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

Mehmet Fatih Bolukbasi1,2, Ankit Gupta1, Sarah Oikemus1, Alan G Derr3, Manuel Garber3,4, Michael H Brodsky1, Lihua Julie Zhu1,3,4 & Scot A Wolfe1,2

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.

The CRISPR-Cas9 genome engineering system is revolutionizing biological sciences owing to its simplicity and efficacy1–3. The most commonly studied Cas9 nuclease (SpCas9) originates from Streptococcus pyogenes4. SpCas9 and its associated guide RNA license a DNA sequence for cleavage on the basis of two stages of sequence interrogation4–8 (Supplementary Fig. 1): (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)4, the primary constraint on Cas9 targeting is the presence of a compatible PAM element4,9,10. The PAM-interacting domain of wild-type SpCas9 preferentially recognizes an NGG element4, although it can inefficiently utilize other PAM sequences (for example, NAG and NGA)9,11. The simplicity of the SpCas9-sgRNA system allows facile editing of genomes in a variety of organisms and cell lines1–3.

The precision of SpCas9 is suboptimal for most gene therapy applications involving editing of a large population of cells12,13. Numerous studies have demonstrated that SpCas9 can cleave the genome at unintended sites9,14–20, with some guides acting at more than 100 off-target sites17. 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 activity17–20. For some guides, off-target sites that differ by up to six nucleotides can be inefficiently cleaved17–20, and bulges can be accommodated in the sgRNA:DNA heteroduplex15. 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 CRISPRseek21,22. 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 (Supplementary Fig. 1) 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 rearrangements17,18,23,24 (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 composition25,26, the use of a pair of Cas9 nickases26–28 or FokI-dCas9 nucleases10,29, inducible assembly of split Cas9 (refs. 30–33), Cas9 PAM variants with enhanced specificity34 and the delivery of Cas9:sgRNA ribonucleoprotein complexes35–37. However, it remains unknown whether these variations can restrict cleavage to a single site in the human genome over a large cohort of treated cells12,38. 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.

We envisioned an improved Cas9 platform in which the precision of target recognition was augmented by the incorporation of pDBDs such as Cys2-His2 zinc-finger proteins39 (ZFPs) or transcription activator–like effectors40 (TALEs) (Fig. 1a and Supplementary Fig. 2). Both of these pDBD platforms can be...
programmed to recognize nearly any sequence in the genome\textsuperscript{39–42}. 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\textsuperscript{39} and TALE nucleases\textsuperscript{40}) or sequence-specific nuclease domains (e.g., megaTAL\textsuperscript{41}). 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\textsuperscript{44,45}. 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.

**RESULTS**

**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\textsuperscript{46}. 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 (Fig. 1b). 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.E.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 (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 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 (Fig. 1c and Supplementary Fig. 5). 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 (Fig. 1d,e). Because of the smaller size of SpCas9-ZFPs relative to SpCas9-TALEs, which confers advantages for certain viral-delivery systems\textsuperscript{47}, we 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 test this, we attenuated the DNA-binding affinity of SpCas9 by independently mutating the key PAM recognition residues (Arg1333 and Arg1335) to either lysine or serine (Fig. 2a and Supplementary Fig. 6). 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\textsuperscript{R1335S}) (Fig. 2b). We found that R1335K (SpCas9\textsuperscript{R1335K}) lacked activity with the NAG PAM even as a SpCas9-ZFP fusion. This prompted a broader assessment of PAM...

---

**Figure 1** Development of an 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 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. (c) 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. (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 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. Genomic regions were PCR amplified, and lesions (indicating cleavage and mutagenic non-homologous end joining) were detected by T7EI assay. NegCT indicates untreated cells. (e) Quantification of lesion frequency from three independent biological replicates performed on different days in HEK293T cells. Error bars indicate 1 s.e.m.
Figure 2 | Attenuating the nuclease activity of SpCas9. (a) Four PAM-interacting amino acids neighboring the NGG PAM (magenta) in the structure of SpCas9 (ref. 7). Arginines at positions 1,333 and 1,335 were mutated to attenuate the DNA-binding affinity of SpCas9. (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. Reporter assays were performed in HEK293T cells. Bars represent means from three independent biological replicates performed on different days. (c) Quantification of average T7EI-based lesion rates at the PLXNB2 locus, which contains an NGG PAM flanking a Zif268-recognition sequence, from three independent biological replicates performed on different days in HEK293T cells (Supplementary Fig. 7). Error bars indicate ± s.e.m.

