Francisella novicida Cas9 interrogates genomic DNA with very high specificity and can be used for mammalian genome editing

Sundaram Acharya[1,5], Arpit Mishra[1,4,5], Deepanjan Paul[1], Asgar Hussain Ansari[1,2], Mohd. Azhar[1,2], Namrata Sharma[1], Meghali Aich[1,2], Dipanjali Sinha[1,2], Saumya Sharma[1,2], Shivani Jain[1], Arjun Ray[1], Souvik Maiti[1,2,3], Debojyoti Chakraborty[1,2,6,*]

[1]CSIR-Institute of Genomics & Integrative Biology, Mathura Road, Delhi- 110025, India

[2]Academy of Scientific & Innovative Research, Anusandhan Bhawan, New Delhi-110001, India

[3]CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune, 411008, India

[4]Present address: University of Washington School of Medicine, Seattle, USA

[5]These authors contributed equally

[6]Lead contact

[*]Correspondence: debojyoti.chakraborty@igib.in (D.C.)

SUMMARY

Genome editing using the CRISPR Cas9 system has been used to manipulate eukaryotic DNA and make precise heritable changes. Although the widely used *Streptococcus pyogenes* Cas9 (SpCas9) and its engineered variants have been efficiently harnessed for numerous gene-editing applications across different platforms, concerns remain, regarding their putative off targeting at multiple loci across the genome. Here we report that *Francisella novicida* Cas9 (FnCas9) shows a very high specificity of binding to its intended targets and negligible binding to off-

target loci. The specificity is determined by its minimal binding affinity with DNA when up to two mismatches to the target sgRNA are present in the sgRNA:DNA heteroduplex. *In vivo*, FnCas9 can be used for NHEJ mediated gene disruption and HDR mediated genomic integration. We propose that FnCas9 can be used for precise therapeutic genome editing.

SIGNIFICANCE STATEMENT

Therapeutic genome editing has been significantly accentuated by the ease and wide applicability of CRISPR Cas9 based gene correction. However, genomic off targeting has persisted as a major setback for its clinical applications for a large number of genetic disorders. Although high fidelity versions of Cas9 have been rationally designed, they still recognize and bind to off targets without cleaving. In this study, we characterize a naturally occurring orthogonal Cas9 from *Francisella novicida* (FnCas9) that shows negligible binding affinity to off targets that differ by more than two mismatches, rendering it highly specific in target recognition. Finally, we show that FnCas9 can direct both HDR and NHEJ mediated DNA repair, further expanding the toolbox for highly specific gene editing.

INTRODUCTION

The process of introducing changes in the DNA of cells using Clustered regularly interspersed short palindromic repeats (CRISPR)-CRISPR-associated (Cas) proteins has emerged as a powerful technique in molecular biology with potentially far reaching applications in gene therapy (Hille et al., 2018; Hsu et al., 2014; Huang et al., 2018; Knott et al., 2018; Komor et al., 2017; Singh and Chakraborty et al., 2016). The method involves harnessing the prokaryotic type II CRISPR-Cas protein Cas9, which in complex with a single guide RNA (sgRNA), can be directed to any region in the DNA upstream of a protospacer adjacent motif (PAM) with which the sgRNA sequence finds a match (Cong et al., 2013;

Jinek et al., 2012; Sternberg et al., 2014). Upon double stranded cleavage at the target site, the endogenous repair machinery of the cell can be utilized to make nucleotide changes in the DNA (Hille and Charpentier, 2016; Maeder and Gersbach, 2016; Singh et al., 2016; van der Oost et al., 2014). Although several Cas9 proteins recognizing different PAM sequences have so far been reported in literature, only a subset of these have been characterized and have demonstrated genome-editing ability in eukaryotic cells (Adli, 2018; Cebrian-Serrano and Davies, 2017; Esvelt et al., 2013; Komor et al., 2017; Najm et al., 2018).

Cas9 from *Francisella novicida* (FnCas9) is one of the largest Cas9 orthologs and has been shown to predominantly interact with the 5'-NGG-3' PAM motif in DNA (Hirano et al. 2016). The crystal structure of FnCas9 ribonucleoprotein (RNP) in complex with target DNA has revealed both conserved and divergent features of interaction that is unique among the Cas9 enzymes studied. Unlike SpCas9, FnCas9 does not form a bilobed structure, has a different sgRNA scaffold and has been implicated in RNA targeting (Hirano et al., 2016; Price et al., 2015). By structure guided protein engineering, FnCas9 can be made to recognize a 5'-YG-3' PAM (Hirano et al., 2016). Although the protein can efficiently cleave DNA substrates *in vitro*, its *in vivo* activity at several genomic loci is significantly diminished as compared to SpCas9 (Chen et al., 2017; Hirano et. al, 2016).

FnCas9 has been reported to have high intrinsic specificity for its targets, most notably by tolerating only a single mismatch in the sgRNA at the 5' position of the PAM distal end (Chen et al., 2017). This is in stark contrast with SpCas9, which has shown variable levels of off targeting due to tolerance of mismatches predominantly in the 'non-seed' region in the sgRNA, wherever these are encountered in the genome (Hsu et al., 2013). To what extent FnCas9 mediates this high specificity of target interrogation is not known and whether these properties can be harnessed for highly specific genome editing at a given DNA loci has not been investigated. The distinct structural attributes of FnCas9 and

its low tolerance for mismatches led us to investigate its DNA interaction properties and role in genome editing.

RESULTS

FnCas9 shows sequence dependent DNA binding affinity and cleavage kinetics

The interaction of Cas9 to its substrate occurs in a multi-step process involving PAM binding, target unwinding and DNA:RNA heteroduplex formation (Shibata et al., 2017; Singh et al., 2016). As several components in this process are dependent on enzyme conformation, we speculated that FnCas9 might possess DNA interrogation parameters that are intrinsically linked to its structure. To dissect the binding affinity of FnCas9 RNP complex with target DNA, we used MicroScale Thermophoresis (MST) to determine the dissociation constant (K_d) of FnCas9:target interaction. To perform binding affinity measurements, we generated recombinant catalytically *dead* (inactive) FnCas9 (dFnCas9) tagged with green fluorescent protein (GFP) and confirmed its inability to perform *in vitro* cleavage of a DNA substrate (Figure S1A,B). Electrophoretic gel mobility shift experiments (EMSA) concluded that FnCas9 and its sgRNA sequence interaction reaches saturation at a molar ratio of approximately 1:1.5, similar to that reported for SpCas9 binding with its sgRNA (Anders et al., 2015, Sternberg et al., 2014; Figure 1A).

Next, we investigated the binding of dFnCas9: sgRNA RNP complex with two different DNA substrates (*Vegfa3* and *Emx1*_1) and observed a K_d of 150.5 ± 36.5 nM and 77.7 ± 17.2 nM respectively (Figure 1B). To compare the binding of dSpCas9 to the same targets, we generated recombinant dSpCas9-GFP and interrogated the substrates using MST. We observed lower K_d of substrate binding for dSpCas9 for each of the targets (49.9 nM ± 9.4 nM and 10.9 ± 6.2 nM respectively) suggesting that in comparison with dFnCas9, dSpCas9 might have a generally higher affinity for the same substrate sequence (Figure 1B). Notably,

K_d of SpCas9 binding to its substrates as reported in literature using other techniques (such as EMSA, Beacon assays or Active site titration assays) also fall in a similar range (≤ 100nM) as observed under our experimental conditions, suggesting concordance between the different methods used in determining binding affinity (Chen et al., 2017; Mekler et al., 2017; Raper et al., 2018; Richardson et al., 2016; Sternberg et al., 2014). We tested two more substrate sequences (*Emx1_2*, *c-Myc*) with dFnCas9 and once again observed a wide distribution of dissociation constants (K_d 435.4 ± 48.2 nM and 47.9 + 36.9 nM respectively, Figure S1C). Collectively, this suggests that FnCas9 shows a diverse range of binding affinities with different DNA substrates and its affinity for targets is generally lower than that of SpCas9. We then proceeded to investigate if these differences in binding affinity might lead to variability in cleavage efficiencies across different substrates. To investigate substrate cleavage, we purified wild type FnCas9 and SpCas9 and performed *in vitro* cleavage (IVC) assays. When different DNA substrates were incubated with equal amounts of FnCas9 RNP, we observed that cleavage reached completion at different time points (30 min – 2h) for these substrates, suggesting that the sequence of the target might affect the rate of cleavage; SpCas9 however, did not show such variability with any of the targets (Fig 1C). Interestingly, when we took a substrate that showed gradual completion of cleavage (*c-MYC*) and performed IVC with increasing molar concentrations of RNP, we observed that substrate cleavage completion could be shifted to an earlier time-point upon incubation with higher concentration of FnCas9 RNP (Figure S1D). Thus, FnCas9 cleavage efficiency varies proportionally to RNP concentration indicating that it might act as a single turnover enzyme, a feature reported earlier for SpCas9 (Sternberg et al., 2014, Figure S1D). Collectively, these results suggest that FnCas9 has a generally high threshold for substrate recognition and even in the case of complete match of crRNA with its target, exhibits different cleavage kinetics when encountering different target sequences.

More than one mismatch in sgRNA:DNA heteroduplex abolishes FnCas9 cleavage

The specificity of a genome editing protein is guided by a balance between its affinity for target and ability to discriminate off targets. A recent report had suggested that FnCas9 shows higher intrinsic specificity than SpCas9 to its target by showing less tolerance to single mismatches at certain sgRNA positions (Chen et al., 2017). We investigated the *in vitro* cleavage efficiency of both SpCas9 and FnCas9 by systematically changing every base in a given substrate and observed that whereas SpCas9 cleaved nearly equally at all mismatched positions, FnCas9 was less tolerant to single mismatches and was particularly stringent at base positions 19, 18, 17 and 16 at the PAM distal end (Figure 2A) where considerable abolishment of cleavage was observed (~ 56 ± 7%). Interestingly, engineered highly specific Cas9 variants that are able to prevent cleavage at off targets containing PAM distal mismatches adopt a conformational structure that renders the HNH DNA cleavage domain inactive (Chen et al., 2017). Recent biophysical studies using single molecule fluorescence resonance energy transfer (smFRET) have revealed a highly dynamic conformation of spCas9 where allosteric interactions between the PAM distal end and the HNH domain of the Cas9 enzyme renders a cleavage-impaired conformationally 'closed' configuration upon encountering mismatches close to the 5' end of the sgRNA (Osuka et al., 2018; Yang et al., 2018). To dissect if a greater stringency of target recognition could be achieved in case of FnCas9 by increasing the number of mismatches at the 5' end of the substrate, we selected two well-studied loci EMX1 and VEGFA3, amplified their genomic off targets with 2 and 3 mismatches at the non-seed region (PAM distal end) and interrogated the *in vitro* cleavage efficiency of FnCas9. Remarkably, FnCas9 was unable to cleave the substrate in the presence of 2 or 3 mismatches for both loci suggesting that it is extremely specific in target recognition, particularly when the mismatches occur together in the PAM distal region (Figure 2B). SpCas9 however, was able to cleave both the mismatched substrates (Figure 2B). Previous studies have

highlighted the importance of defined mutations in the REC3 domain of highly specific engineered versions of Cas9 in determining target specificity by allosterically regulating the HNH domain from adopting a cleavage-competent form (Chen et al., 2017, Casini et al., 2018). However, the different engineered variants showed similar binding affinity for their off-targets as wild type SpCas9, even though they did not cleave these targets suggesting that they probably remain bound to off-targets in a cleavage incompetent state (Chen et al., 2017). We asked if FnCas9 too shows similar properties and investigated the binding affinity for off-targets using MST. Strikingly, FnCas9 showed negligible to no binding affinity for substrates having two mismatches where no cleavage was observed suggesting that it either interacts extremely weakly or is evicted from the substrate following off-target interrogation (Figure 2C, S2A). SpCas9 however, showed a strong binding affinity for both on and off targets as reported previously (Chen et al., 2017) (Figure 2C, S2A). Taken together these results indicate that FnCas9 has a fundamentally distinct outcome of off-target recognition and binding as compared to SpCas9 and its engineered derivatives.

