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

Mehmet Fatih Bolukbasi[1,2], **Ankit Gupta**[1], **Sarah Oikemus**[1], **Alan G. Derr**[3], **Manuel Garber**[3,4], **Michael H. Brodsky**[1], **Lihua Julie Zhu**[1,3,4], and **Scot A. Wolfe**[1,2]

[1]Department of Molecular, Cell and Cancer Biology, University of Massachusetts Medical School, Worcester, MA

[2]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA

[3]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, MA

[4]Department of Molecular Medicine, University of Massachusetts Medical School, Worcester, MA

Abstract

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

Correspondence should be addressed to S.A.W. (; Email: scot.wolfe@umassmed.edu)

Accession Codes
Illumina Sequencing data has been submitted to the Sequence Read Archive: PRJNA294793

Introduction

The CRISPR-Cas9 genome engineering system is revolutionizing biological sciences due to its simplicity and efficacy[1–3]. The most commonly studied Cas9 nuclease originates from *Streptococcus pyogenes* (SpCas9)[4]. SpCas9 and its associated guide RNA license a DNA sequence for cleavage based on two stages of sequence interrogation[4–8] (Supplementary Fig. 1): 1) compatibility of the PAM element with the specificity of the PAM-interacting domain, and 2) complementarity of the guide RNA sequence with the target site. Because it is straightforward to program Cas9 to cleave a desired target site through incorporation of a complementary single guide RNA (sgRNA)[4], the primary constraint on Cas9 targeting is the presence of a compatible PAM element[4,9,10]. The PAM-interacting domain of wild-type SpCas9 preferentially recognizes a nGG element[4], although it can inefficiently utilize other PAM sequences (e.g. nAG, nGA)[9,11]. The simplicity of the SpCas9-sgRNA system allows facile editing of genomes in a variety of organisms and cell lines[1–3].

The precision of SpCas9 is sub-optimal for most gene therapy applications involving editing of a large population of cells[12,13]. Numerous studies have demonstrated that SpCas9 can cleave the genome at unintended sites[9,14–20], with some guides acting at more than 100 off-target sites[17]. Recent genome-wide analyses of SpCas9 precision indicate that the majority of genomic loci that differ at two nucleotides from the guide RNA sequence and a subset of genomic loci that differ at three nucleotides are cleaved with moderate activity[17–20]. For some guides, off-target sites that differ by up to six nucleotides can be inefficiently cleaved[17–20] and bulges can be accommodated within the sgRNA:DNA heteroduplex[15]. 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[21,22]. We found that the vast majority of guides (~98% in exons and ~99% in promoters) have one or more off-target sites with 3 or fewer mismatches (Supplementary Fig. 1), and thus are likely to have some level of off-target activity. Because off-target breaks have the potential to cause both local mutagenesis and genomic rearrangements (e.g. segmental deletions, inversions and translocations)[17,18,23,24], the resulting collateral damage from SpCas9 treatment could have adverse consequences in therapeutic applications.

Reduced off-target cleavage rates have been reported with several modifications to the structure or delivery of the CRISPR-Cas9 system. Examples include: changing guide sequence length and composition[25,26]; employing a pair of Cas9 nickases[26–28] or FokI-dCas9 nucleases[10,29]; inducible assembly of split Cas9[30–33]; Cas9 PAM variants with enhanced specificity[34]; and delivery of Cas9:sgRNA ribonucleoprotein complexes[35–37]. However, it remains unknown whether these variations can restrict cleavage to a single site within the human genome over a large cohort of treated cells[12,38]. In addition, some of the most promising approaches (e.g. paired nickases or dimeric FokI-dCas9) restrict the

targetable sequence space by requiring the proximity of two sequences compatible with Cas9 recognition.

We envisioned an improved Cas9 platform, where the precision of target recognition would be augmented by the incorporation of a programmable DNA-binding domain (pDBD), such as Cys_2-His_2 zinc finger protein (ZFPs)[39] or transcription-activator like effectors (TALEs)[40] (Fig. 1a and Supplementary Fig. 2). Both of these pDBD platforms can be programmed to recognize nearly any sequence within the genome[39–42]. Indeed, pDBDs have been employed with great success as targeting domains for programmable nucleases by incorporating non-specific FokI nuclease domain (ZFNs[39] and TALENs[40]) or sequence-specific nuclease domains (e.g. megaTAL[43]). One favorable characteristic of pDBDs is their inherent modularity whereby specificity and affinity can be rationally tuned by adjusting the number and composition of incorporated modules and the linkage between modules[44,45]. Here, we demonstrate that the fusion of a pDBD to a mutant SpCas9 with attenuated DNA-binding affinity generates a chimeric nuclease with 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.

Results

Defining the properties of the SpCas9-pDBD framework

To define the parameters necessary for the function of a SpCas9-pDBD chimera, we assayed cleavage of a Cas9:sgRNA target site with a suboptimal nAG PAM using a plasmid reporter assay[46]. We examined the ability of a pDBD fusion (ZFP or TALE) to SpCas9 to enhance nuclease activity when the pDBD binding sites are located at different positions and orientations relative to the Cas9 target site (Fig. 1b). In pilot experiments, the most robust activity was observed using a C-terminal fusion of a ZFP or a TALE to SpCas9 when the pDBD binding sites are 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 a nAG PAM target to a level comparable to wild-type SpCas9 activity on a nGG PAM (Fig. 1b) while being expressed at similar levels (Supplementary Fig. 3). SpCas9-pDBD nuclease activity remained dependent on the length of the guide sequence (Supplementary Fig. 4), confirming that the chimera retains 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 displayed high activity for nAG, nGA, nGC as well as the standard nGG PAM (Fig. 1c and Supplementary Fig. 5). Accounting for reverse complements of the functional PAM elements, the SpCas9-pDBD chimeras can recognize seven of the 16 possible dinucleotide sequence combinations. The increased targeting range for SpCas9-pDBDs was also observed at genomic target sites (Fig. 1d, e). Because of the smaller size of SpCas9-ZFPs relative to SpCas9-TALEs - conferring advantages for certain viral delivery systems[47] - we have focused primarily on SpCas9-ZFP chimeras for the immediate development of this platform.

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 achieve this, we attenuated the DNA-binding affinity of SpCas9 by independently mutating the key PAM recognition residues (Arg1333 and Arg1335)[7] to either Lysine or Serine (Fig. 2a and Supplementary Fig 6). In the plasmid reporter assay, all four mutations reduced 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]) (Fig. 2b). Interestingly, 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[7] (*i.e.* R1333 mutants prefer nnG PAMs, whereas the R1335K mutant prefers nGn PAMs; Supplementary Fig. 6). 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]) displayed only background activity, which could be restored to wild type levels in the presence of a ZFP fusion (Fig. 2c,d and Supplementary Fig. 7 and 8). To confirm that the ZFP-dependent restoration of activity is general, we assessed the nuclease activity of three additional SpCas9[MT3]-ZFP fusions, two of which restore nuclease function (Supplementary Fig. 9 and Supplementary Table 1). Thus, altering the affinity of PAM recognition through mutation generates SpCas9 variants that are dependent on the attached pDBD for efficient function. This pDBD dependence establishes a third stage of target site licensing for our SpCas9[MT3]-pDBDs, which should increase their precision.

