# RNA-dependent RNA targeting by CRISPR-Cas9

Steven C. Strutt[1], Rachel M. Torrez[1,2], Emine Kaya[1,3], Oscar A. Negrete[4] and Jennifer A. Doudna[1, 5-8]

[1]Department of Molecular and Cell Biology, University of California, Berkeley, CA

[2]Current address: Department of Medicinal Chemistry, University of Michigan, Ann Arbor MI

[3]Current address: Global Blood Therapeutics, 400 E. Jamie Court, South San Francisco, CA

[4]Sandia National Laboratories, Biotechnology and Bioengineering Department, Livermore, CA

[5]Howard Hughes Medical Institute

[6]Department of Chemistry, University of California, Berkeley, CA

[7]Innovative Genomics Institute, University of California, Berkeley, CA

[8]MBIB Division, Lawrence Berkeley National Laboratory, Berkeley, CA

Contact: doudna@berkeley.edu

Running title: RNA targeting by CRISPR-Cas9

**Abstract**

Double-stranded DNA (dsDNA) binding and cleavage by Cas9 is a hallmark of type II CRISPR-Cas bacterial adaptive immunity. All known Cas9 enzymes are thought to recognize DNA exclusively as a natural substrate, providing protection against DNA phage and plasmids. Here we show that Cas9 enzymes from both subtypes II-A and II-C can recognize and cleave single-stranded RNA (ssRNA) by an RNA-guided mechanism that is independent of a protospacer-adjacent motif (PAM) sequence in the target RNA. RNA-guided RNA cleavage is programmable and site-specific, and we find that this activity can be exploited to reduce infection by single-stranded RNA phage *in vivo*. We also demonstrate that Cas9 can direct PAM-independent repression of gene expression in bacteria. These results indicate that a subset of Cas9 enzymes have the ability to act on both DNA and RNA target sequences, and suggest the potential for use in programmable RNA targeting applications.

**Introduction**

  Prokaryotic clustered regularly interspaced short palindromic repeat (CRISPR)-CRISPR-associated (Cas) systems provide immunity against plasmids and bacteriophage by using foreign DNA stored as CRISPR spacer sequences together with Cas nucleases to stop infection (Wright et al. 2016; Mohanraju et al. 2016). One such nuclease, Cas9 of the type II systems, employs a CRISPR RNA (crRNA) and a trans-activating crRNA (tracrRNA) to target spacer-complementary regions (protospacers) on the foreign genetic element to guide double-stranded DNA cleavage (Jinek et al. 2012). A protospacer adjacent motif (PAM) must also be present for the Cas9-RNA complex to bind and cleave DNA (Jinek et al. 2012; Gasiunas et al. 2012; Anders et al. 2014; Szczelkun et al. 2014). Combining the crRNA and tracrRNA into a chimeric, single-guide RNA (sgRNA) simplified the system for widespread adoption as a versatile genome editing technology (Jinek et al. 2012).

  To date, both genetic and biochemical data support the conclusion that *in vivo*, Cas9 is exclusively a DNA-targeting enzyme. Nonetheless, multiple studies have harnessed Cas9 for RNA targeting under specific circumstances. For example, the *S. pyogenes* Cas9 (SpyCas9) can be supplied with a short DNA oligo containing the PAM sequence (a PAMmer) to induce single-stranded RNA (ssRNA) binding and cutting (O'Connell et al. 2014; Nelles et al. 2016). More recently, it was demonstrated that SpyCas9 could be used to target repetitive RNAs and repress translation in certain mRNAs in the absence of a PAMmer (Liu et al. 2016; Batra et al. 2017). A different Cas9 homolog from *Francisella novicida* (FnoCas9) has been implicated in degradation of a specific mRNA but through a mechanism independent of RNA-based cleavage (Sampson et al. 2013). Together with evidence that some Cas9 homologs can target single-stranded DNA substrates under some conditions (Ma et al. 2015; Zhang et al.

2015), these studies raised the possibility that certain Cas9 enzymes might have intrinsic RNA-guided RNA cleavage activity.

To determine whether evolutionarily divergent Cas9 homologs have a native capacity for programmable RNA targeting, we compared biochemical behavior of enzymes from the three major Cas9 subtypes. This analysis revealed that certain type II-A and II-C Cas9s can bind and cleave single-stranded RNA sequences with no requirement for a PAM or PAMmer. Furthermore, we found that this activity can inhibit gene expression and confer moderate protection against infection by ssRNA phage through a mechanism reminiscent of RNA-guided DNA targeting. These results establish the utility of Cas9 for facile RNA-guided RNA targeting and suggest that this activity may have biological relevance in bacteria.


**Results**

**Cas9 catalyzes PAM-independent RNA-guided RNA cleavage**

To assess whether divergent Cas9 enzymes can catalyze binding to and cleavage of RNA substrates by a mechanism distinct from that of double-stranded DNA cleavage, we tested homologs from the three major subtypes of Cas9 proteins for their ability to cleave single-stranded RNA *in vitro* (Figure 1A, B; Figure 1-figure supplement 1A-C*)*. When programmed with a cognate sgRNA, *S. aureus* Cas9 (SauCas9) and *C. jejuni* Cas9 (CjeCas9) direct cleavage of RNA in the absence of a PAMmer (Figure 1; Figure 1-figure supplement 1). No RNA cleavage was detected using SpyCas9, which requires a PAMmer for efficient RNA cleavage *in vitro* (O'Connell et al. 2014), or using *F. novicida* Cas9 (FnoCas9). While the cleavage efficiencies for both SauCas9 and CjeCas9 are indistinguishable (Figure 1-figure supplement 1D), we focused on the activity of SauCas9 due to the abundance of mechanistic and structural data for this

enzyme (Nishimasu et al. 2015; Friedland et al. 2015; Ran et al. 2015; Kleinstiver et al. 2015).

RNA cleavage activity and products were similar to those of canonical Cas9-mediated DNA cleavage activity *in vitro*. RNA targeting by SauCas9 requires the presence of a guide RNA and a catalytically-active protein, as both apo protein lacking the guide RNA and a catalytically inactive mutant (D10A and N580A) do not cleave RNA (Figure 1-figure supplement 2A). Furthermore, addition of EDTA to chelate divalent metal ions abolished RNA cleavage, verifying that divalent metal ions are necessary for catalysis. As with DNA substrates (Sternberg et al. 2014), incubation of SauCas9 with an excess of RNA target demonstrated that cleavage is single-turnover (Figure 1-figure supplement 2B, C). Hydrolysis mapping of the cleavage product revealed that the predominant RNA cleavage site is shifted by one nucleotide compared to the site of DNA cleavage (Garneau et al. 2010; Jinek et al. 2012; Gasiunas et al. 2012) (Figure 1-figure supplement 2D, E). The shift is consistent with that observed for PAM-dependent SpyCas9 RNA-cleavage (O'Connell et al. 2014) and is likely due to the more compact geometry of an RNA-RNA helix relative to an RNA-DNA hybrid helix (Wang et al. 1982).

SauCas9 targets ssRNA in the absence of a PAMmer, a contrast to SpyCas9 targeting of ssRNA (O'Connell et al. 2014). Testing SauCas9 *in vitro* ssRNA cleavage in the presence of a PAMmer (30x molar excess over ssRNA target) revealed that turn-over was two-fold slower than the reaction with only target ssRNA (Figure 1C, Figure 1-figure supplement 3C). SauCas9 ssRNA cleavage conducted in the presence of a non-complementary, control DNA oligo did not yield a similar reduction in cleavage rate (Figure 1-figure supplement 3C), indicating that the complementary PAMmer impairs RNA cleavage activity. Consistent with cleavage being guide-dependent, single-stranded RNA that is not complementary to the sgRNA is not cleaved (Figure 1 and Figure 1-

figure supplement 3). Double-stranded RNA (dsRNA) is also not a substrate for SauCas9.

Given that Cas9 proteins are active with different length guide RNA segments (~20 - 24 nt) (Chylinski et al. 2013; Ran et al. 2015; Friedland et al. 2015; Kim et al. 2017), we tested whether longer guide segments might enhance ssRNA targeting activity. Increasing the length of the targeting region of the guide up to 23 nt results in tighter binding and more efficient cleavage (Figure 1-figure supplement 4), mirroring the preference for longer guides for DNA cleavage (Ran et al. 2015; Friedland et al. 2015). Extending the guide strand complementarity to the target beyond 23 nt did not increase RNA target binding or cleavage efficiency, indicating that 23 nt is the optimal length for *in vitro* binding and targeting applications. The apparent dissociation constant ($K_{d,app}$) of the SauCas9-sgRNA complex (23 nt targeting region) for the ssRNA target is 1.8 ± 0.09 nM (Figure 1-figure supplement 4D), which is ~5x weaker than the 0.34 ± 0.03 nM binding affinity measured for a dsDNA substrate of the same sequence.


**Cleavage efficiency is impaired by duplex regions in target RNA**

We noted that SauCas9-catalyzed ssRNA cleavage is limited to ~30% fraction cleaved (Figure 1-figure supplement 3), compared to >80% fraction cleaved for ssDNA and dsDNA targets.  Greater thermodynamic stability of RNA secondary structures, relative to those in ssDNA (Bercy and Bockelmann 2015), might occlude SauCas9-sgRNA binding to an ssRNA target sequence, a possibility that we tested using a panel of partially duplexed RNA substrates (Figure 2). Previously, introduction of a short segment of mismatched base pairs to mimic partially unwound dsDNA substrates was shown to enhance the ability of type II-C Cas9s (including CjeCas9) to unwind and cleave dsDNA (Ma et al. 2015). Here, we found that duplex-RNA substrates containing a 2- or 6-base pair mismatched segment located near the 5' or 3' end of the 23 nt guide

6

RNA region of the sgRNA could not be cleaved (Figure 2A-C, substrates 5, 6, 10, and 11). However, when the unpaired region was increased to 12-base pairs, SauCas9 was able to cleave the target strand. There was a slight cleavage preference for RNA substrates in which the 12-base pair mismatched segment is located near the 5′ end of the guide sequence of the sgRNA (Figure 2A-C, substrates 7 and 12).

