biol* **21**, 1238-1252 (2014).
384. Zhang, J., Kobert, K., Flouri, T. & Stamatakis, A. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**, 614-620 (2014).
385. Blankenberg, D. et al. Manipulation of FASTQ data with Galaxy. *Bioinformatics* **26**, 1783-1785 (2010).
386. Zhu, L.J. et al. GUIDEseq: a bioconductor package to analyze GUIDE-Seq datasets for CRISPR-Cas nucleases. *BMC Genomics* **18**, 379 (2017).
387. Nishimasu, H. et al. Crystal Structure of Staphylococcus aureus Cas9. *Cell* **162**, 1113-1126 (2015).
388. Stankunas, K. et al. Rescue of degradation-prone mutants of the FK506-rapamycin binding (FRB) protein with chemical ligands. *Chembiochem* **8**, 1162-1169 (2007).
389. Liang, F.S., Ho, W.Q. & Crabtree, G.R. Engineering the ABA plant stress pathway for regulation of induced proximity. *Sci Signal* **4**, rs2 (2011).
390. Polstein, L.R. & Gersbach, C.A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat Chem Biol* **11**, 198-200 (2015).
391. Murugan, K., Babu, K., Sundaresan, R., Rajan, R. & Sashital, D.G. The Revolution Continues: Newly Discovered Systems Expand the CRISPR-Cas Toolkit. *Mol Cell* **68**, 15-25 (2017).

392. Cornu, T.I., Mussolino, C. & Cathomen, T. Refining strategies to translate genome editing to the clinic. *Nat Med* **23**, 415-423 (2017).
393. Dagdas, Y.S., Chen, J.S., Sternberg, S.H., Doudna, J.A. & Yildiz, A. A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9. *Sci Adv* **3**, eaao0027 (2017).
394. Dever, D.P. & Porteus, M.H. The changing landscape of gene editing in hematopoietic stem cells: a step towards Cas9 clinical translation. *Curr Opin Hematol* **24**, 481-488 (2017).
395. Andrews, S. FastQC A Quality Control tool for High Throughput Sequence Data. www.bioinformatics.babraham.ac.uk Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/. (Accessed: 8 May 2017)
396. Dabney, J. & Meyer, M. Length and GC-biases during sequencing library amplification: a comparison of various polymerase-buffer systems with ancient and modern DNA sequencing libraries. *Biotechniques* **52**, 87-94 (2012).
397. Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods* **9**, 72-74 (2011).
398. Paquet, D. et al. Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9. *Nature* **533**, 125-129 (2016).
399. Chen, F. et al. Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. *Nat Commun* **8**, 14958 (2017).
400. Gantz, V.M. & Bier, E. Genome editing. The mutagenic chain reaction: a method for converting heterozygous to homozygous mutations. *Science* **348**, 442-444 (2015).
401. Nelson, C.E. & Gersbach, C.A. Engineering Delivery Vehicles for Genome Editing. *Annu Rev Chem Biomol Eng* **7**, 637-662 (2016).
402. Iwai, H., Zuger, S., Jin, J. & Tam, P.H. Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. *FEBS Lett* **580**, 1853-1858 (2006).
403. Muona, M., Aranko, A.S., Raulinaitis, V. & Iwai, H. Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro. *Nat Protoc* **5**, 574-587 (2010).
404. Gantz, V.M. & Bier, E. The dawn of active genetics. *Bioessays* **38**, 50-63 (2016).
405. Godfray, H.C.J., North, A. & Burt, A. How driving endonuclease genes can be used to combat pests and disease vectors. *BMC Biol* **15**, 81 (2017).
406. Hryhorowicz, M., Zeyland, J., Slomski, R. & Lipinski, D. Genetically Modified Pigs as Organ Donors for Xenotransplantation. *Mol Biotechnol* (2017).
407. Yang, L. et al. Genome-wide inactivation of porcine endogenous retroviruses (PERVs). *Science* **350**, 1101-1104 (2015).
408. Niu, D. et al. Inactivation of porcine endogenous retrovirus in pigs using CRISPR-Cas9. *Science* **357**, 1303-1307 (2017).
409. Fogarty, N.M.E. et al. Genome editing reveals a role for OCT4 in human embryogenesis. *Nature* **550**, 67-73 (2017).

410. Ma, H. et al. Correction of a pathogenic gene mutation in human embryos. *Nature* **548**, 413-419 (2017).
411. Tang, L. et al. CRISPR/Cas9-mediated gene editing in human zygotes using Cas9 protein. *Mol Genet Genomics* **292**, 525-533 (2017).
412. Egli, D. et al. Inter-homologue repair in fertilized human eggs? *bioRxiv* (2017).
413. Chandrasegaran, S., Bullen, C.K. & Carroll, D. Genome editing of human embryos: to edit or not to edit, that is the question. *J Clin Invest* **127**, 3588-3590 (2017).
414. Isaac, R.S. et al. Nucleosome breathing and remodeling constrain CRISPR-Cas9 function. *Elife* **5** (2016).
415. Horlbeck, M.A. et al. Nucleosomes impede Cas9 access to DNA in vivo and in vitro. *Elife* **5** (2016).
416. Staahl, B.T. et al. Efficient genome editing in the mouse brain by local delivery of engineered Cas9 ribonucleoprotein complexes. *Nat Biotechnol* **35**, 431-434 (2017).
417. Bessis, N., GarciaCozar, F.J. & Boissier, M.C. Immune responses to gene therapy vectors: influence on vector function and effector mechanisms. *Gene Ther* **11 Suppl 1**, S10-17 (2004).
418. Gaskell, G. et al. Public views on gene editing and its uses. *Nat Biotechnol* **35**, 1021-1023 (2017).
419. Musunuru, K., Lagor, W.R. & Miano, J.M. What Do We Really Think About Human Germline Genome Editing, and What Does It Mean for Medicine? *Circ Cardiovasc Genet* **10** (2017).
420. Handyside, A.H., Kontogianni, E.H., Hardy, K. & Winston, R.M. Pregnancies from biopsied human preimplantation embryos sexed by Y-specific DNA amplification. *Nature* **344**, 768-770 (1990).
421. Traeger-Synodinos, J. Pre-implantation genetic diagnosis. *Best Pract Res Clin Obstet Gynaecol* **39**, 74-88 (2017).
422. Greenfield, A. et al. Assisted reproductive technologies to prevent human mitochondrial disease transmission. *Nat Biotechnol* **35**, 1059-1068 (2017).

APPENDIX

Appendix 1.1 Primer information for T7EI assays in Chapter II		
Target name	**5p Oligo Sequence**	**3p Oligo Sequence**
LRTM2	CTAGAAAGGGTAAGTGCCCCCTTGT	CCACAACTGTCACCTGTCCCTGG
KANK3	CTGCACTCACCTCTGTGAGGCAGG	GGAGTGGGGAGGAAGAGGGAACAG
TGM2	GCGATGAATGTTGCTCATGTTCTGCTTC	CTTCTGTGGACCCATCAGTGGTCTC
PLXNB2	GACAGACATAGGATCTGGGAACTTGC	GTAGCATCTTCCTCTGAGCAATACCG
TS2	AGCCCAGAAGTTGGACGAAAAGT	AGCCCCAGCTACCACCTCCT
OT2-2	TCCTGTCACAATTCCCTGAACTC	CTAAACTCTGACTGGGCCTCCA
TS3	GTGTGCAGACGGCAGTCACTAGG	CGTTCCCTCTTTGCTAGGAATATTGAAG
OT3-2	TGTCATCGTTATCGCTCATTTCCTAC	GCCGTCTGTTAGAGGGACAAGAACTA
TS4	CTCAGTCTTCCCATCAGGCTCTCAG	ACCCTAGTCATTGGAGGTGACATCG
DNAJC6	CTTAAATATTCTGCCGCAGTTGG	GAAAGCGGGTGAAAAGAAAGG
PLXDC2	ATCCCGGTGGTTACTGTAGCTGT	ATTGAAATTGCCCAGAAAACACA
GPRC5B	CGAGTCTTGTTTACACACGCAGA	TTAAGGTTGAGAGTGGGCTGAAC

Appendix 1.2 Deep Sequencing primers used in Chapter II	
Target name	**5p Oligo Sequence**
TS2	ctacacgacgctcttccgatctAGCCCAGAAGTTGGACGAAAAGT
OT2-1	ctacacgacgctcttccgatctCTGTCCTCTCCACATCCTCTCTG
OT2-2	ctacacgacgctcttccgatctAATATTCCTGGCCAAGTCGATTCC
OT2-6	ctacacgacgctcttccgatctTTACTCTTGAGATTGGAACGGGAAA
OT2-9	ctacacgacgctcttccgatctAAACAGCCAGAAGAGAACATCCAC
OT2-15	ctacacgacgctcttccgatctAATTGGGCAAAGCAAAAACCAG
OT2-17	ctacacgacgctcttccgatctTGCTGTCTAAACTTCCATTTCTCCA
OT2-19	ctacacgacgctcttccgatctGAACCCCTAGCCCAAACTGAGATA
OT2-23	ctacacgacgctcttccgatctGGCTCCTCGGTGCTCTGGTC
OT2-24	ctacacgacgctcttccgatctCTCATCCTTGTATCAGCTGCCTTC
TS2BZFP-OT1	ctacacgacgctcttccgatctACCACAGGAGTTCCGTCCTCTAAG
TS2BZFP-OT2	ctacacgacgctcttccgatctCAGTCCCAATAACAGTGATTCCAT
TS3	ctacacgacgctcttccgatctGTCTTCGAGAGTGAGGACGTGTG
OT3-2	ctacacgacgctcttccgatctCTGCAGTGAGGAGGTGGTTCTT
OT3-3	ctacacgacgctcttccgatctCTTTTCCTGTGGAACAACCAGAC
OT3-4	ctacacgacgctcttccgatctCTGTGTACACTGGTGGGCAGGAG
OT3-17	ctacacgacgctcttccgatctTTGCACAAATAGGCGCTTAATAAAT
OT3-18	ctacacgacgctcttccgatctCATATCTGTCACCACACAGTTACCA
OT3-19	ctacacgacgctcttccgatctATTGCGCGATCTTGAGTGAC
OT3-20	ctacacgacgctcttccgatctGGAGAAGCAAGTACCTTCTTTACAA
OT3-21	ctacacgacgctcttccgatctGCAGCCACACGACACTGAGG
OT3-22	ctacacgacgctcttccgatctGTGTGCCTCGTTTGGCTTCT
OT3-24	ctacacgacgctcttccgatctACAAATGGATGGGTGGGAGGTT
OT3-29	ctacacgacgctcttccgatctCGTGGGGCTAAAGTTGTATATGTG
OT3-30	ctacacgacgctcttccgatctGTAATGGGAGTAAGTGGACGTGTG
OT3-39	ctacacgacgctcttccgatctCTTGGAGAAGGCAGCTCGTC
VEGF3_New1	ctacacgacgctcttccgatctCCAGGGTCCCTTACTCCTTTTT
VEGF3_New2	ctacacgacgctcttccgatctGCACTGGGGAAAGAGAAGAGAAG
VEGF3_New3	ctacacgacgctcttccgatctTCCTCATAATTTATCAACAAACACAAA
TS3ZFP-OT1	ctacacgacgctcttccgatctGCTTACTCCCAGTGTGGGGATG
TS3ZFP-OT2	ctacacgacgctcttccgatctGACTGAGGGTGGAGGGGAAG
TS3ZFP-OT3	ctacacgacgctcttccgatctGCTCCAGCACCTCAATTTCCTTC
TS4	ctacacgacgctcttccgatctGCAGAAGCTGGAGGAGGAAGG
OT4-1	ctacacgacgctcttccgatctAAGAATAGGGGCTTATGGCATGG
OT4-2	ctacacgacgctcttccgatctTCCCCATCTTCCAGTTCCATAGC
OT4-3	ctacacgacgctcttccgatctCCACTTGGAGAGTCAGAGGTCACA
OT4-4	ctacacgacgctcttccgatctCATTCCTCCTGAGGGAAAATAAA

OT4-5	ctacacgacgctcttccgatctCAGCTCTGGCCTGGGAAAATA
OT4-15	ctacacgacgctcttccgatctGTTGGGGTCATGTGTGGTCCT
OT4-20	ctacacgacgctcttccgatctTTAGTGAGGAGCAGCTTTACAGG
OT4-22	ctacacgacgctcttccgatctTACCAAACAGCCCAGTTTTCTCC
OT4-26	ctacacgacgctcttccgatctACCGAGCTCAGGGTCTCTGC
OT4-52	ctacacgacgctcttccgatctCAGAGGCTGAAGAGGAAGACCAG
OT4-53	ctacacgacgctcttccgatctTGCAAATAGAGCCCTTTATTCAT
TS4ZFP-OT1	ctacacgacgctcttccgatctCCAGCTTAGCGTTTGTTTTCCTG
TS4ZFP-OT2	ctacacgacgctcttccgatctACTAGGGTGGGAGCCCAGAGAT
Target name	**3p Oligo Sequence**
TS2	agacgtgtgctcttccgatctAGCCCCAGCTACCACCTCCT
OT2-1	agacgtgtgctcttccgatctATGGCCTATGGGTTGAAAACAGT
OT2-2	agacgtgtgctcttccgatctCTGAAATGACCCCCAAGGGATAC
OT2-6	agacgtgtgctcttccgatctCGTGAAACTCCACACTTATCTACGC
OT2-9	agacgtgtgctcttccgatctTTGCTCTTTGTTGAGAGGGAAAGA
OT2-15	agacgtgtgctcttccgatctCGGAGCTGTTTCTGCTGACTGT
OT2-17	agacgtgtgctcttccgatctAACAGATGAGCCACAACCCTGTT
OT2-19	agacgtgtgctcttccgatctTTTGTGCTGATCTGTGGGTATCCT
OT2-23	agacgtgtgctcttccgatctGGAGTTAAGGGTGTCTCCGAGGT
OT2-24	agacgtgtgctcttccgatctCTGCGCTCTTCTTTCCTCTTCAA
TS2BZFP-OT1	agacgtgtgctcttccgatctGCCTCCAGGAGTGGGAGTGG
TS2BZFP-OT2	agacgtgtgctcttccgatctAGGTAGCTGGGGGAGGCACT
TS3	agacgtgtgctcttccgatctTCTATTGGAATCCTGGAGTGACC
OT3-2	agacgtgtgctcttccgatctGGGAAGTCACCGACAACAACAAG
OT3-3	agacgtgtgctcttccgatctCATGCATTCATGCACATATACCAC
OT3-4	agacgtgtgctcttccgatctCCCATTTCTCCTTTGAGGTTCATC
OT3-17	agacgtgtgctcttccgatctAGATAGCAGGTAAGGAGGGGTGTC
OT3-18	agacgtgtgctcttccgatctGGAATCTAATGTATGGCATGGTG
OT3-19	agacgtgtgctcttccgatctAGCAGCAGGAGCCTAGTACCAC
OT3-20	agacgtgtgctcttccgatctCCCATGCTTTTCTTGTGATAGTGA
OT3-21	agacgtgtgctcttccgatctGTCCACACACAGCGTCTTCC
OT3-22	agacgtgtgctcttccgatctAATGGTGAAGACAAGCACCTCAT
OT3-24	agacgtgtgctcttccgatctCATTGACCAATTCATTCACGCATC
OT3-29	agacgtgtgctcttccgatctGAGGTTAAATCCCTCCTCCACAG
OT3-30	agacgtgtgctcttccgatctTGGGTACAGCTGCAGACACAG
OT3-39	agacgtgtgctcttccgatctAAAGGTGGAAAGCATGATGAGG
VEGF3_New1	agacgtgtgctcttccgatctCAGGACAGAACACTTGGGTAAGTGAA
VEGF3_New2	agacgtgtgctcttccgatctTTTGGTGATGCTTGTTCATGTTG
VEGF3_New3	agacgtgtgctcttccgatctTTTATTTCCATGCATACAAACGTG