Assessing the precision of SpCas9-ZFP fusions

Next we compared the precision of SpCas9-ZFPs to SpCas9 using sgRNAs with previously defined off-target sites[14,25]. We programmed three different four-finger ZFPs to recognize 12 base pair sequences neighboring the TS2, TS3 or TS4 sgRNA target sites for use as SpCas9[MT3]-ZFP fusions (Fig. 3a). The activity of SpCas9, SpCas9[MT3] and SpCas9[MT3]-ZFP[TS2:TS3:TS4] with the corresponding sgRNA was compared at each target site. In all cases SpCas9[MT3] dramatically decreased cleavage efficiencies, which were restored by the cognate ZFP fusion (Fig. 3b). The activity of SpCas9[MT3]-ZFP was dependent on the presence of both a cognate sgRNA and ZFP (Fig. 3c). Consistent with the dependence on ZFP binding, truncation of one zinc finger from either end of ZFP[TS3] reduced the activity of SpCas9[MT3]-ZFP[TS3] at the TS3 target site, and the removal of two zinc fingers abrogated activity (Supplementary Fig. 10). The additional stage of target site licensing supplied by the pDBD dramatically increased the precision of SpCas9[MT3]-ZFP[TS3] relative to wild type SpCas9; lesion rates at the most active off-target site (OT3-2) for sgRNA[TS3] were 22% by T7EI assay with wild type Cas9, but were undetectable with SpCas9[MT3]-ZFP[TS3] (Fig. 3c). We also programmed two TALE arrays to target SpCas9[MT3] to TS3 and TS4 (TALE[TS3] and TALE[TS4]). Nuclease activity at the TS3 site but not TS4 can be restored by the related SpCas9[MT3]-TALE fusion (Supplementary Fig. 11).

To examine the catalytic tolerance of the SpCas9^{MT3}-ZFP^{TS3}:sgRNA complex to mismatches between the guide and a target sequence, we utilized a set of guides that progressively shift blocks of 2 base mismatches from the 5′ to the 3′ end of the guide sequence. SpCas9^{MT3}-ZFP^{TS3} has a lower tolerance for mismatches between the guide and target site relative to SpCas9^{WT}, whereas SpCas9^{WT}-ZFP^{TS3} appears to modestly increase the tolerance for mismatches (Fig. 3d and Supplementary Table 2). SpCas9^{MT3}-ZFPs also exhibit reduced activity with truncated sgRNAs[25] (Supplementary Fig. 12), consistent with the requirement for a higher degree of guide-target site complementarity to achieve efficient cleavage.

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$^{TS2:TS3:TS4}$ (ref. [14,25]), as well as several additional genomic loci that have favorable ZFP$^{TS2:TS3:TS4}$ recognition sites and have some complementarity to the TS2, TS3 or TS4 guide sequences (43 total; Supplementary Tables 3 and 4). 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 (Fig. 4a). In most cases, utilizing SpCas9^{MT3}-ZFP$^{TS2:TS3:TS4}$ reduced lesion rates at off-target sites to background levels resulting in improvements in the Specificity Ratio of up to 150-fold (Fig. 4b). Only one off-target site (OT2-2), which has a neighboring sequence that is similar to the expected ZFP^{TS2} recognition sequence (Supplementary Fig. 13), still displays high lesion rates. One other site (OT2-6), displays some residual activity both for SpCas9^{MT3} and SpCas9^{MT3}-ZFP^{TS2} that is above the background error rate within our sequencing data. Overall, these data demonstrate a dramatic enhancement in precision for SpCas9^{MT3}-ZFPs relative to 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 display high activity at the target site (Fig. 4c). However, these truncations reduced or eliminated off-target activity at OT2-2, reflecting a profound improvement in the precision of SpCas9^{MT3}-ZFP^{TS2} (Fig. 4c and Supplementary Fig. 14). Similarly, utilization of a ZFP (TS2*) that recognizes an alternate sequence neighboring the TS2 guide target site also abolishes off-target activity at OT2-2, confirming that cleavage by SpCas9^{MT3} at this off-target site is ZFP dependent (Fig. 4c & Supplementary Fig. 4). Given the improvements in precision realized by 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 employed GUIDE-seq[17] to provide an unbiased assessment of the propensity for SpCas9^{MT3}-ZFPs to cleave at alternate off-target sites within the genome. Using a modified version of the original protocol and bioinformatics pipeline, we assessed genome-wide DSB induction by SpCas9 and the SpCas9^{MT3}-ZFP$^{TS2:TS3:TS4}$ (see Methods). This analysis

reveals a dramatic enhancement of the precision of the SpCas9^{MT3}-ZFPs for all three target sites (Fig. 5 and Supplementary Table 5). For SpCas9^{MT3}-ZFP^{TS3} and SpCas9^{MT3}-ZFP^{TS4} we did not detect nuclease dependent-oligonucleotide capture at any site besides the target site. For SpCas9^{MT3}-ZFP^{TS2}, which retains two active off-target sites that overlap with SpCas9, there is a dramatic reduction in cleavage activity at all of the alternate sequences. In addition there is one new weak off-target site (OTG2-42) for SpCas9^{MT3}-ZFP^{TS2}. These data demonstrate that the utilization of the SpCas9^{MT3}-ZFP fusion reduces cleavage at wild type SpCas9 off-target sites without generating a new class of highly active ZFP-mediated off-target sites.

Discussion

Our analysis of the activity of SpCas9-pDBD chimeras provides important 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 efficient utilization 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 likely reflects the bypass of a kinetic barrier to R-loop formation that follows PAM recognition, as proposed by Seidel and colleagues[6]. We believe that the pDBD tethering of SpCas9 achieves activity at a target site containing a sub-optimal PAM by increasing the effective concentration of SpCas9 around the target site and hence, stabilizing the SpCas9-PAM interaction[48]. 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 can overcome this barrier at sub-optimal PAMs. Our data also support an allosteric licensing mechanism, as described by Doudna and colleagues[5], which likely restricts 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 (Fig. 3d and Supplementary Fig. 12) further supports the interplay between PAM recognition and guide complementarity in the licensing of nuclease activity.