Interestingly, the 23-base pair mismatched segment RNA substrates ('Bubble' substrates 8 and 9) are targeted more efficiently than their ssRNA counterparts (substrates 1 and 2) (Figure 2C). We measured the binding affinity of all substrates and found that both the 23-base pair mismatched segment RNA and ssRNA substrates are bound with similar affinity (Figure 2D). Furthermore, the apparent difference in cleavage efficiency was not due to the presence of a double-stranded PAM sequence, as mutating the PAM region does not impair cleavage (Figure 2C, compare substrates 8 and 9). We hypothesize that RNA containing a mismatched segment presents a more accessible substrate to the Cas9-sgRNA complex due to stable annealing between the ends of the non-target and target strands, whereas the ssRNA substrate alone has ends that are predicted to stabilize a conformation that is partially structured and therefore inaccessible (Figure 2-figure supplement 1A).

An alternative hypothesis to explain the limited cleavage of ssRNA substrates is that SauCas9 enzyme inactivation occurs over the course of the reaction, even with SauCas9 protein-sgRNA (ribonucleoprotein, RNP) present in 10-fold excess relative to the ssRNA substrate. To test this, we spiked reactions with fresh SauCas9 protein alone or SauCas9 RNP after reactions reached equilibrium; however, we did not observe an increase in the amount of ssRNA cleavage (Figure 2-figure supplement 1B, C). We also tested whether the SauCas9 RNP was able to cleave a second ssRNA substrate that was added to the reaction after it reached completion (Figure 2-figure supplement 1D, E). After 1hr of incubation, the addition of a second target ssRNA complementary to the

7

guide RNA resulted in a burst of cleavage activity, whereas a non-complementary ssRNA substrate did not stimulate cleavage. The second target ssRNA is cleaved to a comparable extent to that observed when this second target was the only substrate in the reaction (Figure 2-figure supplement 1D, E, compare reactions 1 and 3). These observations suggest that SauCas9 RNP is still competent and available for cleavage at the end of the reaction and that a property intrinsic to the ssRNA substrate is the limiting factor. We propose that the observed difference in cleavage extents for various RNA substrates reflects the fraction of molecules that are structurally accessible for cleavage by the SauCas9 RNP.

**SauCas9 confers *in vivo* protection against RNA phage**

Based on the biochemical ability of SauCas9 RNP to bind and cleave ssRNA substrates, we wondered whether this activity might provide protection against RNA phage infection in bacteria. To test this, we generated a plasmid library encoding sgRNAs containing guide sequences complementary to the genome of MS2, a single-stranded RNA phage that can infect *E. coli*. A subset of these sgRNAs contained scrambled guide sequences that should not target MS2, providing negative controls. Another sgRNA subset included single-nucleotide mismatches introduced at each position of a target sequence to test for mismatch sensitivity in ssRNA recognition. This plasmid library, comprising 18,114 sgRNAs, was co-transformed into *E. coli* along with a vector encoding a catalytically active version of SauCas9 and the population of transformants was subjected to infection by bacteriophage MS2 (Figure 3A). The experiment was performed in biological triplicate and included an untreated control population and two experimental conditions (multiplicities of infection (MOIs) of 10 and 100). After selection, plasmids were recovered from surviving colonies and sequenced (Figure 3A).

We identified between 131 and 166 sgRNAs that were significantly enriched (false discovery rate (FDR)-adjusted p-value < 0.05) in the two different MS2 infection conditions (Figure 3B). The majority of these sgRNAs were perfectly complementary to the MS2 genome, and only three and five control sgRNAs (out of 708 total control sgRNAs) for the MOI-10 and -100 conditions, respectively, were enriched (Figure 3B). The lengths of enriched guide sequences were skewed towards shorter targeting lengths (Figure 3-figure supplement 1A, left); however, this likely reflects bias in the cloned input library since the ratio between the enriched guide sequences and those of the library without phage selection are similar (Figure 3-figure supplement 1A, right). When comparing the degree of enrichment between the different guide lengths, the 23-nt guide segment sgRNAs were preferentially enriched over those of shorter length (Figure 3C), consistent with the *in vitro* observation that longer guides are more efficient for directing ssRNA cleavage (Figure 1-figure supplement 4C). To assess whether there was any sequence bias within the enriched guides, we aligned guide sequences of all lengths at their 3' end. These alignments showed no specific sequence bias in the enriched guides relative to those in the unselected library (Figure 3-figure supplement 1B). This is consistent with the crystal structure of an SauCas9-sgRNA-DNA bound complex which revealed the absence of base-specific contacts of Cas9 to the target strand (Nishimasu et al. 2015).

Strikingly, mapping enriched guide sequences onto the MS2 genome showed that enriched sgRNAs were clustered at specific regions, which were consistent across both experimental conditions (Figure 3D; Figure 3-figure supplement 1C, D). Together with our biochemical data suggesting that SauCas9 cannot bind or cleave structured RNAs (Figure 3), we interpret these targeting "hotspots" to be regions of low structural complexity. It is important to note that sgRNAs containing different guide segment lengths overlap at these regions, possibly indicating that increases in targeting efficiency

9

due to guide length are secondary to target accessibility to the Cas9 RNP. We mapped the enriched guide sequences onto the published secondary structure of the MS2 genome determined through cryoelectron microscopy (Dai et al. 2017) (Figure 3-figure supplement 2). Guides targeted not only single-stranded, accessible regions but also those that form apparently stable secondary structures. The structure of the MS2 genome was determined on the intact phage particle, however, and may not represent the RNA structure(s) relevant to the infection stage during which SauCas9-mediated protection is crucial.

Highly enriched sgRNAs from the screen were confirmed for their ability to confer protection against MS2 phage infection through a soft-agar plaque assay. Reconstitution of SauCas9 with a targeting guide confers approximately a ten-fold protection against the RNA phage (Figure 3E, F). No protection was observed in the absence of an sgRNA or SauCas9 protein. Scrambling the sequence of the guide also abrogates protection, confirming that sequence complementary is necessary for phage elimination. Guide segments of all lengths tested (20-23 nts) conferred protection to a similar level (Figure 3-figure supplement 3A, B), consistent with the result from the MS2 screen that guide segments of all lengths were enriched in 'hotspot' regions (Figure 3D; Figure 3-figure supplement 1C). Two 'control' guides were enriched in both the MOI-10 and -100 treatments. Interestingly, both guides conferred protection but their scrambled counterparts did not (Figure 3-figure supplement 3C, D). Whereas a possible off-target binding site was found for one guide (#14238) within the MS2 genome (Figure 3-figure supplement 3E), it remains unclear how guide #14210 confers protection. Possibly this sgRNA acts by targeting an *E. coli* host factor that is necessary for infection.

Screening against the MS2 genome was also used to test the effect of single-nucleotide mismatches on SauCas9's targeting ability. We computed an average fold change (between phage treated and untreated samples) for all sgRNAs that contained a

mismatch at the same position, and obtained average values for mismatches at each position across the guide. We observed a pronounced gradient of increasing guide stringency with length. On average, short guides were less sensitive to mismatches, while mismatches in longer sgRNAs led to decreased recovery compared to control samples (Figure 3-figure supplement 4A, B). Previous work and models suggest that shorter guide segments should be more sensitive to mismatches and lead to higher fidelity Cas9 targeting (Fu et al. 2014; Bisaria et al. 2017). Further study is needed to thoroughly examine this unexpected pattern of RNA-targeting stringency, as one shortcoming of this experiment is that mismatched guides were not designed, *a priori*, to recognize accessible parts of the MS2 genome. Nevertheless, despite potential noise introduced in this analysis due to guide segments that target inaccessible MS2 regions, we observe an interesting correlation between mismatches in the MS2 screen and *in vitro* biochemical cleavage assays for the sgRNA with a 23 nt guide segment sequence (Figure 3-figure supplement 4C, D). The first few nucleotides in the 'seed' region (guide 3´ end proximal) are sensitive to mismatches, while a central region of sensitivity is also observed, similar to previously demonstrated regions of sensitivity for SpyCas9 DNA cleavage (Cong et al. 2013; Jiang et al. 2013; Fu et al. 2016; Gorski et al. 2017).

**SauCas9 represses gene expression in *E. coli***

An efficient RNA-targeting Cas9 could serve as an important tool in regulating gene expression *in vivo*. To test the ability of SauCas9 to mediate repression of host gene expression, we targeted dSauCas9 and dSpyCas9 RNPs to a GFP reporter sequence encoded in the *E. coli* chromosome (Qi et al. 2013). Catalytically inactive versions of Cas9 were used to prevent cleavage of the bacterial chromosome when targeting a site adjacent to a PAM. As expression of Cas9 and sgRNA exerts metabolic stress on *E. coli*, GFP fluorescence values were normalized by the $OD_{600}$ value to

11

account for differences in cell growth between cultures (Oakes et al. 2016). When using

sgRNAs designed to recognize a sequence in the GFP gene adjacent to the appropriate

PAM for SauCas9 (NNGRRT) or SpyCas9 (NGG), GFP expression is significantly

reduced (Figure 4A) consistent with CRISPR-interference (CRISPRi) (Qi et al. 2013;

Gilbert et al. 2014).  When sgRNAs were designed to recognize GFP sequences not

flanked by a PAM, dSauCas9 but not dSpyCas9 was able to repress GFP expression.

The SauCas9-mediated GFP repression was dependent on sgRNAs that target the

coding strand; sgRNAs that recognize the non-coding strand did not result in reduced

GFP expression (Figure 4-figure supplement 1A). The length of the targeting sequence

*in vivo* corroborates *in vitro* data, with longer guides working more efficiently (Figure 4B).

Different guide sequences display variable efficiencies of targeting. We tiled

sgRNAs across the GFP mRNA sequence to test the robustness of dSauCas9 to

repress GFP expression (Figure 4C). As no sites are adjacent to PAM sequences, all

repression presumably occurs on the mRNA level. The efficiency of dSauCas9-mediated

GFP repression varied according to the target sequence, with some dSauCas9 RNPs

reducing GFP signal to 15-30% of that observed in the presence of the sgRNA alone

(Figure 4C, GFP2 and 6) and others showing no ability to repress GFP expression

(GFP7 and 9). Electrophoretic mobility shift assays support the conclusion that

repression is not occurring at the dsDNA level by promiscuous PAM binding (Figure 4-

figure supplement 1B). Repression is largely equivalent between catalytically active and

inactive forms of SauCas9 (Figure 4-figure supplement 1C), suggesting that binding of

the Cas9-sgRNA complex to the mRNA is sufficient for repression and consistent with *in*

*vitro* data showing that the enzyme does not catalyze multiple-turnover RNA cleavage.