Intrigued by the ability of FnCas9 to discern mismatches in the target with high accuracy, we looked at the protein structure to understand the basis for this specificity. The interaction of FnCas9 with its substrate is predominantly mediated by an expanded REC3 and the associated REC2 domains that are structurally distinct from those found in SpCas9 or *Staphylococcus aureus* Cas9 (SaCas9) (Hirano et al., 2016). In the crystal structure of FnCas9, the REC3 domain was reported to contain a structural zinc ion in a coordination sphere consisting of 4 cysteine residues (Hirano et al., 2016). Notably, complete abolishment of FnCas9 activity was observed when *in vitro* cleavage reaction was supplanted with trace amounts of Zn^{2+} (1mM) suggesting that the binding of the metal ion inhibits FnCas9 mediated cleavage even in the presence of Mg^{2+} (Figure S2B). Upon analyzing the crystal structure of FnCas9 RNP in complex with target DNA, we observed that the FnCas9 has a much higher electrostatic potential than SpCas9 and interacts with several bases both on the PAM distal and PAM proximal ends of the substrate (Figure S2C). We speculate that these

interactions might be necessary for FnCas9 to adopt a cleavage competent state and in the presence of at least 2 mismatches, the complex dissociates. In fact, the electrostatic contributions of FnCas9 on each of the 20 bases in the substrate are higher than that of SpCas9, possibly leading to stringent interrogation and subsequent dissociation upon mismatch recognition (Figure S2C). We asked if these features could make FnCas9 extend its specificity for off-target discrimination at both the PAM proximal and distal ends. To test this, we designed substrates carrying the human *HBB* sequence and introduced two mutations at different positions throughout the substrate. We observed that cleavage activity went down drastically, suggesting that FnCas9 has very low tolerance for double mismatches along most of the sgRNA:DNA duplex, while SpCas9 activity remained largely unaffected (Figure S3A). This discrimination of off-targets is not correlated with distance between the two mismatches (Figure S3A). We also inquired if off-targets can be distinguished by introducing double mismatches in the sgRNA sequence. To this end, we targeted the human *HBB* substrate using sgRNAs having double mismatches at various positions in the PAM proximal and distal ends and observed that complete loss of cleavage activity could be effected with mismatch combinations both at PAM proximal and distal ends. (Figure S3B). We propose that rationally designing sgRNAs with defined mismatches that account for these cleavage association outcomes can bring very high specificity to FnCas9 based genome editing experiments with minimum to no off-target activity.

FnCas9 cleaves genomic targets 9 bp away from PAM

Intrigued by the high specificity of substrate recognition under *in vitro* conditions, we next investigated if FnCas9 can function *in vivo* as a genome-editing agent. To investigate cellular entry and DNA binding, we generated a construct expressing recombinant FnCas9 tagged with green fluorescent protein (GFP), Nuclear Localization Signal (NLS) and Human influenza hemagglutinin (HA) and confirmed its activity by *in vitro* cleavage (Figure S4A). Targeting this construct to

telomeric repeats in mouse Neuro2A (N2A) cells by electroporation showed a proportion of cells with nuclear dots suggesting that FnCas9 can access the nucleus in such cells (Figure S4B). We then investigated if FnCas9 can bind to genomic DNA by targeting an integrated GFP sequence in mouse N2A cells by chromatin immunoprecipitation (ChIP) using an antibody against the HA peptide. We transfected a plasmid expressing FnCas9 followed by *in vitro* transcribed sgRNAs targeting the GFP sequence and observed up to 100-fold enrichment of the protein at the target after 24 hours post sgRNA transfection (Figure 3A) thus confirming that FnCas9 was able to access its target site on DNA. Since FnCas9 can enter the nucleus and localize to DNA, we next investigated if it could introduce insertions/deletions (indels) upon binding to DNA. We selected the human c-MYC and the EMX1 loci and designed sgRNAs targeting these regions. We obtained comparable number of indels at the c-MYC locus for both SpCas9 and FnCas9 (~ 7%) and a higher number of indels at the EMX1 locus for SpCas9 (~6.5%) as compared to FnCas9 (~ 2.5%) suggesting that depending on the sequence targeted, FnCas9 might show variability in cleavage efficiency, an extrapolation from our biophysical studies. The outcome of double stranded break was similar in both SpCas9 and FnCas9 as reflected by no significant differences in insertion or deletion events (Figure S4C). However, while most of the indels in SpCas9 treated samples occurred ~3 nucleotides upstream of the PAM as has been reported previously, the majority of FnCas9 indel events were detected ~9 nucleotides upstream of the PAM (Figure 3C). We speculate that the cleavage site of FnCas9 maybe influenced by the relative positions of the RUVC and HNH nuclease domains, which makes contacts with DNA at positions distinct from SpCas9 (Figure S2C). Interestingly, an earlier report had measured FnCas9 *in vivo* activity and observed differences in cleavage efficiency across multiple genomic loci and chromatin inaccessibility was identified as a possible determinant of reduced activity at certain loci (Chen et al., 2017). To determine if the large size and unique domain conformation of FnCas9 might hinder its genomic recruitment, we investigated the transient binding of apo-dFnCas9 and apo-dSpCas9 to PAM sites across the genome in HEK293T cells using ChIP

followed by deep sequencing (Figure 3D). We uncovered 3075 peaks for dFnCas9 and 2071 peaks for dSpCas9 with 318 overlapping peaks suggesting that dFnCas9, like dSpCas9, can access a comparable number of sites on genomic DNA (Figure 3E). Whether these peaks represent 'open' or 'closed' chromatin would require further experimentation. The higher number of peaks observed in case dFnCas9 might be due to a wider spectrum of PAM recognition (NGG and to a lesser extent NGA and NAG) as compared to dSpCas9 (NGG) (Hirano et al., 2016). Interestingly, despite dFnCas9 being a considerably larger enzyme (1629 amino acids) as compared to dSpCas9 (1368 amino acids), the median peak width of dFnCas9 ChIP peaks (245 bp) was nearly identical to that of dSpCas9 (243 bp) (Figure S4D). Although the exact footprint of any protein on DNA cannot be determined accurately by ChiP-Seq reads, we speculate that both proteins probably occupy a similar binding space on genomic DNA. Taken together, our results suggest that the larger size of FnCas9 does not hinder its DNA accessibility as a genome-editing agent.

FnCas9 can generate successful HDR mediated genome editing outcomes

Since HDR mediated genome engineering regulates an overarching component of therapeutic genome editing, we interrogated if FnCas9 can integrate a GFP reporter in a targeted fashion. We transfected HEK293T cells with FnCas9 expressing plasmid, *in vitro* transcribed sgRNAs and a plasmid template containing out-of-frame GFP flanked by sequences bearing homology sites to the mammalian *FASN* locus and looked for GFP expression in these cells. We observed 0.89 % GFP positive cells, 72h after sgRNA and template transfection suggesting that FnCas9 mediated DNA cleavage can introduce HDR (Figure 3F). Under identical transfection conditions, SpCas9 gave 0.83% GFP positive cells (Figure S4E). FACS sorted GFP positive cells retained green fluorescence 11 days after transfection confirming stable integration further validated by genotyping PCR (Figure 3G). We conclude that FnCas9 can be successfully used for mammalian genome editing.

DISCUSSION

Precise genome engineering depends on the ability of the genome-editing agent to interrogate a given DNA loci with high specificity and minimum off targeting (Wang et al., 2016). The bacterial type II CRISPR Cas system has been efficiently exploited to edit a variety of genomes and has recently gained prominence for gene therapy applications, predominantly for monogenic disorders like beta thalassemia and sickle cell anemia (Lee et al., 2015; Traxler et al., 2016). As one of the first Cas9 proteins described, SpCas9 has been highly successful in gene editing in diverse biological systems but reports of off-targeting remain a concern, particularly those of large deletions and complex rearrangements (Kosicki et al., 2018). In this study, we show that FnCas9, a Type II orthogonal Cas9 protein shows very high specificity to its target and negligible affinity to off-targets thereby increasing the repertoire of naturally occurring Cas9 proteins that can be used for therapy.

Binding studies of FnCas9 with its intended targets and off-targets reveal that FnCas9 has a fundamentally different mode of negotiating off-targets with PAM distal mismatches. Whereas highly specific engineered domains of SpCas9 remain in an associated ‘cleavage incompetent’ state with its off targets (Chen et al., 2017), FnCas9 shows negligible affinity to such substrates possibly by eviction from the target upon encountering mismatches. Since the REC3 domain, which determines this specificity in the engineered variants is structurally unique in FnCas9, further biophysical studies aimed at visualizing the mode of action of this domain might elucidate the exact nature of target discrimination.

Systematic dissection of FnCas9 cleavage across different substrates reveal that its activity is largely abolished in the presence of 2 mismatches, most prominently at the PAM distal end. This understanding sets up a framework for FnCas9 sgRNA design algorithms with a high degree of specificity by rationally introducing mismatches in sgRNA sequences at defined positions that can distinguish between the target and a very similar off-target. Along similar lines,

the 2 mismatch off-target discrimination can potentially allow FnCas9 to accurately distinguish single nucleotide polymorphisms (SNPs) in a given DNA by rationally designed mismatches in the sgRNA, eventually adding to the expanding scope of Cas9 based disease detection platforms (Sashital, 2018).

Although FnCas9 has a large size that can potentially deter its genomic interrogation to various sites, we observed successful *in vivo* genome editing at two genomic loci interrogated suggesting that it can access endogenous targets. Along the same lines, genome wide binding of apo-dFnCas9 showed similar binding propensity as apo-dSpCas9 suggesting that chromatin accessibility is probably not compromised due to the larger size of the protein. Overlaying DNAse I hypersensitivity data with FnCas9 binding sites will give clues regarding the enrichment of the protein in 'open' and 'closed' genomic targets.

Finally, we show that FnCas9 can be used for HDR based modification of a desired target in the genome. This lays the foundation for including this protein to currently available genome editing agents for potential gene therapy applications. The bulky nature of the protein will necessitate shorter engineered versions that will be more apt for delivery, potentially by incorporating associated domains from smaller Cas9 orthologs yet retaining the on-target specificity of FnCas9. Due to the modularity of the Cas9 enzyme, such chimeric constructs would need to exploit the steadily evolving functional link between REC3 and substrate mismatch discrimination.

Our studies indicate that FnCas9 might dissociate from its mismatched substrates rather rapidly resulting in minimal affinity with off-targets, a property that is strikingly dissimilar from other rationally designed Cas9 proteins that are highly specific (Chen et al., 2017). Remaining associated with off-target loci might result in transcriptional repression in these regions that can confound on-target activity and lead to cleavage-independent gene expression outcomes. Although it remains to be tested, FnCas9, due to its low binding affinity to off-targets, might present a more attractive scenario for genome editing where even long term persistence of the FnCas9 inside the cell might not induce off-targeting

because of the high threshold for DNA binding when mismatches to the targets is encountered.

FnCas9, a type II-B Cas system has evolved separately from the II-A and II-C systems comprising of the majority of the commonly used Cas9 proteins used for genome editing (Chylinski et al., 2014). Its distinct structural attributes leading to high targeting specificity hints at the existence of more members of the II-B family that can be similarly explored for mammalian genome editing. Further, FnCas9 can be structurally engineered to render it competent for base editing, an attractive therapeutic application for monogenic disorders where the current generation of base editors have so far shown variability in off-targeting (Kim et al., 2017).