We find that mutations to the SpCas9 PAM interacting domain introduce a third stage of licensing (pDBD site recognition) for efficient target site cleavage within the SpCas9^{MT}-pDBD system (Fig. 5e). The weakened interaction between mutant Cas9 and the PAM sequence now necessitates increased effective concentration for nuclease function that is achieved by the high affinity interaction of the tethered pDBD with its target site. This combination dramatically improves precision as assessed using targeted deep-sequencing and GUIDE-seq analysis. Compared with the previous GUIDE-seq analysis of the TS2, TS3 and TS4 targets for SpCas9, we detect five, three and three of the top 5 off-target sites that were previously described[17]. The discrepancy between these studies could be due to our lower sequencing depth, the use of an alternate cell line, or different delivery methods. Nonetheless, from our analysis we can exclude the presence of a new class of highly active off-target sites that are generated by the fusion of the ZFP to Cas9.

This system has important advantages over other previously described Cas9 variant systems that improve precision[10],[25–30]. The SpCas9^{MT}-pDBD system increases the targeting range of the nuclease by expanding the repertoire of highly active PAM sequences. This is in contrast to dimeric systems (dual nickases or FokI-dCas9 nucleases) that have a more restricted targeting range due to 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 also 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 have the optimal balance of specificity and affinity. In the case of SpCas9^{MT3}-ZFP^{TS3} this proved prudent (Supplementary Fig. 10). However for SpCas9^{MT3}-ZFP^{TS2} improved precision was achieved by utilizing a three finger ZFP, demonstrating the flexibility provided by modular pDBDs. (See the Supplementary Discussion for more details on ZFP design for Cas9-ZFPs and 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 tuning the pDBD, 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 Cas9 orthologs from other species that have superior characteristics for gene therapy applications (e.g. more compact Cas9 nucleases[49],[50] 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 provides a potent platform for the optimization of nuclease-based reagents that cleave a single target site in the human genome.

Methods

Plasmid Constructs

Our SpCas9-pDBD experiments employed the following plasmids: All sgRNAs are expressed via a U6 promoter from pLKO1-puro[51]. All SpCas9 and SpCas9-DBD fusions are expressed via pCS2-Dest gateway plasmid under chicken beta-globin promoter[52]. ZFPs were assembled as gBlocks (Integrated DNA Technologies) from finger modules based on previously described recognition preferences[53],[54]. ZFPs were cloned into a pCS2-Dest-SpCas9 plasmid backbone cloned thorough BspEI and XhoI sites. TALEs were assembled via golden gate assembly[55] into our JDS TALE plasmids[56]. Assembled TALEs were cloned into BbsI digested pCS2-Dest-SpCas9-TALEntry backbone through Acc65I and BamHI sites. Sequences of the SpCas9-pDBDs are listed in Supplementary Figure 15, and these plasmids will be deposited at addgene for distribution to the community. Plasmid reporter assays of nuclease activity utilized the restoration of GFP activity through SSA-mediated repair of an inactive GFP construct using the M427 plasmid developed by the Porteus laboratory[46]. SpCas9 target sites were cloned into plasmid M427 via ligation independent methods following SbfI digestion. Mutations in the PAM interacting domain of SpCas9 were generated by cassette mutagenesis.

Cell Culture Assay

Human Embryonic Kidney (HEK293T) cells were obtained from our collaborator M. Green (UMass Medical School) are cultured in high glucose DMEM with 10% FBS and 1% Penicillin and Streptomycin (Gibco) in a 37°C incubator with 5% CO_2. These cells were not verified nor tested for mycoplasma contamination. For transient transfection, we used early to mid-passage cells (passage number 5–25). Approximately 1.6×10^5 cells are transfected with 50 ng SpCas9-pDBD-expressing plasmid, 50 ng sgRNA expressing plasmid and 100 ng mCherry plasmid via Polyfect transfection reagent (Qiagen) in 24-well format according to the manufacturer's suggested protocol. For SSA-reporter assay, 150 ng M427 SSA-reporter plasmid is also included in the co-transfection mix.

Western Blot

HEK293T cells are transfected with 500 ng Cas9 and 500 ng sgRNA expressing plasmid in a 6-well plate by Lipofectamine 3000 transfection reagent (Invitrogen) according to manufacturer's suggested protocol. 48 hours after transfection, cells are harvested and lysed with 100 µl RIPA buffer. 8ul of cell lysate is used for electrophoresis and blotting. The blots are probed with anti-HA (Sigma #H9658) and anti alpha-tubilin (Sigma #T6074) primary antibodies; then HRP conjugated anti-mouse IgG (Abcam #ab6808) and anti-rabbit IgG secondary antibodies, respectively. Visualization employed Immobilon Western Chemiluminescent HRP substrate (EMD Millipore #WBKLS0100).

Flow cytometry Reporter Assay

48 hours post-transfection cells are trypsinized and harvested into a microcentrifuge tube. Cells are centrifuged at 500*g for 2 minutes, washed once with 1 x PBS, recentrifuged at 500*g for 2 minutes and resuspended in 1 x PBS for flow cytometry (Becton Dickonson FACScan). For FACS analysis, 10000 events are counted from each sample. To minimize the effect of differences in the efficiency of transfection among samples, cells are initially gated for mCherry-expression, and the percentage of EGFP expressing cells (nuclease positive events) are quantified within mCherry positive cells. All of the experimental replicates are performed in triplicate on different days with data reported as mean values with error bars indicating the standard error of the mean.

Genomic targeting analysis with T7EI

72 hours post-transfection cells are harvested and genomic DNA is extracted via DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's suggested protocol. 50ng input DNA is PCR-amplified using "T7EI primers" that are specific for each genomic region (Supplementary Table 4) with Phusion High Fidelity DNA Polymerase (New England Biolabs): (98°C, 15s; 67°C, 25s; 72°C, 18s) for 30 cycles. 10 µl of a PCR product is hybridized and treated with 0.5µl T7 Endonuclease I (New England Biolabs) in 1 x NEB Buffer2 for 45 minutes[57]. The samples are run on a 2.5% agarose gel and quantified with ImageJ software[58]. Indel percentages are calculated as previously described[57]. Experiments for T7EI analysis are performed in triplicate on different days with data reported as mean values with standard error of the mean.

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

For generation of each amplicon, we used two-step PCR amplification approach 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 complementary tails to the TruSeq adaptor sequences (Supplementary Table 4). 50 ng input DNA is PCR amplified with Phusion High Fidelity DNA Polymerase (New England Biolabs): (98°C, 15s; 67°C 25s; 72°C 18s) for 30 cycles. 5 μl of each PCR reaction is gel-quantified by ImageJ[58] against a reference ladder and equal amounts from each genomic locus PCR are pooled for each treatment group (15 different treatment groups). The pooled PCR products from each group are run on a 2% agarose gel and the DNA from the expected product size (between 100 and 200 bp) is extracted and purified via QIAquick Gel Extraction Kit (Qiagen). In the second step, the purified pool from each treatment group was amplified with a "universal forward primer and an indexed reverse primer" to reconstitute the TruSeq adaptors (Supplementary Table 4). 2ng of input DNA is PCR amplified with Phusion High Fidelity DNA Polymerase (New England Biolabs): (98°C, 15s; 61°C, 25s; 72°C, 18s) for 9 cycles. 5 μl of each PCR reaction is gel-quantified by ImageJ[58], and then equal amounts of the products from each treatment group are mixed and run on a 2% agarose gel. Full-size products (~250bp in length) are gel-extracted and purified via QIAquick Gel Extraction Kit (Qiagen). The purified library was deep sequenced using a paired-end 150 bp MiSeq run.