TS3ZFP-OT1	agacgtgtgctcttccgatctTAGATGCTGAAGTGGGCCTAGCA
TS3ZFP-OT2	agacgtgtgctcttccgatctGCTCCAACACCTGACATTTTCCT
TS3ZFP-OT3	agacgtgtgctcttccgatctCACTGGGGAGCTGTCGTACAGAA
TS4	agacgtgtgctcttccgatctCATCGATGTCCTCCCCATTG
OT4-1	agacgtgtgctcttccgatctTTTTTCTGAGGGCTGCTACCTG
OT4-2	agacgtgtgctcttccgatctCAGCTTCCTGCAGTGAGAAAGGA
OT4-3	agacgtgtgctcttccgatctCATTCATGGAGGGGCACAGAAG
OT4-4	agacgtgtgctcttccgatctTGGTTTACATTTTATTAATGGCTTACA
OT4-5	agacgtgtgctcttccgatctCACACCAGCAATGCTCTCGTC
OT4-15	agacgtgtgctcttccgatctCTGCTTGACCAGGCCATTTGTA
OT4-20	agacgtgtgctcttccgatctCCCATTTCTTCAAATCTCCTTGC
OT4-22	agacgtgtgctcttccgatctTCACCTCCTTTATTGCAGATTGCT
OT4-26	agacgtgtgctcttccgatctCAGAGACCTAGAAGCCAGCCAAT
OT4-52	agacgtgtgctcttccgatctTTGGCCCCAGTCTCTCTTCTATG
OT4-53	agacgtgtgctcttccgatctTGGCCTTTGTAGGAAAACACCA
TS4ZFP-OT1	agacgtgtgctcttccgatctCTGGCTGAGGGCATAACTCAAG
TS4ZFP-OT2	agacgtgtgctcttccgatctCATGTGCACACACACGTATTTCG

Appendix 1.3 Deep Sequencing primers used in Chapter III	
Target Site	**5p Oligo Sequence**
PMPCA	ctacacgacgctcttccgatctGAACTACTACACTCCCGACCG
VEGFA_TS2	ctacacgacgctcttccgatctAGCCCAGAAGTTGGACGAAAAGT
OTD_TS2_1	ctacacgacgctcttccgatctTGCAGCTGCCTCTGCCTCTA
OTD_TS2_2	ctacacgacgctcttccgatctGATACCAGCATGGGCTACCCA
OTD_TS2_3	ctacacgacgctcttccgatctACCCCTAGCCCAAACTGAGATATTAAC
OTD_TS2_4	ctacacgacgctcttccgatctTTACTCTTGAGATTGGAACGGGAAATC
OTD_TS2_5	ctacacgacgctcttccgatctCTGTCCTCTCCACATCCTCTCTG
OTD_TS2_6	ctacacgacgctcttccgatctTGCTGTCTAAACTTCCATTTCTCCA
OTD_TS2_7	ctacacgacgctcttccgatctGCAAAGCAAAAACCAGGGCCAG
OTD_TS2_8	ctacacgacgctcttccgatctGAGTTCCCAACCTTTTTGACACTAG
OTD_TS2_9	ctacacgacgctcttccgatctCAGAGCAGCATCTGCTATTTC
OTD_TS2_10	ctacacgacgctcttccgatctCTCCTGAGCAAATGGCAAAC
OTD_TS2_11	ctacacgacgctcttccgatctTCCAGGACCTGAGATTCCATT
OTD_TS2_12	ctacacgacgctcttccgatctGGGACAGAAACAAGCCCTTT
OTD_TS2_13	ctacacgacgctcttccgatctAAACAAATGAGAGAGCATACCTG
OTD_TS2_14	ctacacgacgctcttccgatctGAGTGCAGTGGCACGATCT
OTD_TS2_15	ctacacgacgctcttccgatctTCTCGTGGAAGTGAATGAGG
OTD_TS2_16	ctacacgacgctcttccgatctAATCACTGTGCTCTCCCT
VEGFA_TS3	ctacacgacgctcttccgatctGTCTTCGAGAGTGAGGACGTGTG
OTD_TS3_1	ctacacgacgctcttccgatctTGCAGTGAGGAGGTGGTTCTTG
OTD_TS3_2	ctacacgacgctcttccgatctTGTCACCACACAGTTACCACCTT
OTD_TS3_3	ctacacgacgctcttccgatctACACTGGTGGGCAGGAGCA
OTD_TS3_4	ctacacgacgctcttccgatctGTGTTCTCATGAGATCTGATGGT
OTD_TS3_5	ctacacgacgctcttccgatctTCTCCACCATGCCTCTGGCAT
OTD_TS3_6	ctacacgacgctcttccgatctCAGCCGGCTTTGTAGCTTTT
OTD_TS3_7	ctacacgacgctcttccgatctACTGCAGGCAAGCTGTCAAG
OTD_TS3_8	ctacacgacgctcttccgatctTGCCCCCAGGAGAGTGGCT
OTD_TS3_9	ctacacgacgctcttccgatctGCCAAGATCACGCCATTTGCACT
OTD_TS3_10	ctacacgacgctcttccgatctGCCAAGGGTATTCATAAATTCTTC
OTD_TS3_11	ctacacgacgctcttccgatctGACCATATATAGAGACCATATATATAGAC
OTD_TS3_12	ctacacgacgctcttccgatctTGACGAGGACCCTGGAG
OTD_TS3_13	ctacacgacgctcttccgatctGGCAATTATTCTGTCAAGATTTTAAAGG
OTD_TS3_15	ctacacgacgctcttccgatctACCGAGGAGGTGGGGGTGT
OTD_TS3_16	ctacacgacgctcttccgatctATGGAATAGTTTATTCGTGTGTGT
OTD_TS3_17	ctacacgacgctcttccgatctCATTCTGCTGTTGTGAGACTTAT
OTD_TS3_18	ctacacgacgctcttccgatctGACCACATCCAGTGCCTGA
OTD_TS3_19	ctacacgacgctcttccgatctGTCTCTTGGGAGTGGTTTTGAGCA

OTD_TS3_20	ctacacgacgctcttccgatctTTGGAATTTTGGGCTTTATATTTGA
OTD_TS3_21	ctacacgacgctcttccgatctTCACGTGATCCTCCCCTTTCAGAC
OTD_TS3_22	ctacacgacgctcttccgatctTCACTTGAGCTCACAGGTTTG
OTD_TS3_24	ctacacgacgctcttccgatctGAGTAGGAAAGGTATGTCTTGTTCCT
Target Site	**3p Oligo Sequence**
PMPCA	agacgtgtgctcttccgatctAGTGTACTGGGCCACAGATCTG
VEGFA_TS2	agacgtgtgctcttccgatctAGCCCCAGCTACCACCTCCT
OTD_TS2_1	agacgtgtgctcttccgatctACTATTGCACACCTGGGTCCT
OTD_TS2_2	agacgtgtgctcttccgatctAATATTCCTGGCCAAGTCGATTCC
OTD_TS2_3	agacgtgtgctcttccgatctGTGCTGATCTGTGGGTATCCTTTTTAG
OTD_TS2_4	agacgtgtgctcttccgatctTGCAACTTAACTTACGTGAAACTCCACA
OTD_TS2_5	agacgtgtgctcttccgatctATGGCCTATGGGTTGAAAACAGT
OTD_TS2_6	agacgtgtgctcttccgatctAACAGATGAGCCACAACCCTGTT
OTD_TS2_7	agacgtgtgctcttccgatctAGCTGTTTCTGCTGACTGTCGGT
OTD_TS2_8	agacgtgtgctcttccgatctAAATGTAATGTGCTCGAATCATCCT
OTD_TS2_9	agacgtgtgctcttccgatctGCACAATGATTGGTCATCTGTTA
OTD_TS2_10	agacgtgtgctcttccgatctGTTCTGGGCTGGACACTGAG
OTD_TS2_11	agacgtgtgctcttccgatctGACAAGCTGCCAGGAACAAA
OTD_TS2_12	agacgtgtgctcttccgatctTGGCACCATAGGCATAATCTG
OTD_TS2_13	agacgtgtgctcttccgatctGGGATGCTTGAAGATGGAAAG
OTD_TS2_14	agacgtgtgctcttccgatctTAGCCTGGCATAGTGGTGGT
OTD_TS2_15	agacgtgtgctcttccgatctCAGTCGACAGTGACACAATTC
OTD_TS2_16	agacgtgtgctcttccgatctCGGATAAGATGGTCTTGGATTG
VEGFA_TS3	agacgtgtgctcttccgatctTCTATTGGAATCCTGGAGTGACC
OTD_TS3_1	agacgtgtgctcttccgatctGAAGTCACCGACAACAACAAGCC
OTD_TS3_2	agacgtgtgctcttccgatctGGGAATCTAATGTATGGCATGG
OTD_TS3_3	agacgtgtgctcttccgatctCATTTCTCCTTTGAGGTTCATC
OTD_TS3_4	agacgtgtgctcttccgatctAGGCCTCGGGAAACTTACAATCATG
OTD_TS3_5	agacgtgtgctcttccgatctCAGAGATGTGTGGCTGTGTCA
OTD_TS3_6	agacgtgtgctcttccgatctGAACAAGCTGCTGGCTTTCC
OTD_TS3_7	agacgtgtgctcttccgatctACACACCCACACACCCTCAC
OTD_TS3_8	agacgtgtgctcttccgatctCTCATCAGTGCTGTGTTCGCCT
OTD_TS3_9	agacgtgtgctcttccgatctGGAAGAAGATCAACTAAGCAGTTATAAGC
OTD_TS3_10	agacgtgtgctcttccgatctGTATCTGTCTCTGTCTCTCTCTCC
OTD_TS3_11	agacgtgtgctcttccgatctAGTGCCATGCTATTTACATTCTGTCT
OTD_TS3_12	agacgtgtgctcttccgatctCTCGGGACAGGCAAGTT
OTD_TS3_13	agacgtgtgctcttccgatctGGCAACTTCCAACTCTTTATTTCCGA
OTD_TS3_15	agacgtgtgctcttccgatctCCTACCTGGTGGCCCTGTG
OTD_TS3_16	agacgtgtgctcttccgatctGAAGGAAGTTCTCAGGACACACAAG

OTD_TS3_17	agacgtgtgctcttccgatctGTACATTGACTGAAATAAGGATGGT
OTD_TS3_18	agacgtgtgctcttccgatctCTCGCATGGCACACTCAG
OTD_TS3_19	agacgtgtgctcttccgatctGAAAGTTCACTCTGGCCCCAGTG
OTD_TS3_20	agacgtgtgctcttccgatctAGTACCCAACAATGCCCATATC
OTD_TS3_21	agacgtgtgctcttccgatctGCTGAGGCTGGAGCATTGTTTGAG
OTD_TS3_22	agacgtgtgctcttccgatctCTCAAGCAATCCTCCCTCAA
OTD_TS3_24	agacgtgtgctcttccgatctAGTCAAAAGGTCATTTGCAGGGTC

Appendix 1.4 Deep Sequencing primers used in Chapter IV	
Target Site	**Sequence**
5p_DS_OT_B7_1	ctacacgacgctcttccgatctTGTAAAATGCACACAAGGTTCG
5p_DS_OT_B7_2	ctacacgacgctcttccgatctGACAGAGCTTTCAGGCCAACA
5p_DS_OT_B7_3	ctacacgacgctcttccgatctTCAGCCAACCGTAAGACTCAGA
5p_DS_OT_B7_4	ctacacgacgctcttccgatctACCAATTCAGGGATGGTGAGG
5p_DS_OT_B7_5	ctacacgacgctcttccgatctATTCCATCATGCTGCCTCCTT
5p_DS_OT_B7-10_1	ctacacgacgctcttccgatctTCCTTAGATTGTTTTCTTCCTCATGTT
5p_DS_OT_B7-10_2	ctacacgacgctcttccgatctAAAGGAGGACCAGAGGTGGACT
5p_DS_OT_B7-10_3	ctacacgacgctcttccgatctCCTCAAAAAGGAAGGAACTGGA
5p_DS_OT_B7-10_4	ctacacgacgctcttccgatctAATATTTGCTGACGGCCTTCTC
5p_DS_OT_B7-10_5	ctacacgacgctcttccgatctGCCTCCCAAAGTGTTGGATTT
5p_DS_OT_B9_1	ctacacgacgctcttccgatctCCTGGCTAATACTTTGGGGAGT
5p_DS_OT_B9_2	ctacacgacgctcttccgatctGTGGACTGCATCTCAGGAAAGA
5p_DS_OT_B9_3	ctacacgacgctcttccgatctAGCTGATCAATCCCCAATGTCT
5p_DS_OT_B9_4	ctacacgacgctcttccgatctGCACTGCTCTTTGGAAAACTACC
5p_DS_OT_B9_5	ctacacgacgctcttccgatctTGGAACACAGGTCTTCCTACCC
5p_DS_OT_B9_6	ctacacgacgctcttccgatctGAAGAGGTGGAAGGTGCTGCT
5p_DS_OT_B9_7	ctacacgacgctcttccgatctGCTGATCAATCCCCAATGTCTC
5p_DS_OT_B9_8	ctacacgacgctcttccgatctCGTGTGTGAAAGCCATGTTGA
5p_DS_OT_B9_9	ctacacgacgctcttccgatctCAGGACCACAGTCCTCTGCTG
5p_DS_OT_B9_10	ctacacgacgctcttccgatctTGGATTTTCCATTTGAAGCACA
5p_DS_OT_B10_1	ctacacgacgctcttccgatctCATGTGTTCATTGACATACAGGACA
5p_DS_OT_B10_2	ctacacgacgctcttccgatctTCAGGCTGCTTGACCTTTCTG
5p_DS_OT_B10_3	ctacacgacgctcttccgatctGCCACATTTTCTGAAGCATCC
5p_DS_OT_B10_4	ctacacgacgctcttccgatctGTGAGGGGAGATGGAATCGAG
5p_DS_OT_B10_5	ctacacgacgctcttccgatctGGGTGGTTTTCTCATCCCACT
5p_DS_OT_B11_1	ctacacgacgctcttccgatctCTCAGTGGCAGCAGATCCTGT
5p_DS_OT_B11_2	ctacacgacgctcttccgatctTCTCCATTTTCTAAGCGCAACA
5p_DS_OT_B11_3	ctacacgacgctcttccgatctAGCAGATATGGGACCTGAAAGG
5p_DS_OT_B11_4	ctacacgacgctcttccgatctACTCATGCTTGTTTTGGGAGGA
5p_DS_OT_B11_5	ctacacgacgctcttccgatctTTAGGTACCCCTGTGAATAAATCC
5p_DS_OT_B11_6	ctacacgacgctcttccgatctCCATTTGTGTCCTTCTCTCTCA
5p_DS_OT_B11_7	ctacacgacgctcttccgatctATGTGGCCAAGAGAGCACCT
5p_DS_OT_B11_8	ctacacgacgctcttccgatctCAGTCTGGTCTCAGCTTTGGTC
5p_DS_OT_B11_9	ctacacgacgctcttccgatctGCAGGGATGGAAAAGAGAAGAC
5p_DS_OT_B11_10	ctacacgacgctcttccgatctAGCCAAGGAGAAATTGAGGTGA
5p_DS_OT3-4	ctacacgacgctcttccgatctACACTGGTGGGCAGGAGCA
5p_DS_OT3-18	ctacacgacgctcttccgatctTGTCACCACACAGTTACCACCTT