ACKNOWLEDGMENTS

This work was supported by Council for Scientific and Industrial Research (CSIR) Mission Mode Project grant HCP0008 to D.C. and S.M. The *myc* construct was a kind gift from the lab of Shantanu Chaudhury, CSIR IGIB and the FASN construct used for HDR experiments was generously shared by Mihail Sarov, MPICBG, Dresden. The authors wish to thank the lab of Frank Buchholz and Beena Pillai for constructs and cell lines used in the study and Debasis Dash, CSIR IGIB, Sambit Dalui, CSIR IICB, and all members of S.M. and D.C. labs for helpful discussions and feedback. The support of Manish Kumar, Imaging facility CSIR IGIB and Nanotemper, Bangalore (MST experiments) is also gratefully acknowledged.

AUTHOR CONTRIBUTIONS

D.C., S.M., S.A., A.M., D.P., and M.Az. designed the experiments. S.A., A.M., D.P., M.Az, N.S., M.Ai., D.S., S.S. and S.J. performed experiments and collected data, A.H.A., D.P., A.M. and A.R. analyzed data, D.C. wrote the manuscript. All authors reviewed the manuscript and approved the conclusions.

DECLARATION OF INTERESTS

The authors declare no conflict of interest.

REFERENCES

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. Nature communications *9*, 1911.

Anders, C., Niewoehner, O., and Jinek, M. (2015). In Vitro Reconstitution and Crystallization of Cas9 Endonuclease Bound to a Guide RNA and a DNA Target. Methods in enzymology *558*, 515-537.

Casini, A., Olivieri, M., Petris, G., Montagna, C., Reginato, G., Maule, G., Lorenzin, F., Prandi, D., Romanel, A., Demichelis, F.*, et al.* (2018). A highly specific SpCas9 variant is identified by in vivo screening in yeast. Nature Biotechnology *36*, 265.

Cebrian-Serrano, A., and Davies, B. (2017). CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. Mammalian genome : official journal of the International Mammalian Genome Society *28*, 247-261.

Chen, F., Ding, X., Feng, Y., Seebeck, T., Jiang, Y., and Davis, G.D. (2017). Targeted activation of diverse CRISPR-Cas systems for mammalian genome editing via proximal CRISPR targeting. Nature communications *8*, 14958.

Chen, J.S., Dagdas, Y.S., Kleinstiver, B.P., Welch, M.M., Sousa, A.A., Harrington, L.B., Sternberg, S.H., Joung, J.K., Yildiz, A., and Doudna, J.A. (2017). Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. Nature *550*, 407-410.

Chylinski, K., Makarova, K.S., Charpentier, E., and Koonin, E.V. (2014). Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Research *42*, 6091-6105.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A.*, et al.* (2013). Multiplex Genome Engineering Using

CRISPR/Cas Systems. Science *339*, 819-823.

Esvelt, K.M., Mali, P., Braff, J.L., Moosburner, M., Yaung, S.J., and Church, G.M. (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. Nature methods *10*, 1116-1121.

Hille, F., and Charpentier, E. (2016). CRISPR-Cas: biology, mechanisms and relevance. Philosophical Transactions of the Royal Society B: Biological Sciences *371*, 20150496.

Hille, F., Richter, H., Wong, S.P., Bratovic, M., Ressel, S., and Charpentier, E. (2018). The Biology of CRISPR-Cas: Backward and Forward. Cell *172*, 1239-1259.

Hirano, H., Gootenberg, J.S., Horii, T., Abudayyeh, O.O., Kimura, M., Hsu, P.D., Nakane, T., Ishitani, R., Hatada, I., Zhang, F.*, et al.* (2016). Structure and Engineering of Francisella novicida Cas9. Cell *164*, 950-961.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. Cell *157*, 1262-1278.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology *31*, 827-832.

Huang, C.H., Lee, K.C., and Doudna, J.A. (2018). Applications of CRISPR-Cas Enzymes in Cancer Therapeutics and Detection. Trends in cancer *4*, 499-512.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821.

Kim, D., Lim, K., Kim, S.T., Yoon, S.H., Kim, K., Ryu, S.M., and Kim, J.S. (2017). Genome-wide target specificities of CRISPR RNA-guided programmable deaminases. Nature biotechnology *35*, 475-480.

Knott, G.J., and Doudna, J.A. (2018). CRISPR-Cas guides the future of genetic engineering. Science *361*, 866-869.

Komor, A.C., Badran, A.H., and Liu, D.R. (2017). CRISPR-Based Technologies for the Manipulation of Eukaryotic Genomes. Cell *168*, 20-36.

Kosicki, M., Tomberg, K., and Bradley, A. (2018). Repair of double-strand breaks

induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. Nature Biotechnology *36*, 765.

Lee, C., Bao, G., Porteus, M.H., Cornu, T., Miccio, A., Cradick, T., Cathomen, T., Lundberg, A., and Mavilio, F. (2015). Gene Editing with Crispr-Cas9 for Treating Beta-Hemoglobinopathies. Blood *126*, 3376-3376.

Maeder, M.L., and Gersbach, C.A. (2016). Genome-editing Technologies for Gene and Cell Therapy. Molecular therapy : the journal of the American Society of Gene Therapy *24*, 430-446.

Mekler, V., Minakhin, L., and Severinov, K. (2017). Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation. Proceedings of the National Academy of Sciences of the United States of America *114*, 5443-5448.

Najm, F.J., Strand, C., Donovan, K.F., Hegde, M., Sanson, K.R., Vaimberg, E.W., Sullender, M.E., Hartenian, E., Kalani, Z., Fusi, N.*, et al.* (2018). Orthologous CRISPR-Cas9 enzymes for combinatorial genetic screens. Nature biotechnology *36*, 179-189.

Osuka, S., Isomura, K., Kajimoto, S., Komori, T., Nishimasu, H., Shima, T., Nureki, O., and Uemura, S. (2018). Real-time observation of flexible domain movements in CRISPR-Cas9. The EMBO journal *37*.

Price, A.A., Sampson, T.R., Ratner, H.K., Grakoui, A., and Weiss, D.S. (2015). Cas9-mediated targeting of viral RNA in eukaryotic cells. Proceedings of the National Academy of Sciences of the United States of America *112*, 6164-6169.

Raper, A.T., Stephenson, A.A., and Suo, Z. (2018). Functional Insights Revealed by the Kinetic Mechanism of CRISPR/Cas9. Journal of the American Chemical Society *140*, 2971-2984.

Richardson, C.D., Ray, G.J., DeWitt, M.A., Curie, G.L., and Corn, J.E. (2016). Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nature Biotechnology *34*, 339.

Sashital, D.G. (2018). Pathogen detection in the CRISPR-Cas era. Genome medicine *10*, 32.

Shibata, M., Nishimasu, H., Kodera, N., Hirano, S., Ando, T., Uchihashi, T., and

Nureki, O. (2017). Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. Nature communications *8*, 1430.

Singh, A., Chakraborty, D., and Maiti, S. (2016). CRISPR/Cas9: a historical and chemical biology perspective of targeted genome engineering. Chemical Society reviews *45*, 6666-6684.

Singh, D., Sternberg, S.H., Fei, J., Doudna, J.A., and Ha, T. (2016). Real-time observation of DNA recognition and rejection by the RNA-guided endonuclease Cas9. Nature communications *7*, 12778.

Sternberg, S.H., Redding, S., Jinek, M., Greene, E.C., and Doudna, J.A. (2014). DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. Nature *507*, 62.

Traxler, E.A., Yao, Y., Wang, Y.-D., Woodard, K.J., Kurita, R., Nakamura, Y., Hughes, J.R., Hardison, R.C., Blobel, G.A., Li, C.*, et al.* (2016). A genome-editing strategy to treat β-hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition. Nature Medicine *22*, 987.

van der Oost, J., Westra, E.R., Jackson, R.N., and Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR–Cas systems. Nature Reviews Microbiology *12*, 479.

Wang, H., La Russa, M., and Qi, L.S. (2016). CRISPR/Cas9 in Genome Editing and Beyond. Annual Review of Biochemistry *85*, 227-264.

Yang, M., Peng, S., Sun, R., Lin, J., Wang, N., and Chen, C. (2018). The Conformational Dynamics of Cas9 Governing DNA Cleavage Are Revealed by Single-Molecule FRET. Cell reports *22*, 372-382.

FIGURE LEGENDS

Figure 1. FnCas9 shows different affinity and cleavage kinetics for substrates as compared to SpCas9.

(A) Left, representative EMSA showing binding of purified dFnCas9-GFP at different concentrations to its sgRNA sequence (250 nM). Right, quantification of protein: sgRNA binding. Error bars represent SEM (3 independent experiments).

(B) MST results showing binding of dFnCas9-GFP (left) and dSpCas9-GFP (right) to two substrates, *Vegfa3* (top) and *Emx1_*1 (bottom) expressed as normalized fluorescence units (y-axis) with respect to varying concentrations of purified substrate (x-axis). The substrate sequences are indicated in the box with the PAM shown in Green. Error bars represent SEM (2 independent experiments).

(C) *In vitro* cleavage assay showing activity of WT SpCas9 and WT FnCas9 (50 nM Cas9: 150 nM sgRNA) on indicated substrates (linearized plasmids, 250ng) at different time points (1 min – 2h). Top band represent uncleaved target while the two bottom bands represent cleaved products. Quantification is shown on the right, y-axis represent percentage cleavage. Error bars represent SD (3 independent experiments).

Figure 2. More than 1 mismatch in target can abolish FnCas9 cleavage activity due to loss of DNA binding.

(A) Mismatches at different positions along the substrate alter FnCas9 cleavage outcomes while retaining SpCas9 activity. Representative *in vitro* cleavage outcome of HBB substrate (purified PCR product) showing mismatches at a single position in every lane (indicated in red, other

bases remaining unchanged) interrogated with FnCas9 (top) or SpCas9 (bottom) is shown, PAM is indicated in green. Top band represent uncleaved target while the two bottom bands represent cleaved products. Quantification for each reaction is shown below the gel images. Error bars represent SD (3 independent experiments).

(B) FnCas9 does not tolerate 2 or more mismatches in the substrate at the PAM distal end. *In vitro* cleavage outcomes of EMX1 (top) and VEGFA3 (bottom) targets interrogated by WT FnCas9 and SpCas9 are shown (right). On the left, the sequences with 2 or more mismatches (indicated as red bases) are presented. PAM is indicated in green. Data representative of 3 independent experiments.

(C) MST results showing binding outcomes of dFnCas9-GFP (left) and dSpCas9-GFP (right) to *VEGFA3* substrate expressed as normalized fluorescence units (y-axis) with respect to varying concentrations of purified substrate (x-axis). The substrate sequences are indicated in the box with the PAM shown in Green. Error bars represent SEM (2 independent experiments).

Figure 3. FnCas9 can perform both NHEJ and HDR mediated genome editing in mammalian cells

(A) Chromatin Immunoprecipitation (ChIP) of FnCas9 showing binding of FnCas9 RNP complex to endogenous GFP target in mouse Neuro 2A cell line. Fold enrichment (percentage of input) is shown on y-axis. A non-targeting region is taken as negative control. Error bars represent S.E.M (3 PCR replicates), data representative of two independent experiments.

(B) Indel events (expressed as percentage of total reads) obtained by amplicon sequencing upon SpCas9 or FnCas9 targeting *EMX1* (left) or *c-MYC* (right) in HEK293T cells. Error bars represent SEM (3 independent experiments). Student's t-test p-values are presented as *** <0.005.