Figure 3 | SpCas9mt3-ZFP chimeras have improved precision. (a) Sequences of TS2, TS3 and TS4 for the SpCas9:sgRNAs described by Joung and colleagues14,25. 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, SpCas9mt3 and SpCas9mt3-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 (ref. 25) for various SpCas9-chimera: sgRNA combinations. The activity at the target site for SpCas9mt3-ZFP was dependent on the cognate sgRNA and ZFP, where SpCas9mt3-ZFPmt3 was able to discriminate between TS3 and OT3-2. Magenta arrowheads indicate T7EI cleavage products consistent with target-site lesions. (d) Genomic target-site cleavage activity by SpCas9, SpCas9WT-ZFPmt3 and SpCas9mt3-ZFPmt3 in response to dinucleotide mismatches placed at different positions in the guide sequence targeting the TS3 site (Supplementary Table 2). 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. SpCas9mt3-ZFPmt3 demonstrated superior discrimination relative to SpCas9 for dinucleotide mismatches in the sgRNA recognition sequence. Error bars indicate ± s.e.m.

specificity for the three active SpCas9-ZFP mutants, which revealed a preference for alternate PAMs that preserved the remaining arginine-guanine interaction7 (i.e., R1333 mutants preferred NGN PAMs, whereas the R1335K mutant preferred NGN PAMs; Supplementary Fig. 6). The activity of each SpCas9 mutant was also characterized on compatible genomic target sites with an NGG PAM. R1333K (SpCas9mt3) retained independent activity on a subset of target sequences, whereas R1333S (SpCas9mt3s) and R1333K (SpCas9mt3k) showed only background activity, which could be restored to wild-type levels in the presence of a ZFP fusion (Fig. 2c and Supplementary Figs. 7 and 8). To confirm that the ZFP-dependent restoration of activity was general, we assessed the nuclease activity of three additional SpCas9mt3-ZFP fusions, two of which restored nuclease function (Supplementary Fig. 9 and Supplementary Table 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 SpCas9mt3-pDBDs, which we believed would increase their precision.

Assessing the precision of SpCas9mt3-ZFP fusions

Next we compared the precision of SpCas9mt3-ZFPs to that of SpCas9 using sgRNAs with previously defined off-target sites14,25. 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 SpCas9mt3-ZFP fusions (Fig. 3a). The activity of SpCas9, SpCas9mt3 and SpCas9mt3-ZFP with the corresponding sgRNA was compared at each target site. In all cases SpCas9mt3 dramatically decreased the cleavage efficiency, which was restored by the cognate ZFP fusion (Fig. 3b). The activity of SpCas9mt3-ZFP was dependent on the presence of both a cognate sgRNA and a cognate ZFP (Fig. 3c). Consistent with the dependence on ZFP binding, truncation of one zinc finger from either end of ZFPmt3 decreased the activity of SpCas9mt3-ZFPmt3 at TS3, 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 SpCas9MT3-ZFPTS2 relative to that of wild-type SpCas9; lesion rates for sgRNA5TS3 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 SpCas9MT3-ZFPTS2 (Fig. 3c). We also programmed two TALE arrays to target SpCas9MT3 to TS3 and TS4 (TALETS3 and TALETS4). Nuclease activity at TS3, but not at TS4, was restored by the related SpCas9MT3-TALE fusion (Supplementary Fig. 11).