Sequences from each genomic locus within a specific index were identified based on a perfect match to the final 11 bp of the proximal genomic primer used for locus amplification (Supplementary Table 6). Insertions or deletions in the SpCas9 target region were defined based on 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)[59], where there were typically 33 bp between these elements in the unmodified locus (Supplementary Table 6). Distances that were greater than expected were binned as "insertions (I)", and distances that were shorter were binned as "deletions (D)". Reads that did not contain the suffix sequence were marked as undefined (U). For some loci the background sequencing error rate is high. For example for OT2-1 a homopolymer sequence in the guide region leads to a high error rate.

All statistical analyses were performed using R, a system for statistical computation and graphics[60]. Log odd ratios of lesion were calculated for the on-target and off-target sites of each individual Cas9 treatment group vs. the untreated control for each of the three independent experiments. T-test was applied to assess whether the log odd ratio is significantly different from 0, i.e., whether there is a significant difference in lesion odds between each individual Cas9 treatment group and the untreated control for the on-target and off-target sites. Odds ratios and their 99% confidence intervals were obtained by taking exponent 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, p-values were adjusted using the Benjamini-Hochberg (BH) method[61]. Only loci that have significant BH-adjusted p-values in the combined data for the treatment group relative to the control are considered significant.

GUIDE-Seq off-target analysis for SpCas9-pDBDs

We performed GUIDE-Seq with some modifications to the original protocol[17]. Importantly, there is an error in the original publication with regards to the GSP1 and GSP2 primer sets, which listed incompatible combinations. It was necessary to properly assort the primer sets for the positive (+) and negative (−) strands to get 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[17] in the following manner: In a 24-well format, HEK293T cells are transfected with 250 ng Cas9, 150 ng sgRNA, 50 ng GFP, and 10 pmol of annealed GUIDE-Seq oligonucleotide using Lipofectamine 3000 transfection reagent (Invitrogen) according to manufacturer's suggested protocol. 48 hours post-transfection, genomic DNA was extracted via DNeasy Blood and Tissue kit (Qiagen) according to the manufacturer's suggested protocol. Library preparations are done with original adaptors according to protocols described by the Joung laboratory[17], where each library was barcoded for pooled sequencing. The barcoded, purified libraries were deep sequenced as a pool using two paired-end 150 bp MiSeq runs.

Reads containing the identical molecular index and identical starting 8 bp elements on the Read1 were pooled into one unique read. The initial 30 bp and the final 50 bp of the unique Read2 sequences were clipped for removal of the adapter sequence and low quality sequences and then mapped to the human genome (hg19) using Bowtie2. Peaks containing mapped unique reads were identified using a pile-up program ESAT (http://garberlab.umassmed.edu/software/esat/) using a window of 25 bp with a 15 bp overlap. Neighboring windows that are on different strands of the genome and less than 50 bp apart were merged using Bioconductor package ChIPpeakAnno[62,63]. Peaks that were present with multiple different guides (hotspots[17]) or do not contain unique reads for both sense and anti-sense libraries[17] were discarded. The remaining peaks were searched for sequence elements that were complementary to the nuclease target site using CRISPRseek[21]. Only peaks that harbor a sequence with less than 7 mismatches to the target site were considered potential off-target sites. These regions are reported in Supplementary Table 5 and the number of reads from the sense and the antisense libraries were combined into the final read number.

CRISPRseek analysis of potential off-target site for SpCas9 sgRNAs

Human hg19 exon and promoter sequences were fetched using Bioconductor packages ChIPpeakAnno[62,63] and TxDb.Hsapiens.UCSC.hg19.knownGene. A subset of 16500 exons and 192 promoter sequences of 2 kb each were selected for sgRNA searching and genome-wide off target analysis using Bioconductor package CRISPRseek[21,22] using the default

settings (both nGG and nAG PAMs are allowed) except BSgenomeName = BSgenome.Hsapiens.UCSC.hg19, annotateExon = FALSE, outputUniqueREs = FALSE, exportAllgRNAs = "fasta" and fetchSequence = FALSE. After excluding sgRNAs with on-target or/and off-targets in the haplotype blocks, there are 124793 unique sgRNAs from exon sequences and 55687 unique gRNA from promoter sequences included in the analysis. Each guide was binned based on either the off-target site with the fewest number of mismatches to the guide sequence or the sum of the off-target scores for the top 10 off-target sites. The fraction of guides in each bin for exons or promoters is displayed as a pie chart.

Reproducibility—No statistical methods were used to predetermine sample size, and the investigators were not blinded to allocation during experiments and outcome assessment.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Doudna JA, Charpentier E. Genome editing. The new frontier of genome engineering with CRISPR-Cas9. Science. 2014; 346:1258096–1258096. [PubMed: 25430774]

2. Sander JD, Joung JK. CRISPR-Cas systems for editing, regulating and targeting genomes. Nature biotechnology. 2014; 32:347–355.

3. Hsu PD, Lander ES, Zhang F. Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell. 2014; 157:1262–1278. [PubMed: 24906146]

4. Jinek M, et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science. 2012; 337:816–821. [PubMed: 22745249]

5. Sternberg SH, Redding S, Jinek M, Greene EC, Doudna JA. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature. 2014; 507:62–67. [PubMed: 24476820]

6. Szczelkun MD, et al. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. Proceedings of the National Academy of Sciences. 2014; 111:9798–9803.

7. Anders C, Niewoehner O, Duerst A, Jinek M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. Nature. 2014; 513:569–573. [PubMed: 25079318]

8. Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. Science. 2015; 348:1477–1481. [PubMed: 26113724]

9. Hsu PD, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology. 2013; 31:827–832.

10. Tsai SQ, et al. Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nature biotechnology. 2014; 32:569–576.

11. Zhang Y, et al. Comparison of non-canonical PAMs for CRISPR/Cas9-mediated DNA cleavage in human cells. Sci Rep. 2014; 4:5405. [PubMed: 24956376]

12. Gabriel R, von Kalle C, Schmidt M. Mapping the precision of genome editing. Nature biotechnology. 2015; 33:150–152.

13. Ledford H. CRISPR, the disruptor. Nature. 2015; 522:20–24. [PubMed: 26040877]

14. Fu Y, et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature biotechnology. 2013; 31:822–826.

15. Lin Y, et al. CRISPR/Cas9 systems have off-target activity with insertions or deletions between target DNA and guide RNA sequences. Nucleic Acids Research. 2014; 42:7473–7485. [PubMed: 24838573]

16. Pattanayak V, et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nature biotechnology. 2013; 31:839–843.

17. Tsai SQ, et al. GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nature biotechnology. 2015; 33:187–197.

18. Frock RL, et al. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nature biotechnology. 2015; 33:179–186.

19. Kim D, et al. Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nature Methods. 2015; 12:237–243. [PubMed: 25664545]

20. Wang X, et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nature biotechnology. 2015; 33:175–178.

21. Zhu LJ, Holmes BR, Aronin N, Brodsky MH. CRISPRseek: A Bioconductor Package to Identify Target-Specific Guide RNAs for CRISPR-Cas9 Genome-Editing Systems. PLoS ONE. 2014; 9:e108424. [PubMed: 25247697]

22. Zhu LJ. Overview of guide RNA design tools for CRISPR-Cas9 genome editing technology. Frontiers in Biology. 2015; 10:289–296.