While we speculate that the Cas9-RNP blocks the ribosome directly (either at initiation or

during elongation), our data do not rule out the possibility that Cas9 is otherwise

destabilizing the mRNA transcript through an unknown mechanism.

Together our biochemical and *in vivo* data support a model in which SauCas9 can readily bind and cleave bacteriophage RNA and mRNA sequences that are exposed and unstructured (Figure 4D). Regions that form strong structures are inaccessible to SauCas9 RNP binding, thereby preventing cleavage or repression activity. As Cas9 cleavage activity is limited by target accessibility, we expect that RNA occluded by RNA-binding proteins would also be recalcitrant to cleavage.

**Discussion**

Investigation of CRISPR-Cas9 has focused on its function as a double-stranded DNA endonuclease, while the ability of diverse homologs to cleave natural RNA substrates has remained unexplored. Here, we present evidence that type II-A and type II-C Cas9 enzymes can catalyze programmable and PAM-independent single-stranded RNA cleavage. Focusing on SauCas9, we show that this enzyme can be employed both biochemically and in cells to cleave RNA and regulate genes on both the transcriptional and translational level in parallel by accounting for target site PAM proximity. Importantly, SauCas9 ssRNA scission requires only an sgRNA and does not need a PAMmer, thereby simplifying applications (Nelles et al. 2015) and facilitating delivery to cells as a pre-assembled RNP (Zuris et al. 2015; Mout et al. 2017)

The RNA-targeting capability of SauCas9 and related Cas9 enzymes offers the advantage of repressing viruses whose lifecycles do not involve a DNA genome or intermediate, thereby rendering them inaccessible to Cas9-mediated DNA cleavage. We demonstrated that SauCas9 could be programmed to confer protection to *E. coli* against MS2, an RNA bacteriophage with no DNA intermediate. Whether RNA-based viral repression by Cas9 occurs in natural systems is not known, but seems possible based on our results. DNA cleavage by SauCas9 remains more rapid than RNA cleavage, indicating that DNA-targeting is probably the biologically preferred method for phage and

plasmid interference. However, Cas9 activity on RNA is PAM-independent and may mitigate the effects of PAM-escape mutants that would evade DNA-level interference (Deveau et al. 2008), thus acting as an additional line of defense.

Intriguingly, 'hotspots' of preferential targeting emerged when tiling guides across the genome, but these sites were devoid of sequence bias. In conjunction with *in vitro* cleavage data of partially structured RNAs, we suggest that SauCas9 cleavage efficiency is inversely related to structural complexity of the RNA target. As an alternative to the current approach of screening multiple sgRNAs for activity, experimental knowledge about RNA structure, such as SHAPE-seq data (Loughrey et al. 2014), would simplify target identification for viral targeting and repression experiments. Nevertheless, future work will concentrate on understanding the structural constraints on RNA targeting and methods to improve Cas9 access to duplex RNA regions.

SauCas9 holds promise for a range of RNA targeting applications. We showed that SauCas9 could repress gene expression in *E. coli*. Repression of the reporter occurs in the absence of the PAM and is specific for targeting of the coding strand. Recently, the Type VI CRISPR-Cas system effector, Cas13, has been proposed and demonstrated to target RNA (Shmakov et al. 2015; Abudayyeh et al. 2016; East-Seletsky et al. 2016). 'Activated' Cas13 exhibits robust *trans* cleavage of RNAs(Abudayyeh et al. 2016; East-Seletsky et al. 2016; Smargon et al. 2017). While RNA-cleavage by SauCas9 is single-turnover and kinetically less robust than that of Cas13, Cas9 does not cleave RNAs indiscriminately and lends itself to targeting of specific transcripts. A programmable Cas9 capable of repressing genes on the RNA level has potential advantages over CRISPRi DNA-based techniques (Qi et al. 2013; Gilbert et al. 2014). For example, isoform-specific targeting of different transcripts originating from the same transcription start site or resulting from alternative splicing events might be possible. More broadly, due to its intrinsic ssRNA-binding activity,

14

SauCas9 may have utility as a platform for directing other effector proteins to specific RNA molecules, such as proteins or domains that up-regulate translation or RNA base-modifying enzymes for site-specific epigenetic modification of RNAs.

## Materials and methods

Strains and plasmids used in this study can be found in the Key Resources Table while all nucleic acids sequences and substrates are listed in Supplemental File 1.

### Phylogenetic tree construction and RNA folding

Cas9 homolog sequences were obtained from Chylinski and colleagues (Chylinski et al. 2014). A structure-guided alignment was produced using PROMALS3D (Pei et al. 2008) and a maximum-likelihood tree was inferred using PHYML3.0 (Guindon et al. 2010). The structure of the pUC ssRNA target was predicted using Mfold (Zuker 2003).

### Protein purification

All proteins were expressed as His-MBP fusions (Addgene vector #29706) in *E. coli* strain BL21(DE3). Cells were grown to an $OD_{600}$ of 0.6-0.8, induced with 0.4M IPTG, and then incubated overnight at 16°C with shaking. Proteins were purified using Superflow Ni-NTA affinity resin (Qiagen, Valencia, CA), followed by a HiTrap HP Heparin column (GE Healthcare, Pittsburgh, PA)  and gel filtration on a Superdex S200 (GE Healthcare, Pittsburgh, PA), as previously described (Jinek et al. 2012). Cas9 protein sequences can be found in Supplemental File 1.

### Oligonucleotide purification and radiolabeling

DNA oligonucleotides were synthesized by IDT (Coralville, IA). Target RNAs and sgRNAs were transcribed *in vitro* as previously described (Sternberg et al. 2012). DNA targets and *in vitro* transcribed RNAs were gel purified by 7M urea denaturing PAGE. Target RNAs and DNAs were 5´ end-labeled with [γ-P32-ATP] by treatment with PNK

(NEB, Ipswich, MA). T1 sequencing and hydrolysis ladders were prepared according to manufacturer's directions (Ambion, Grand Island, NY). A list of all sgRNAs and targets can be found in Supplemental File 1.

**In vitro cleavage assays**

Cas9 was reconstituted with equimolar sgRNA in 1x cleavage buffer (20 mM Tris-HCl – pH 7.5, 200 mM KCl, 1 mM TCEP, 5% glycerol, 5 mM MgCl2) for 10 min at 37°C, then immediately placed on ice. Cleavage reactions were conducted with 1 nM target and 10 nM reconstituted Cas9-sgRNA in 1x cleavage buffer unless otherwise noted. Structured RNA substrates were prepared by annealing two separate *in vitro* transcribed RNAs. The target strand was annealed with 10-fold excess of the non-target strand to ensure that all target is complexed prior to the cleavage reaction. Reactions were incubated at 37°C for the indicated time and quenched in Heparin-EDTA buffer (10 µg/ml heparin, 25 mM EDTA) at 25°C for 5 min. Reactions were diluted with 2x Formamide loading buffer and incubated at 95°C for 5 min prior to separation on a 15% denaturing 7M urea PAGE gel. Gels were dried overnight and exposed to a phosphor imaging screen (Amersham/GE Healthcare, Pittsburgh, PA). Results were visualized on a Typhoon (GE Healthcare, Pittsburgh, PA) and quantified in ImageQuantTL (v8.1, GE Healthcare, Pittsburgh, PA). The cleaved fraction of total signal was calculated independently for three separate experiments and were fit with a one-phase exponential decay model in Prism7 (GraphPad Software, La Jolla, CA).

**Filter binding and electrophoretic mobility shift assays**

Binding reactions consisted of 750 nM catalytically inactive SauCas9 reconstituted with sgRNA to the final concentrations indicated. Radiolabeled target RNA was added to a final concentration of 1nM and the reactions were incubated at 37°C for one hour. Bound probe was separated from unbound using a three-filter system on a vacuum manifold (Rio 2012). Membranes were allowed to dry prior to phosphor imaging and

quantification. EMSAs were performed in the presence of 300 nM dSauCas9 and 1 nM

radiolabeled target strand DNA pre-annealed in the presence of 10x non-target strand.

Complexes were incubated at 37°C for one hour prior to separation on 6% non-

denaturing PAGE. Gels were dried prior to phosphor imaging. Three independent

experiments were performed and the fraction of bound out of total signal was calculated

in ImageQuantTL. Binding isotherms were determined in Prism7 using a one-site binding

model.

**MS2 screen and plaque assay**

All guides of length 20-23 nt antisense to the MS2 bacteriophage genome were

synthesized (CustomArray Inc., Bothell, WA) and cloned into a guide expression vector

(Oakes et al. 2016) modified with the SauCas9 sgRNA scaffold. XL1-Blue *E. coli* cells

with a vector containing a tetracycline-inducible wtSauCas9 construct were made

electrocompetent and transformed with the MS2-guide plasmid library in triplicate.

Approximately 1 x $10^6$ transformants were grown for 30 min at 37°C with shaking prior to

addition of antibiotics and 10nM anhydrotetracycline (aTc) (Sigma, St. Louis, MO) for

protein induction. After an additional 30 min of growth, cultures were split into three

equal pools and treated with none, 3.3 x $10^6$, or 3.3 x $10^7$ MS2 bacteriophage. After

three hours of infection, cells were plated on LB-agar supplemented with antibiotics and

incubated at 37°C for 16 hours. Plates were scraped with LB and plasmids were isolated

using a MidiPrep kit (Qiagen, Valencia, CA), according to the manufacturer's protocol.

High-throughput sequencing libraries were prepared by PCR amplification of the variable

region of the guide plasmid. Dual unique-molecular identifiers (UMIs), included to

separate true single-nucleotide mismatches, as well as duplicates, from PCR artifacts

(Kou et al. 2016), were incorporated during a single round of PCR. Excess UMIs were

removed by ExoI digestion (Thermo Scientific, Waltham, MA) prior to library amplification

and barcoding. Individual guides (Supplemental File 1) were cloned using

oligonucleotides synthesized by IDT and co-transformed into XL1-Blue *E. coli* cells with the SauCas9 vector. Resistance to MS2 bacteriophage was conducted using a soft-agar overlay method (Abudayyeh et al. 2016) and plaque forming units (PFUs) were calculated. To minimize variability in plaquing efficiency, the same phage dilutions were used for all experiments.