5p_DS_OT3-20	ctacacgacgctcttccgatctTCATGGAAGAATGCAAAGGAGA
5p_DS_OT3-2	ctacacgacgctcttccgatctTGCAGTGAGGAGGTGGTTCTTG
5p_DS_OT3-3	ctacacgacgctcttccgatctTTCCTGTGGAACAACCAGACAC
5p_DS_OT3_Sp_new	ctacacgacgctcttccgatctTTTGAAGTCCGTGCTTGAATGA
5p_DS_OT3-21	ctacacgacgctcttccgatctGCCAGCACCCTTGACGTCTG
5p_DS_OTD3-6	ctacacgacgctcttccgatctACAAGCTGCTGGCTTTCCTAAG
5p_DS_OT3-1	ctacacgacgctcttccgatctACTGCAGGCAAGCTGTCAAG
5p_DS_OT3-17	ctacacgacgctcttccgatctGGAAGTGCCTTGCACAAATAGG
3p_DS_OT_B7_1	agacgtgtgctcttccgatctCGGACAGTGAGGAGGATAGGAC
3p_DS_OT_B7_2	agacgtgtgctcttccgatctCATTTGGGCTTGATGCTTGAG
3p_DS_OT_B7_3	agacgtgtgctcttccgatctTTTGCCAATCTGAAGGTCCAA
3p_DS_OT_B7_4	agacgtgtgctcttccgatctGGTGCCTTTCACACTGTGGATAC
3p_DS_OT_B7_5	agacgtgtgctcttccgatctACACTGAAGGAAGCCCTGCAC
3p_DS_OT_B7-10_1	agacgtgtgctcttccgatctGGCCATCAGTCATCTGCAATTT
3p_DS_OT_B7-10_2	agacgtgtgctcttccgatctACCATGGCTCTGATGATACCAA
3p_DS_OT_B7-10_3	agacgtgtgctcttccgatctAGCATTAGACTCTGGGCTGAGG
3p_DS_OT_B7-10_4	agacgtgtgctcttccgatctAAACGCCTTCTTTCTTCACCTG
3p_DS_OT_B7-10_5	agacgtgtgctcttccgatctTGGCCATGTTTAGAATTGAAGAGA
3p_DS_OT_B9_1	agacgtgtgctcttccgatctCCCCAGAGTCACTCAAACAACA
3p_DS_OT_B9_2	agacgtgtgctcttccgatctACACTCCAGGAGGCAGTTAAGG
3p_DS_OT_B9_3	agacgtgtgctcttccgatctGAGGATTTCATCATGCTGGTTG
3p_DS_OT_B9_4	agacgtgtgctcttccgatctGCAGGGGAGTTTGGTGTACATT
3p_DS_OT_B9_5	agacgtgtgctcttccgatctGTAAGGATCTTGCTCCCCACAG
3p_DS_OT_B9_6	agacgtgtgctcttccgatctCAACAGGGGCATGGAGAACT
3p_DS_OT_B9_7	agacgtgtgctcttccgatctGGATTTCATCATGCTGGTTGG
3p_DS_OT_B9_8	agacgtgtgctcttccgatctGGCCATGGTCAAAACCCTTAT
3p_DS_OT_B9_9	agacgtgtgctcttccgatctCTCCTGGCACTGCTGCTGCT
3p_DS_OT_B9_10	agacgtgtgctcttccgatctCCAGGAAGAGTTTGGTTTTCC
3p_DS_OT_B10_1	agacgtgtgctcttccgatctCCTTTCCACCTGTAAGCACAGA
3p_DS_OT_B10_2	agacgtgtgctcttccgatctCAAGAAGGGTGGTAGGGATGG
3p_DS_OT_B10_3	agacgtgtgctcttccgatctTTGTTCAATGAAAGAGAACCCACA
3p_DS_OT_B10_4	agacgtgtgctcttccgatctTCCTGCATCTGAACCCACATT
3p_DS_OT_B10_5	agacgtgtgctcttccgatctCTGCAGTGGACCAGTTGTGC
3p_DS_OT_B11_1	agacgtgtgctcttccgatctAAATGACTGAAGCCCAAACTCC
3p_DS_OT_B11_2	agacgtgtgctcttccgatctCCCAGCTGCTCTTTCTTTTCTC
3p_DS_OT_B11_3	agacgtgtgctcttccgatctCAGGGTCATTGGTGAGTCAGAG
3p_DS_OT_B11_4	agacgtgtgctcttccgatctTTTCCCAGCACACCATATAGGA
3p_DS_OT_B11_5	agacgtgtgctcttccgatctAACCCAGTGGTAATTGAGCAAA
3p_DS_OT_B11_6	agacgtgtgctcttccgatctCCCTGGCTATGTTTCTGTCGAT

3p_DS_OT_B11_7	agacgtgtgctcttccgatctCCCTTCCTTGCATTAATGGATTT
3p_DS_OT_B11_8	agacgtgtgctcttccgatctGTTACTTCTGGGCCCATTTCTG
3p_DS_OT_B11_9	agacgtgtgctcttccgatctTTAGACGGCTTCAAGGTCATGT
3p_DS_OT_B11_10	agacgtgtgctcttccgatctGGGAGGGAAGAGTCTGAGGAG
3p_DS_OT3-4	agacgtgtgctcttccgatctCATTTCTCCTTTGAGGTTCATC
3p_DS_OT3-18	agacgtgtgctcttccgatctGGGAATCTAATGTATGGCATGG
3p_DS_OT3-20	agacgtgtgctcttccgatctCCCCATGCTTTTCTTGTGATAG
3p_DS_OT3-2	agacgtgtgctcttccgatctGAAGTCACCGACAACAACAAGCC
3p_DS_OT3-3	agacgtgtgctcttccgatctACACAGAGACACCCCACACACT
3p_DS_OT3_Sp_new	agacgtgtgctcttccgatctCCCAACACCTACATCTACCCACA
3p_DS_OT3-21	agacgtgtgctcttccgatctCGAATTCCGTCCACACACAG
3p_DS_OTD3-6	agacgtgtgctcttccgatctGGTGAGGAGCAACGAGACGTTA
3p_DS_OT3-1	agacgtgtgctcttccgatctACACACCCACACACCCTCAC
3p_DS_OT3-17	agacgtgtgctcttccgatctGTTGGTTTGGAAGGCTGTCATT

Appendix 2.1 GUIDE-seq data in Chapter 2							
SpCas9^{WT}							
Name	**Sequence**	**Alignment to target**	**# of mismatches**	**chromosome**	**Strand**	**Start**	**Peak Score**
OTG2-1	CTGCCCCCCCACCCCGCCACTGG	CTG....C.....A.	5	chrX	+	149932511	133
OTG2-2	GCCCCCACCCACCCCGCCTCTGG	.C....AC.....	3	chr18	-	21359552	114
OTG2-3	GGGCCCCTCCACCCCGCCTCTGG	.GG.....	2	chr11	-	31817476	108
OTG2-4	ATTCCCCCCCACCCCGCCTCAGG	ATT....C.....	4	chr2	+	242214589	88
OTG2-5	CTACCCCTCCACCCCGCCTCCGG	CTA.....	3	chr5	+	6715101	76
OTG2-6	TGCCCCCCCCACCCCACCTCTGG	TG.....C.....A....	4	chr16	-	56963422	60
OTG2-7	TACCCCCCACACCCCGCCTCTGG	T......CA.....	3	chr17	-	4358745	54
TS2	GACCCCCTCCACCCCGCCTCCGG	.....	0	chr6	-	43738555	47
OTG2-8	GACCCCCCCCACCCCGCCCCCGG	.......C.....C.	2	chr15	-	33286100	45
OTG2-9	TGCCCCTCCCACCCCGCCTCTGG	TG....TC.....	4	chr17	-	40044750	45
OTG2-10	ACACCCCCCCACCCCGCCTCAGG	ACA....C.....	4	chr9	-	103599642	40
OTG2-11	TGCCCCCCCCACCCCGCCCCCGG	TG.....C.....C.	4	chrX	+	129040623	37
OTG2-12	CCGCCCCTCCACCCCGCCACTGG	CCG.....A.	4	chr17	-	55740522	28
OTG2-13	ATCCCCCTCCACCCCCACCCCTGG	AT.....A..C.	4	chr11	+	12308917	27
OTG2-14	GACCCCTCCCACCCCCGACTCCGG	......TC.....A...	3	chr9	+	27338857	25
OTG2-15	GCTTCCCTCCACCCCCGCATCCGG	.CTT.....A..	4	chr11	+	71948787	25
OTG2-16	CCTCCCCCACACCCCGCATCCGG	CCT....CA.....A..	6	chr1	+	151031868	24
OTG2-17	TACCCCCCCCACCCCCGCCACAGG	T......C.....A.	3	chr11	+	13948353	22
OTG2-18	CCCCCCCCCCCCCCCCGCCTCAGG	CC.....C..C.	4	chr13	-	100546982	22
OTG2-19	AGGCCCCCACACCCCGCCTCAGG	AGG....CA.....	5	chr4	+	1496240	21
OTG2-20	CTCCCCCTCCACCCCCACCTCCAG	CT.....A....	3	chr6	-	31462820	21
OTG2-21	CTCCCCCGCCACCCCGCCCCAGG	CT.....G.....C.	4	chr17	-	16954845	18
OTG2-22	AGCCCCCACCTCCCCGCCTCGGG	AG.....A..T.	4	chr22	+	43684478	18

OTG2-23	GTACCCCACCACCC CGCCCCAGG	.TA....A.... ......C.	4	chr8	+	14482 2944	18
OTG2-24	ACCCCCCCCCGCCC CGCCCCCGG	AC.....C..G. ......C.	5	chr11	+	73458 481	17
OTG2-25	CCACCCCCCCACCC CGCCCCTGG	CCA....C.... ......C.	5	chr20	-	10913 954	16
OTG2-26	CTCCCCACCCACCC CGCCTCAGG	CT....AC....	4	chr4	+	38537 610	15
OTG2-27	CCACCCCCCCACCC CGCCCCAGG	CCA....C.... ......C.	5	chr2	+	17057 3333	15
OTG2-28	GTCCCCCTCCTCCC CACCTCCGG	.T........T. ...A....	3	chrX	+	15283 7111	14
OTG2-29	GACTCCCTCCGCCC CGCTTCCAG	...T......G. .....T..	3	chr5	+	17958 8285	13
OTG2-30	CTACCCCTCCACCC CGACTCGGA	CTA......... ....A...	4	chr18	+	55101 343	12
OTG2-31	CCCCCCCCCCACCC CGCCCCCGG	CC.....C.... ......C.	4	chr2	-	12924 4195	7
OTG2-32	GCCCCCCACCACCC CACCTCGGG	.C.....A.... ...A....	3	chr19	+	13122 171	6
OTG2-33	CCCCCCCACCACCC CGCCCCGGG	CC.....A.... ......C.	4	chr8	-	18041 539	6
OTG2-34	GACCCCCCCCACCC CACCCCAGG	.......C.... ...A..C.	3	chr6	+	26470 606	6
OTG2-35	CCCCACCCCCACCC CGCCTCAGG	CC..A..C....	4	chr10	-	11629 4249	5
OTG2-36	TACCCCCTCCACCC CCCTCCAGG	T.......... ...C.TC.	4	chr17	+	30391 257	4
OTG2-37	GTCCCCTCCCACCC CGCCTCCAG	.T....TC....	3	chr14	+	90403 304	4
OTG2-38	GACCCCTCACACCC CGCCCCTGG	......TCA... ......C.	4	chr7	-	95319 670	4
OTG2-39	CCCCCACCCCACCC CGCCTCCAG	CC...A.C....	4	chr9	+	13066 3069	4
OTG2-40	GACCCTCCCCACCC CACCCCTGG	.....T.C.... ...A..C.	4	chr1	+	22609 3591	4
OTG2-41	CATACCCCCCACCC CGCCCCGGG	C.TA...C.... ......C.	5	chr4	-	88419 10	3
TS3	GGTGAGTGAGTGTG TGCGTGTGG	....	0	chr6	+	43737 454	64
OTG3-1	TGTGGGTGAGTGTG TGCGTGAGG	T...G.....	2	chr5	+	11543 4659	38
OTG3-2	AGTGAGTGAGTGTG TGTGTGGGG	A.......... ....T...	2	chr14	-	65102 435	26
OTG3-3	GGTGAGTGAGTGTG TGTGTGAGG	....T...	1	chr14	+	10602 9015	18
OTG3-4	AGAGAGTGAGTGTG TGCATGAGG	A.A........ .....A..	3	chr5	-	89440 962	7
OTG3-5	AGTGAGTGAGTGTG AGTGCGGGG	A.......... ..A.T.C.	4	chr8	-	67579 421	6
OTG3-6	TGTGAGTGAGTGTG TGTGTGTGA	T.......... ....T...	2	chr8	-	22932 496	5

SpCas9^{MT3}-ZFP							
Name	**Sequence**	**Alignment to target**	**# of mismatches**	**chromosome**	**Strand**	**Start**	**Peak Score**
OTG2-3	GGGCCCCTCCACCCCGCCTCTGG	.GG.....	2	chr11	-	31817476	26
OTG2-5	CTACCCCTCCACCCCGCCTCCGG	CTA.....	3	chr5	+	6715101	7
TS2	GACCCCCTCCACCCCGCCTCCGG	....	0	chr6	-	43738555	27
OTG2-10	ACACCCCCCCACCCCGCCTCAGG	ACA....C....	4	chr9	-	103599642	2
OTG2-20	CTCCCCCTCCACCCCACCTCCAG	CT..... ...A....	3	chr6	-	31462820	2
TS3	GGTGAGTGAGTGTGTGCGTGTGG	....	0	chr6	+	43737454	53
OTG3-2	AGTGAGTGAGTGTGTGTGTGGGG	A..... ....T...	2	chr14	-	65102435	1

Appendix 2.2 GUIDE-seq data in Chapter 3

SpCas9^{WT}

Name	Sequence	Alignment to target	# of mismatch	Chromosome	Strand	Start position	Peak Score
TS2	GACCCCCTCCACC CCGCCTCCGG	....	0	chr6	-	43738556	128
OTD_TS2_1	CTGCCCCCCCACC CCGCCACTGG	CTG....C.... .....A.	5	chrX	+	149932512	448
OTD_TS2_2	GGGCCCCTCCACC CCGCCTCTGG	.GG.....	2	chr11	-	31817477	353
OTD_TS2_3	ATTCCCCCCCACC CCGCCTCAGG	ATT....C....	4	chr2	+	242214590	354
OTD_TS2_4	CTACCCCTCCACC CCGCCTCCGG	CTA.....	3	chr5	+	6715102	323
OTD_TS2_5	GACCCCCCCCACC CCGCCCCCGG	.....C.... .....C.	2	chr15	-	33286101	206
OTD_TS2_6	ACACCCCCCCACC CCGCCTCAGG	ACA....C....	4	chr9	-	103599643	202
OTD_TS2_7	TACCCCCCACACC CCGCCTCTGG	T.....CA...	3	chr17	-	4358746	170
OTD_TS2_8	TGCCCCCCCCACC CCACCTCTGG	TG.....C.... ...A....	4	chr16	-	56963423	146
OTD_TS2_9	TGCCCCCTCCACC CCACCTCTGA	TG..... ...A....	3	chr6	+	110521018	17
	CTACCCCTCCACC CCGACTCGGA	CTA..... ....A...	4	chr18	+	55101344	14
	GCCCCCACCCACC CCGCCTCTGG	.C....AC....	3	chr18	-	21359553	174
	CTCCCCCTCCACC CCACCTCCAG	CT..... ...A....	3	chr6	-	31462821	43
	CCCCCCCCCCCCC CCGCCTCAGG	CC.....C..C.	4	chr13	-	100546983	149
	CTCCCCACCCACC CCGCCTCAGG	CT....AC....	4	chr4	+	38537611	67
	CCCCCCCCCCACC CCGCCCTCGG	CC.....C.... .....CT	5	chr11	+	19798641	8
	TACCCCCCCCACC CCGCCACAGG	T.....C.... .....A.	3	chr11	+	13948354	133
	TGCCCCCCCCACC CCGCCCCCGG	TG.....C.... .....C.	4	chrX	+	129040624	120
	CCGCCCCTCCACC CCGCCACTGG	CCG..... .....A.	4	chr17	-	55740523	108
	TGCCCCTCCCACC CCGCCTCTGG	TG....TC....	4	chr17	-	40044751	97
	GACCCCTCCCACC CCGACTCCGG	.....TC.... ....A...	3	chr9	+	27338858	92
	GTACCCCACCACC CCGCCCCAGG	.TA....A.... .....C.	4	chr8	+	144822945	84
	GCTTCCCTCCACC CCGCATCCGG	.CTT..... .....A..	4	chr11	+	71948788	77