23. Brunet E, et al. Chromosomal translocations induced at specified loci in human stem cells. Proceedings of the National Academy of Sciences. 2009; 106:10620–10625.

24. Lee HJ, Kim E, Kim JS. Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Research. 2010; 20:81–89. [PubMed: 19952142]

25. Fu Y, Sander JD, Reyon D, Cascio VM, Joung JK. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nature biotechnology. 2014; 32:279–284.

26. Cho SW, et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Research. 2014; 24:132–141. [PubMed: 24253446]

27. Ran FA, et al. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell. 2013; 154:1380–1389. [PubMed: 23992846]

28. Mali P, et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nature biotechnology. 2013; 31:833–838.

29. Guilinger JP, Thompson DB, Liu DR. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nature biotechnology. 2014; 32:577–582.

30. Zetsche B, Volz SE, Zhang F. A split-Cas9 architecture for inducible genome editing and transcription modulation. Nature biotechnology. 2015; 33:139–142.

31. Nihongaki Y, Kawano F, Nakajima T, Sato M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. Nature biotechnology. 2015; 33:755–760.

32. Wright AV, et al. Rational design of a split-Cas9 enzyme complex. Proceedings of the National Academy of Sciences. 2015; 112:2984–2989.

33. Davis KM, Pattanayak V, Thompson DB, Zuris JA, Liu DR. Small molecule-triggered Cas9 protein with improved genome-editing specificity. Nat Chem Biol. 2015; 11:316–318. [PubMed: 25848930]

34. Kleinstiver BP, et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015; 523:481–485. [PubMed: 26098369]

35. Kim S, Kim D, Cho SW, Kim J, Kim JS. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. Genome Research. 2014; 24:1012–1019. [PubMed: 24696461]

36. Ramakrishna S, et al. Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. Genome Research. 2014; 24:1020–1027. [PubMed: 24696462]

37. Zuris JA, et al. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. Nature biotechnology. 2015; 33:73–80.

38. Tsai SQ, Joung JK. What's changed with genome editing? Cell Stem Cell. 2014; 15:3–4. [PubMed: 24996161]

39. Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD. Genome editing with engineered zinc finger nucleases. Nat Rev Genet. 2010; 11:636–646. [PubMed: 20717154]

40. Joung JK, Sander JD. TALENs: a widely applicable technology for targeted genome editing. Nat Rev Mol Cell Biol. 2013; 14:49–55. [PubMed: 23169466]

41. Persikov AV, et al. A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. Nucleic Acids Research. 2015; 43:1965–1984. [PubMed: 25593323]

42. Hubbard BP, et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. Nature Methods. 2015; doi: 10.1038/nmeth.3515

43. Boissel S, et al. megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. Nucleic Acids Research. 2014; 42:2591–2601. [PubMed: 24285304]

44. Khalil AS, et al. A synthetic biology framework for programming eukaryotic transcription functions. Cell. 2012; 150:647–658. [PubMed: 22863014]

45. Meckler JF, et al. Quantitative analysis of TALE-DNA interactions suggests polarity effects. Nucleic Acids Research. 2013; 41:4118–4128. [PubMed: 23408851]

46. Wilson KA, Chateau ML, Porteus MH. Design and Development of Artificial Zinc Finger Transcription Factors and Zinc Finger Nucleases to the hTERT Locus. Mol Ther Nucleic Acids. 2013; 2:e87. [PubMed: 23612114]

47. 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. 2010; 138:485–498. [PubMed: 20084428]

48. Klemm JD, Pabo CO. Oct-1 POU domain-DNA interactions: cooperative binding of isolated subdomains and effects of covalent linkage. Genes & Development. 1996; 10:27–36. [PubMed: 8557192]

49. Hou Z, et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proceedings of the National Academy of Sciences. 2013; 110:15644–15649.

50. Ran FA, et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature. 2015; 520:186–191. [PubMed: 25830891]

51. Kearns NA, et al. Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development. 2014; 141:219–223. [PubMed: 24346702]

52. Villefranc JA, Amigo J, Lawson ND. Gateway compatible vectors for analysis of gene function in the zebrafish. Dev Dyn. 2007; 236:3077–3087. [PubMed: 17948311]

53. Gupta A, et al. An optimized two-finger archive for ZFN-mediated gene targeting. Nature Methods. 2012; 9:588–590. [PubMed: 22543349]

54. Zhu C, et al. Using defined finger-finger interfaces as units of assembly for constructing zinc-finger nucleases. Nucleic Acids Research. 2013; 41:2455–2465. [PubMed: 23303772]

55. Cermak T, et al. Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. Nucleic Acids Research. 2011; 39:e82–e82. [PubMed: 21493687]

56. Kok FO, Gupta A, Lawson ND, Wolfe SA. Construction and application of site-specific artificial nucleases for targeted gene editing. Methods Mol Biol. 2014; 1101:267–303. [PubMed: 24233786]

57. Gupta A, et al. Targeted chromosomal deletions and inversions in zebrafish. Genome Research. 2013; 23:1008–1017. [PubMed: 23478401]

58. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. Nature Methods. 2012; 9:671–675. [PubMed: 22930834]

59. Gupta A, Meng X, Zhu LJ, Lawson ND, Wolfe SA. Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. Nucleic Acids Research. 2011; 39:381–392. [PubMed: 20843781]

60. Ihaka R, Gentleman R. R: A Language for Data Analysis and Graphics. Journal of Computational and Graphical Statistics. 1996; 5:299–314.

61. 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. 1995; 57:289–300.