**MS2 survival and mismatch analysis**

After applying a low-pass filter, reads were trimmed using cutadapt v. 1.14 (Martin 2011) and paired-end overlapping reads were merged using pandaseq (Masella et al. 2012) for error correction. Reads were mapped to the MS2 genome with bowtie2 v2.3.0 (Langmead and Salzberg 2012) using the "very-sensitive" option and de-duplicated based on the dual-UMI (Smith et al. 2017). Feature counts were obtained using HTSeq-count (Anders et al. 2015). Differential expression was calculated using standard pipelines implemented in "edgeR" (Robinson et al. 2010; McCarthy et al. 2012). Significantly enriched guides were defined as those with an FDR-corrected p-value < 0.05. Guides with a positive fold-change compared to the control were mapped to the MS2 genome and visualized using the "Sushi" package (Phanstiel et al. 2014). To examine for nucleotide composition bias, sequences of guides with a significant positive enrichment were aligned at the 3´ end (PAM-proximal) and motifs were analyzed using the WebLogo server (Crooks et al. 2004). The distribution of $\log_2$ fold-change values of significantly enriched guides plotted as box and whisker plots in Prism. The secondary structure of the MS2 genome was obtained from (Dai et al. 2017) and reads were mapped and visualized in Forna (Kerpedjiev et al. 2015). $\log_2$ fold-change values of single-nucleotide mismatch (SNP) guides for each treatment were partitioned by length and averaged at each position. High-throughput sequencing data accompanying this paper are available through the Sequencing Read Archive under the BioProject accession number PRJNA413805.

### *E. coli in vivo* GFP repression

Based on the system outlined previously, SauCas9 was cloned into a tetracycline-inducible vector, while individual guides are under control of a constitutive promoter (Oakes et al. 2016). Plasmids were transformed into an *E. coli* strain with a GFP reporter gene integrated into the chromosome (Qi et al. 2013). Cultures were grown in M9 medium supplemented with 0.4% w/v glucose to mid-log phase and diluted to an $OD_{600}$ of 0.05 prior to transfer to a Tecan Microplate reader (Tecan Systems, San Jose, CA). Protein expression was induced with 10 nM aTc. GFP and $OD_{600}$ were measured every ten minutes for at least 18 hours. Curves of GFP expression over time were fit with a logistic growth model in Prism. At 80% of the maximum value, or at least after 16 hours of growth, the GFP signal was normalized by cell density at $OD_{600}$. To account for effects of guide and protein expression, GFP/$OD_{600}$ was normalized to a null guide or null protein culture, respectively. As expression of different guides change GFP expression levels, the ratio between normalized RNP and guide values was taken to allow comparison of RNP-based repression across different guides. All experiments were conducted in triplicate and all graphing and quantitative analyses were conducted in Prism. Guide and target sequences can be found in Supplemental File 1.

## References

Abudayyeh OO, Gootenberg JS, Konermann S, Joung J, Slaymaker IM, Cox DBT, Shmakov S, Makarova KS, Semenova E, Minakhin L, et al. 2016. C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353**: aaf5573.

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

Anders S, Pyl PT, Huber W. 2015. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**: 166–169.

Batra R, Nelles DA, Pirie E, Blue SM, Marina RJ, Wang H, Chaim IA, Thomas JD, Zhang N, Nguyen V, et al. 2017. Elimination of Toxic Microsatellite Repeat Expansion RNA by RNA-Targeting Cas9. *Cell* **170**: 899–912.e10.

Bercy M, Bockelmann U. 2015. Hairpins under tension: RNA versus DNA. *Nucleic Acids Research* **43**: 9928–9936.

Bisaria N, Jarmoskaite I, Herschlag D. 2017. Lessons from Enzyme Kinetics Reveal Specificity Principles for RNA-Guided Nucleases in RNA Interference and CRISPR-Based Genome Editing. *Cell Syst* **4**: 21–29.

Chylinski K, Le Rhun A, Charpentier E. 2013. The tracrRNA and Cas9 families of type II CRISPR-Cas immunity systems. *RNA Biol* **10**: 726–737.

Chylinski K, Makarova KS, Charpentier E, Koonin EV. 2014. Classification and evolution of type II CRISPR-Cas systems. - PubMed - NCBI. *Nucleic Acids Research* **42**: 6091–6105. https://www.ncbi.nlm.nih.gov/pubmed/24728998?dopt=Abstract.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819–823.

Crooks GE, Hon G, Chandonia J-M, Brenner SE. 2004. WebLogo: a sequence logo generator. *Genome Res* **14**: 1188–1190.

Dai X, Li Z, Lai M, Shu S, Du Y, Zhou ZH, Sun R. 2017. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **541**: 112–116.

Deveau H, Barrangou R, Garneau JE, Labonté J, Fremaux C, Boyaval P, Romero DA, Horvath P, Moineau S. 2008. Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. *Journal of Bacteriology* **190**: 1390–1400.

East-Seletsky A, O'Connell MR, Knight SC, Burstein D, Cate JHD, Tjian R, Doudna JA. 2016. Two distinct RNase activities of CRISPR-C2c2 enable guide-RNA processing and RNA detection. *Nature* **538**: 270–273.

Friedland AE, Baral R, Singhal P, Loveluck K, Shen S, Sanchez M, Marco E, Gotta GM, Maeder ML, Kennedy EM, et al. 2015. Characterization of Staphylococcus aureus Cas9: a smaller Cas9 for all-in-one adeno-associated virus delivery and paired nickase applications. *Genome Biol* **16**: 257.

Fu BXH, St Onge RP, Fire AZ, Smith JD. 2016. Distinct patterns of Cas9 mismatch tolerance in vitro and in vivo. *Nucleic Acids Research* **44**: 5365–5377.

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

Garneau JE, Dupuis M-È, Villion M, Romero DA, Barrangou R, Boyaval P, Fremaux C, Horvath P, Magadán AH, Moineau S. 2010. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67–71.

Gasiunas G, Barrangou R, Horvath P, Siksnys V. 2012. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* **109**: E2579–86.

Gilbert LA, Horlbeck MA, Adamson B, Villalta JE, Chen Y, Whitehead EH, Guimaraes C, Panning B, Ploegh HL, Bassik MC, et al. 2014. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell* **159**: 647–661.

Gorski SA, Vogel J, Doudna JA. 2017. RNA-based recognition and targeting: sowing the seeds of specificity. *Nat Rev Mol Cell Biol* **18**: 215–228.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**: 307–321.

Hirano H, Gootenberg JS, Horii T, Abudayyeh OO, Kimura M, Hsu PD, Nakane T, Ishitani R, Hatada I, Zhang F, et al. 2016. Structure and Engineering of Francisella novicida Cas9. - PubMed - NCBI. *Cell* **164**: 950–961. https://www.ncbi.nlm.nih.gov/pubmed/26875867.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31**: 827–832.

Jiang F, Zhou K, Ma L, Gressel S, Doudna JA. 2015. STRUCTURAL BIOLOGY. A Cas9-guide RNA complex preorganized for target DNA recognition. - PubMed - NCBI. *Science* **348**: 1477–1481. https://www.ncbi.nlm.nih.gov/pubmed/26113724?dopt=Abstract.

Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* **31**: 233–239.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. - PubMed - NCBI. *Science* **337**: 816–821. https://www.ncbi.nlm.nih.gov/pubmed/22745249?dopt=Abstract.

Kerpedjiev P, Hammer S, Hofacker IL. 2015. Forna (force-directed RNA): Simple and effective online RNA secondary structure diagrams. *Bioinformatics* **31**: 3377–3379.

Kim E, Koo T, Park SW, Kim D, Kim K, Cho H-Y, Song DW, Lee KJ, Jung MH, Kim S, et al. 2017. In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat Commun* **8**: 14500.

Kleinstiver BP, Prew MS, Tsai SQ, Nguyen NT, Topkar VV, Zheng Z, Joung JK. 2015. Broadening the targeting range of Staphylococcus aureus CRISPR-Cas9 by modifying PAM recognition. *Nat Biotechnol* **33**: 1293–1298.

Kou R, Lam H, Duan H, Ye L, Jongkam N, Chen W, Zhang S, Li S. 2016. Benefits and Challenges with Applying Unique Molecular Identifiers in Next Generation Sequencing to Detect Low Frequency Mutations. ed. J. Wang. *PLoS ONE* **11**: e0146638.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359.

Liu Y, Chen Z, He A, Zhan Y, Li J, Liu L, Wu H, Zhuang C, Lin J, Zhang Q, et al. 2016. Targeting cellular mRNAs translation by CRISPR-Cas9. *Sci Rep* **6**: srep29652. https://www.nature.com/articles/srep29652.

Loughrey D, Watters KE, Settle AH, Lucks JB. 2014. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Research* **42**: e165–e165.

Ma E, Harrington LB, O'Connell MR, Zhou K, Doudna JA. 2015. Single-Stranded DNA Cleavage by Divergent CRISPR-Cas9 Enzymes. *Molecular Cell* **60**: 398–407.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnetjournal* **17**: 10.

Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. 2012. PANDAseq: paired-end assembler for illumina sequences. *BMC Bioinformatics* **13**: 31.

McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**: 4288–4297.

Mohanraju P, Makarova KS, Zetsche B, Zhang F, Koonin EV, van der Oost J. 2016. Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**: aad5147.

Mout R, Ray M, Yesilbag Tonga G, Lee Y-W, Tay T, Sasaki K, Rotello VM. 2017. Direct Cytosolic Delivery of CRISPR/Cas9-Ribonucleoprotein for Efficient Gene Editing. *ACS Nano* **11**: 2452–2458.

Nelles DA, Fang MY, Aigner S, Yeo GW. 2015. Applications of Cas9 as an RNA-programmed RNA-binding protein. *Bioessays* **37**: 732–739.

Nelles DA, Fang MY, O'Connell MR, Xu JL, Markmiller SJ, Doudna JA, Yeo GW. 2016. Programmable RNA Tracking in Live Cells with CRISPR/Cas9. *Cell* **165**: 488–496. http://www.cell.com/cell/abstract/S0092-8674(16)30204-5.

Nishimasu H, Le Cong, Yan WX, Ran FA, Zetsche B, Li Y, Kurabayashi A, Ishitani R, Zhang F, Nureki O. 2015. Crystal Structure of Staphylococcus aureus Cas9. *Cell* **162**: 1113–1126. http://www.cell.com/abstract/S0092-8674(15)01020-X.