(C) Position of SpCas9 and FnCas9 mediated cleavage in substrate. Graphs showing the percentage of indels along every position in the substrate sequence for *EMX1* sgRNA (left) and *c-MYC* sgRNA (right) represented as bars. Dotted colored zone represents the bases with highest number of indels between which SpCas9 or FnCas9 possibly cleaves. Error bars represent SEM (3 independent experiments). PAM is indicated in green and bases are labeled as nucleotides away from PAM.

(D) Schematic for ChIP-seq with apo-dSpCas9 or apo-dFnCas9 in HEK293T cells. Cells were harvested 36h after transfection with respective plasmids and ChIP-seq was performed after pull-down using an antibody against HA.

(E) Overlap of ChIP-seq peaks obtained using dSpCas9 or dFnCas9.

(F) HDR expressed as percentage of GFP positive cells in HEK293T cells upon transfection with different plasmid/sgRNA/donor combinations as indicated. Dotted line represents background present from experiments where at least one of the components was excluded. Error bars represent SD (3 independent experiments). Welch's t-test p-values are presented as * < 0.05.

(G) HEK293T cells expressing GFP 8 d after sorting GFP positive cells following FnCas9 based HDR, indicating stable integration in the genome.

(H) Genotyping PCR confirming targeted GFP integration at the FASN locus in sorted GFP negative or positive HEK293T cells after FnCas9 mediated HDR. Forward and reverse primers target the left and right homology arms in the FASN locus flanking the GFP sequence and produce either a long (886 bp) or short (143 bp) PCR product for integrated or non-integrated products respectively.

Figure S1. FnCas9 is a single turnover enzyme

(A) Coomassie gel showing different purified proteins used in the study.

(B) *In vitro* cleavage assay showing mutation of nuclease domains in FnCas9 abolishes its cleavage activity.

(C) MST results showing binding outcomes of dFnCas9-GFP to *EMX1*_2 and *c-MYC* substrates expressed as normalized fluorescence units (y-axis) with respect to varying concentrations of purified substrate (x-axis). The substrate sequences are indicated in the box with the PAM shown in Green. Error bars represent SEM (2 independent experiments).

(D) *In vitro* cleavage assays showing outcome of *c-MYC* substrate cleavage (2.5nM) with varying FnCas9 RNP concentrations (as indicated) across different time points. Note that cleavage activity increases with increasing concentration of FnCas9 RNP complex.

Figure S2. FnCas9 shows unique structural attributes for substrate binding and cleavage

(A) MST results showing binding outcomes of dFnCas9-GFP (left) and dSpCas9-GFP (right) to *EMX1*_1 off-target substrates expressed as normalized fluorescence units (y-axis) with respect to varying concentrations of purified substrate (x-axis). The substrate sequences are indicated in the box with the PAM shown in Green. Red denotes the mismatches in the target. Error bars represent SEM (2 independent experiments).

(B) Activity of FnCas9 is abolished in the presence of Zn^{2+} ions. *In vitro* cleavage assay showing FnCas9 is active on a substrate (linearized plasmid pCAGSS_GFP_RFP) only in the presence of Mg^{2+} and absence of Zn^{2+}.

(C) Crystal structures of FnCas9 (PDB 5B2O, top) and SpCas9 (PDB 5F9R, bottom) RNP in complex with substrate DNA in ribbon (left) and electrostatic potential surface representation (right). PAM is indicated in red and individual domains are color-coded. Heat-map of electrostatic contribution is indicated on the right. Note the higher positive electrostatic

potential of FnCas9 domains (indicated by dotted circles) interacting with substrate. REC3 domain in each enzyme has been denoted with dotted circles.

(D) Quantification of electrostatic potential contribution of FnCas9 (top) and SpCas9 (bottom) on substrate DNA nucleotides. Nucleotide positions are indicated below from PAM proximal to PAM distal end (left to right). PAM is denoted in green.

Figure S3. Cleavage activity of FnCas9 is affected in the presence of 2 mismatches in the substrate or the sgRNA

(A) 2 Mismatches at different positions along the substrate alter FnCas9 cleavage outcomes while retaining SpCas9 activity. Representative *in vitro* cleavage outcome of HBB substrate sequence showing mismatches at different positions (indicated in red) interrogated with FnCas9 (top) or SpCas9 (bottom) is shown, PAM is indicated in green. Top band represent uncleaved target while the two bottom bands represent cleaved products. Quantification for each reaction is shown below the gel images. Error bars represent SD (3 independent experiments).

(B) Representative *in vitro* cleavage outcome of HBB sgRNA sequence showing mismatches at different positions (indicated in red) interrogated with FnCas9 is shown, position of PAM with respect to the sequence orientation is indicated in green. Top band represent uncleaved target while the two bottom bands represent cleaved products. Quantification for each reaction is shown below the gel image. Error bars represent SD (3 independent experiments).

Figure S4. FnCas9 can mediate *in vivo* genome editing

(A) Schematic showing the plasmid for expressing HA-NLS-FnCas9 cloned in pET GFP LIC vector used for electroporation (left). Confirmation of *in vitro* activity of the RNP complex (right).

(B) Representative Immunofluorescence image showing the nuclear localization of FnCas9-NLS-GFP protein (shown in green) with sgRNA (as RNP complex) against telomeric repeats transfected in mouse ES cells. DNA is counterstained with DAPI and represented in blue. Scale bar 5 μm.

(C) Outcomes of FnCas9 mediated genomic cleavage showing insertions/deletions (indels) per event at two loci *EMX1* (left) and *c-MYC* (right) detected after amplicon sequencing. Each point represents an individual indel event obtained from 3 independent experiments.

(D) Distribution of peak width from ChIP sequencing of apo-dFnCas9 or apo-dSpCas9 binding to the human genome. Dotted line represents the median peak width across all samples. Peak width bins (100 bp) is represented on the x-axis and density of peaks is denoted on the y-axis.

(I) HDR expressed as percentage of GFP positive cells in HEK293T cells upon transfection with different plasmid/sgRNA/donor combinations as indicated. Dotted line represents background present from experiments where at least one of the components was excluded. Student's t-test p-values are presented as * < 0.05, *** <0.001.

MATERIALS AND METHODS

Plasmid Construction:

The gene encoding full length *Franciscella novicida* (Fn) Cas9 nuclease residues 1-1629 bp) was PCR amplified using PX408 (Addgene 68705) as template and cloned in pET-SUMO Champion vector (Thermo Fisher Scientific) and pET-His6-GFP-TEV-LIC vector (Addgene 29663) following restriction enzyme based cloning and ligation independent cloning (LIC) respectively. Catalytically inactive FnCas9 double mutants were generated on pET-His6-FnGFP-TEV-LIC (This paper) and on PX408 (Addgene 68705) plasmid backbones by QuickChange II site directed mutagenesis kit (Agilent) following manufacturer's protocol with some modifications. The sequences were confirmed by DNA sequencing.

Protein purification:

The proteins used in this study were purified as reported previously (Jinek et al., 2012). Briefly, plasmids for different Cas9 proteins were expressed in *Escherichia coli* Rosetta2 (DE3) (Novagen). The protein expressing Rosetta2 (DE3) cells were cultured at 37℃ in LB medium (supplemented with 50mg/l kanamycin) until OD_{600} reached 0.6 and protein expression was induced by addition of 0.5mM isopropyl β-D-thiogalactopyranoside (IPTG). The Rosetta2 (DE3) cells were further cultured at 18℃ overnight and harvested by centrifugation. The *E.coli* cells were resuspended in lysis buffer (20mM Tris-Cl, pH8, 500mM NaCl, 5% glycerol, 1mM DTT) supplemented with 1X PIC (Roche), 100ug/ml lysozyme and lysed by sonication and centrifuged. The lysate was affinity purified by Ni-NTA beads

(Roche) and the eluted protein was further purified by size-exclusion chromatography on HiLoad Superdex 200 16/600 column (GE Healthcare) in 20 mM HEPES pH 7.5, 150 mM KCl, 10% glycerol, 1mM DTT. The concentration of purified proteins were measured by Pierce BCA protein assay kit (Thermo Fisher Scientific). The purified proteins were stored at -80℃ until further use.

In case of 6X His-MBP-dSpCas9, the 6X His-MBP tag was removed by incubating the affinity bound protein with PreScission Protease overnight in cleavage buffer (50mM Tris-Cl, pH8, 150mM NaCl, 1mM EDTA, 1mM DTT). The cleaved Cas9 protein was separated from fusion tag on HiLoad Superdex 200 16/600 column (GE Healthcare) in 20 mM HEPES pH 7.5, 150 mM KCl, 10% glycerol, 1mM DTT.

***In vitro* transcription:**

In vitro transcribed sgRNAs were synthesized usig MegaScript T7 Transcription kit (Thermo Fisher Scientific) using T7 promoter containing template as substrates. IVT reactions were incubated overnight at 37℃ followed by NucAway spin column (Thermo Fisher Scientific) purification. IVT sgRNAs were stored at -20℃ until further use.

***In vitro* Cleavage assay (IVC):**

Cas9-sgRNA complex was reconstituted at 25℃ for 10 mins in reaction buffer (20 mM HEPES, pH7.5, 150mM KCl, 1mM DTT, 10% glycerol, 10mM $MgCl_2$). Linearized plasmid was incubated in reaction buffer with reconstituted RNP complex at 37℃. The reaction was stopped by adding 25 mM EDTA final concentration and then treated with 1ul of 20mg/ml Proteinase K in 15 ul reaction for 30 mins at 55℃. The reaction products were resolved in EtBr-stained 1% agarose gel, visualized by Syngene UV transilluminator and quantified by ImageJ.

Microscale thermophoresis (MST) assay:

MST was performed as reported previously (Dong et al., 2016; Wienken et al., 2010). Briefly, dSpCas9-GFP and dFnCas9-GFP proteins were complexed with PAGE purified respective IVT sgRNAs and that IVT sgRNAs were purified by 12% Urea-PAGE. The binding affinities of the Cas9 proteins and sgRNA RNP complexes towards different genomic loci like c-Myc promoter, EMX1, VEGFA3 were calculated using Monolith NT. 115 (NanoTemper Technologies GmbH, Munich, Germany). RNP complex (Protein:sgRNA molar ratio,1:1) was reconstituted at 25℃ for 10 mins in reaction buffer (20 mM HEPES, pH7.5, 150mM KCl, 1mM DTT, 10mM $MgCl_2$). HPLC purified 30 bp dsDNA (IDT) of different genomic loci with varying concentrations (ranging from 0.09nM to 30μM) were incubated with RNP complex at 37^0 C temperature for 30 min in reaction buffer. The sample was loaded into NanoTemper standard treated capillaries and measurements were performed at 25°C using 20% LED power and 40% MST power. Data analyses were done using NanoTemper analysis software.

Electrophoretic mobility shift assay (EMSA):

EMSA was performed following the protocol reported earlier (Anders et al., 2015). Catalytically inactive dead (D11A-H96A FnCas9) Cas9-sgRNA complex was reconstituted at 25℃ for 10 mins in reaction buffer (20 mM HEPES, pH7.5, 150mM KCl, 1mM DTT, 10% glycerol, 10mM $MgCl_2$). 250nM sgRNA was incubated with different concentration of reconstituted RNP complex at 37℃ for 1hr. The substrates were resolved by 8% Native polyacrylamide gel electrophoresis (0.5X TBE buffer with 2Mm $MgCl_2$) run at 4℃. DNA was visualized by Typhoon FLA 7000 (GE Healthcare) and quantified by ImageJ.