To examine the catalytic tolerance of the SpCas9MT3-ZFPTS3; 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. SpCas9MT3-ZFPTS3 had a lower tolerance for mismatches between the guide and target site than did wild-type SpCas9 (SpCas9WT), whereas SpCas9WT-ZFPTS3 seemed to have a modestly increased tolerance for mismatches (Fig. 3d and Supplementary Table 2). SpCas9MT3-ZFPs also exhibited decreased activity with truncated sgRNAs25 (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 sgRNAsTS2/TS3/TS4 (refs. 14,25), as well as several additional genomic loci with favorable ZFPPTS2/TS3/TS4 recognition sites and some complementarity to the TS2, TS3 or TS4 guide sequences (43 total; Supplementary Tables 3 and 4). We compared the nuclease activity of SpCas9, SpCas9MT3, SpCas9WT-ZFPPTS2/TS3/TS4 and SpCas9MT3-ZFPPTS2/TS3/TS4 at these off-target sites and found that SpCas9MT3-ZFPPTS2/TS3/TS4 dramatically increased the precision of target-site cleavage (Fig. 4a). In most cases, using SpCas9MT3-ZFPPTS2/TS3/TS4 decreased lesion rates at off-target sites to background levels, resulting in up to 150-fold improvements in the specificity ratio (Fig. 4b). Only one off-target site (OT2-2), which had a neighboring sequence similar to the expected ZFPPTS2 recognition sequence (Supplementary Fig. 13), still had high lesion rates. One other site (OT2-6) showed some residual activity for both SpCas9MT3 and SpCas9MT3-ZFPPTS2 that was above the background error rate in our sequencing data. Overall, these data demonstrate a dramatic enhancement in precision for SpCas9MT3-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 SpCas9MT3-ZFPPTS2 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 (Fig. 4c). However, these truncations decreased or eliminated off-target activity at OT2-2, reflecting a profound improvement in the precision of SpCas9MT3-ZFPPTS2 (Fig. 4c and Supplementary Fig. 14). Similarly, utilization of a ZFP (TS2*) that recognized

**Figure 4** 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) (Supplementary Table 3) assayed by deep-sequencing of PCR products spanning each genomic locus for SpCas9WT (blue), SpCas9MT3 (light blue), SpCas9WT-ZFP (red) and SpCas9MT3-ZFP (pink). (b) Specificity ratio of SpCas9MT3-ZFP relative to that of SpCas9WT 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 SpCas9WT and SpCas9MT3-ZFPTS2 variants that altered the number of zinc fingers or changed them completely (TS2*). The binding site for ZFPPTS2* 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 (Supplementary Fig. 14). Error bars indicate ± s.e.m.
an alternate sequence neighboring the TS2 guide target site also abolished off-target activity at OT2-2, confirming that cleavage by SpCas9MT3 at this off-target site was ZFP dependent (Fig. 4c and Supplementary Fig. 4). 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 even 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)\(^{17}\) to provide an unbiased assessment of the propensity for SpCas9MT3-ZFPs to cleave at alternate off-target sites in the genome. Using modified versions of the original protocol and bioinformatics pipeline, we assessed genome-wide double-stranded break induction by SpCas9 and SpCas9MT3-ZFPT52,\(^{2}\) (Online Methods). This analysis revealed dramatically enhanced precision of the SpCas9MT3-ZFPs compared with that of SpCas9 for all three target sites (Fig. 5 and Supplementary Table 5). For SpCas9MT3-ZFPT53 and SpCas9MT3-ZFPT54, we did not detect nuclease-dependent oligonucleotide capture at any site other than the target site. For SpCas9MT3-ZFPT52, 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 SpCas9MT3-ZFPT52. These data demonstrate that use of the SpCas9MT3-ZFP fusion decreased 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 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\(^6\). 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\(^{48}\). 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 will be able to overcome this barrier at suboptimal PAMs. Our data also support an allosteric licensing mechanism, as described by Doudna and colleagues, 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\textsuperscript{MT3}, ZFP\textsuperscript{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 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\textsuperscript{MT}, pDBD system (Fig. 5e). 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\textsuperscript{17}. 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\textsuperscript{10,25–30}. The SpCas9\textsuperscript{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\textsuperscript{MT3}, ZFP\textsuperscript{TS3}, this proved prudent (Supplementary Fig. 10). However, for SpCas9\textsuperscript{MT3}, ZFP\textsuperscript{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 provided in the Supplementary Discussion. 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\textsuperscript{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 provide a potent platform for the optimization of nuclease-based reagents that cleave a single target site in the human genome.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. Illumina Sequencing data have been submitted to the Sequence Read Archive under accession number SRP063361.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank M. Porteus (Stanford Medicine, Stanford, California, USA) for GFP reporter vector M427, N. Rhind for the use of his FACS machine, D. Kinter and the UMass Medical School Deep Sequencing Core for their assistance with the Illumina sequencing, and E. Sontheimer for insightful discussions. All new reagents described in this work have been deposited with the nonprofit plasmid-distribution service Addgene. This work was supported by the US National Institutes of Health (grant R01AI117839 to S.A.W. and J. Luban, grant U01HG007910 to M.G. and J. Luban, and grant R01HL93766 to S.A.W. and N. Lawson).