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O. 2014. Crystal structure of Cas9 in complex with guide RNA and target DNA. - PubMed - NCBI. *Cell* **156**: 935–949. https://www.ncbi.nlm.nih.gov/pubmed/24529477?dopt=Abstract.

Oakes BL, Nadler DC, Flamholz A, Fellmann C, Staahl BT, Doudna JA, Savage DF. 2016. Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. *Nat Biotechnol* **34**: 646–651.

O'Connell MR, Oakes BL, Sternberg SH, East-Seletsky A, Kaplan M, Doudna JA. 2014. Programmable RNA recognition and cleavage by CRISPR/Cas9. *Nature* **516**: 263–266. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4268322/.

Pei J, Kim B-H, Grishin NV. 2008. PROMALS3D: a tool for multiple protein sequence and structure alignments. *Nucleic Acids Research* **36**: 2295–2300.

Phanstiel DH, Boyle AP, Araya CL, Snyder MP. 2014. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics* **30**: 2808–2810.

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA. 2013. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**: 1173–1183.

Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, et al. 2015. In vivo genome editing using Staphylococcus aureus Cas9. - PubMed - NCBI. *Nature* **520**: 186–191. https://www.ncbi.nlm.nih.gov/pubmed/25830891.

Rehmsmeier M, Steffen P, Hochsmann M, Giegerich R. 2004. Fast and effective prediction of microRNA/target duplexes. *RNA* **10**: 1507–1517.

Rio DC. 2012. Filter-binding assay for analysis of RNA-protein interactions. *Cold Spring Harb Protoc* **2012**: 1078–1081.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.

Sampson TR, Saroj SD, Llewellyn AC, Tzeng Y-L, Weiss DS. 2013. A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. *Nature* **497**: 254–257.

Shmakov S, Abudayyeh OO, Makarova KS, Wolf YI, Gootenberg JS, Semenova E, Minakhin L, Joung J, Konermann S, Severinov K, et al. 2015. Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems. *Molecular Cell* **60**: 385–397.

Smargon AA, Cox DBT, Pyzocha NK, Zheng K, Slaymaker IM, Gootenberg JS, Abudayyeh OA, Essletzbichler P, Shmakov S, Makarova KS, et al. 2017. Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. *Molecular Cell* **65**: 618–630.e7.

Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res* **27**: 491–499.

Sternberg SH, Haurwitz RE, Doudna JA. 2012. Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. *RNA* **18**: 661–672.

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

Szczelkun MD, Tikhomirova MS, Sinkunas T, Gasiunas G, Karvelis T, Pschera P, Siksnys V, Seidel R. 2014. Direct observation of R-loop formation by single RNA-guided Cas9 and Cascade effector complexes. *Proc Natl Acad Sci USA* **111**: 9798–9803.

Wang AH, Fujii S, van Boom JH, van der Marel GA, van Boeckel SA, Rich A. 1982. Molecular structure of r(GCG)d(TATACGC): a DNA--RNA hybrid helix joined to double helical DNA. *Nature* **299**: 601–604.

Wright AV, Nuñez JK, Doudna JA. 2016. Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell* **164**: 29–44.

Yamada M, Watanabe Y, Gootenberg JS, Hirano H, Ran FA, Nakane T, Ishitani R, Zhang F, Nishimasu H, Nureki O. 2017. Crystal Structure of the Minimal Cas9 from Campylobacter jejuni Reveals the Molecular Diversity in the CRISPR-Cas9 Systems. *Molecular Cell* **65**: 1109–1121.e3. https://www.ncbi.nlm.nih.gov/pubmed/28306506.

Zhang Y, Rajan R, Seifert HS, Mondragón A, Sontheimer EJ. 2015. DNase H Activity of Neisseria meningitidis Cas9. *Molecular Cell* **60**: 242–255.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research* **31**: 3406–3415.

Zuris JA, Thompson DB, Shu Y, Guilinger JP, Bessen JL, Hu JH, Maeder ML, Joung JK, Chen Z-Y, Liu DR. 2015. Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* **33**: 73–80.
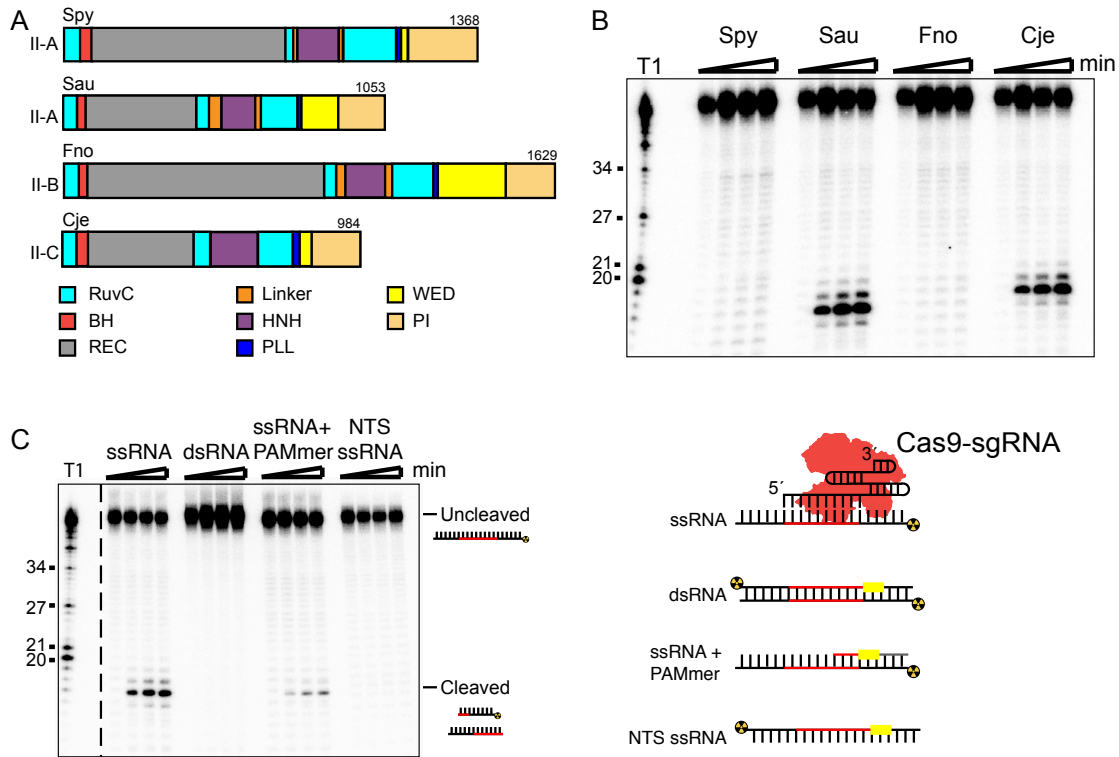
**Figures**



**Figure 1: SauCas9 cleaves single-stranded RNA without a PAMmer.**

(A) Schematic of Cas9 proteins tested for sgRNA mediated RNA cleavage. RuvC, RuvC nuclease domain; BH, bridge-helix; REC, recognition domain; HNH, HNH nuclease domain; PLL, phosphate-lock loop; WED, wedge domain; PI, PAM-interacting domain. Adapted from (Nishimasu et al. 2014; 2015; Hirano et al. 2016; Yamada et al. 2017). (B) Representative *in vitro* cleavage of ssRNA by Cas9-sgRNA RNP complexes of homologs in (A). Radiolabeled pUC target RNA was incubated with Cas9 RNP at 37˚C and time points were taken at 0, 10, 30, and 60 min. Full time course is presented in Figure 1-figure supplement 1B. T1 indicates size markers generated by RNase T1 digestion of ssRNA target. Size in nucleotides is indicated on the left. (C) (Right) *In vitro* cleavage assay of various RNA substrates (Left). Full time course is presented in Figure 1-figure supplement 3A.
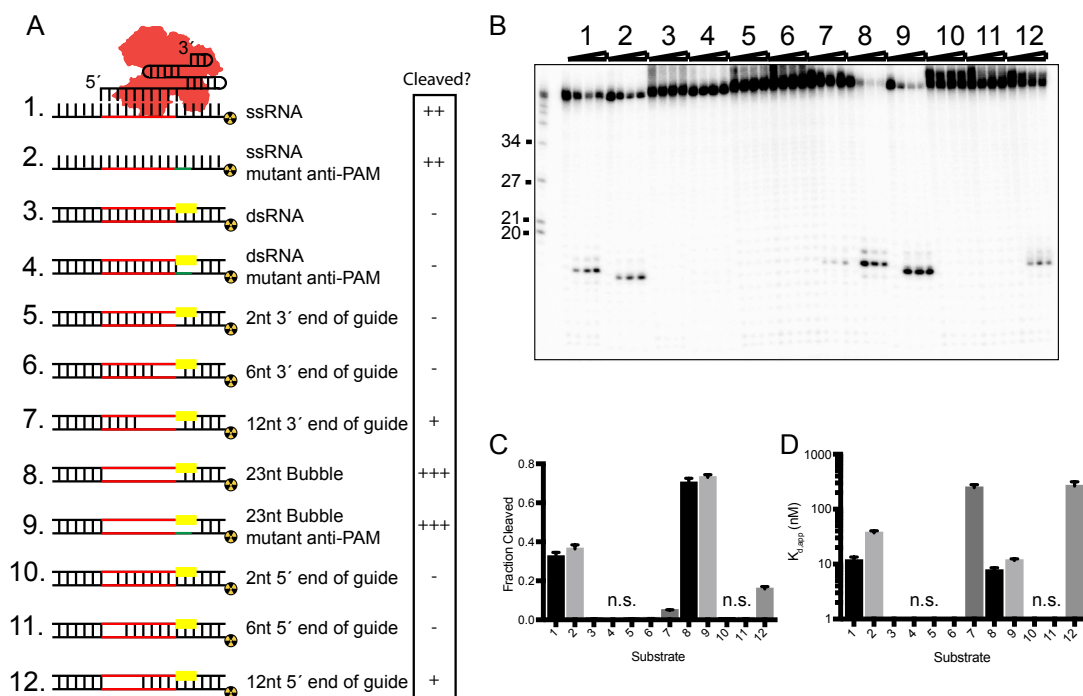
**Figure 2: *In vitro* RNA cleavage is impaired by strong secondary structure.**