	CCCCCCCCCCACC CCGCCCCCGG	CC.....C.... ......C.	4	chr2	-	12924 4196	67
	ATCCCCCTCCACC CCACCCCTGG	AT..... ...A..C.	4	chr11	+	12308 918	65
	ACCCCCCCCCGCC CCGCCCCCGG	AC.....C..G. ......C.	5	chr11	+	73458 481	58
	ACCCCCCCCCCCC CCGCCCCCGG	AC.....C..C. ......C.	5	chr2	+	24857 577	47
	CAACCCCCCCACC CCGCTTCAGG	C.A....C.... .....T..	4	chr3	+	14039 8784	47
	CCACCCCCCCACC CCGCCCCTGG	CCA....C.... ......C.	5	chr20	-	10913 955	41
	CACTCCCCCCACC CCGCCCCAGG	C..T...C.... ......C.	4	chr9	-	12613 8175	41
	GGACCCTCCCACC CCACCTCAAG	.GA...TC.... ...A....	5	chr13	-	10062 1456	40
	AGGCCCCCACACC CCGCCTCAGG	AGG....CA...	5	chr4	+	14962 41	39
	CCACCCCCCCACC CCGCCCCAGG	CCA....C.... ......C.	5	chr2	+	17057 3334	35
	GACCCCCCCCACC CCACCCCAGG	.....C.... ...A..C.	3	chr6	+	26470 607	33
	CTCCCCCGCCACC CCGCCCCAGG	CT.....G.... ......C.	4	chr17	-	16954 846	27
	CATACCCCCCACC CCGCCCCGGG	C.TA...C.... ......C.	5	chr4	-	88419 11	27
	GACCCCTCCCTCC CCACCTCAGG	......TC..T. ...A....	4	chr1	-	51442 186	26
	AAGACCCCCCACC CCGCCCCAGG	A.GA...C.... ......C.	5	chr19	+	45952 038	26
	CCCCCCCCCCCCC CCGCCTCCGG	CC.....C..C.	4	chr22	+	50884 786	26
	CCCCACCCCCACC CCGCCTCAGG	CC..A..C....	4	chr10	-	11629 4250	25
	CTACCCCCACTCC CCGCCTCCGG	CTA....CA.T.	6	chr10	+	10282 1501	24
	TCCACCCCCCACC CCGCCCCGGG	TC.A...C.... ......C.	5	chr7	+	10493 3939	24
	GACCCCTCACACC CCGCCCCTGG	......TCA... ......C.	4	chr7	-	95319 671	22
	GTCCCTCCCCACC CCGCCTTGGG	.T...T.C.... .....T	4	chr22	+	40102 384	21
	GCCCCCCACCACC CCACCTCGGG	.C.....A.... ...A....	3	chr19	+	13122 172	20
	CGCCCTCCCCACC CCGCCTCCGG	CG...T.C....	4	chr10	+	13514 9931	19
	AGACCCCCCCACC CCACCCCAGG	AGA....C.... ...A..C.	6	chr20	-	35991 127	19
	GACCCCACCCACC CCGCCGCAGG	......AC.... ......G.	3	chr12	+	13252 9350	17
	AGCCCCCCCCTCC CCGCCCCAGG	AG.....C..T. ......C.	5	chr14	+	10460 1192	16

	CTCCCCCCCCACC CCGTCCCCGG	CT.....C.... ....T.C.	5	chr2	-	12907 7573	16
	CACCCCCCCCCCC CCGCCCCCGG	C......C..C. ......C.	4	chr17	+	79881 137	15
	GGACCCCCCCCCC CCGCCCCCGG	.GA....C..C. ......C.	5	chr3	-	13187 297	15
	GCCCCCCCCCACC CCGCAGCTGG	.C.....C.... .....AG.	4	chr7	-	11179 73	15
	GGACCCCGACGCC CCGCCTCAGG	.GA....GA.G.	5	chr3	-	31574 996	14
	CAGTCCCCCCACC CCACCTCTGG	C.GT...C.... ...A....	5	chr10	+	72538 201	13
	CATCCCCCCCACC CCACCCCGGG	C.T....C.... ...A..C.	5	chr11	-	46141 849	13
	ACACCCCCCCACC CCACCCCAGG	ACA....C.... ...A..C.	6	chr2	-	22153 5826	12
	AGCCCCCACCTCC CCGCCTCGGG	AG.....A..T.	4	chr22	+	43684 479	12
	CTCACCCCCCACC CCACCTCTGG	CT.A...C.... ...A....	5	chr11	-	26862 43	11
	ACCCCCCCCCCAC CCGCCCCCGG	AC.....C..CA ......C.	6	chr6	-	27382 04	11
	GACCCTGTCCACC CCACCTCAGG	.....TG..... ...A....	3	chr1	+	87532 1	10
	GACCCGCCCCGCC CCGCCTCTGG	.....G.C..G.	3	chr1	+	11714 529	10
	GACCCTCCCCACC CCACCCCTGG	.....T.C.... ...A..C.	4	chr1	+	22609 3591	10
	AGGCCCCCCCGCC CCGCCTCAGG	AGG....C..G.	5	chr11	-	37466 1	10
	CCCCCCCACCACC CCGCCCCGGG	CC.....A.... ......C.	4	chr8	-	18041 540	10
	GGCCCTCTCCACT CCACCTCAGG	.G...T...... T..A....	4	chr1	-	17873 8721	9
	AACCTCCCCCACC CCACCCCAGG	A...T..C.... ...A..C.	5	chr17	-	46103 841	9
	AGACCCCCCCACC CCACCCCAGG	AGA....C.... ...A..C.	6	chr21	+	47440 541	9
	GCCGCCCCCCACT CCGCCTCCGG	.C.G...C.... T.......	4	chr4	-	14376 7367	9
	CCCCCCCCCCACC CCGCCCCGAG	CC.....C.... ......C.	4	chr12	+	22487 644	8
	CTCCCCCCCCTCC CCGCCTCGGG	CT.....C..T.	4	chr5	+	13902 8240	8
	CCGCCCCCCCACC CCGCCGCCGG	CCG....C.... ......G.	5	chr5	+	17087 8053	8
	CCCCCCTCCCACC CCGCCTCTAG	CC....TC....	4	chr9	+	37465 348	8
	CCCCCCCCCCGCC CCGCCTCCAG	CC.....C..G.	4	chr10	+	10272 9240	7
	CACCCCCTCCCCT CCGCCTCAGG	C.........C. T.......	3	chr11	-	61321 433	7

	CATTCCCCCCACC CCACCTCAGG	C.TT...C.... ...A....	5	chr12	-	26025089	7
	CCCCCCCCCCCCC CCGCCCCAGG	CC.....C..C. ......C.	5	chr13	-	52341910	7
	GTCCCCTCCCACC CCGCCTCCAG	.T....TC....	3	chr14	+	90403305	7
	GAGCTCCCCCACC CCGCCCCGGG	..G.T..C.... ......C.	4	chr19	-	14142969	7
	AGCCCCCCCCACC CCGCCCCTGG	AG.....C.... ......C.	4	chr19	-	28341692	7
	AACACCCCCCACC CCACCCCGGG	A..A...C.... ...A..C.	5	chr19	-	51915513	7
	CCCCCCCCCCCCC CCGCCCCCGG	CC.....C..C. ......C.	5	chr4	+	54259934	7
	GTCCCCCTCCTCC CCACCTCCGG	.T........T. ...A....	3	chrX	+	152837112	7
	AACCCCACCCACC CCATCTCAGG	A.....AC.... ...AT...	5	chr1	-	237195675	6
	CGGACCCCCCACC CCGCCCCCAG	CGGA...C.... ......C.	6	chr8	+	66934423	6
	TACCCCCTCCACC CCGCTCCAGG	T.......... .....TC.	3	chr17	+	28975224	5
	AAGCCCCCCCACC CCGCCCCCGA	A.G....C.... ......C.	4	chr19	-	15311180	5
	GGCCCCACCCACC CCGCCGCAGG	.G....AC.... ......G.	4	chr19	-	30335983	5
	CCGCCCCCCCATC CCGCCCCAGG	CCG....C...T ......C.	6	chr20	-	62575811	5
	GCCTCCCCCCACC CAGCCTCGGG	.C.T...C.... ..A.....	4	chr6	-	36853819	5
	GGCCCCCTCCTCC TCGCCTCTGG	.G........T. .T......	3	chr1	+	9884874	4
	GACCCCCTCCACC CCACTCCTGG	....A.TC.	3	chr1	+	19506 1768	4
	GGTCCCCCCCACC CCACCTCCAG	.GT....C.... ...A....	4	chr11	-	7099172	4
	AACCCCACCCACC CCGCCCTGGG	A.....AC.... ......CT	5	chr11	-	17662650	4
	GACCCCCCCCACC CCGCCCCGAG	.......C.... ......C.	2	chr11	-	10143 7790	4
	CCGCCCCTGCACC CAGCCTCCGG	CCG.....G... ..A.....	5	chr12	+	13328 7132	4
	ACTCCCCTCCACC CCGGCTCGGG	ACT......... ....G...	4	chr17	-	62167048	4
	CTCTCCCCCCTCC CCGCCCCCGG	CT.T...C..T. ......C.	6	chr17	+	80013580	4
	TGCCCCCTCCACC CCGACCCTGG	TG......... ....A.C.	4	chr19	-	49908283	4
	CGGCCTCCCCACC CCGCCCCCGG	CGG..T.C.... ......C.	6	chr7	-	10013 7234	4
	AACCCCACCCACC CCACCCCCGG	A.....AC.... ...A..C.	5	chr8	-	10451 2392	4

	GGGCCCCCCTACC CCGCCCCCGG	.GG....C.T.. ......C.	5	chr9	+	13898 7551	4
	TCTCCACCCCACC CCGCCCCTGG	TCT..A.C.... ......C.	6	chr1	-	78245 430	3
	TGGCCCCTCCGCC CCACCTCTGG	TGG.......G. ...A....	5	chr1	+	12942 56	3
	TCCCCCCCCCACC CCGCTTTGGG	TC.....C.... .....T.T	5	chr11	-	64012 937	3
	GCCCACCCCCCCC CCCCCCCCAG	.C..A..C..C. ...C..C.	6	chr11	-	64604 177	3
	GCCACCCACCACC CCACCTCAGG	.C.A...A.... ...A....	4	chr12	+	57604 316	3
	GAGCCACTGCACC CAGCCAACAG	..G..A..G... ..A...AA	6	chr13	+	76825 276	3
	AGTAGCCCCCACC CCGCCTCGGG	AGTAG..C....	6	chr16	-	69166 705	3
	TTTCCCCCCCACC CCAACTCAGG	TTT....C.... ...AA...	6	chr17	+	71474 64	3
	AACCCACCCCACC CCACCTCGGA	A....A.C.... ...A....	4	chr19	-	55362 87	3
	TTCTCCCTCCTCC CCGCCTCGGG	TT.T......T.	4	chr19	-	42806 795	3
	GACATCCCCCACC CCGCCCCCAG	...AT..C.... ......C.	4	chr20	+	45571 881	3
	CTTCCCCCCCGCC CCGCCCCGGG	CTT....C..G. ......C.	6	chr4	+	43892 69	3
	GTCCCCCTCCAGC CCGCCCCTGG	.T.........G ......C.	3	chr5	-	77269 272	3
	TGCACCCCCCACC CCGCCCCTGG	TG.A...C.... ......C.	5	chr5	+	13286 0359	3
	GCCCCGCCCCACC CCACCCCCGG	.C...G.C.... ...A..C.	5	chr6	-	13814 822	3
	AACCACCCCCACC CCACCCCAGG	A...A..C.... ...A..C.	5	chr7	+	50765 512	3
	AAGCCCCCCCACC CCGCCCCGGG	A.G....C.... ......C.	4	chr9	-	14026 3439	3
	CCCCCACCCCACC CCGCCTCCAG	CC...A.C....	4	chr9	+	13066 3070	3
	GTCCTCCACCACC CCGCCTCTGG	.T..T..A....	3	chrX	-	11886 5477	3
	CTTTCCCTCCACC CAGCCTCTGG	CTTT........ ..A.....	5	chr15	+	85828 607	2
	AGGCCCTCCCACC CCGCATCAGG	AGG...TC.... .....A..	6	chr16	+	67289 431	2
	CCTGCCCCCCACC CCGCCCCAGG	CCTG...C.... ......C.	6	chr19	+	14943 14	2
	CCTCCCCTCCACC CCGCTCCTGG	CCT......... .....TC.	5	chr2	-	23334 5192	2
	CTCCCTCCCCACC CCACCTCTGG	CT...T.C.... ...A....	5	chr2	+	11699 948	2
	GACACACCCCACC CCACCTCAGG	...A.A.C.... ...A....	4	chr2	+	12744 759	2

	CACCCCCCCCACC CCGTCTTGGG	C.....C.... ....T..T	4	chr2	+	23906 1941	2
	CTTCCCCCACACC CCGCCCCAGG	CTT....CA... .....C.	6	chr3	-	31458 652	2
	TCCCCCCCCCACC CCACCCCCGG	TC.....C.... ...A..C.	5	chr3	+	45348 439	2
	ACCCCACCCCACC CCACCCCAGG	AC...A.C.... ...A..C.	6	chr6	-	14057 990	2
	GTACCCCTCCACC CAGCCCCAGG	.TA..... ..A...C.	4	chr8	+	61544 623	2
SpCas9^{MT3}-ZFP							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS2	GACCCCCTCCACC CCGCCTCCGG	....	0	chr6	-	43738 556	119
OTD_TS2_2	GGGCCCCTCCACC CCGCCTCTGG	.GG.....	2	chr11	-	31817 477	156
OTD_TS2_3	ATTCCCCCCCACC CCGCCTCAGG	ATT....C....	4	chr2	+	24221 4590	4
OTD_TS2_4	CTACCCCTCCACC CCGCCTCCGG	CTA.....	3	chr5	+	67151 02	37
OTD_TS2_6	ACACCCCCCCACC CCGCCTCAGG	ACA....C....	4	chr9	-	10359 9643	5
OTD_TS2_7	TACCCCCCACACC CCGCCTCTGG	T......CA...	3	chr17	-	43587 46	7
OTD_TS2_10	GAGCCCCTGGGGG CCGCCTCTGA	..G.....GGGG G.....	6	chr8	-	37462 351	2
OTD_TS2_11	CAGCCCCTCCTCC AGGCCCCCAG	C.G.....T. .AG...C.	6	chr15	+	78149 388	2
	CTACCCCTCCACC CCGACTCGGA	CTA..... ....A...	4	chr18	+	55101 344	8
	CTCCCCCTCCACC CCACCTCCAG	CT..... ...A....	3	chr6	-	31462 821	2
	AAGCCCCCCCACC CCGCCCCCGA	A.G....C.... ......C.	4	chr19	-	15311 180	7
	GATCCCGCCCTTC CCACCTCTGG	..T...GC..TT ...A....	6	chr1	-	28431 632	2
	CCCCCCCCCCCCC CCGCCCCGGG	CC.....C..C. ......C.	5	chr2	+	22373 1481	2
Split-SpCas9^{WT}-NLS							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS2	GACCCCCTCCACC CCGCCTCCGG	....	0	chr6	-	43738 556	151
OTD_TS2_1	CTGCCCCCCCACC CCGCCACTGG	CTG....C.... ......A.	5	chrX	+	14993 2512	281
OTD_TS2_2	GGGCCCCTCCACC CCGCCTCTGG	.GG.....	2	chr11	-	31817 477	264