62. Zhu LJ, et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics. 2010; 11:237. [PubMed: 20459804]

63. Zhu, LJ. Methods in Molecular Biology. Lee, T-L.; Shui Luk, AC., editors. Vol. 1067. Humana Press; 2013. p. 105-124.
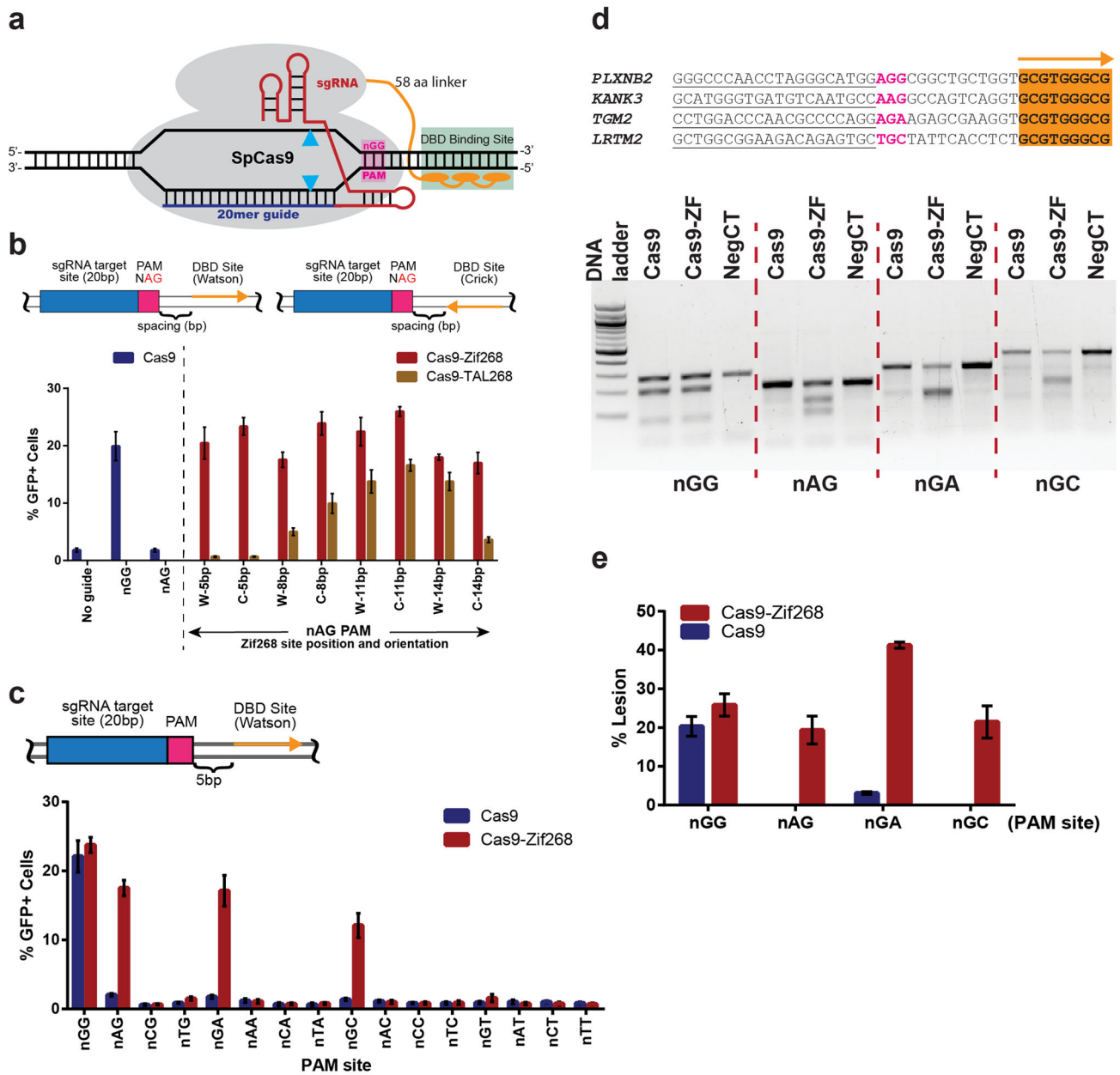
Figure 1.

Development of a SpCas9-pDBD framework. (**a**) Schematic of the SpCas9:sgRNA system fused to a pDBD (orange) that recognizes a binding site 3′ to the PAM. (**b**) (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 is represented by an orange arrow (Watson or Crick). (Bottom) Activity profile of SpCas9 (blue, on an nGG or nAG PAM), SpCas9-Zif268 (red, nAG PAM) or SpCas9-TAL268 (brown, nAG PAM) in the GFP reporter assay on a common sgRNA target site. pDBD site orientation is either Watson (W) or Crick (C), and spacing is 5, 8, 11 or 14 bp from the PAM. (**c**) Activity profile of SpCas9 (blue) or SpCas9-Zif268 (red) in the GFP