AUTHOR CONTRIBUTIONS

M.F.B. and A.G. performed all cell-based experiments. A.G.D., M.G. and L.Z. performed the bioinformatic analysis. S.O. and M.H.B. optimized the GFP reporter assay. M.F.B., A.G., L.J.Z. and S.A.W. directed the research and interpreted results. M.F.B., A.G. and S.A.W. wrote the manuscript with input from all the other authors.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
ARTICLES

12. Gabriel, R., von Kalle, C. & Schmidt, M. Mapping the precision of genome editing. Nat. Biotechnol. 33, 150–152 (2015).

13. Ledford, H. CRISPR, the disruptor. Nature 522, 20–24 (2015).

14. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat. Biotechnol. 31, 822–826 (2013).

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 Res. 42, 7473–7485 (2014).

16. Pattanayak, V. et al. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. 31, 839–843 (2013).

17. 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).

18. Frock, R.L. et al. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat. Biotechnol. 33, 179–186 (2015).

19. Kim, D. et al. Digene-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. Nat. Methods 12, 237–243 (2015).

20. 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).

21. 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).

22. Zhu, L.J. Overview of guide RNA design tools for CRISPR-Cas9 genome editing technology. Front. Biol. 10, 289–296 (2015).

23. Brunet, E. et al. Chromosomal translocations induced at specified loci in human stem cells. Proc. Natl. Acad. Sci. USA 106, 10620–10625 (2009).

24. Lee, H.J., Kim, E. & Kim, J-S. Targeted chromosomal deletions in human cells using zinc finger nucleases. Genome Res. 20, 81–89 (2010).

25. 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).

26. 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).

27. Ran, F.A. et al. Double nicking by RNA-guided CRISPR-Cas9 for enhanced genome editing specificity. Cell 154, 1380–1389 (2013).

28. Mali, P. et al. CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. 31, 833–838 (2013).

29. Gullinger, 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).

30. Zetsche, B., Volz, S.E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. Nat. Biotechnol. 33, 139–142 (2015).

31. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. Nat. Biotechnol. 33, 755–760 (2015).

32. Wright, A.Y. et al. Rational design of a split-Cas9 enzyme complex. Proc. Natl. Acad. Sci. USA 112, 2984–2989 (2015).

33. 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).

34. Kleinstiver, B.P. et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature 523, 481–485 (2015).

35. 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).

36. Ramakrishna, S. et al. Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA. Genome Res. 24, 1020–1027 (2014).

37. 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).

38. Tsai, S.Q. & Joung, J.K. What’s changed with genome editing? Cell Stem Cell 15, 3–4 (2014).

39. 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).

40. Joung, J.K. & Sander, J.D. TALENs: a widely applicable technology for targeted genome editing. Nat. Rev. Mol. Cell Biol. 14, 49–55 (2013).

41. Persikov, A.V. et al. A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. Nucleic Acids Res. 43, 1965–1984 (2015).

42. Hubbard, B.P. et al. Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. Nat. Methods 12, 939–942 (2015).

43. Boissel, S. et al. megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. Nucleic Acids Res. 42, 2591–2601 (2014).