OTD_TS2_3	ATTCCCCCCCACC CCGCCTCAGG	ATT....C....	4	chr2	+	24221 4590	249
OTD_TS2_4	CTACCCCTCCACC CCGCCTCCGG	CTA.....	3	chr5	+	67151 02	271
OTD_TS2_5	GACCCCCCCCACC CCGCCCCCGG	.......C.... ......C.	2	chr15	-	33286 101	176
OTD_TS2_6	ACACCCCCCCACC CCGCCTCAGG	ACA....C....	4	chr9	-	10359 9643	168
OTD_TS2_7	TACCCCCCACACC CCGCCTCTGG	T......CA...	3	chr17	-	43587 46	155
OTD_TS2_8	TGCCCCCCCCACC CCACCTCTGG	TG.....C.... ...A....	4	chr16	-	56963 423	97
OTD_TS2_9	TGCCCCCTCCACC CCACCTCTGA	TG..... ...A....	3	chr6	+	11052 1018	15
OTD_TS2_12	AACCCCTTCTTCC CAGCCTCGGA	A.....T..TT. ..A.....	5	chr6	-	63921 631	15
OTD_TS2_13	AACCCCTTCTTCC CAGCCTCGGA	A.....T..TT. ..A.....	5	chr20	+	13561 41	2
	CTACCCCTCCACC CCGACTCGGA	CTA..... ....A...	4	chr18	+	55101 344	9
	GCCCCCACCCACC CCGCCTCTGG	.C....AC....	3	chr18	-	21359 553	163
	CTCCCCCTCCACC CCACCTCCAG	CT..... ...A....	3	chr6	-	31462 821	25
	CCCCCCCCCCCCC CCGCCTCAGG	CC.....C..C.	4	chr13	-	10054 6983	144
	CTCCCCACCCACC CCGCCTCAGG	CT....AC....	4	chr4	+	38537 611	37
	CCCCCCCCCCACC CCGCCCTCGG	CC.....C.... ......CT	5	chr11	+	19798 641	5
	TACCCCCCCCACC CCGCCACAGG	T......C.... ......A.	3	chr11	+	13948 354	73
	TGCCCCCCCCACC CCGCCCCCGG	TG.....C.... ......C.	4	chrX	+	12904 0624	108
	CCGCCCCTCCACC CCGCCACTGG	CCG..... ......A.	4	chr17	-	55740 523	76
	TGCCCCTCCCACC CCGCCTCTGG	TG....TC....	4	chr17	-	40044 751	90
	GACCCCTCCCACC CCGACTCCGG	......TC.... ....A...	3	chr9	+	27338 858	89
	GTACCCCACCACC CCGCCCCAGG	.TA....A.... ......C.	4	chr8	+	14482 2945	66
	GCTTCCCTCCACC CCGCATCCGG	.CTT..... .....A..	4	chr11	+	71948 788	73
	CCCCCCCCCCACC CCGCCCCCGG	CC.....C.... ......C.	4	chr2	-	12924 4196	46
	ATCCCCCTCCACC CCACCCCTGG	AT..... ...A..C.	4	chr11	+	12308 918	77
	ACCCCCCCCCGCC CCGCCCCCGG	AC.....C..G. ......C.	5	chr11	+	73458 481	69
	ACCCCCCCCCCCC CCGCCCCCGG	AC.....C..C. ......C.	5	chr2	+	24857 577	17

	CAACCCCCCCACC CCGCTTCAGG	C.A....C.... .....T..	4	chr3	+	14039 8784	57
	CCACCCCCCCACC CCGCCCCTGG	CCA....C.... ......C.	5	chr20	-	10913 955	20
	CACTCCCCCCACC CCGCCCCAGG	C..T...C.... ......C.	4	chr9	-	12613 8175	116
	GGACCCTCCCACC CCACCTCAAG	.GA...TC.... ...A....	5	chr13	-	10062 1456	8
	AGGCCCCCACACC CCGCCTCAGG	AGG....CA...	5	chr4	+	14962 41	54
	CCACCCCCCCACC CCGCCCCAGG	CCA....C.... ......C.	5	chr2	+	17057 3334	23
	GACCCCCCCCACC CCACCCCAGG	.......C.... ...A..C.	3	chr6	+	26470 607	21
	CTCCCCCGCCACC CCGCCCCAGG	CT.....G.... ......C.	4	chr17	-	16954 846	19
	CATACCCCCCACC CCGCCCCGGG	C.TA...C.... ......C.	5	chr4	-	88419 11	42
	GACCCCTCCCTCC CCACCTCAGG	......TC..T. ...A....	4	chr1	-	51442 186	14
	AAGACCCCCCACC CCGCCCCAGG	A.GA...C.... ......C.	5	chr19	+	45952 038	20
	CCCCCCCCCCCCC CCGCCTCCGG	CC.....C..C.	4	chr22	+	50884 786	30
	CCCCACCCCCACC CCGCCTCAGG	CC..A..C....	4	chr10	-	11629 4250	13
	CTACCCCCACTCC CCGCCTCCGG	CTA....CA.T.	6	chr10	+	10282 1501	22
	TCCACCCCCCACC CCGCCCCGGG	TC.A...C.... ......C.	5	chr7	+	10493 3939	20
	GACCCCTCACACC CCGCCCCTGG	......TCA... ......C.	4	chr7	-	95319 671	20
	GTCCCTCCCCACC CCGCCTTGGG	.T...T.C.... ....T	4	chr22	+	40102 384	10
	GCCCCCCACCACC CCACCTCGGG	.C.....A.... ...A....	3	chr19	+	13122 172	14
	CGCCCTCCCCACC CCGCCTCCGG	CG...T.C....	4	chr10	+	13514 9931	29
	AGACCCCCCCACC CCACCCCAGG	AGA....C.... ...A..C.	6	chr20	-	35991 127	16
	GACCCCACCCACC CCGCCGCAGG	......AC.... ......G.	3	chr12	+	13252 9350	5
	AGCCCCCCCCTCC CCGCCCCAGG	AG.....C..T. ......C.	5	chr14	+	10460 1192	12
	CTCCCCCCCCACC CCGTCCCCGG	CT.....C.... ....T.C.	5	chr2	-	12907 7573	4
	GGACCCCCCCCCC CCGCCCCCGG	.GA....C..C. ......C.	5	chr3	-	13187 297	12
	GCCCCCCCCCACC CCGCAGCTGG	.C.....C.... .....AG.	4	chr7	-	11179 73	4
	GGACCCCGACGCC CCGCCTCAGG	.GA....GA.G.	5	chr3	-	31574 996	24

	CAGTCCCCCCACC CCACCTCTGG	C.GT...C.... ...A....	5	chr10	+	72538201	35
	CATCCCCCCCACC CCACCCCGGG	C.T....C.... ...A..C.	5	chr11	-	46141849	12
	ACACCCCCCCACC CCACCCCAGG	ACA....C.... ...A..C.	6	chr2	-	221535826	6
	AGCCCCCACCTCC CCGCCTCGGG	AG.....A..T.	4	chr22	+	43684479	18
	CTCACCCCCCACC CCACCTCTGG	CT.A...C.... ...A....	5	chr11	-	2686243	5
	ACCCCCCCCCCAC CCGCCCCCGG	AC.....C..CA ......C.	6	chr6	-	2738204	5
	GACCCTGTCCACC CCACCTCAGG	.....TG..... ...A....	3	chr1	+	875321	10
	GACCCGCCCCGCC CCGCCTCTGG	.....G.C..G.	3	chr1	+	11714529	7
	GACCCTCCCCACC CCACCCCTGG	.....T.C.... ...A..C.	4	chr1	+	226093591	3
	AGGCCCCCCCGCC CCGCCTCAGG	AGG....C..G.	5	chr11	-	374661	24
	CCCCCCCACCACC CCGCCCCGGG	CC.....A.... ......C.	4	chr8	-	18041540	13
	AACCTCCCCCACC CCACCCCAGG	A...T..C.... ...A..C.	5	chr17	-	46103841	8
	AGACCCCCCCACC CCACCCCAGG	AGA....C.... ...A..C.	6	chr21	+	47440541	8
	GCCGCCCCCCACT CCGCCTCCGG	.C.G...C.... T.......	4	chr4	-	143767367	6
	CCCCCCCCCCACC CCGCCCCGAG	CC.....C.... ......C.	4	chr12	+	22487644	6
	CTCCCCCCCCTCC CCGCCTCGGG	CT.....C..T.	4	chr5	+	139028240	5
	CCGCCCCCCCACC CCGCCGCCGG	CCG....C.... ......G.	5	chr5	+	170878053	7
	CCCCCCTCCCACC CCGCCTCTAG	CC....TC....	4	chr9	+	37465348	10
	CCCCCCCCCCGCC CCGCCTCCAG	CC.....C..G.	4	chr10	+	102729240	8
	CACCCCCTCCCCT CCGCCTCAGG	C.........C. T.......	3	chr11	-	61321433	5
	CATTCCCCCCACC CCACCTCAGG	C.TT...C.... ...A....	5	chr12	-	26025089	5
	CCCCCCCCCCCCC CCGCCCCAGG	CC.....C..C. ......C.	5	chr13	-	52341910	2
	GTCCCCTCCCACC CCGCCTCCAG	.T....TC....	3	chr14	+	90403305	14
	GAGCTCCCCCACC CCGCCCCGGG	..G.T..C.... ......C.	4	chr19	-	14142969	2
	AACACCCCCCACC CCACCCCGGG	A..A...C.... ...A..C.	5	chr19	-	51915513	2
	CCCCCCCCCCCCC CCGCCCCCGG	CC.....C..C. ......C.	5	chr4	+	54259934	8

	GTCCCCCTCCTCC CCACCTCCGG	.T........T. ...A....	3	chrX	+	15283 7112	3
	AACCCCACCCACC CCATCTCAGG	A.....AC.... ...AT...	5	chr1	-	23719 5675	9
	CGGACCCCCCACC CCGCCCCCAG	CGGA...C.... ......C.	6	chr8	+	66934 423	5
	GGCCCCACCCACC CCGCCGCAGG	.G....AC.... ......G.	4	chr19	-	30335 983	5
	CCGCCCCCCCATC CCGCCCCAGG	CCG....C...T ......C.	6	chr20	-	62575 811	4
	GCCTCCCCCCACC CAGCCTCGGG	.C.T...C.... ..A.....	4	chr6	-	36853 819	3
	AACCCCACCCACC CCGCCCTGGG	A.....AC.... ......CT	5	chr11	-	17662 650	2
	GACCCCCCCCACC CCGCCCCGAG	.......C.... ......C.	2	chr11	-	10143 7790	4
	ACTCCCCTCCACC CCGGCTCGGG	ACT..... ....G...	4	chr17	-	62167 048	8
	CGGCCTCCCCACC CCGCCCCCGG	CGG..T.C.... ......C.	6	chr7	-	10013 7234	7
	AACCCCACCCACC CCACCCCCGG	A.....AC.... ...A..C.	5	chr8	-	10451 2392	9
	GGGCCCCCCTACC CCGCCCCCGG	.GG....C.T.. ......C.	5	chr9	+	13898 7551	10
	TCTCCACCCCACC CCGCCCCTGG	TCT..A.C.... ......C.	6	chr1	-	78245 430	11
	TTCTCCCTCCTCC CCGCCTCGGG	TT.T......T.	4	chr19	-	42806 795	3
	GACATCCCCCACC CCGCCCCCAG	...AT..C.... ......C.	4	chr20	+	45571 881	8
	CTTCCCCCCCGCC CCGCCCCGGG	CTT....C..G. ......C.	6	chr4	+	43892 69	6
	GTCCCCCTCCAGC CCGCCCCTGG	.T.........G ......C.	3	chr5	-	77269 272	3
	TGCACCCCCCACC CCGCCCCTGG	TG.A...C.... ......C.	5	chr5	+	13286 0359	5
	AACCACCCCCACC CCACCCCAGG	A...A..C.... ...A..C.	5	chr7	+	50765 512	3
	AAGCCCCCCCACC CCGCCCCGGG	A.G....C.... ......C.	4	chr9	-	14026 3439	5
	GTCCTCCACCACC CCGCCTCTGG	.T..T..A....	3	chrX	-	11886 5477	7
	CTCCCTCCCCACC CCACCTCTGG	CT...T.C.... ...A....	5	chr2	+	11699 948	2
	GACACACCCCACC CCACCTCAGG	...A.A.C.... ...A....	4	chr2	+	12744 759	2
	CTTCCCCCACACC CCGCCCCAGG	CTT....CA... ......C.	6	chr3	-	31458 652	7
	CCACCCCTCCACC CTGCTTCGGG	CCA..... ..T..T..	5	chr1	+	15667 5969	3
	TTCCCACTCCATC CCCCTTCTGG	TT...A.....T ...C.T..	6	chr1	+	23232 7590	4

	GAACTCATACACC ACACCTCTGG	..A.T.A.A... .A.A....	6	chr10	+	10350 6362	2
	AGTCCCCCCCACC CCACCACAGG	AGT....C.... ...A..A.	6	chr10	+	12604 2244	2
	GGCTCCCTCCGCC CCGCCCCGGG	.G.T......G. ......C.	4	chr11	-	65479 763	2
	AACCCCTTCTTCC CAACCTCAGA	A.....T..TT. ..AA....	6	chr13	-	47079 419	2
	TCCCCACCCCGCC CCGCCTCTGG	TC...A.C..G.	5	chr15	-	78857 719	13
	TATCCTCCCCACC CCGCCCCGGG	T.T..T.C.... ......C.	5	chr16	-	57974 232	3
	CGCCCACCCCACC CCACCTCAGG	CG...A.C.... ...A....	5	chr17	+	58404 852	3
	CCCCACCCCCACC CCATCTCTGG	CC..A..C.... ...AT...	6	chr19	-	18131 307	2
	AGTCCCATCCACC CCGCCTAAGG	AGT...A..... ....A	5	chr19	-	18633 477	2
	GCTCCCCCCCACC CCGCCCCCAG	.CT....C.... ......C.	4	chr19	+	42785 69	2
	GGCCGCCTCCGCC CCAGCTGCAG	.G..G.....G. ...AG..G	6	chr19	+	81309 30	3
	CCCCCCCCTCACC CCGCCCCTGG	CC.....CT... ......C.	5	chr19	+	45324 565	3
	GACCTCCGCCTCC CAGGTTCAAG	....T..G..T. ..A.GT..	6	chr21	-	27175 337	2
	GACCCCCTTCACC CCACCTATGG	........T... ...A...A	3	chr22	+	23260 188	2
	GACTCCCTCCGCC CCGCTTCCAG	...T......G. .....T..	3	chr5	+	17958 8285	4
	CTGCCCCCCCACC CCAACTCAGG	CTG....C.... ...AA...	6	chr7	+	66023 157	3
	GTCCCTACCCACC CCGCCCCTGG	.T...TAC.... ......C.	5	chr7	+	14296 0402	5
	CCCCCCCCCCCCC CCGCCCCAGG	CC.....C..C. ......C.	5	chr9	+	12888 6485	3
Split-SpCas9^{MT3}-ZFP^{TS2}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS2	GACCCCCTCCACC CCGCCTCCGG	....	0	chr6	-	43738 556	126
OTD_TS2_2	GGGCCCCTCCACC CCGCCTCTGG	.GG.....	2	chr11	-	31817 477	148
OTD_TS2_3	ATTCCCCCCCACC CCGCCTCAGG	ATT....C....	4	chr2	+	24221 4590	12
OTD_TS2_4	CTACCCCTCCACC CCGCCTCCGG	CTA.....	3	chr5	+	67151 02	87
OTD_TS2_5	GACCCCCCCCACC CCGCCCCCGG	.......C.... ......C.	2	chr15	-	33286 101	3
OTD_TS2_6	ACACCCCCCCACC CCGCCTCAGG	ACA....C....	4	chr9	-	10359 9643	10