Cell culture and transfection:

HEK 293T, Neuro2A and Neuro2A GFP cells were grown in DMEM media supplemented with high glucose (Invitrogen), 2mM GlutaMax, 10% FBS (Invitrogen), 1X antibiotic and antimycotic (Invirogen) at 37℃ in 5% CO_2. Mouse embryonic stem cells, mESCs (R1/E) were cultured in DMEM media (Invitrogen) supplemented with 20% Pansera, 1% PenStrep (Invitrogen), 0.1% B-ME, 1%

NEAA (Invitrogen) and 40ug/mL LIF (MPICBG). Transfections of mammalian cells were performed using Lipofectamine 3000 Reagent (Invitrogen) following manufacturer's protocol.

ChIP assay:

Neuro2AGFP and HEK293T cells on 10cm dishes were transfected with 50µg of dSpCas9 (Addgene 61355) and dFnCas9 (This paper) plasmids and after 24hrs 50µg respective IVT sgRNAs (only for Neuro2AGFP) were transfected. Cells were harvested 12 hours post sgRNA transfection. Target cells for ChIP assays were cross-linked using 1% formaldehyde (Sigma) followed by vortexing and incubating with rotation at room temperature for 10 min and subsequently quenched by addition of 125mM glycine. Harvested cells were washed two-times by ice-cold PBS. The cells were spun down and the pellet was resuspended in SDS lysis buffer and freshly added PIC so that 1 ml lysis buffer should have 10 million cells and stored at -80°C until further use. Sonication (30 sec on and 30 sec off for 20 cycles, high) was done in DNA LoBind tubes (Eppendorf) in a Bioruptor (Diagenode). Lysate was centrifuged at 12000g for 10 min at 4°C. 1% of the total volume was kept as input. Protein G Dynabeads (Invitrogen) were blocked with Salmon sperm DNA and rotated for 2 hours at RT. The beads were washed with antibody binding and washing buffer followed by incubation of beads with 10µg of Anti-HA antibody (Abcam, ab9110) and incubated at 4°C for 5 hrs. Beads were washed twice and incubated with the lysate overnight at 4°C. Beads were washed the next day followed by sample elution and reverse crosslinking was proceeded for both input and IP samples using 10% SDS, 1M NaHCO3 and 5M NaCl at 65°C for 5 hrs with shaking followed by RNAse A treatment at 37°C for 30 min and proteinase K treatment in an incubator shaker at 45°C for 2 hrs. DNA purification was performed with QIAquick PCR Purification Kit (Qiagen). 5 ng ChIP DNA was taken forward for ChIP library preparation using TruSeq ChIP library kit following manufacturer's protocol (Illumina, USA). 100 bp paired end sequencing was performed on illumna HiSeq platform at MedGenome Labs Pvt. Ltd. (Bangalore, India).

ChIP-Seq analysis:

Raw sequencing reads were mapped to the human reference genome GRCh38 using bowtie2 (Langmead and Salzberg, 2012). Peaks were called over input sample using MACS2. Model-based Analysis of ChIP-Seq was performed (MACS) with default parameters (Zhang et al., 2008) Overlap between the dFnCas9 and dSpCas9 chip peaks were calculated using bedtools with default settings (Quinlan and Hall, 2010)

Amplicon sequencing:

HEK293T cells on 12 well dishes were transfected with 2µg of SpCas9 (Addgene 62988) and FnCas9 (Addgene 68705) plasmids and 24hrs post-transfection, respective IVT sgRNAs were transfected. Cells were harvested 24 hrs post sgRNA transfection. Only Cas9 plasmids without sgRNA were considered as controls. GeneJET Genomic DNA Purification Kit (Thermo Fischer Scientific) was used to isolate genomic DNA. PCR primers were designed flanking the predicted double stranded break site. The DNA samples were quantified using Qubit dsDNA HS Assay kit (Invitrogen, Q32853). These were also loaded onto 1% agarose gel for qualitative check. The amplicon products were further taken for library preparation using NEB Next Ultra DNA Library Prep Kit for Illumina (NEB, Cat# E7370L). First, the amplicons were end repaired and mono-adenylated at 3' end in a single enzymatic reaction. Next, NEB hairpin-loop adapters are ligated to the DNA fragments in a T4-DNA ligase-based reaction. Following ligation, the loop containing Uracil is linearized using USER Enzyme (a combination of UDG and Endo VIII), to make it available as a substrate for PCR based indexing in the next step. During PCR, barcodes were incorporated using unique primers for each of the samples thereby enabling multiplexing. Library preparation and sequencing was done at MedGenome Labs Pvt. Ltd. (Bangalore, India).

Immunofluorescence:

Nucleofection was done on mESCs (R1/E) using Nucleofector kit (Lonza). 2x10^6 cells were seeded on 6 well dish and nucleofection was done following manufacturer's protocol. Immunofluorescence was performed essentially following the same protocol as mentioned earlier (Liu et al., 2015) Imaging was done in Nikon Eclipse Ti-ε Inverted Microscope system (Epi-TIRF module) under 60X objective.

Homology Directed Repair (HDR):

In HEK 293T cells, 2ug Cas9 plasmids (SpCas9, Addgene 62988 and FnCas9, Addgene 68705) were transfected and 24 hrs Cas9 post-transfection, 1ug GFP-containing Donor plasmid and 1ug IVT sgRNA were transfected and cells were kept in presence of 1uM SCR7 molecule (Sigma-Aldrich). Cells were harvested 72 hrs post sgRNA transfection and HDR read out was taken by BD FACS Melody.

Comparative structural study between Fn- and SpCas9

Electrostatic analysis: For the purpose of understanding the underlying electrostatic contributions in the Fn- and SpCas9 structures, we used the standalone version of PDB2PQR (PMID: 28836357, 17488841) to assign charge and radius parameters for CHARMM force field to the Fn- (PDBID: 5B2O) and Sp- (PDBID: 5F9R) crystal structures.

The electrostatics was calculated using Adaptive Poisson-Boltzmann Solver (APBS) (PMID: 11517324). APBS uses iterative solvers to solve the nonlinear algebraic equations resulting from the discretized Poisson-Boltzmann equation with a fixed error tolerance of 10^-6. The values are coloured in the range of -10 to 35 k_b T e_c^{-1} where

- k_b is Boltzmann's constant: $1.3806504 \times 10^{-23}$ J K^{-1}
- T is the temperature of your calculation in K
- e_c is the charge of an electron: $1.60217646 \times 10^{-19}$ C

The images were generated using UCSF Chimera tool.

Contact analysis: To study the interaction between the nucleotide and protein, atomic contacts were calculated using a 4Å cutoff. The graph was generated using MATLAB.

Quantification and statistical analysis:

Gel image quantification was done using ImageJ. Statistical analysis was done in R environment. Statistical information, including the number of replicates are indicated in corresponding figure legends.

Data Analysis:

For Amplicon sequencing, determination of Indel frequency was performed using CRISPR-DAV (Wang et al., 2017) followed by statistical analysis in R.

Cleavage site prediction was done using CRIPRESSO2 (Clement et al., 2018) with the following parameters : 5bp quantification windows around predicted cleavage site (-3 bp for SpCas9 and -10bp for FnCas9).

Data availability

ChIP-seq data is available in NCBI SRA with id SRP119217. Primers are listed in Supplementary Table 1, ChIP-seq coordinates are listed in Supplementary Table 2.

References:

Anders, C., Niewoehner, O., and Jinek, M. (2015). In Vitro Reconstitution and Crystallization of Cas9 Endonuclease Bound to a Guide RNA and a DNA Target. Methods Enzymol *558*, 515-537.

Dong, D., Ren, K., Qiu, X., Zheng, J., Guo, M., Guan, X., Liu, H., Li, N., Zhang, B., Yang, D.*, et al.* (2016). The crystal structure of Cpf1 in complex with CRISPR RNA. Nature *532*, 522-526.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science *337*, 816-821.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat Methods *9*, 357-359.

Liu, J., Gaj, T., Yang, Y., Wang, N., Shui, S., Kim, S., Kanchiswamy, C.N., Kim, J.S., and Barbas, C.F., 3rd (2015). Efficient delivery of nuclease proteins for genome editing in human stem cells and primary cells. Nat Protoc *10*, 1842-1859.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841-842.

Wang, X., Tilford, C., Neuhaus, I., Mintier, G., Guo, Q., Feder, J.N., and Kirov, S. (2017). CRISPR-DAV: CRISPR NGS data analysis and visualization pipeline. Bioinformatics *33*, 3811-3812.

Wienken, C.J., Baaske, P., Rothbauer, U., Braun, D., and Duhr, S. (2010). Protein-binding assays in biological liquids using microscale thermophoresis. Nat Commun *1*, 100.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W.*, et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.

Clement, K., Rees, H., Canver, M., and Pinello, L. (2018). Analysis and comparison of genome editing using CRISPResso2. bioRxiv, 392217