(A) Schematic representation of structured RNA targets for *in vitro* cleavage assays. Symbols on right indicate relative level of cleavage activity for each substrate: "-" , no cleavage; "+", low cleavage; "++", medium cleavage; "+++" high cleavage. (B) Representative cleavage assay of partially-duplexed RNA targets diagrammed in (A). T1 indicates size markers generated by RNase T1 digestion of ssRNA target. Size in nucleotides is indicated on the left. (C and D) Fraction of target cleaved (C) and $K_{d,app}$ (D) for substrates diagrammed in (A). Fits were determined in Prism using a single-exponential decay and a one-site binding model, respectively. Bars represent mean ± S.D. (n=3). N.s. denotes no significant cleavage or binding.
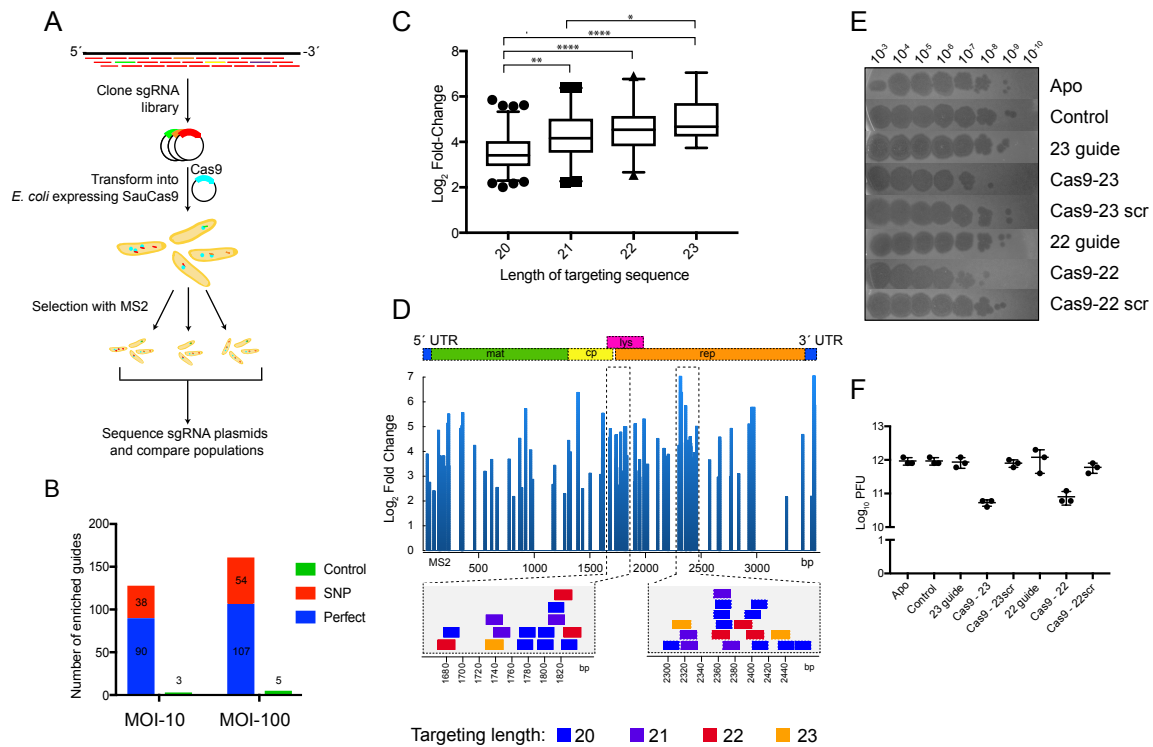
**Figure 3: SauCas9 confers *in vivo* protection against an RNA phage.**

(A) Overview of MS2 targeting screen. Guides tiled against the library were cloned into sgRNA expression plasmids and co-transformed into *E. coli* with a plasmid containing wild-type SauCas9 under inducible control. Plasmids from surviving colonies after MS2 selection were recovered and sequenced. For more detail, see Methods. (B) Number of guides with significant positive enrichment from three biological experiments. SNP, guides with single-nucleotide mismatch. (C) Box and whiskers plot of average $\log_2$ fold-change of perfect MS2 guides by length. Whiskers represent 5% and 95% values with outliers graphed as points. *$p < 0.05$, **$p < 0.01$, ****$p < 0.0001$, by one-way ANOVA. (D) (Upper) $\log_2$ fold-change of guides with an FDR-corrected p-value < 0.05 mapped to the MS2 genome for MOI-100 treatment. Schematic of MS2 genome is provided above. (Lower) Individual guides mapped to highlighted regions of MS2 genome. Other graphs for MOI-10 and -100 treatments are presented in Figure 3-figure supplement 1. (E) Representative plaque assay of SauCas9 *in vivo* protection. *E. coli* containing constructs on the right are spotted with various phage dilutions as indicated. Scr signifies that the targeting portion of the guide has been scrambled to serve as a non-targeting control. (F) Relative plaque forming units (PFU) (mean ± S.D., n=3) from results in (E). More guides and controls are presented in Figure 3 – figure supplement 3.
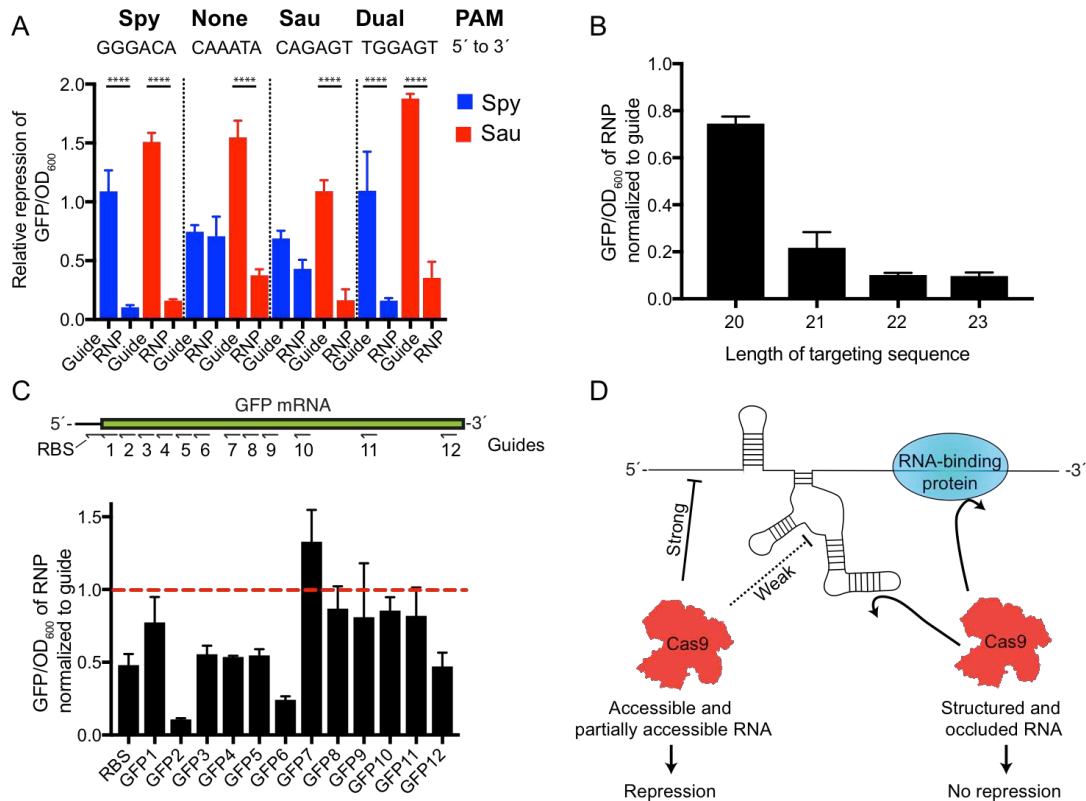
27

**Figure 4: SauCas9 repression of a GFP reporter *in vivo*.**

(A) Comparison of dSpy and dSauCas9 to repress GFP expression on the DNA and RNA level. GFP signal is normalized to $OD_{600}$ to control for difference in cell density between samples. $GFP/OD_{600}$ ratios for guide alone and RNP are normalized to values for a non-targeting guide vector and an Apo protein control, respectively. Target sites were chosen to be adjacent to PAM sites for Spy, Sau, both, or neither as indicated. Note: the slight GFP repression observed with dSpyCas9 using the target sequence adjacent to the Sau PAM (CAGAGT) likely results from the ability of SpyCas9 to use an NAG PAM, albeit with reduced efficiency (Hsu et al. 2013). ****$p < 0.0001$ by one-way ANOVA. (B) Relative expression of GFP using guides with different length targeting sequences. Target site here is the GFP2 sequence chosen for its robust targeting activity. (C) (Upper) Diagram of targeting sequences across the GFP mRNA and ribosome binding site (RBS). (Lower) Relative expression of GFP of SauCas9 RNP normalized to sgRNA alone for targeting sequences across the GFP reporter. Dashed red line indicates that the sgRNA alone is as efficient as the RNP for GFP repression. (A-C) Bars represent mean ± S.D. (n=3). (D) Model for observed SauCas9 ssRNA targeting activity. We propose that accessible RNA is cleaved or repressed efficiently while structured and protein-bound RNA is not targeted by SauCas9.
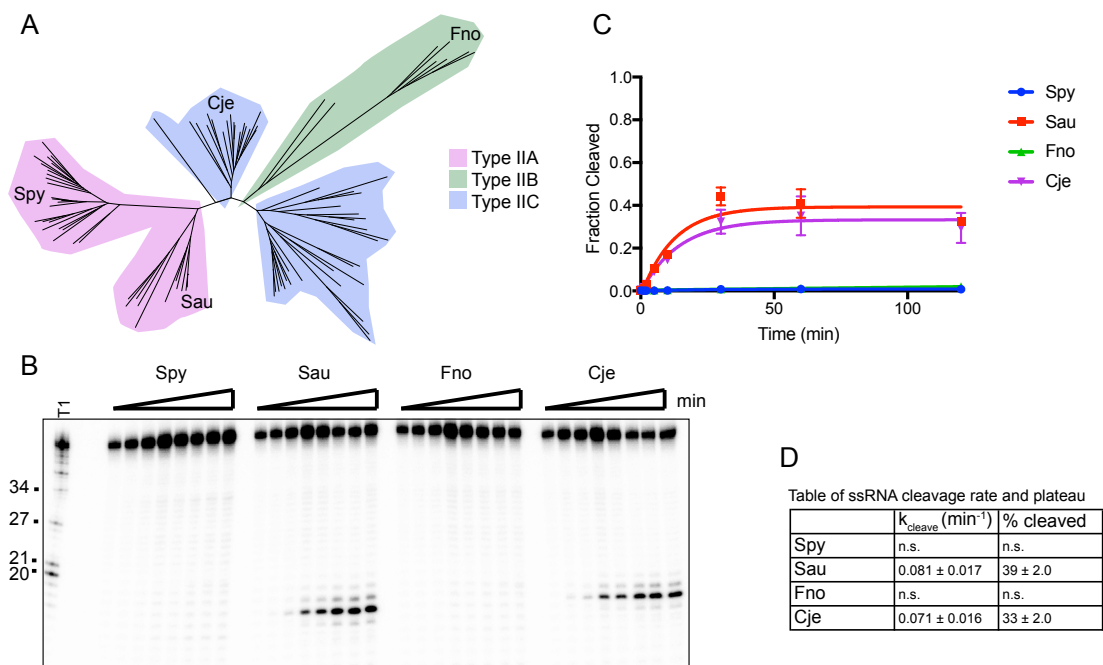
**Figure Supplements**