OTD_TS2_7	TACCCCCCACACC CCGCCTCTGG	T.....CA...	3	chr17	-	43587 46	8
OTD_TS2_8	TGCCCCCCCCACC CCACCTCTGG	TG.....C.... ...A....	4	chr16	-	56963 423	12
OTD_TS2_9	TGCCCCCTCCACC CCACCTCTGA	TG..... ...A....	3	chr6	+	11052 1018	4
OTD_TS2_12	AACCCCTTCTTCC CAGCCTCGGA	A.....T..TT. ..A.....	5	chr6	-	63921 631	20
OTD_TS2_13	AACCCCTTCTTCC CAGCCTCGGA	A.....T..TT. ..A.....	5	chr20	+	13561 41	3
OTD_TS2_14	AACCCCCGCCTCC CGGGTTCAAG	A......G..T. ..G.GT..	6	chr1	+	11725 7633	3
	CTACCCCTCCACC CCGACTCGGA	CTA..... ....A...	4	chr18	+	55101 344	10
	GCCCCCACCCACC CCGCCTCTGG	.C....AC....	3	chr18	-	21359 553	3
	CTCCCCCTCCACC CCACCTCCAG	CT..... ...A....	3	chr6	-	31462 821	7
	GACCCTGTCCACC CCACCTCAGG	.....TG..... ...A....	3	chr1	+	87532 1	4
	GCCCACGTCCACC CTGACCCCGA	.C..A.G..... ..T.A.C.	6	chr11	+	16639 57	2
	GACCCTCTCTATC CCCTCTCAGA	.....T...T.T ...CT...	5	chr19	-	46053 661	2
	GGCGCCCCCGCCC CCGCCCCCGG	.G.G...C.GC. ......C.	6	chr5	-	52405 377	2

SpCas9^{MT3}-DD-ZFP^{TS2}

Name	Sequence	Alignment to target	# of mismatch	Chromosome	Strand	Start position	Peak Score
TS2	GACCCCCTCCACC CCGCCTCCGG	....	0	chr6	-	43738 556	121
OTD_TS2_2	GGGCCCCTCCACC CCGCCTCTGG	.GG.....	2	chr11	-	31817 477	54
OTD_TS2_3	ATTCCCCCCCACC CCGCCTCAGG	ATT....C....	4	chr2	+	24221 4590	11
OTD_TS2_4	CTACCCCTCCACC CCGCCTCCGG	CTA.....	3	chr5	+	67151 02	85
OTD_TS2_5	GACCCCCCCCACC CCGCCCCCGG	.......C.... ......C.	2	chr15	-	33286 101	3
OTD_TS2_6	ACACCCCCCCACC CCGCCTCAGG	ACA....C....	4	chr9	-	10359 9643	8
OTD_TS2_7	TACCCCCCACACC CCGCCTCTGG	T.....CA...	3	chr17	-	43587 46	6
OTD_TS2_8	TGCCCCCCCCACC CCACCTCTGG	TG.....C.... ...A....	4	chr16	-	56963 423	7
OTD_TS2_9	TGCCCCCTCCACC CCACCTCTGA	TG..... ...A....	3	chr6	+	11052 1018	13
OTD_TS2_12	AACCCCTTCTTCC CAGCCTCGGA	A.....T..TT. ..A.....	5	chr6	-	63921 631	4
OTD_TS2_15	CCCCCCCCCCCCC CCGCCATAGA	CC.....C..C. ......AT	6	chr6	-	16356 9986	2

OTD_TS2_16	CCCCCCCACCACA CCCCCTGGAG	CC.....A.... A..C...G	6	chr15	+	91964829	2
	CTACCCCTCCACC CCGACTCGGA	CTA..... ....A...	4	chr18	+	55101344	10
	GCCCCCACCCACC CCGCCTCTGG	.C....AC....	3	chr18	-	21359553	8
	CTCCCCCTCCACC CCACCTCCAG	CT..... ...A....	3	chr6	-	31462821	4
	CCCCCCCCCCCCC CCGCCTCAGG	CC.....C..C.	4	chr13	-	10054 6983	2
	CTCCCCACCCACC CCGCCTCAGG	CT....AC....	4	chr4	+	38537611	2
	CCCCCCCCCCACC CCGCCCTCGG	CC.....C.... ......CT	5	chr11	+	19798641	2
Split-SpCas9^{WT}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS2	GACCCCCTCCACC CCGCCTCCGG	....	0	chr6	-	43738556	83
OTD_TS2_1	CTGCCCCCCCACC CCGCCACTGG	CTG....C.... ......A.	5	chrX	+	14993 2512	118
OTD_TS2_2	GGGCCCCTCCACC CCGCCTCTGG	.GG.....	2	chr11	-	31817477	138
OTD_TS2_3	ATTCCCCCCCACC CCGCCTCAGG	ATT....C....	4	chr2	+	24221 4590	143
OTD_TS2_4	CTACCCCTCCACC CCGCCTCCGG	CTA.....	3	chr5	+	67151 02	147
OTD_TS2_5	GACCCCCCCCACC CCGCCCCCGG	.......C.... ......C.	2	chr15	-	33286101	88
OTD_TS2_6	ACACCCCCCCACC CCGCCTCAGG	ACA....C....	4	chr9	-	10359 9643	119
OTD_TS2_7	TACCCCCCACACC CCGCCTCTGG	T......CA...	3	chr17	-	43587 46	41
OTD_TS2_8	TGCCCCCCCCACC CCACCTCTGG	TG.....C.... ...A....	4	chr16	-	56963423	40
OTD_TS2_9	TGCCCCCTCCACC CCACCTCTGA	TG..... ...A....	3	chr6	+	11052 1018	2
OTD_TS2_12	AACCCCTTCTTCC CAGCCTCGGA	A.....T..TT. ..A.....	5	chr6	-	63921631	12
	CTACCCCTCCACC CCGACTCGGA	CTA..... ....A...	4	chr18	+	55101344	3
	GCCCCCACCCACC CCGCCTCTGG	.C....AC....	3	chr18	-	21359553	47
	CTCCCCCTCCACC CCACCTCCAG	CT..... ...A....	3	chr6	-	31462821	4
	CCCCCCCCCCCCC CCGCCTCAGG	CC.....C..C.	4	chr13	-	10054 6983	60
	CTCCCCACCCACC CCGCCTCAGG	CT....AC....	4	chr4	+	38537611	16
	CCCCCCCCCCACC CCGCCCTCGG	CC.....C.... ......CT	5	chr11	+	19798641	7

	TACCCCCCCCACC CCGCCACAGG	T.....C.... ......A.	3	chr11	+	13948354	35
	TGCCCCCCCCACC CCGCCCCCGG	TG.....C.... ......C.	4	chrX	+	129040624	41
	CCGCCCCTCCACC CCGCCACTGG	CCG......... ......A.	4	chr17	-	55740523	22
	TGCCCCTCCCACC CCGCCTCTGG	TG....TC....	4	chr17	-	40044751	26
	GACCCCTCCCACC CCGACTCCGG	......TC.... ....A...	3	chr9	+	27338858	43
	GTACCCCACCACC CCGCCCCAGG	.TA....A.... ......C.	4	chr8	+	144822945	16
	GCTTCCCTCCACC CCGCATCCGG	.CTT........ .....A..	4	chr11	+	71948788	24
	CCCCCCCCCCACC CCGCCCCCGG	CC.....C.... ......C.	4	chr2	-	129244196	24
	ATCCCCCTCCACC CCACCCCTGG	AT.......... ...A..C.	4	chr11	+	12308918	16
	ACCCCCCCCCGCC CCGCCCCCGG	AC.....C..G. ......C.	5	chr11	+	73458481	15
	ACCCCCCCCCCCC CCGCCCCCGG	AC.....C..C. ......C.	5	chr2	+	24857577	16
	CAACCCCCCCACC CCGCTTCAGG	C.A....C.... .....T..	4	chr3	+	140398784	11
	CCACCCCCCCACC CCGCCCCTGG	CCA....C.... ......C.	5	chr20	-	10913955	5
	CACTCCCCCCACC CCGCCCCAGG	C..T...C.... ......C.	4	chr9	-	126138175	12
	GGACCCTCCCACC CCACCTCAAG	.GA...TC.... ...A....	5	chr13	-	100621456	3
	AGGCCCCCACACC CCGCCTCAGG	AGG....CA...	5	chr4	+	1496241	21
	CCACCCCCCCACC CCGCCCCAGG	CCA....C.... ......C.	5	chr2	+	170573334	7
	GACCCCCCCCACC CCACCCCAGG	.......C.... ...A..C.	3	chr6	+	26470607	10
	CTCCCCCGCCACC CCGCCCCAGG	CT.....G.... ......C.	4	chr17	-	16954846	11
	CATACCCCCCACC CCGCCCCGGG	C.TA...C.... ......C.	5	chr4	-	8841911	8
	GACCCCTCCCTCC CCACCTCAGG	......TC..T. ...A....	4	chr1	-	51442186	4
	AAGACCCCCCACC CCGCCCCAGG	A.GA...C.... ......C.	5	chr19	+	45952038	5
	CCCCCCCCCCCCC CCGCCTCCGG	CC.....C..C.	4	chr22	+	50884786	11
	CCCCACCCCCACC CCGCCTCAGG	CC..A..C....	4	chr10	-	116294250	2
	CTACCCCCACTCC CCGCCTCCGG	CTA....CA.T.	6	chr10	+	102821501	4
	TCCACCCCCCACC CCGCCCCGGG	TC.A...C.... ......C.	5	chr7	+	104933939	5

	GACCCCTCACACC CCGCCCCTGG	......TCA... ......C.	4	chr7	-	95319 671	10
	GCCCCCCACCACC CCACCTCGGG	.C.....A.... ...A....	3	chr19	+	13122 172	2
	CGCCCTCCCCACC CCGCCTCCGG	CG...T.C....	4	chr10	+	13514 9931	8
	AGACCCCCCCACC CCACCCCAGG	AGA....C.... ...A..C.	6	chr20	-	35991 127	4
	AGCCCCCCCCTCC CCGCCCCAGG	AG.....C..T. ......C.	5	chr14	+	10460 1192	7
	GGACCCCCCCCCC CCGCCCCCGG	.GA....C..C. ......C.	5	chr3	-	13187 297	3
	GGACCCCGACGCC CCGCCTCAGG	.GA....GA.G.	5	chr3	-	31574 996	8
	CAGTCCCCCCACC CCACCTCTGG	C.GT...C.... ...A....	5	chr10	+	72538 201	9
	ACACCCCCCCACC CCACCCCAGG	ACA....C.... ...A..C.	6	chr2	-	22153 5826	4
	GACCCTGTCCACC CCACCTCAGG	.....TG..... ...A....	3	chr1	+	87532 1	3
	GACCCGCCCCGCC CCGCCTCTGG	.....G.C..G.	3	chr1	+	11714 529	2
	CCCCCCCACCACC CCGCCCCGGG	CC.....A.... ......C.	4	chr8	-	18041 540	4
	GGCCCTCTCCACT CCACCTCAGG	.G...T...... T..A....	4	chr1	-	17873 8721	2
	AGACCCCCCCACC CCACCCCAGG	AGA....C.... ...A..C.	6	chr21	+	47440 541	4
	CCCCCCTCCCACC CCGCCTCTAG	CC....TC....	4	chr9	+	37465 348	3
	GAGCTCCCCCACC CCGCCCCGGG	..G.T..C.... ......C.	4	chr19	-	14142 969	3
	CCCCCCCCCCCCC CCGCCCCCGG	CC.....C..C. ......C.	5	chr4	+	54259 934	2
	CGGACCCCCCACC CCGCCCCCAG	CGGA...C.... ......C.	6	chr8	+	66934 423	3
	TCTCCACCCCACC CCGCCCCTGG	TCT..A.C.... ......C.	6	chr1	-	78245 430	2
	TCCCCCCCCCACC CCGCTTTGGG	TC.....C.... .....T.T	5	chr11	-	64012 937	3
	TTCTCCCTCCTCC CCGCCTCGGG	TT.T......T.	4	chr19	-	42806 795	2
	AAGCCCCCCCACC CCGCCCCGGG	A.G....C.... ......C.	4	chr9	-	14026 3439	2
	GACACCTCCCACC CCTTCACAGA	...A..TC.... ...TT.A.	6	chr1	+	23078 3394	2
	GACCTCCACCTCC CATGCTCAAG	....T..A..T. ..ATG...	6	chr15	+	29694 599	2
	GACCCCCCAGCCC CCGCCTCAGG	.......CAGC.	4	chr19	+	10163 747	2
SpCas9^{WT}							

Name	Sequence	Alignment to target	# of mismatch	Chromosome	Strand	Start position	Peak Score
TS3	GGTGAGTGAGTGT GTGCGTGTGG	....	0	chr6	+	43737454	520
OTD_TS3_1	AGTGAGTGAGTGT GTGTGTGGGG	A..... ....T...	2	chr14	-	65569153	198
OTD_TS3_2	TGTGGGTGAGTGT GTGCGTGAGG	T...G.....	2	chr5	+	11543 4659	121
OTD_TS3_3	GCTGAGTGAGTGT ATGCGTGTGG	.C..... .A.....	2	chr22	-	37662818	102
OTD_TS3_4	AGAGAGTGAGTGT GTGCATGAGG	A.A..... .....A..	3	chr5	-	89440963	59
OTD_TS3_5	TGTGAGTGAGTGT GTGTGTGTGA	T..... ....T...	2	chr8	-	22932497	18
OTD_TS3_6	TGTGGGTGAGTGT GTGCGTGAGA	T...G.....	2	chr9	-	23824548	8
OTD_TS3_7	GGTGAGTGAGTGT GTGTGTGAGG	....T...	1	chr14	+	10602 9015	7
	GCTATTTGTGTGT GTGGGTGAAG	.C.ATT..T... ....G...	6	chr13	+	66489436	2
	ACTGTGTGAGTGT GTGCGTGAGG	AC..T.....	3	chr19	-	40561861	2
	GGTGAGTGTGTGT GTGCATGTGG	.....T... .....A..	2	chr2	-	17746 3420	2
	GATGAGTGTGTGT GTGTGTGAGG	.A.....T... ....T...	3	chr4	-	62067613	2
	TGTGGATGTCTGT GTGTGTGGGG	T...GA..TC.. ....T...	6	chr5	-	16840 2464	2
	GATGAGTGAGTGA GTGAGTGGGG	.A..... A...A...	3	chr6	-	15707 8321	2
	GGTTAGTGAGAAG CTGTGTGGGA	...T.....AA GC..T...	6	chr9	-	13324 4355	2
	AGGGAGTGAGTGT GAGAGTGCGG	A.G..... ..A.A...	4	chr6	+	14445 8274	3
	GGTGTGTGAGTGT GTGTGTGTGA	....T..... ....T...	2	chr7	+	39341109	3
	GTAGAGTGAGTGT GTGTGTGTGG	.TA..... ....T...	3	chr8	+	48997789	3
	GGTGAGTGAGTGA GTGAGTGAGG	..... A...A...	2	chr10	-	10937 8061	4
	GGTGAGTCAGTGT GTGAGTGAGG	.....C.... ....A...	2	chr2	-	73317044	4
	AGTGAGTGAGTGT GAGTGCGGGG	A..... ..A.T.C.	4	chr8	-	67579421	4
	AGTGAGTGAGTGT GTGTGTGTGA	A..... ....T...	2	chr16	+	12264596	5
	TGTGAGTGAGTGT GTGTGTGTGA	T..... ....T...	2	chr19	+	6109016	6
	AGTGTGTGAGTGT GTGCGTGTGG	A...T.....	2	chr20	-	20178278	8