Figure 1

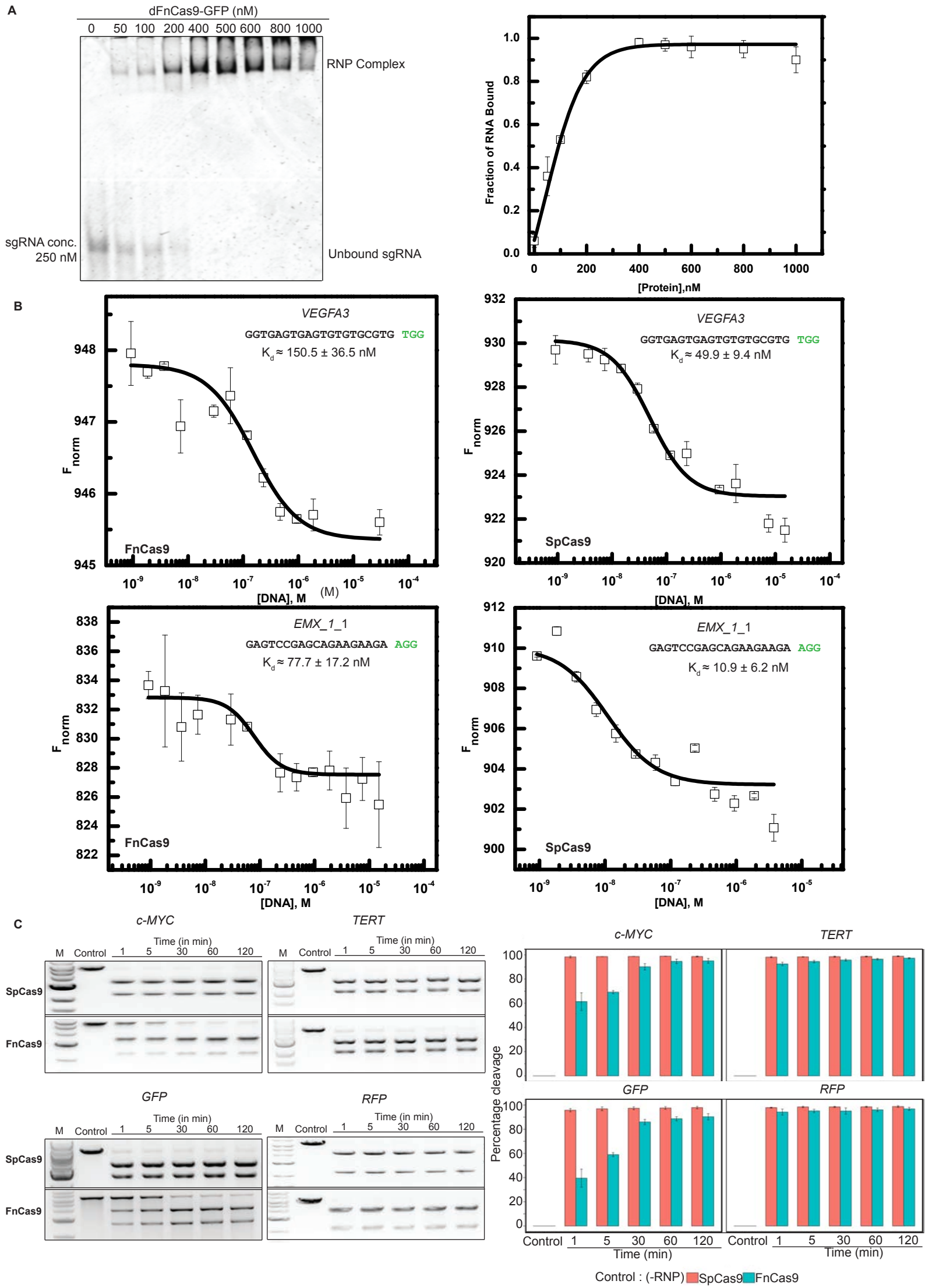

A
dFnCas9-GFP (nM)
0 50 100 200 400 500 600 800 1000
RNP Complex
sgRNA conc. 250 nM
Unbound sgRNA
Fraction of RNA Bound
[Protein],nM
B
VEGFA3
GGTGAGTGAGTGTGTGCGTG TGG
Kd ≈ 150.5 ± 36.5 nM
FnCas9
Fnorm
[DNA], M (M)
VEGFA3
GGTGAGTGAGTGTGTGCGTG TGG
Kd ≈ 49.9 ± 9.4 nM
SpCas9
EMX_1_1
GAGTCCGAGCAGAAGAAGA AGG
Kd ≈ 77.7 ± 17.2 nM
FnCas9
EMX_1_1
GAGTCCGAGCAGAAGAAGA AGG
Kd ≈ 10.9 ± 6.2 nM
SpCas9
C
c-MYC
TERT
GFP
RFP
M Control 1 5 30 60 120
Time (in min)
SpCas9
FnCas9
Percentage cleavage
Control : (-RNP) SpCas9 FnCas9

Figure 2

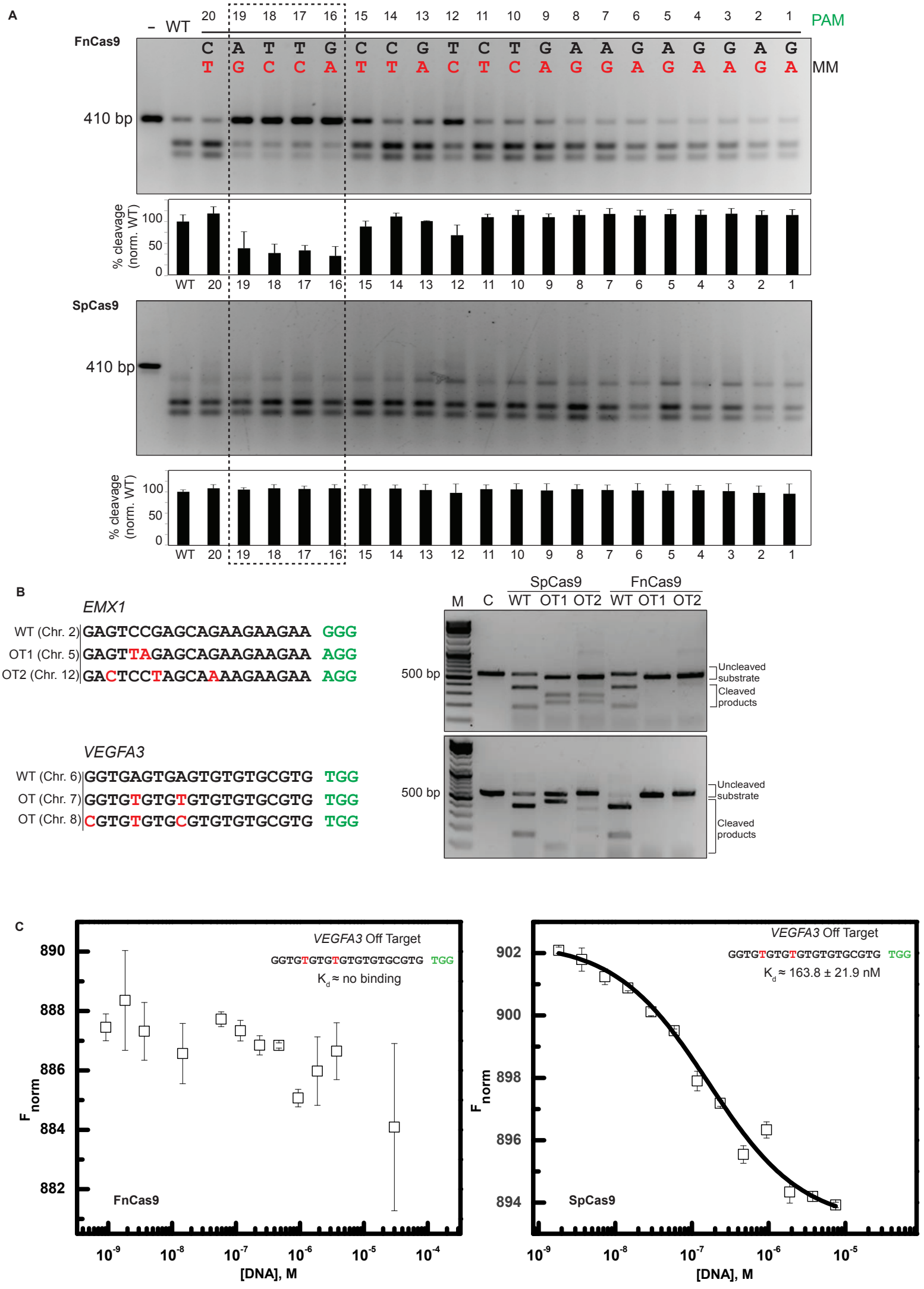

A
FnCas9
SpCas9
410 bp
PAM
MM
% cleavage (norm. WT)
B
EMX1
WT (Chr. 2) GAGTCCGAGCAGAAGAAGAA GGG
OT1 (Chr. 5) GAGTTAGAGCAGAAGAAGAA AGG
OT2 (Chr. 12) GACTCCTAGCAAAAGAAGAA AGG
VEGFA3
WT (Chr. 6) GGTGAGTGAGTGTGTGCGTG TGG
OT (Chr. 7) GGTGTGTGTGTGTGTGCGTG TGG
OT (Chr. 8) CGTGTGTGCGTGTGTGCGTG TGG
M C SpCas9 WT OT1 OT2 FnCas9 WT OT1 OT2
500 bp
Uncleaved substrate
Cleaved products
C
VEGFA3 Off Target
GGTGTGTGTGTGTGTGCGTG TGG
Kd ≈ no binding
FnCas9
Fnorm
[DNA], M
VEGFA3 Off Target
GGTGTGTGTGTGTGTGCGTG TGG
Kd ≈ 163.8 ± 21.9 nM
SpCas9

Figure 3

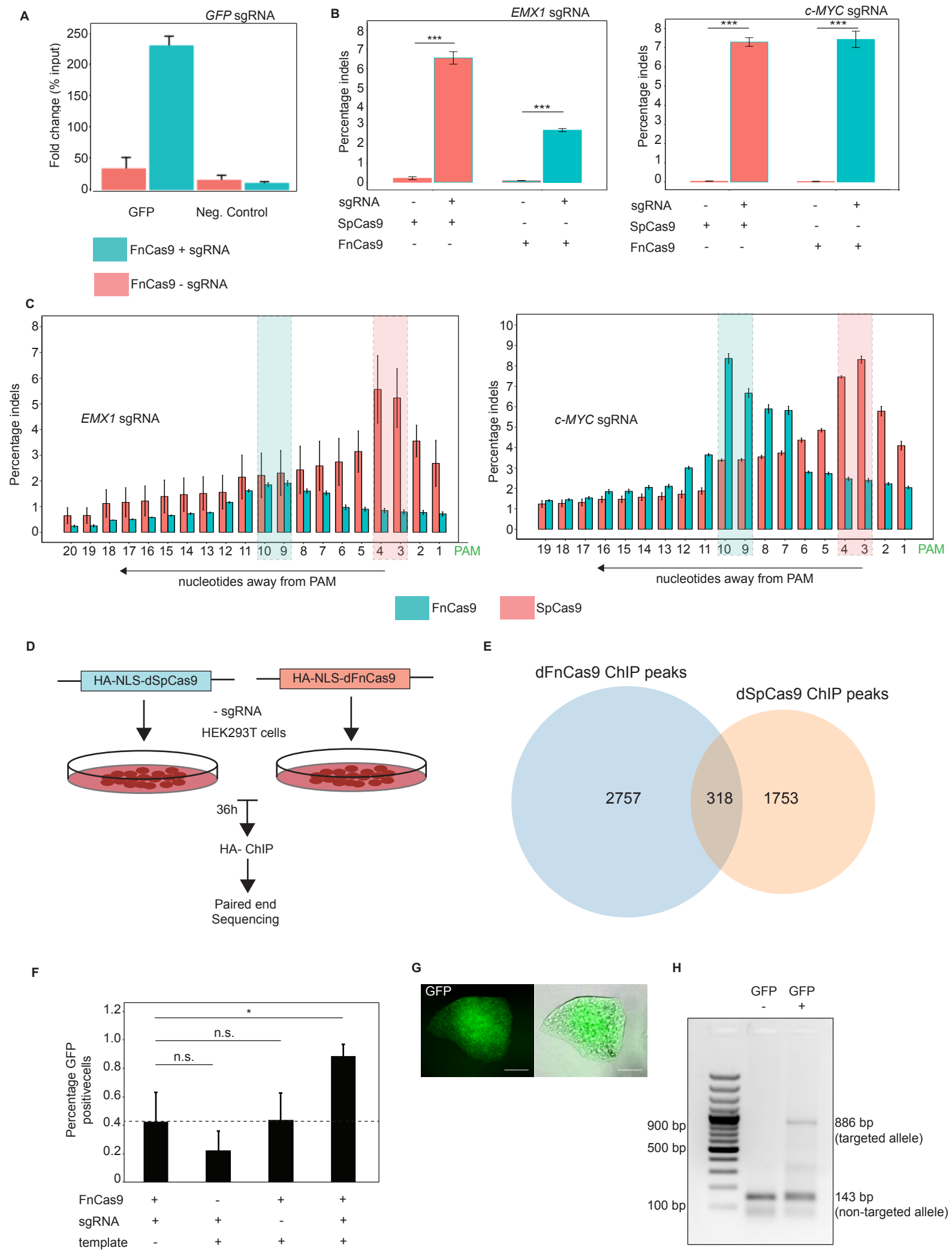

A
GFP sgRNA
Fold change (% input)
GFP
Neg. Control
FnCas9 + sgRNA
FnCas9 - sgRNA
B
EMX1 sgRNA
Percentage indels
sgRNA
SpCas9
FnCas9
c-MYC sgRNA
C
EMX1 sgRNA
c-MYC sgRNA
Percentage indels
nucleotides away from PAM
PAM
FnCas9
SpCas9
D
HA-NLS-dSpCas9
HA-NLS-dFnCas9
- sgRNA
HEK293T cells
36h
HA- ChIP
Paired end
Sequencing
E
dFnCas9 ChIP peaks
dSpCas9 ChIP peaks
2757
318
1753
F
Percentage GFP positivecells
n.s.
n.s.
*
FnCas9
sgRNA
template
G
GFP
H
GFP
GFP
900 bp
500 bp
100 bp
886 bp
(targeted allele)
143 bp
(non-targeted allele)