28

**Figure 1 – Figure Supplement 1: RNA is cleaved by SauCas9 and CjeCas9**

(A) Phylogenetic tree of Cas9 homologs assayed for ssRNA cleavage activity. Tree was generated using homologs gathered from (Chylinski et al. 2014). Only homologs tested for activity are highlighted as leaves on the tree. Clades are colored by Cas9 sub-type. (B) Representative *in vitro* cleavage gel for ssRNA targeting by various Cas9 homologs in (A). Target used for cleavage was the pUC ssRNA. Time points are 0, 1, 2, 5, 10, 30, 60, and 120 mins. T1 RNase digest size fragments are given on the left. (C) Quantification of fraction cleaved in (B). Fit was determined in Prism using a single-exponential decay model. Error bars represent the mean ± S.D. (n=3). (D) Apparent pseudo-first order fit parameters of the data in (C) where '% cleaved' indicates the fraction of substrate cleaved when the reaction plateaus (mean ± S.D.).
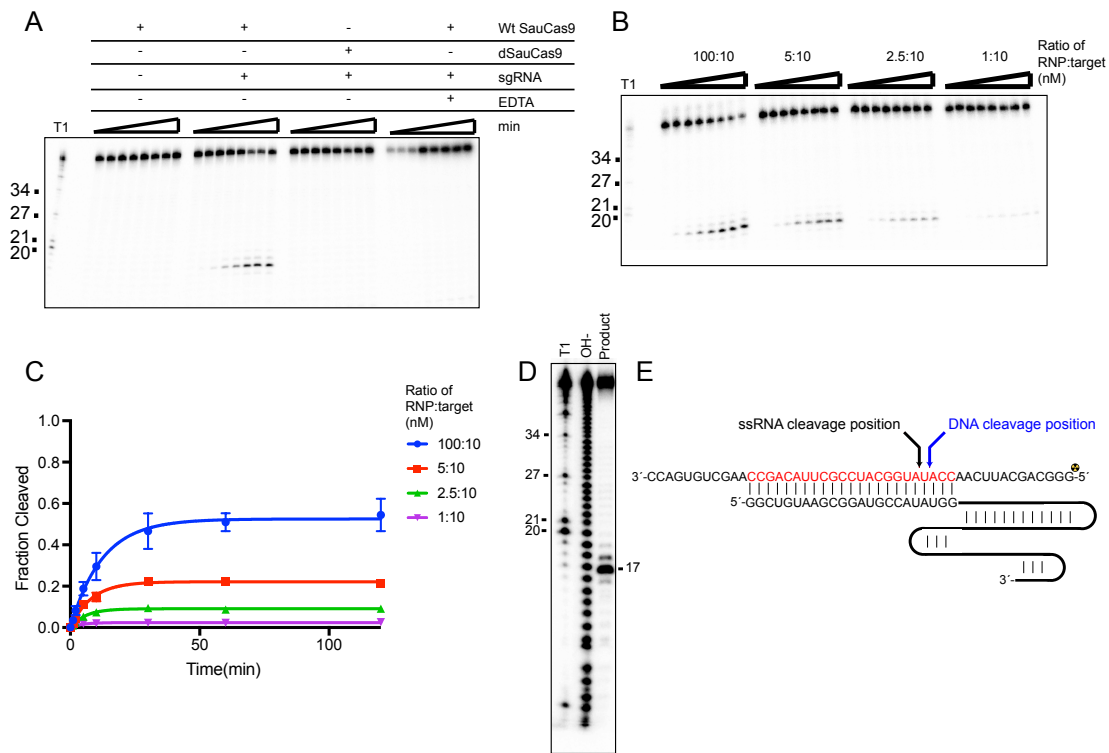
**Figure 1 – Figure Supplement 2: ssRNA cleavage is similar to canonical dsDNA cleavage by Cas9**

(A) *In vitro* SauCas9 cleavage assay of ssRNA. Reactions were incubated with wild-type (Wt SauCas9) or catalytically-inactive dSauCas9 (D10A and N580A) in the presence or absence of sgRNA as indicated above the reactions. EDTA was included at 25mM where applicable. (B) SauCas9 ssRNA cleavage is single-turnover. SauCas9 RNP was incubated with the RNA target in the various ratios indicated. (A and B) Time points are 0, 1, 2, 5, 10, 30, 60, and 120 mins. T1 RNase digest size fragments are given on the left. Target used for cleavage was the pUC ssRNA. (C) Graphical representation of ssRNA fraction cleaved of reactions in (B). Fit was determined in Prism using a single-exponential decay model. Error bars represent the mean ± S.D. (n=3). (D) Mapping of SauCas9 ssRNA cleavage site. Reaction products from a 2-hr incubation of SauCas9 RNP with the pUC ssRNA target were separated on a 15% denaturing PAGE gel with a hydrolysis and T1 digest ladder to determine exact site of the major cleavage product. (E) Diagram of canonical DNA cleavage position and ssRNA cleavage position as determined in (D).
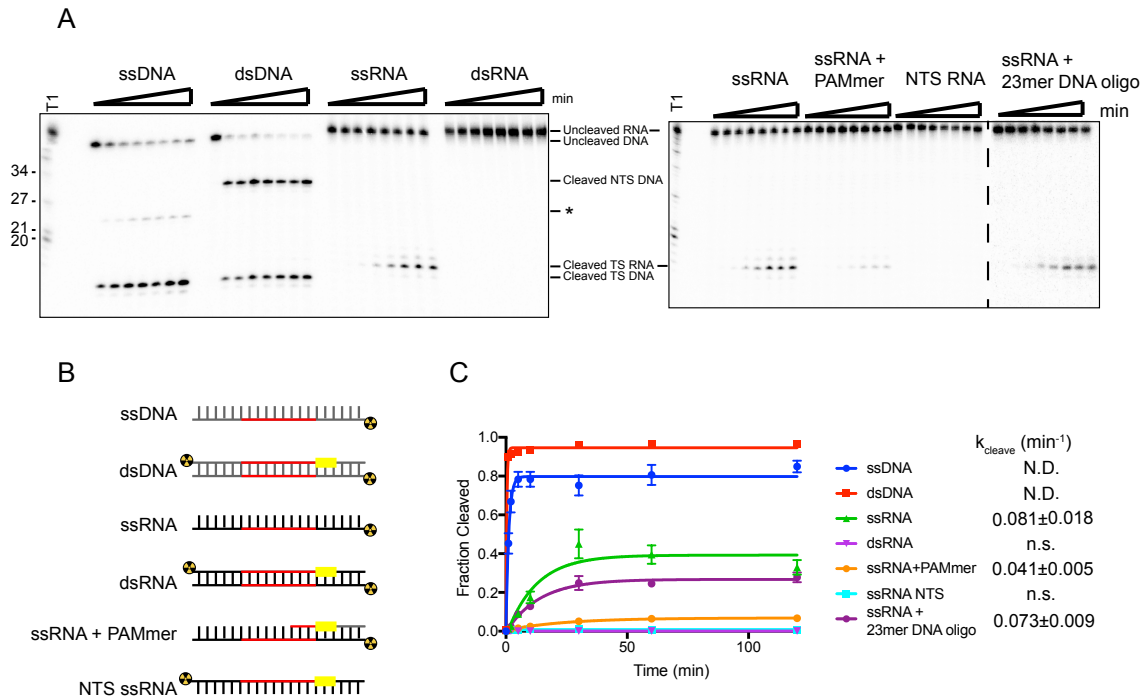
**Figure 1 – Figure Supplement 3: SauCas9 cleavage of different nucleic acid substrates.**

(A) Representative cleavage assay of nucleic acid substrates diagramed in (B) by SauCas9. Asterisk denotes an off-target cleavage site. Time points are 0, 1, 2, 5, 10, 30, 60, and 120 mins. T1 RNase digest size fragments are given on the left. (C) Quantification of results in (A). Fit was determined in Prism using a single-exponential decay model. Error bars represent the mean ± S.D. (n=3). Apparent pseudo-first order rate constant ($k_{cleave}$ ± S.D.) is given to the right of the substrate legend. N.D. indicates that an accurate rate cannot be determined due to the reaction reaching completion before the second time point. N.s., not significant.
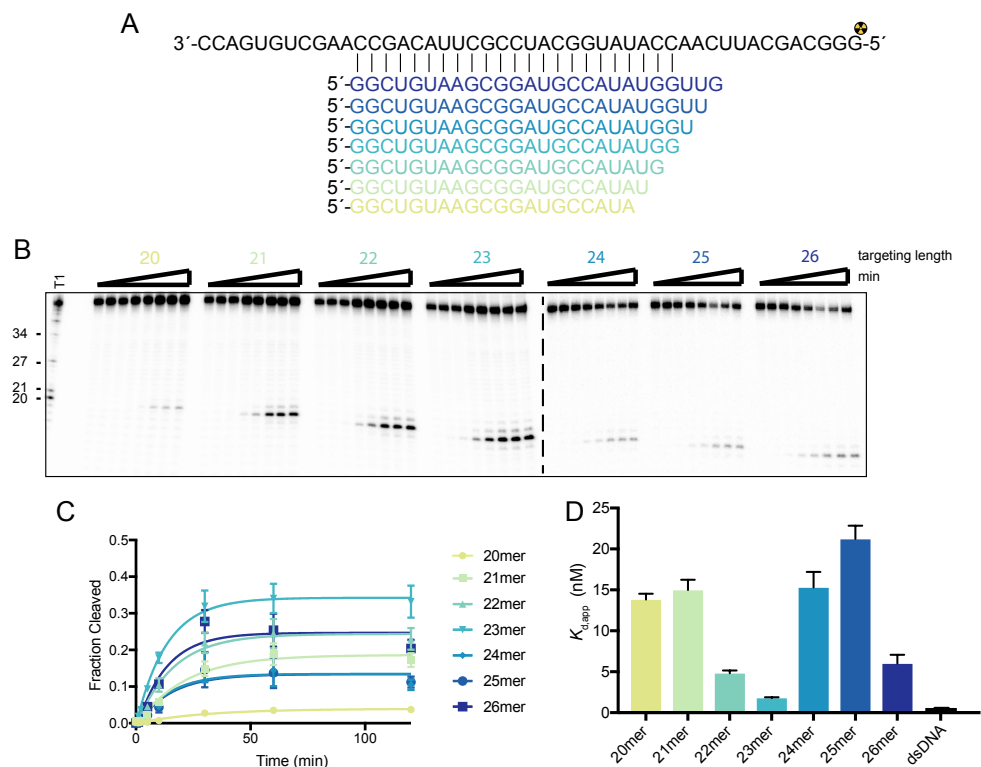
**Figure 1 – Figure Supplement 4: SauCas9 prefers a complementary region of 23nt for binding and cleavage.**