	GGTGAGAGAGTGT GTGCACGGGG	......A..... .....AC.	3	chr22	+	43939280	8
	AGTGAATGAGTGT GTGTGTGTGG	A....A..... ....T...	3	chr3	-	19399 3878	9
	GTTGAGTGAATGT GTGCGTGAGG	.T.......A..	2	chr10	-	98760 582	10
	AGTGGGTGAGTGA GTGCGTGCGG	A...G..... A.....	3	chr11	+	68851 122	11
	GGTGAGAGAGTGT GTGCGTAGGA	......A..... ....A	2	chr16	-	74898 114	12
	AGCGAGTGGGTGT GTGCGTGGGG	A.C.....G...	3	chr14	+	74353 480	14
	TGTGAGTGAGTGT GTGTATGGGG	T..... ....TA..	3	chr5	-	29367 373	16
	TGTGAGTAAGTGT GTGTGTGTGG	T......A.... ....T...	3	chr14	-	62078 767	17
SpCas9^{MT3}-ZFP							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS3	GGTGAGTGAGTGT GTGCGTGTGG	....	0	chr6	+	43737454	140
OTD_TS3_8	GGTGAGAGCGAGG GTTGGTGTGG	......A.C.A. G..TG...	6	chr11	-	57254546	5
OTD_TS3_9	TGTGTGTGTGTGT GTGTTTTGAG	T...T...T... ....TT.T	6	chr1	-	29261945	4
	ATAGAGTGTGGGT GTGGGTGTGG	ATA.....T.G. ....G...	6	chr10	+	71425390	3
	TGAGAGAGAGAGT TTGCCTGAGG	T.A...A...A. .T...C..	6	chr16	-	73573395	2
SpCas9^{MT3}-TAL							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS3	GGTGAGTGAGTGT GTGCGTGTGG	....	0	chr6	+	43737454	175
OTD_TS3_12	GGTGAGGGAGTAT CGGGGAGTGA	......G....A .CG.G.A.	6	chr7	-	2520194	30
	TGTGTGTGTGTGT GTGTGTTTGG	T...T...T... ....T..T	5	chr14	-	92959300	2
	AGAGAGTGAGTGT GTGGCAGAAG	A.A..... ....GCA.	5	chr2	-	4917103	2
Split-SpCas9^{WT}-NLS							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS3	GGTGAGTGAGTGT GTGCGTGTGG	....	0	chr6	+	43737454	231
OTD_TS3_1	AGTGAGTGAGTGT GTGTGTGGGG	A..... ....T...	2	chr14	-	65569153	62

OTD_TS3_2	TGTGGGTGAGTGT GTGCGTGAGG	T...G.....	2	chr5	+	11543 4659	30
OTD_TS3_3	GCTGAGTGAGTGT ATGCGTGTGG	.C..... .A.....	2	chr22	-	37662 818	50
OTD_TS3_4	AGAGAGTGAGTGT GTGCATGAGG	A.A..... .....A..	3	chr5	-	89440 963	17
OTD_TS3_5	TGTGAGTGAGTGT GTGTGTGTGA	T..... ....T...	2	chr8	-	22932 497	5
OTD_TS3_6	TGTGGGTGAGTGT GTGCGTGAGA	T...G.....	2	chr9	-	23824 548	2
	AGCGAGTGGGTGT GTGCGTGGGG	A.C.....G...	3	chr14	+	74353 480	9
	TGTGAGTAAGTGT GTGTGTGTGG	T......A.... ....T...	3	chr14	-	62078 767	6
	AGTGGGTGAGTGA GTGCGTGCGG	A...G..... A.....	3	chr11	+	68851 122	6
	AGTGTGTGAGTGT GTGCGTGTGG	A...T.....	2	chr20	-	20178 278	6
	AGTGAATGAGTGT GTGTGTGTGG	A....A..... ....T...	3	chr3	-	19399 3878	4
	AGTGAGTGAGTGA GTGAGTGAGG	A..... A...A...	3	chr4	+	58326 591	4
	GATGAGTGTGTGT GTGTGTGAGG	.A.....T... ....T...	3	chr4	-	62067 613	3
	TGTGAGTGAGTGT GTGTATGGGG	T..... ....TA..	3	chr5	-	29367 373	2
	GGTGAGTGAGTGA GTGAGTGAGG	..... A...A...	2	chrX	-	41726 212	2
Split-SpCas9^{MT3}-ZFP^{TS3}-NLS							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS3	GGTGAGTGAGTGT GTGCGTGTGG	....	0	chr6	+	43737 454	239
OTD_TS3_1	AGTGAGTGAGTGT GTGTGTGGGG	A..... ....T...	2	chr14	-	65569 153	6
OTD_TS3_15	AGTGCGTGTGTGC GTGCGCCAGG	A...C...T... C.....CC	6	chr16	+	67283 108	4
OTD_TS3_16	TGTGTGTGTGTGT GTGTGTGTGA	T...T...T... ....T...	4	chr8	+	41273 263	3
OTD_TS3_6	TGTGGGTGAGTGT GTGCGTGAGA	T...G.....	2	chr9	-	23824 548	2
OTD_TS3_10	TGTGTGTGTCTGT GTGTGTGTGA	T...T...TC.. ....T...	5	chr14	+	98411 510	2
OTD_TS3_17	TGTGTGTGTGTGT GTGTGTGTGA	T...T...T... ....T...	4	chrX	-	11651 7683	2
OTD_TS3_18	GGTGGCCGCATGG GTGCGTGAGG	....GCC.CA.. G.....	6	chr16	+	30845 953	2
OTD_TS3_19	AGTGACTGAATGG GTGAGGGGAG	A....C...A.. G...A.G.	6	chr2	-	20421 958	2
OTD_TS3_20	TGTGTGTGCATGT GTGTTTGGAG	T...T...CA.. ....TT..	6	chr2	+	13999 2438	2

OTD_TS3_21	GATGTGGTGGTGTGTGCCTGTAG	.A..T.GTG.....C..	6	chr4	-	110094379	2
SplitCas9^{MT3}-TALE^{TS3}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS3	GGTGAGTGAGTGTGTGCGTGTGG	....	0	chr6	+	43737454	99
OTD_TS3_13	TGTGTGTGTGTGTGTGCATTAAG	T...T...T..... .A.T	5	chr3	-	68073944	3
SpCas9^{MT3}-DD-ZFP^{TS3}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS3	GGTGAGTGAGTGTGTGCGTGTGG	....	0	chr6	+	43737454	145
OTD_TS3_11	TGTGTGTATGTGTGTCTGTGTAG	T...T..AT.....CT...	6	chr9	-	24270033	8
OTD_TS3_1	AGTGAGTGAGTGTGTGTGTGGGG	A..... ....T...	2	chr14	-	65569153	5
OTD_TS3_10	TGTGTGTGTCTGTGTGTGTGA	T...T...TC.. ....T...	5	chr14	+	98411510	3
OTD_TS3_6	TGTGGGTGAGTGTGTGCGTGAGA	T...G.....	2	chr9	-	23824548	2
OTD_TS3_25	TCTGGGTGTGTGTCTGTGTGGGG	TC..G...T.....C..T...	6	chr2	-	43351368	2
SpCas9^{MT3}-DD-TALE^{TS3}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS3	GGTGAGTGAGTGTGTGCGTGTGG	....	0	chr6	+	43737454	175
OTD_TS3_14	GGGCAGTGTGTGTGTGGCGGGGG	..GC....T.....GCG.	6	chr16	+	79840707	9
OTD_TS3_10	TGTGTGTGTCTGTGTGTGTGA	T...T...TC.. ....T...	5	chr14	+	98411510	7
OTD_TS3_22	GGTGGGTGGGGGTGGGGGGGGGG	....G...G.G...G.G.G.	6	chr21	+	40664046	7
OTD_TS3_1	AGTGAGTGAGTGTGTGTGTGGGG	A..... ....T...	2	chr14	-	65569153	4
OTD_TS3_5	TGTGAGTGAGTGTGTGTGTGA	T..... ....T...	2	chr8	-	22932497	3
OTD_TS3_23	GATGAGTGTGGCTGAGGGTGGGG	.A.....T.GC..A.G...	6	chr14	+	76796338	2
OTD_TS3_24	GGAGAGTGGGTTTTTGGCTGAGA	..A.....G..T.T..GC..	6	chr20	-	47227171	2
	GGAGAGAGAGAGAGGGGGGTGGGG	..A...A...A.A.G.G...	6	chr4	+	112044802	2
	TGTGTGTGTGTATGTGTGTGA	T...T...T..A....T...	5	chr6	+	143392920	2

	TGGGATTGGGGGT GTGTGTGGGG	T.G..T..G.G. ....T...	6	chr8	-	27807865	2
Split-SpCas9^{WT}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS3	GGTGAGTGAGTGT GTGCGTGTGG	....	0	chr6	+	43737454	195
OTD_TS3_1	AGTGAGTGAGTGT GTGTGTGGGG	A..... ....T...	2	chr14	-	65569153	31
OTD_TS3_3	GCTGAGTGAGTGT ATGCGTGTGG	.C..... .A.....	2	chr22	-	37662818	19
OTD_TS3_4	AGAGAGTGAGTGT GTGCATGAGG	A.A..... .....A..	3	chr5	-	89440963	14
OTD_TS3_2	TGTGGGTGAGTGT GTGCGTGAGG	T...G.....	2	chr5	+	115434659	10
	GGTGGGTAAGTGG GAGCTTGGGG	....G..A.... G.A..T..	5	chr2	-	200628256	4
OTD_TS3_7	GGTGAGTGAGTGT GTGTGTGAGG	....T...	1	chr14	+	106029015	3
	AGTGTGTGAGTGT GTGCGTGTGG	A...T.....	2	chr20	-	20178278	3
	TGTGTGTGTGTGT GTGTGTGTGG	T...T...T... ....T...	4	chr20	+	47519599	3
OTD_TS3_10	TGTGTGTGTCTGT GTGTGTGTGA	T...T...TC.. ....T...	5	chr14	+	98411510	2
	TGTGTGTATGTGT GTGCGAAGGA	T...T..AT... ......AA	6	chr11	-	32522389	2
	TGTGTGTTTGTGT GTGCAAGAGA	T...T..TT... .....AA.	6	chr4	+	119812483	2

Appendix 2.3 GUIDE-seq data in Chapter 4

SpCas9^{WT}

Name	Sequence	Alignment to target	# of mismatch	Chromosome	Strand	Start position	Peak Score
TS3	GGTGAGTGAGTGTGT GCGTGTGG	....	0	chr6	+	43737454	472
OT3-4	GCTGAGTGAGTGTAT GCGTGTGG	.C........ ..A.....	2	chr22	-	37662818	381
OT3-18	TGTGGGTGAGTGTGT GCGTGAGG	T...G.....	2	chr5	+	115434659	305
OT3-20	AGAGAGTGAGTGTGT GCATGAGG	A.A....... ......A..	3	chr5	-	89440963	239
OT3-2	AGTGAGTGAGTGTGT GTGTGGGG	A......... .....T...	2	chr14	-	65569153	218
OT3-3	AGTGTGTGAGTGTGT GCGTGTGG	A...T.....	2	chr20	-	20178278	188
OT3_Sp_new	AGTGGGTGAGTGAGT GCGTGCGG	A...G..... .A.....	3	chr11	+	68851122	186
OT3-21	AGCGAGTGGGTGTGT GCGTGGGG	A.C.....G..	3	chr14	+	74353480	182
OT3-6	TGTGGGTGAGTGTGT GCGTGAGA	T...G.....	2	chr9	-	23824548	117
OT3-17	GTTGAGTGAATGTGT GCGTGAGG	.T.......A.	2	chr10	-	98760582	98
GUIDEseq_HF1	AGTGAATGAGTGTGT GTGTGTGG	A....A..... .....T...	3	chr3	-	193993878	90
Zhang_VEGFA1-OT7	TGTGAGTAAGTGTGT GTGTGTGG	T......A... .....T...	3	chr14	-	62078767	82
	GGTGAGTGTGTGTGT GCATGTGG	........T.. ......A..	2	chr2	-	177463420	74
	GATGAGTGAGTGAGT GAGTGGGG	.A........ .A...A...	3	chr6	-	157078321	53
	GGGGAATGAGTGTGT GCATGGAG	..G..A..... ......A..	3	chr1	+	212639761	51
	GGTGAGTAAGTGTGA GCGTAAGG	.......A... ...A....A	3	chr18	+	74103158	51
OT3-19	ACTGTGTGAGTGTGT GCGTGAGG	AC..T.....	3	chr19	-	40561861	51
	CGCGAGTGAGTGTGT GCGCGGGG	C.C....... ....C.	3	chr3	-	71632803	50
VEGF3_new2	AGCGAGTGAGTGTGT GTGTGGGG	A.C....... .....T...	3	chr9	-	18733629	49
OT3-22	AGGGAGTGACTGTGT GCGTGTGG	A.G......C.	3	chr10	+	130228337	44

	AGTGAGTGAGTGAGT GAGTGAGG	A..... .A...A...	3	chr7	-	15267137 2	34
	TGTGAGTGAGTGTGT GTGTGTGA	T..... .....T...	2	chr8	-	22932497	34
	GGTGAGAGAGTGTGT GCGTAGGA	......A.... ....A	2	chr16	-	74898114	33
	GGTGGGTGAATGGGT GCGTGGGG	....G....A. .G.....	3	chr17	+	19141553	30
	GATGAGTGTGTGTGT GTGTGAGG	.A......T.. .....T...	3	chr4	-	62067613	29
OT3-1	GGTGAGTGAGTGTGT GTGTGAGG	....T...	1	chr14	+	10602901 5	28
	CTGGAGTGAGTGTGT GTGTGTGG	CTG..... .....T...	4	chr19	-	47732486	28
	TGTGAGTGAGTGTGT GTGTGTGA	T..... .....T...	2	chr19	+	6109016	26
Zhang_VEGFA1-OT10	CGTGAGTGAGTGTGT ACCTGGGG	C..... ....A.C..	3	chr20	+	50724388	24
	AGGGAGTGAGTGTGA GAGTGCGG	A.G..... ...A.A...	4	chr6	+	14445827 4	24
VEGF3_new1	GTAGAGTGAGTGTGT GTGTGTGG	.TA..... .....T...	3	chr8	+	48997789	24
	GGTGAGTGAGTGAGT GAGTGAGT	..... .A...A...	2	chr4	+	58326587	23
	GGTGTGTGAGTGAGT GTGTGTAT	....T..... .A...T...	3	chr5	-	29367377	23
	GGAGAGTGAGTGTGT GCATGTGC	..A..... ......A..	2	chr1	+	181557187	22
	TGTGAGTGAGTGTGT GCGTGTGA	T.....	1	chr16	+	80016314	22
	GGGGAATGAGTGTGT GTGTGGGG	..G..A..... .....T...	3	chr3	-	40369878	22
	GGTGTGTGAGTGTGT GTGTGTGG	....T..... .....T...	2	chr22	+	49739984	21
	GGTGAGTCAGTGTGT GAGTGAGG	.......C... .....A...	2	chr2	-	73317044	21
	GGGGGGTGAGTGTGT GTGTGGGG	..G.G..... .....T...	3	chr1	+	32738747	19
	GGTGTGTGAGTGTGT GTGTGTGA	....T..... .....T...	2	chr7	+	39341109	18
	AGTGAGTGAGTGTGA GTGCGGGG	A..... ...A.T.C.	4	chr8	-	67579421	17
	AGTGAATGAGTGTGT GCATGTGA	A....A..... ......A..	3	chrX	+	10561439 8	17
	AGAGAGTGAGTGTGT GTGTTGGG	A.A..... .....T..T	4	chr10	+	10786736 2	16