Figure S1

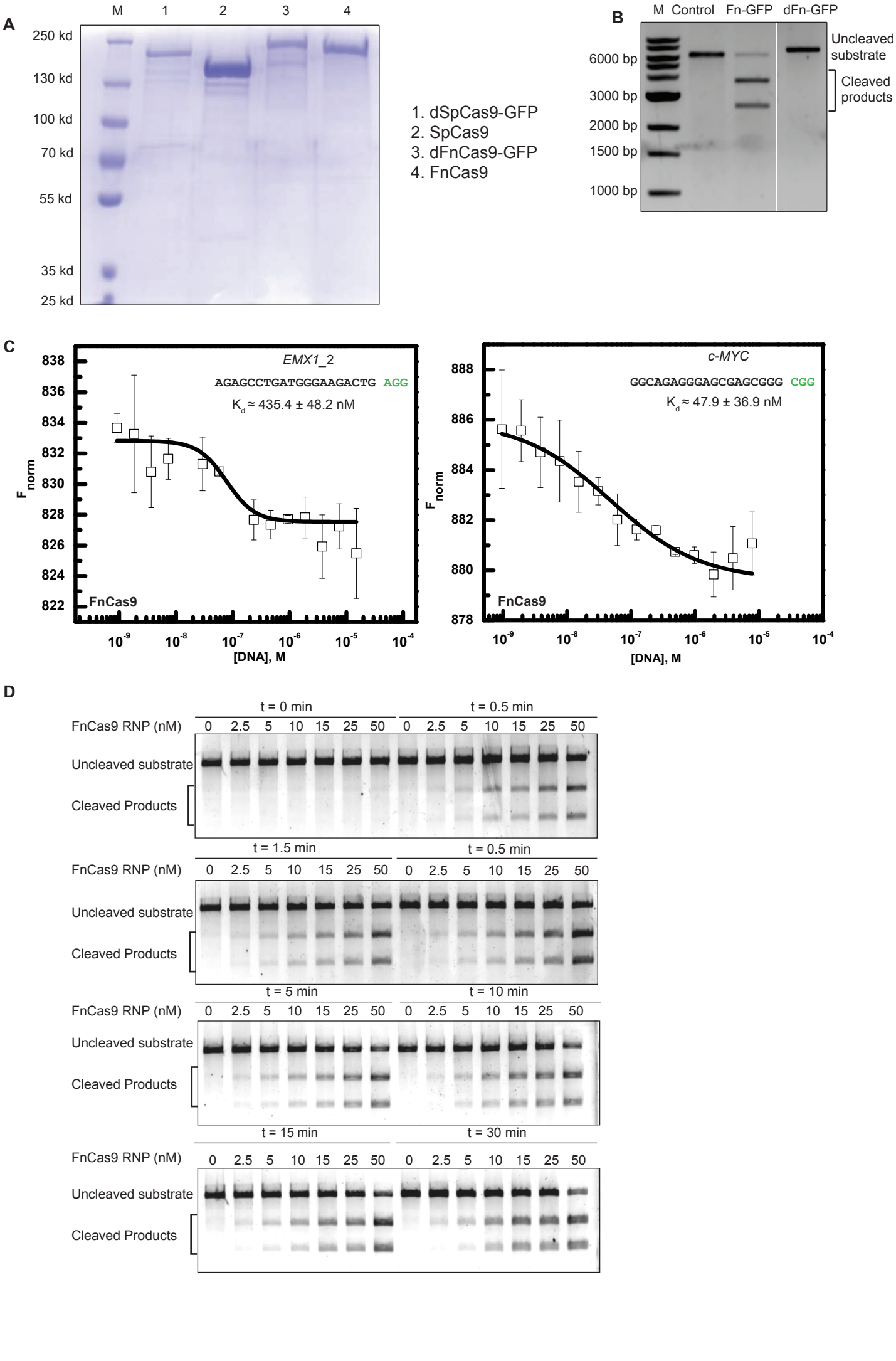
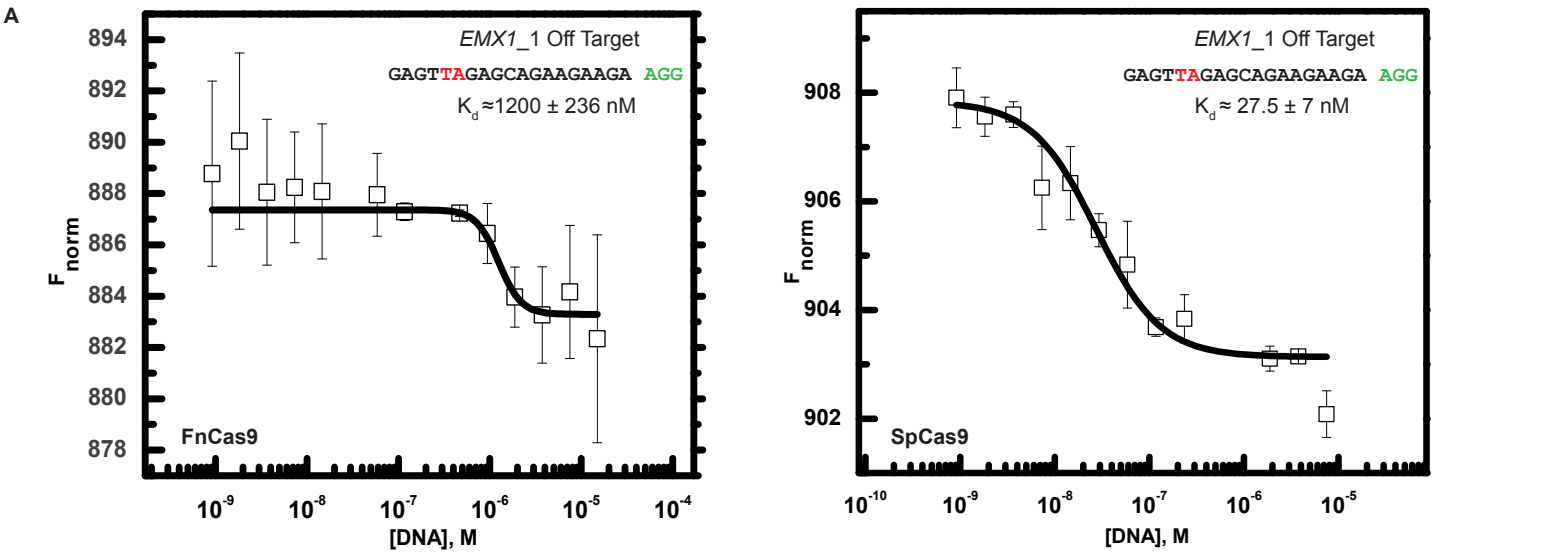
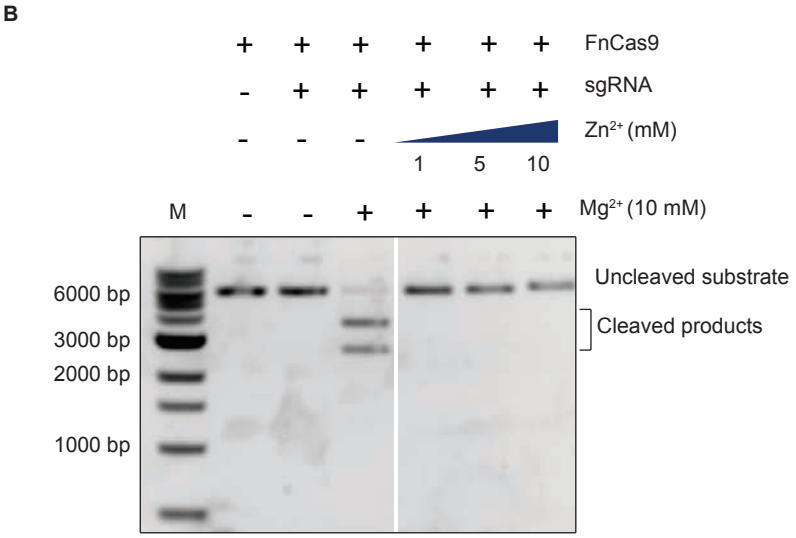
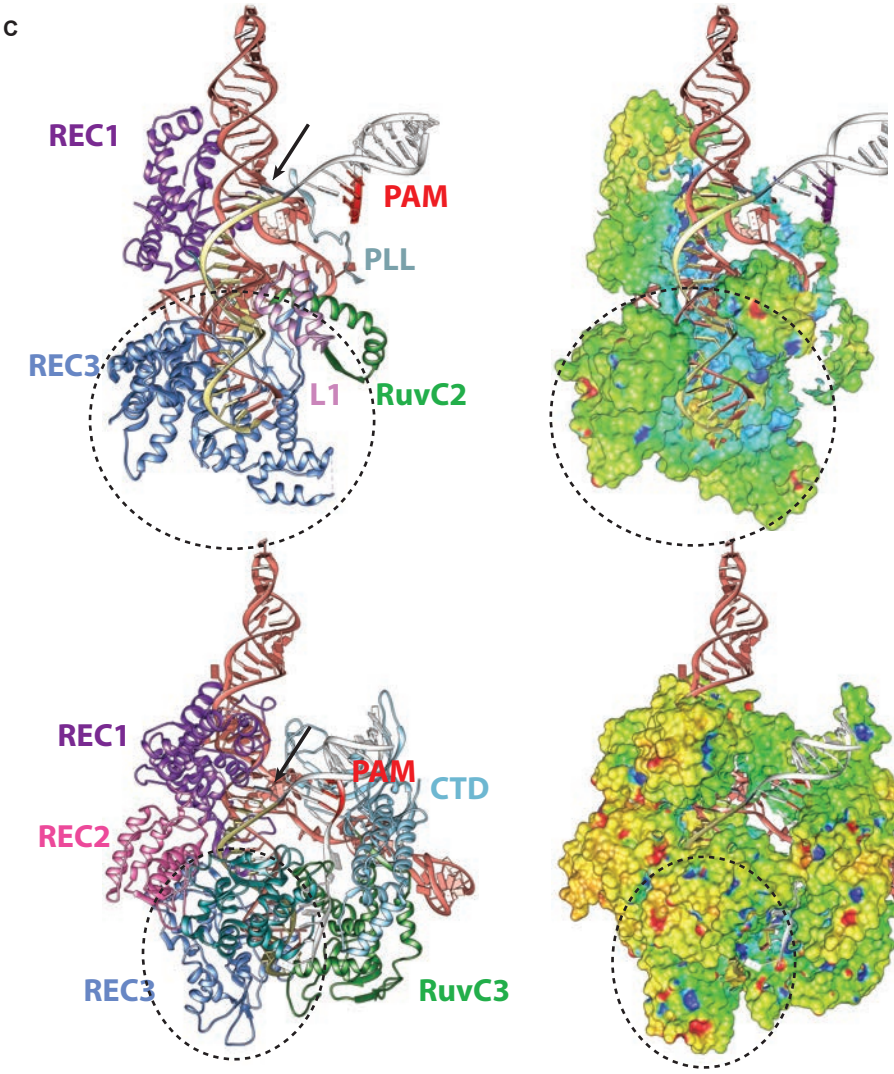

A
M 1 2 3 4
250 kd
130 kd
100 kd
70 kd
55 kd
35 kd
25 kd
1. dSpCas9-GFP
2. SpCas9
3. dFnCas9-GFP
4. FnCas9
B
M Control Fn-GFP dFn-GFP
6000 bp
3000 bp
2000 bp
1500 bp
1000 bp
Uncleaved substrate
Cleaved products
C
EMX1_2
AGAGCCTGATGGGAAGACTG AGG
Kd ≈ 435.4 ± 48.2 nM
FnCas9
Fnorm
[DNA], M
c-MYC
GGCAGAGGGAGCGAGCGGG CGG
Kd ≈ 47.9 ± 36.9 nM
FnCas9
Fnorm
[DNA], M
D
t = 0 min
t = 0.5 min
t = 1.5 min
t = 5 min
t = 10 min
t = 15 min
t = 30 min
FnCas9 RNP (nM)
0 2.5 5 10 15 25 50
Uncleaved substrate
Cleaved Products