44. Khalil, A.S. et al. A synthetic biology framework for programming eukaryotic transcription functions. Cell 150, 647–658 (2012).

45. Meckler, J.F. et al. Quantitative analysis of TALE-DNA interactions suggests polarity effects. Nucleic Acids Res. 43, 4118–4128 (2013).

46. 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).

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 138, 485–498 (2010).

48. 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).

49. Hou, Z. et al. Efficient genome engineering in human pluripotent stem cells using Cas9 from Neisseria meningitidis. Proc. Natl. Acad. Sci. USA 110, 15644–15649 (2013).

50. Ran, F.A. et al. In vivo genome editing using Staphylococcus aureus Cas9. Nature 520, 186–191 (2015).
ONLINE METHODS

Plasmid constructs. Our SpCas9-pDBD experiments used the following plasmids: All sgRNAs were expressed via a U6 promoter from pLKO1-puro51. All SpCas9 and SpCas9-pDBD fusions were expressed via pCS2-Dest gateway plasmid under chicken β-globin promoter52. ZFPs were assembled as gBlocks (Integrated DNA Technologies) on the basis of previously described recognition preferences53,54. ZFPs were cloned into a pCS2-Dest-SpCas9 plasmid backbone cloned through BspEI and XhoI sites. TALEs were assembled via golden gate assembly55 into our JDS TALE plasmids56. Assembled TALEs were cloned into BbsI digested pCS2-Dest-SpCas9-TAL library backbone through Acc65I and BamHI sites. Sequences of the SpCas9-pDBDs are listed in the Supplementary Note, and these plasmids 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 laboratory46. 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.

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% CO2. 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 × 10⁵ 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.

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).

Flow cytometry reporter assay. 48 h after transfection, cells were trypsinized and collected in a microcentrifuge tube. Cells were centrifuged at 500g for 2 min, washed once with 1x PBS, recentrifuged at 500g for 2 min and resuspended in 1x 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 (nuclelease-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.

Genomic targeting analysis with T7EI. 72 h after transfection, cells were harvested and genomic DNA was extracted using 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 (Supplementary Table 4) 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 (ref. 57). The samples were run on a 2.5% agarose gel and quantified with ImageJ software58. Insertion-deletion percentages were calculated as previously described57. Experiments for T7EI analysis were performed in triplicate on different days; data are reported as mean ± s.e.m.

Targeted deep-sequencing-based off-target analysis for SpCas9-pDBDs. For the generation of each ampiclon, 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 adadapter sequences (Supplementary Table 4). 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 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 (Supplementary Table 4). 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, 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 (Supplementary Table 6). 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)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, 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

doi:10.1038/nmeth.3624
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. 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. 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.

GUIDE-seq off-target analysis for SpCas9-pDBDs. We performed GUIDE-seq with some modifications to the original protocol. 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 GATCTCGACGCTCTCCCTGTATTTAATGAATGGATGTGCATATGTTATAAATCCTAGTAAATACCACTATGACAACTCAATTAAAC − Nuclease_off_−_GSP1 GATCTCGACGCTCTCCCTATACC

TGAGTTGTCATATGTTAATAA − Nuclease_off_+_GSP2 CCTCTCTATGGGACGCTCGGTATTAGATGTCATATGTTATAAATACCTAGTAAATACCACTATGACAACTCAATTAAAC − Nuclease_off_−_GSP2 CCTCTCTATGGGACGCTCGGTATTAGATGTCATATGTTATAAATACCTAGTAAATACCACTATGACAACTCAATTAAAC −

In addition, our protocol differed from the published protocol 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, 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. Peaks that were present with multiple different guides (hot spots) or that did not contain unique reads for both sense and antisense libraries were discarded. The remaining peaks were searched for sequence elements that were complementary to the nuclease target site using CRISPRseek. 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 Supplementary Table 5, and the numbers 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 and TxDb. Human 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 with the default settings (both NGG and NAG PAMs were allowed), except BSgenomName = BSgenome. Human 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 (Supplementary Fig. 1).

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