reporter assay on a common target site with different PAM sequences and a neighboring Zif268 site. (**d**) (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 cuts efficiently only at the target site with a nGG PAM, but SpCas9-Zif268 cuts efficiently at additional target sites with nAG, nGA or nGC PAMs. Genomic regions were PCR-amplified, and lesions (indicating cleavage and mutagenic NHEJ) were detected by T7 Endonuclease I (T7EI) assay. (**e**) Quantification of lesion frequencies from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate standard error of the mean.
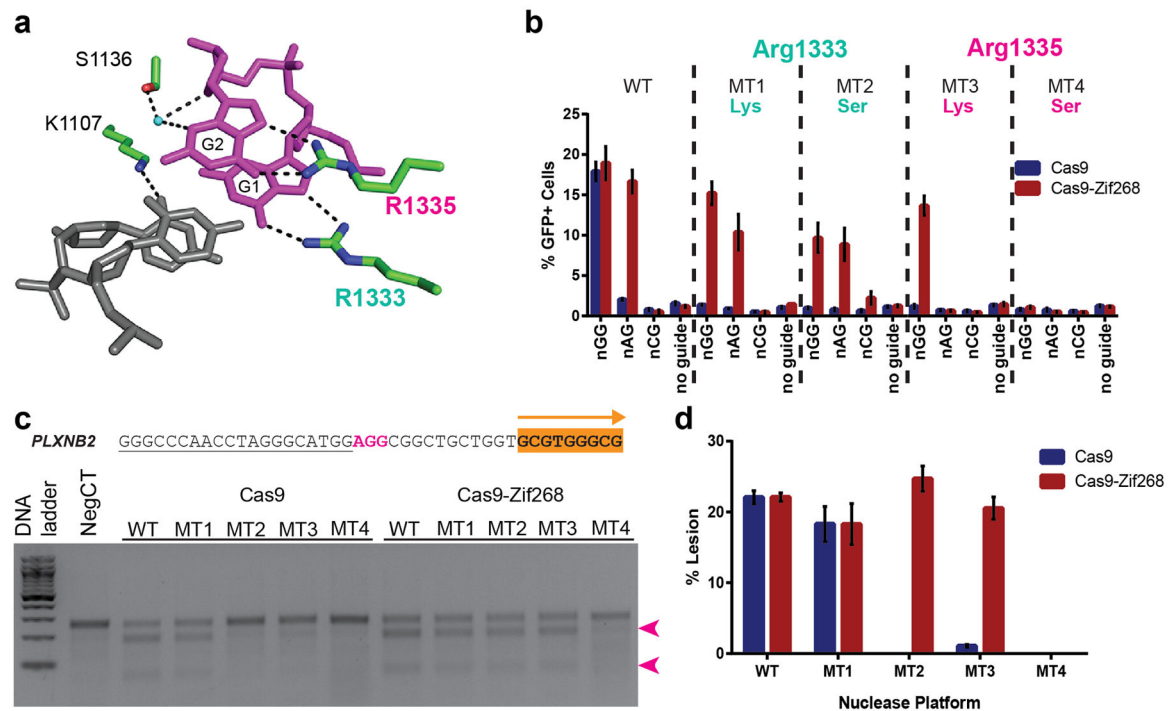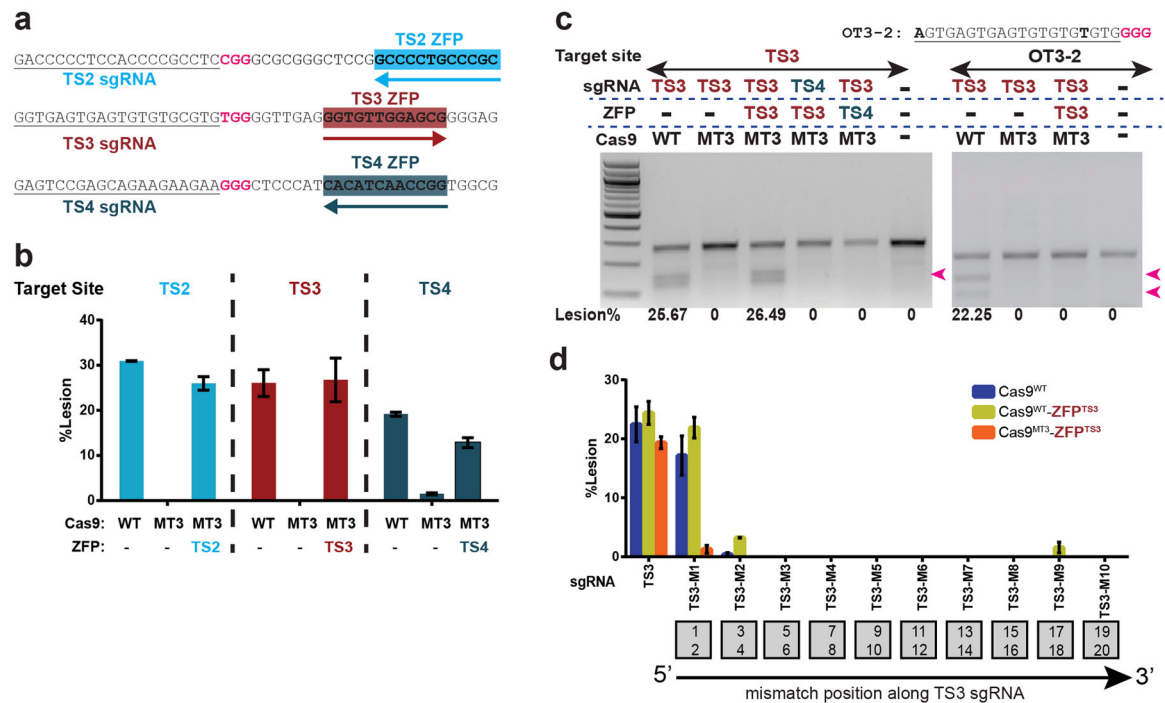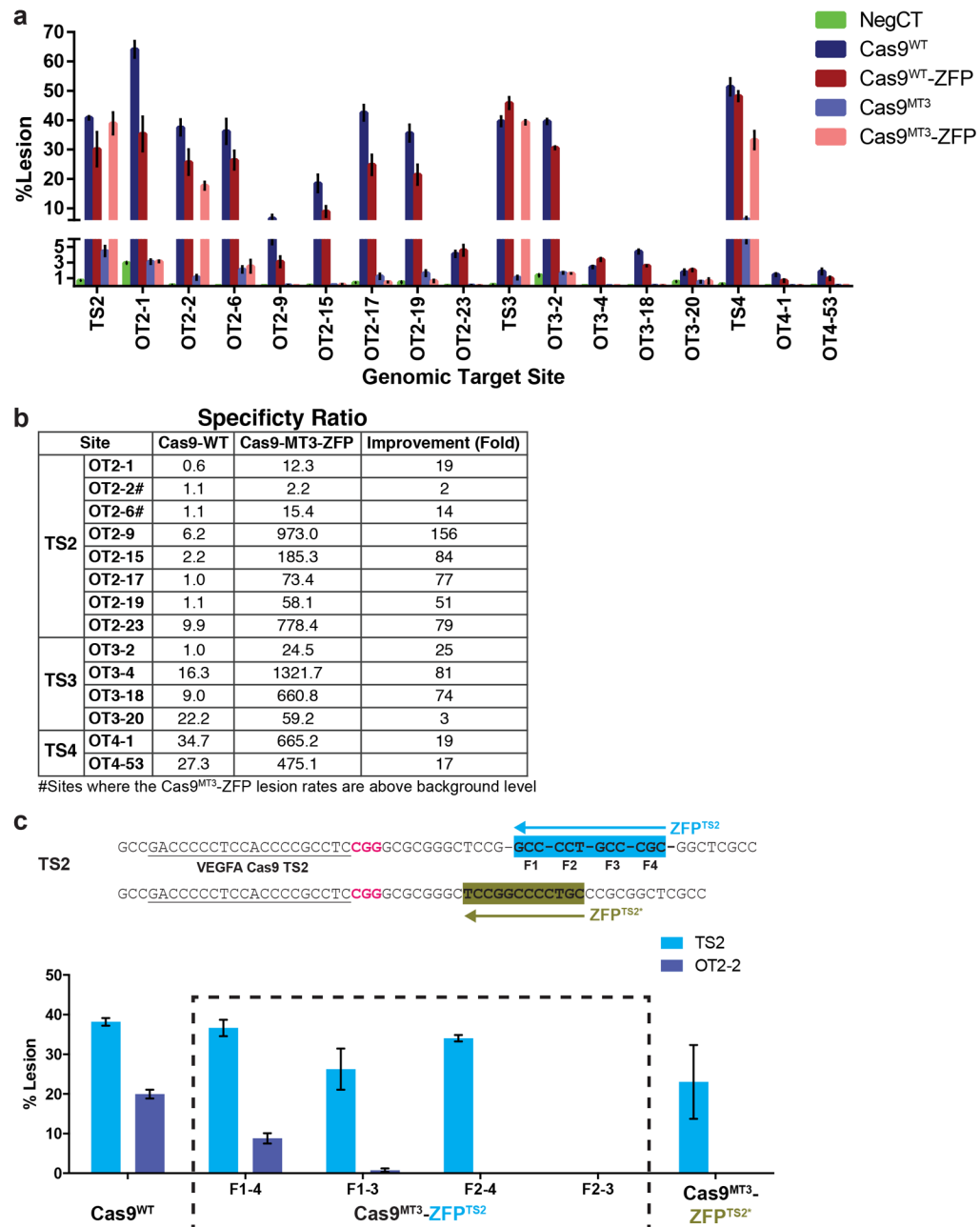
Figure 2.
Attenuating nuclease activity of SpCas9. (**a**) Four PAM-interacting amino acids neighboring the nGG PAM (magenta) in the structure of SpCas9[7]. Arginines at positions 1333 and 1335 were mutated to attenuate DNA-binding affinity of SpCas9. (**b**) Activity profile of SpCas9 (blue) or SpCas9-Zif268 (red) bearing lysine or serine substitutions at positions 1333 or 1335 in the PAM interaction domain in comparison to wild-type (WT) SpCas9. Reporter assays were performed in HEK293T cells. Bar heights represent means from three independent biological replicates performed on different days. Error bars indicate standard error of the mean. (**c**) T7 Endonuclease I (T7EI) assays on PCR products spanning a genomic target site (underlined) with an NGG PAM (magenta) and neighboring Zif268 site (orange) for SpCas9 or SpCas9 mutants with or without a Zif268 fusion. For SpCas9[MT2] & SpCas9[MT3], robust nuclease activity is only observed when Zif268 is fused to the C-terminus. The gel image is representative of T7EI assays at this genomic target site, where cleaved products are noted by magenta arrowheads. (**d**) Quantification of average T7EI-based lesion rates at the *PLXNB2* locus from three independent biological replicates performed on different days in HEK293T cells (Supplementary Fig. 7). Error bars indicate standard error of the mean.