Figure S2


A
EMX1_1 Off Target
GAGTTAGAGCAGAAGAAGA AGG
Kd ≈1200 ± 236 nM
FnCas9
Fnorm
[DNA], M
EMX1_1 Off Target
GAGTTAGAGCAGAAGAAGA AGG
Kd ≈ 27.5 ± 7 nM
SpCas9
Fnorm
[DNA], M


B
FnCas9
sgRNA
Zn2+ (mM)
1 5 10
M
Mg2+ (10 mM)
6000 bp
3000 bp
2000 bp
1000 bp
Uncleaved substrate
Cleaved products


C
REC1
PAM
PLL
REC3
L1
RuvC2
REC1
PAM
CTD
REC2
REC3
RuvC3

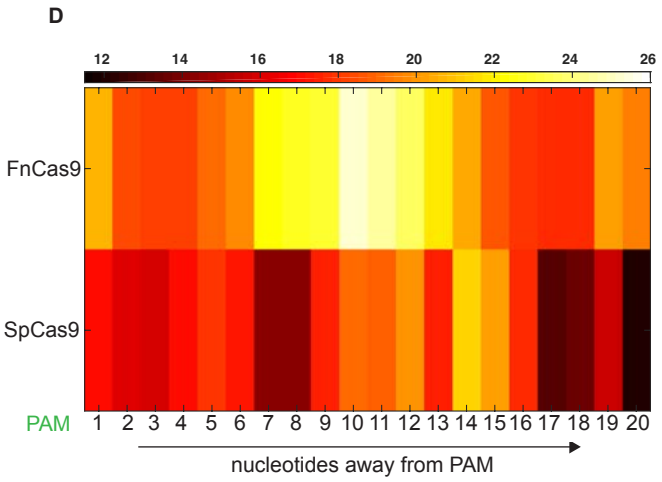

D
12 14 16 18 20 22 24 26
FnCas9
SpCas9
PAM
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
nucleotides away from PAM

Figure S3

A

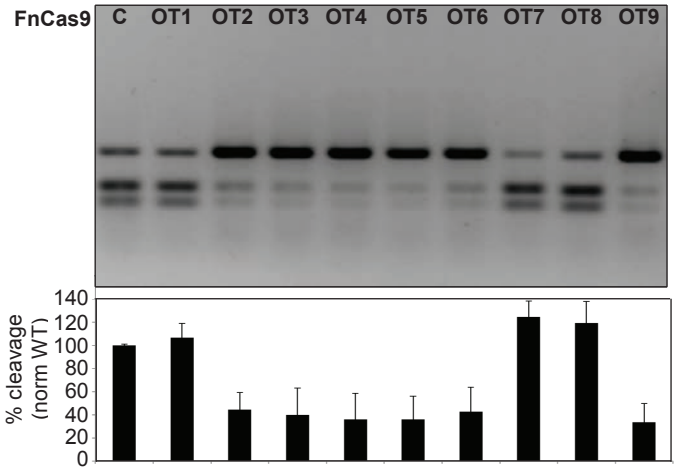
HBB substrate

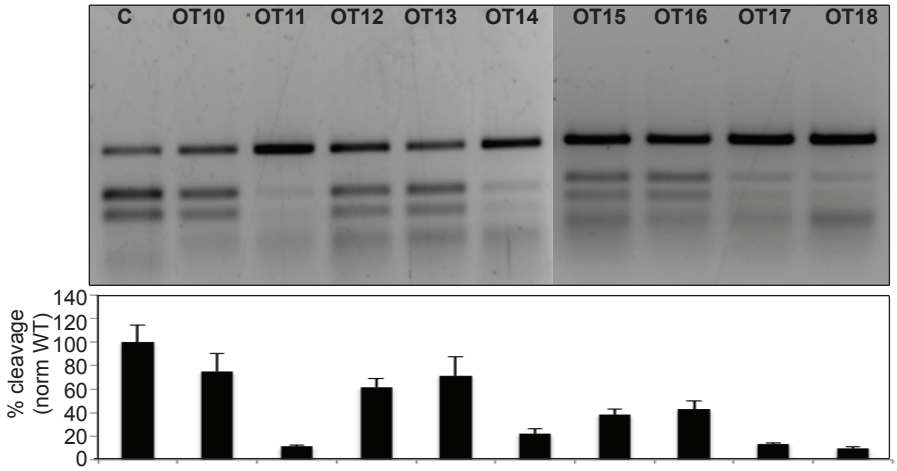
Control	PAM	GAGGAGAAGTCTGCCGTTAC
OT1	PAM	AAGGAGAAGTCTGCCGTTAT
OT2	PAM	GGGGAGAAGTCTGCCGTTGC
OT3	PAM	GAAGAGAAGTCTGCCGTCAC
OT4	PAM	GAGAAGAAGTCTGCCGCTAC
OT5	PAM	GAGGGGAAGTCTGCCATTAC
OT6	PAM	GAGGAAAAGTCTGCTGTTAC
OT7	PAM	GAGGAGGAGTCTGTCGTTAC
OT8	PAM	GAGGAGAGGTCTACCGTTAC
OT9	PAM	GAGGAGAAGCTTGCCGTTAC

B

HBB sgRNA

Control	GTAACGGCAGACTTCTCCTC	PAM
OT10	GTAACGGCAGACTTCCCCCC	PAM
OT11	GTAACGGCAGACTTGTCTTC	PAM
OT12	GTAACGGCAGACTCCTTCTC	PAM
OT13	GTAACGGCAGACCTCCCCTC	PAM
OT14	GTAACGGCAGAGTTGTCCTC	PAM
OT15	GTAATGGTAGACTTCTCCTC	PAM
OT16	GTAGCGACAGACTTCTCCTC	PAM
OT17	GTGACAGCAGACTTCTCCTC	PAM
OT18	GCAATGGCAGACTTCTCCTC	PAM





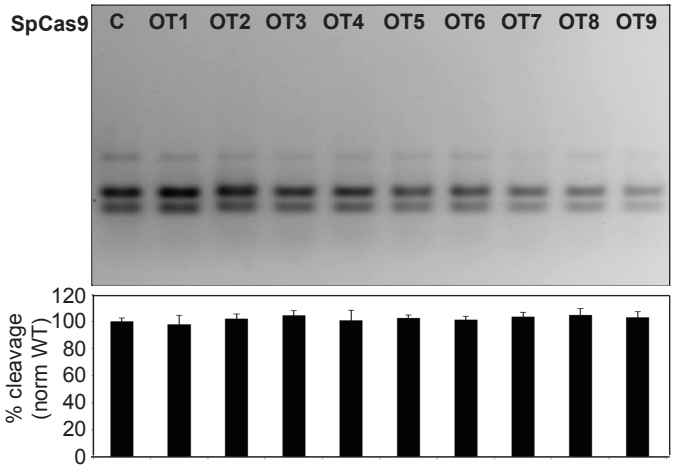
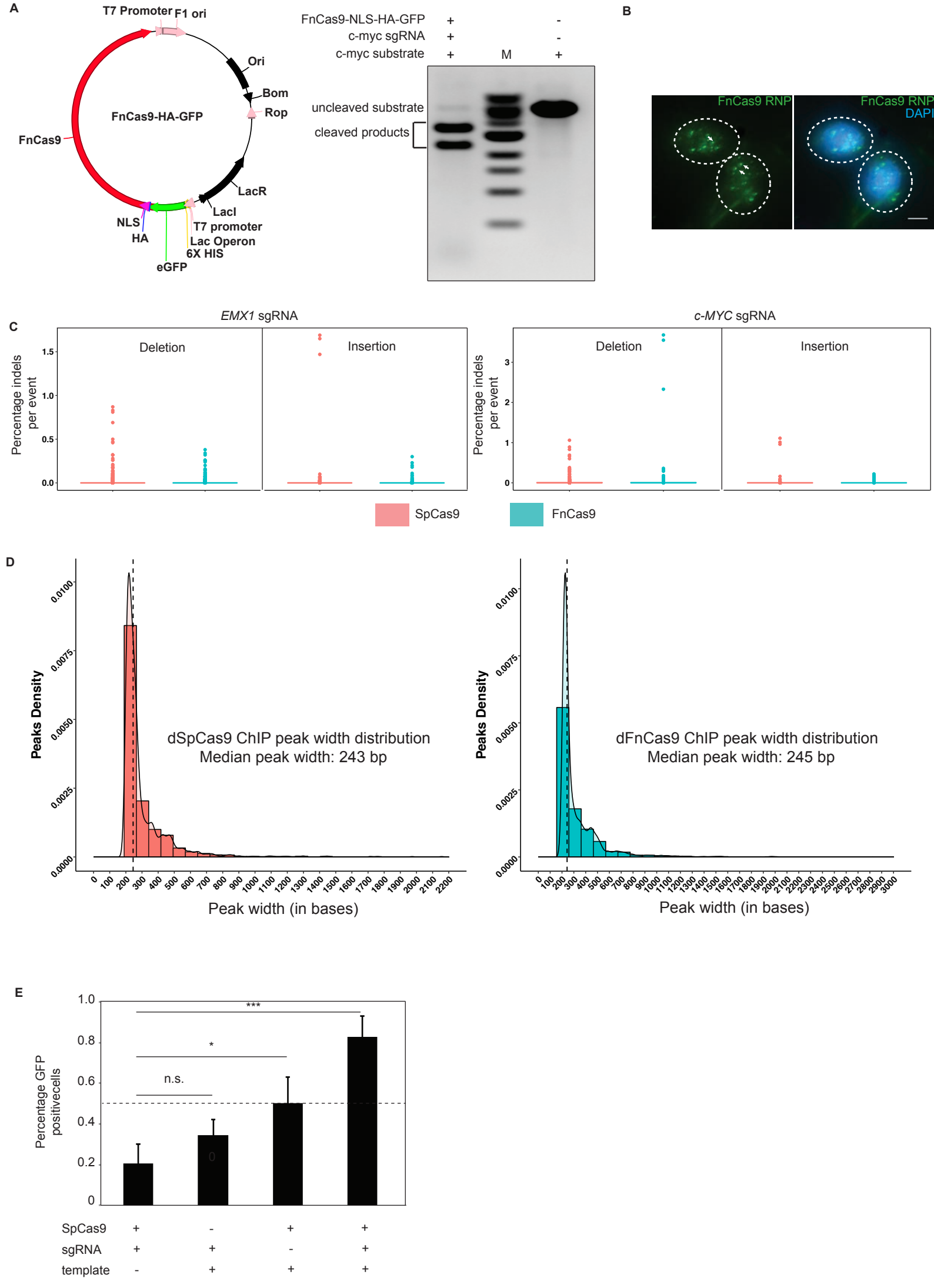


Figure S4


A
T7 Promoter
F1 ori
Ori
Bom
Rop
FnCas9-HA-GFP
FnCas9
LacR
LacI
NLS
HA
T7 promoter
Lac Operon
6X HIS
eGFP
FnCas9-NLS-HA-GFP
c-myc sgRNA
c-myc substrate
M
uncleaved substrate
cleaved products
B
FnCas9 RNP
FnCas9 RNP
DAPI
C
EMX1 sgRNA
c-MYC sgRNA
Deletion
Insertion
Percentage indels per event
SpCas9
FnCas9
D
Peaks Density
dSpCas9 ChIP peak width distribution
Median peak width: 243 bp
dFnCas9 ChIP peak width distribution
Median peak width: 245 bp
Peak width (in bases)
E
Percentage GFP positivecells
n.s.
*
***
SpCas9
sgRNA
template