	TGTGAGTGAGTGTGT GTGTGGAG	T......... .....T...	2	chr19	-	165694816	16
	GGTGAGCAAGTGTGT GTGTGTGG	......CA... .....T...	3	chr2	-	230506235	14
	AGTGAGTGAGTGTGT GTGTGAGA	A......... .....T...	2	chr3	-	51049346	14
	GGTGAGTGAGTGAGT GAGTGAGG	.... .A...A...	2	chrX	-	417262126	14
	TGTGGGTGAGTGTGT GTGTGTGG	T...G..... .....T...	3	chr1	-	483050326	13
	AGTGAGAAAGTGTGT GCATGCGG	A.....AA... ......A..	4	chr2	+	186962086	13
	GGTGAGTAGGTGTGT GTGTGGGG	.......AG.. .....T...	3	chr11	-	76257897	12
	TGTGAGTGAGTGTGT GTGTGTGA	T......... .....T...	2	chr15	-	928642066	12
	GGTGAGTGAGAGTGT GTGTGTGG	.........A .....T...	2	chr5	+	150224704	12
	AGTGAGTGAGTGTGT GTGTGTGA	A......... .....T...	2	chr16	+	122645960	11
	GGTGAGTGCGTGTGT GCGTGCGC	.........C..	1	chr16	+	672831044	11
	GGTGAGTGAGTGAGT GAGTGAGG	.... .A...A...	2	chr10	-	109378061	10
	GGAGAGTGAGTATGT GTGTGTGT	..A....... A....T...	3	chr10	-	57496551	10
	AGTGAGTGAGTGAGT GAGTGAGT	A......... .A...A...	3	chr11	+	791785027	10
	TGTGAGTGAGTGTGT GTGTGTGA	T......... .....T...	2	chr14	+	581184276	10
	ATTGAGTGAGTATGT GTGTGAGG	AT........ A....T...	4	chr11	+	633663256	9
	GGAGAGTGAGTGTGT GTGTGAGA	..A....... .....T...	2	chr5	-	130857596	9
	AGGGTGTGAGTGAGT GCATGTGT	A.G.T..... .A....A..	5	chr5	-	178746537	9
	GGTGTGTGAGTGTGT GCATTGGG	....T..... ......A.T	3	chr6	+	390286259	9
	AATGAGTGAGTGTGT GAGTGAAG	AA........ .....A...	3	chr1	+	116485627	8
	AGTGTGTGTGTGTGT GGTTTTTT	A...T...T.. .....GT.T	6	chr11	+	907627047	8
	GGTGGGTGAGTGAGT GAGTGAGG	....G..... .A...A...	3	chr12	+	131196650	8

	AGAGAGTGAGTGTGT GTGTGAGA	A.A..... .....T...	3	chr1	-	24504846 1	8
	GGTGAATGAGTGTGT GCTCTGGG	.....A..... ......TCT	4	chr16	-	84032640	8
	ACTGAGTGAGTGTGA GTGTGAGG	AC..... ...A.T...	4	chr3	+	194073635	8
	GATGAGTGAGTGTGT GAGTGAGA	.A..... .....A...	2	chr5	-	149501688	8
	TGTGAGAGAGAGTGT GCGTGTGG	T.....A...A	3	chr1	-	47690888	6
	AATGAGTGAGTGTGT GTGTGTGA	AA..... .....T...	3	chr16	+	73585909	6
	TGTGAGTGAGTGTGG GTGTGTGG	T..... ...G.T...	3	chr1	+	236216059	5
OT3-24	CATGAGTGAGTGTGT GGGTGGGG	CA..... .....G...	3	chr19	+	1716775	5
	GGCAGGGAAGTGTGT GCATGGAT	..CAG.GA... ......A..	6	chr5	+	172635548	5
	AGTGTGTGAGTGTGT GAGTGAGG	A...T..... .....A...	3	chr13	-	91706950	4
	AGAGAGTGAGTGTGT GTGTGAGA	A.A..... .....T...	3	chr17	+	35006783	4
	AGAGAGAGAGTGTGT GTGTGGGG	A.A...A.... .....T...	4	chr3	-	169379099	4
	GATGAGCGAGTGTGT GTGTATGG	.A....C.... .....T..A	4	chrX	-	39606143	4
	GGTGAGTGGGTGTGT GTGTTGGG	........G.. .....T..T	3	chr13	+	60683006	3
	TGTGAGTGAGTGGGT GAGTGTGT	T..... .G...A...	3	chr16	-	72877916	3
	TGTGAGTGGGTGTGT GCATGTGG	T.......G.. ......A..	3	chr21	+	44179960	3
	GGTGAGAGAGTGTGT GCACGGGG	......A.... ......AC.	3	chr22	+	43939280	3
SpCas9^{MT3}-dSaCas9							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS3	GGTGAGTGAGTGTGT GCGTGTGG	....	0	chr6	+	43737454	479
OT3-2	AGTGAGTGAGTGTGT GTGTGGGG	A..... .....T...	2	chr14	-	65569153	77
	GGAGAGTGAGTGTGT GCATGTGC	..A..... ......A..	2	chr1	+	181557187	13

OT3-18	TGTGGGTGAGTGTGT GCGTGAGG	T...G.....	2	chr5	+	11543465 9	9
OT3-6	TGTGGGTGAGTGTGT GCGTGAGA	T...G.....	2	chr9	-	23824548	7
SpCas9^{MT3}-dNmCas9							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
TS3	GGTGAGTGAGTGTGT GCGTGTGG	....	0	chr6	+	43737454	84
	AGTGAGTGAGTGTGT GTGTGTGT	A..... .....T...	2	chr3	+	13652162 5	8
	TGTGTGAGAGTGTCT GTGTTCCA	T...T.A.... ..C..T..T	6	chr17	+	5972994	4
	TGTTAGTGAGTGTTT AACTGGCT	T..T..... ..T.AAC..	6	chr9	+	27310251	4
	TGGGAGTGGGAGTGG GGGTGGAG	T.G.....G.A ...G.G...	6	chr2	+	16135053 0	3
SpCas9^{WT}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
BCL11A_TS7,10,11_Sp	GCTAGTCTAGTGCAA GCTAACAG	....	0	chr2	-	60722412	79
NmCas9^{WT}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
BCL11A_TS7_NmCas9	CAGGCTCCAGGAAGG GTTTGGCCTCTGATT	....	0	chr2	-	60722370	256
	GAGGGGAAGGGAAGG GGATGGGGGTGGTTA	G...GGAAG.. .....GA...G	9	chr7	+	21584864	41
	GTGGTGGCAGGATGG CTTGAGCCCAGGCCG	GT..TGG.... .T..C..GA..	9	chr11	+	18962094	3
	ATGACCCCAGGGAGG TTTTTATATGGGACT	AT.A.C..... G...T...TAT	9	chr11	-	66196236	3
SpCas9^{WT}-NmCas9^{WT}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
BCL11A_TS7_NmCas9	CAGGCTCCAGGAAGG GTTTGGCCTCTGATT	....	0	chr2	-	60722370	177
	CAGCCTCCAGAGCAG CTGGGACCACAGGTG	...C.....A GCA.C.GG.A.	9	chr6	+	15199993 8	8

	CAGCCTCCAGGGAGA ACCTCCAGACTGCAG	...C..... G..AACC.CCA	9	chr13	-	84765515	3
	CAGGCTAAAGGTGTG GAGAGGAAAAGGTAA	......AA... TGT..AGA..A	9	chr8	-	88109044	3
SaCas9^{WT}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
BCL11A_TS9_SaCas9	CTTTTATCACAGGCT CCAGGAAGGGT	....	0	chr2	-	60722383	1704
BCL11A_TS11_SaCas9	CCTCTGATTAGGGTG GGGGCGTGGGT	....	0	chr2	-	60722353	1248
BCL11A_TS10_SaCas9	GTTTGGCCTCTGATT AGGGTGGGGGC	....	0	chr2	-	60722359	662
	CCCCTGATTAGGGTG AAGGGCTGGGG	..C..... ....AA..G	4	chr15	+	70872512	380
	CCTCTCACCAGGGTG GGGCTGTGAGT	.....C.CC.. ....CT	5	chr1	-	33605885	374
	CCTTTGAATAAAGTG GGGGCATGGGG	...T...A..A A........	4	chr16	+	55410314	357
	CCTCTCATTAGGGTA GAGACAAGAAT	.....C..... ...A.A.A.	4	chr11	+	99006985	111
	CCTCTGCTCACGGTG GGGGCATGGGG	......C.C.C	3	chr20	+	58297869	71
	TCTCTGACAAGAGAG GAGGCAAGGGT	T......CA.. A.A..A...	6	chr1	-	15131640 3	69
	CCTTTGACTGGGGTG GGGGGTGGGGT	...T...C.G. ....G	4	chr1	+	48703020	54
	CTTCTGCTTAGGGTG GAGGCATGAGG	.T....C.... .....A...	3	chr4	-	88808529	48
	CCTCAGATAAGAGTG GGGACCAGGAG	....A...A.. A......A.	4	chr13	-	51729736	46
	CTTCTGACAAAGATG GGGGCGGGGGT	.T.....CA.A .A.......	5	chr15	+	46221458	36
	CCTCTGAGGAGGGTG AGGGCAGGAGG	.......GG.. ....A....	3	chr9	+	13649857 8	34
	CATCTGATGGGGGTG GGGGCTTGAGG	.A......GG.	3	chr19	+	38932084	26
	CCTTTGAGAGAGGTG GGGGCAGGGGT	...T...GAGA	5	chr4	+	38730629	26
	CTTCTGACCAAGGCA GGGGCTTGAGC	.T.....CC.A ..CA.....	6	chr12	+	54421154	22
	CCTTTGGTTAGGATG GGGGTAGGGGT	...T..G.... .A......T	4	chr1	-	11715319 4	20
	CCTTTGGTTAGGATG GGGGTAGGGGT	...T..G.... .A......T	4	chr2	+	91930605	20
	CCTCTGATTAGGGTT GGGTCGTGGAG	....T...T.	2	chr1	+	23622871 2	19

	CCTCTGGTAGGGGTA GGGGTCAGGGT	......G.AG. ...A....T	5	chr14	-	52389234	18
	CACCTGAAAAAGGTG AGGGCAGGAGT	.AC....AA.A ....A....	6	chr12	-	24479594	17
	ACCCTAATTAGAGGA GGGGCCTGGTA	A.C..A..... A.GA.....	6	chr3	-	25449995	15
	CCTATGATTTGGGTG GGGGCATGAAG	...A.....T.	2	chr8	-	11747581 8	14
	CCTCTGGCATGGGTG AGGGCATGGGC	......GCAT. ....A....	5	chr2	+	67515306	13
	CCTCTCACTAGGGTG AGGGCTGGGCT	.....C.C... ....A....	3	chr12	+	644761	12
	CCTCTGGTCGGGGCG GGGGCGGGGGC	......G.CG. ..C......	4	chr19	-	49841518	12
	TCTCTGACCAGGGTG GGGGCTAGCCA	T......CC..	3	chr1	-	19053367	11
	CCTCAGCTTAGGATG GGGGCTGGAGA	....A.C.... .A.......	3	chr8	-	13977141 1	10
	CCTCTGCATAGGGTG AGGGCTAGATG	......CA... ....A....	3	chr8	-	21428232	10
	CCTCTGACCTTGGAG GGGGCCAGAGT	.......CCTT ..A......	5	chr9	-	11635152 9	10
	CCTCTGAACAAAGTG GGGGCTGGAAC	.......AC.A A........	4	chr8	+	13335142 1	9
	CTGCTGATTAGGGTG AGGGTTAGGGT	.TG........ ....A...T	4	chr10	-	11893032 6	8
	CCTCTGACCAGGGTG GAGTAGTGGGT	.......CC.. .....A.TA	5	chr9	-	37289531	8
	CCTGTGAGTAGAGTG AGGGCAGGAGG	...G...G... A...A....	4	chr11	+	75166399	7
	CCCCTGACTGGGGCG GGGGCGTGGTA	..C....C.G. ..C......	4	chr19	+	36630408	7
	CCCCTGACTAGGGTG GGAGCATGGCA	..C....C... ......A..	3	chr3	-	14963659	7
	CCTCTGACGGGGATG GGGATGGGGGT	.......CGG. .A.....AT	6	chr5	+	61602772	7
	CCACTCATTAGGGTG GAAACAGGAGT	..A..C..... .....AAA.	5	chrX	+	82777389	5
	CCTCTGATTAGGGTG GAGGTGAGAGA	....A..T	2	chr11	+	22217343	4
	CCTCTGTCTAGGGTG GACACTTGAGT	......TC... .....ACA.	5	chr11	+	33189658	4
	TCTCTGACTAGGAAA GGGGCCAGAGT	T......C... .AAA.....	5	chr1	+	34860363	4
	CCTCTGAGTGGGGTG AAGGCCAGGGG	.......G.G. ....AA...	4	chr15	+	50571952	4

	CTTCTGATGAGGGTG GAGGCTGGACA	.T.....G.. .....A...	3	chr16	-	80054318	3
	CCTCGGATTAAAGTG GGGGCTGGGCA	....G.....A A.....	3	chr18	-	56008669	3
	CCCCTCATCATGGTG GAGGCTGGGGT	..C..C..C.T .....A...	5	chr6	+	37572942	3
SpCas9^{WT}-NmCas9^{WT}							
Name	**Sequence**	**Alignment to target**	**# of mismatch**	**Chromosome**	**Strand**	**Start position**	**Peak Score**
BCL11A_TS9_SaCas9	CTTTTATCACAGGCT CCAGGAAGGGT	....	0	chr2	-	60722383	1809
BCL11A_TS9_SpCas9	CCAGGGTCAATACAA CTTTGAAG	....	0	chr2	-	60722434	1809
BCL11A_TS11_SaCas9	CCTCTGATTAGGGTG GGGGCGTGGGT	....	0	chr2	-	60722353	210
BCL11A_TS10_SaCas9	GTTTGGCCTCTGATT AGGGTGGGGGC	....	0	chr2	-	60722359	116
	CCCCTGATTAGGGTG AAGGGCTGGGG	..C..... ....AA..G	4	chr15	+	70872512	100
	CCTTTGAATAAAGTG GGGGCATGGGG	...T...A..A A.....	4	chr16	+	55410314	82
	CCTCTCACCAGGGTG GGGCTGTGAGT	.....C.CC.. ....CT	5	chr1	-	33605885	81
	CCTCTGCTCACGGTG GGGGCATGGGG	......C.C.C	3	chr20	+	58297869	19
	TCAGGGTCAAGACAA CTTTGAGG	T.........G	2	chr3	+	15217629	16
	CTTCTGCTTAGGGTG GAGGCATGAGG	.T....C.... .....A...	3	chr4	-	88808529	15
	CTTTTATCAGAAGTT CCAGGGAGGGT	.........G. A.T.....	3	chrX	+	68234110	12
	CCTCTCATTAGGGTA GAGACAAGAAT	.....C..... ...A.A.A.	4	chr11	+	99006985	12
	CCTCTGAGGAGGGTG AGGGCAGGAGG	.......GG.. ....A....	3	chr9	+	136498578	12
	TCTCTGACAAGAGAG GAGGCAAGGGT	T......CA.. A.A..A...	6	chr1	-	151316403	11
	CCTTTGACTGGGGTG GGGGGTGGGGT	...T...C.G. ....G	4	chr1	+	48703020	11
	CCTCTGACCTTGGAG GGGGCCAGAGT	.......CCTT ..A.....	5	chr9	-	116351529	11
	CCTCAGATAAGAGTG GGGACCAGGAG	....A...A.. A......A.	4	chr13	-	51729736	9
	CCAGGGTCAGCACCT TCTTGGCT	.........GC ..CTTC...	6	chr3	-	51743062	8
	CCTTTGGTTAGGATG GGGGTAGGGGT	...T..G.... .A......T	4	chr1	-	117153194	8

	CTTTTATTGCAGCCT CCAGGTGGAAT	.......TG.. .C.......	3	chr17	+	17826167	6
	AGAGGGCCAATACAA CTTTAGGG	AG....C.... ....A	4	chr10	+	56947264	5
	CCTCAGCTTAGGATG GGGGCTGGAGA	....A.C.... .A.......	3	chr8	-	139771411	4
	CCTTTGGTTGGGGCA GGGGCAGGAAT	...T..G..G. ..CA.....	5	chr6	-	36696690	4
	CTTCTGCGTGGGGGG GGGGGGGGGGT	.T....CG.G. ..G.....G	6	chr12	-	119418907	3
	CATTTGACAAGGGGG GGGGGGGGGAG	.A.T...CA.. ..G.....G	6	chr17	+	91910405	3
	CCCCTGAGCAAGGTG GAGGCTGGATT	..C....GC.A .....A...	5	chr6	-	31852626	3