Figure 3.

SpCas9[MT]-ZFP chimeras have improved precision. (**a**) Sequences of Target Site 2 (TS2), Target Site 3 (TS3) and Target Site 4 (TS4) for the SpCas9:sgRNAs described by Joung and colleagues[14,25]. The 12 bp ZFP binding sites for TS2, TS3 and TS4 are highlighted in cyan, red and teal, respectively, with the arrow indicating the strand that is bound. (**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. Error bars indicate standard error of the mean. (**c**) Representative T7EI assay comparing lesion rates at TS3 and off-target site 2 (OT3-2)[25] for various SpCas9-chimera:sgRNA combinations. The activity at the target site for SpCas9[MT3]-ZFP is dependent on the cognate sgRNA and ZFP, where SpCas9[MT3]-ZFP[TS3] can discriminate between TS3 and OT3-2. (**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 within the guide sequence targeting the TS3 site (Supplementary Table 2). (Top) T7EI assay data from PCR products spanning TS3 site in three independent biological replicates performed on different days in HEK293T cells. Error bars indicate standard error of the mean. (Bottom) Schematic indicating the position of the dinucleotide mismatches across the guide sequence. SpCas9[MT3]-ZFP[TS3] displays superior discrimination to SpCas9 for dinucleotide mismatches in the sgRNA recognition sequence.

Figure 4.
Deep sequencing analysis of SpCas9^{MT3}-ZFP chimera precision. (**a**) Lesion rates for target sites and off-target sites with statistically significant activity (Supplementary Table 3) assayed by deep sequencing PCR products spanning each genomic locus for SpCas9 (blue), SpCas9^{MT3} (light blue), SpCas9^{WT}-ZFP (red) and SpCas9^{MT3}-ZFP (pink). Error bars indicate standard error of the mean. (**b**) Improvement in precision of SpCas9^{MT3}-ZFP relative to SpCas9^{WT} as measured by the relative Specificity Ratio of 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 alter the number of zinc fingers or change them completely (TS2*). The binding site for the 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 impacts target site activity, but it dramatically improves precision. Data are from three independent biological replicates performed on different days in HEK293T cells (Supplementary Figure 14). Error bars indicate standard error of the mean.
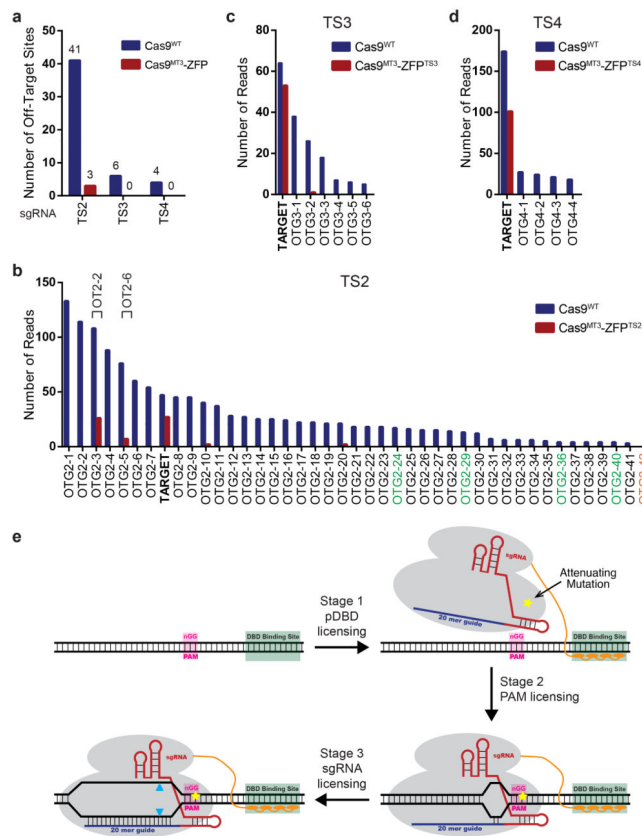
Figure 5.
Genome-wide off-target analysis of SpCas9[MT3]-ZFPs by GUIDE-seq[17]. (**a**) Number of off-target sites with nuclease activity detected for SpCas9[WT] (blue) and SpCas9[MT3]-ZFP (red) with TS2, TS3 and TS4 guides. (**b–d**) Number of unique reads captured by GUIDE-seq for nuclease active sites within the genome (TS2, TS3 or TS4 target site (bold) and off-target sites). Previously defined off-target sites are colored black[14,17] and potential new off-target sites that were identified in this analysis are colored green for SpCas9[WT] or orange for SpCas9[MT3]-ZFP. Some sites (e.g. OGT2-10 & OGT2-20) contain only reads from a single library for SpCas9[MT3]-ZFP, so are not binned as off-target sites in Fig. 5A. Detailed information the sites and counts are presented in Supplementary Table 5. (**e**) Model of the three stages of target site licensing that are necessary for SpCas9[MT3]-pDBD to cleave DNA. Due to the modification of SpCas9 (mutation indicated by yellow star), 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.