(A) Diagram of pUC ssRNA target and regions of complementary for the different length sgRNAs. (B) Representative *in vitro* cleavage assays using sgRNAs with a complementary region to the target of the indicated lengths. Time points are 0, 1, 2, 5, 10, 30, 60, and 120 mins. T1 RNase digest size fragments are given on the left. (C) Quantification of cleavage products from reactions in (B). Length of targeting region of the sgRNA given as *n*-mer. Fit was determined in Prism using a single-exponential decay model. Error bars represent the mean ± S.D. (n=3). (D) Filter binding data for dSauCas9 and the structured RNA substrates were fit in Prism using a one-site binding model and the apparent dissociation constant ($K_{d,app}$) was determined. Bars represent the mean ± S.D. (n=3).
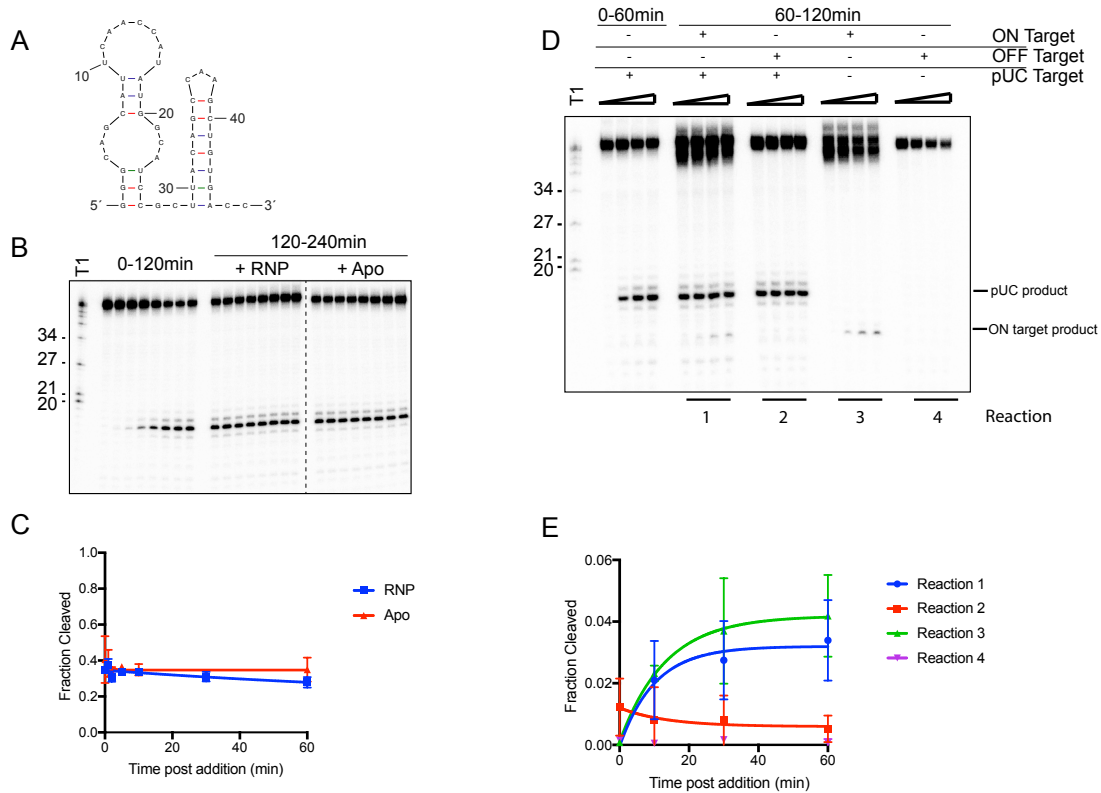
**Figure 2 – Figure Supplement 1: RNA cleavage is limited by the RNA target.**

(A) Predicted secondary structure of target RNA used in this study. (B) *In vitro* cleavage assay of ssRNA with SauCas9 was conducted for 2 hr (time points: 0, 1, 2, 5, 10, 30, 60, 120 mins). The reaction was split and SauCas9-sgRNA RNP or apo SauCas9 were added. The reaction was further incubated at 37°C and additional time points at were taken to check for additional cleavage of the target. Time points were taken at 0, 1, 2, 5, 10, 30, 60, and 120 mins post-RNP/apo SauCas9 addition. (C) Fit for data in (B) was determined in Prism using a single-exponential decay model. Error bars represent the mean ± S.D. (n=3). (D) *In vitro* cleavage assay of two ssRNA targets added sequentially. After 60 min incubation of SauCas9 with the pUC target, another target containing either the same recognition sequence (ON target – reaction 1) or an unrelated sequence (OFF target – reaction 2) were added to the reaction. Cleavage was assayed for an additional 60 mins (time points: 0, 10, 30, 60 min). Reactions containing only the second target (Reactions 3 and 4) were conducted with SauCas9 RNP that was incubated for 60min at 37°C prior to addition to the cleavage reaction. (E) Quantification of cleavage of second target in (D) for time points after addition. Fit was determined in Prism using a single-exponential decay model. Error bars represent the mean ± S.D. (n=3).
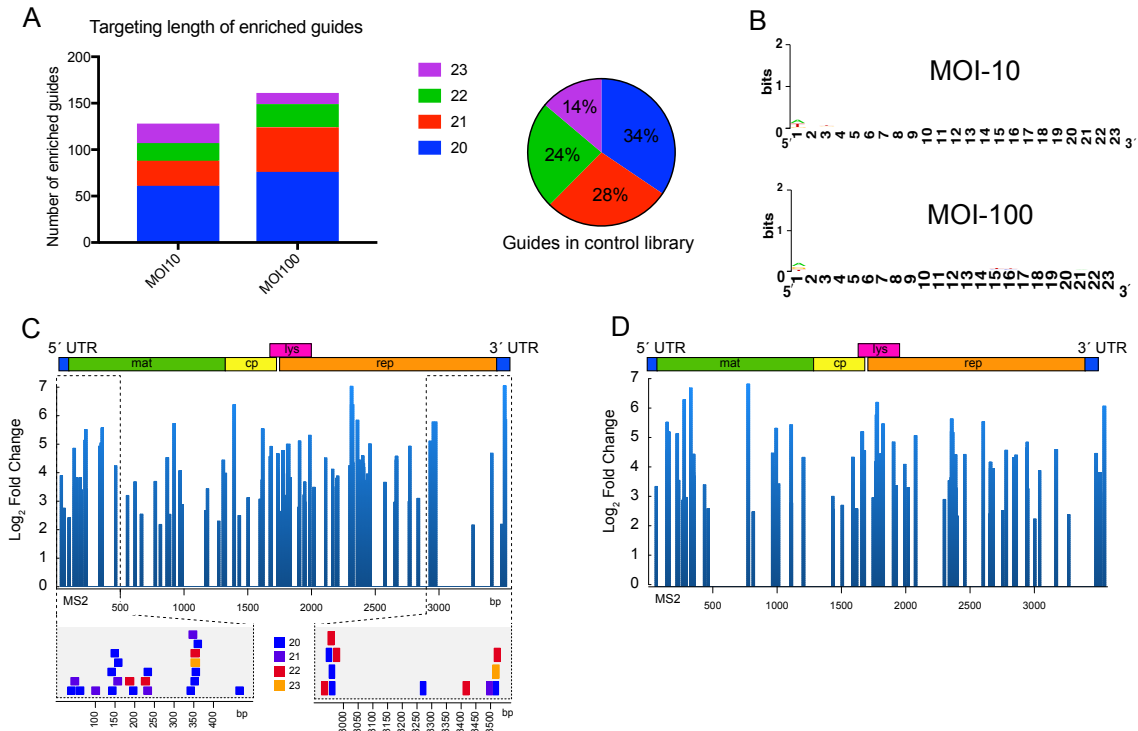
**Figure 3 – Figure Supplement 1: Enriched guides do not display sequence bias and cluster to regions on the MS2 genome.**

(A) (Left) Stacked bar graph of positively enriched guides (FDR-adjusted p-value < 0.05) for perfectly complementary and single-nucleotide mismatch (SNP) guides for a multiplicities of infection (MOI) of 10 and 100. (Right) Percentages of perfect and SNP guides by length averaged across the control libraries (n=3). (B) WebLogo (Crooks et al. 2004) representation of positively enriched guides (perfect complementarity, FDR-adjusted p-value <0.05) for MOI-10 (n=84) and MOI-100 (n=107). Different length guides were aligned at their 3´ end, which contains the pre-ordered 'seed' region (Jiang et al. 2015). (C) (Upper) Log$_2$ fold-change of positively enriched guides (FDR-adjusted p-value < 0.05) mapped to the MS2 genome for MOI-100 treatment. Schematic of MS2 genome is provided above. (Lower) Individual guides mapped to highlighted regions of MS2 genome. (D) Log$_2$ fold-change of guides with an FDR-adjusted p-value < 0.05 mapped to the MS2 genome for MOI-10 treatment.
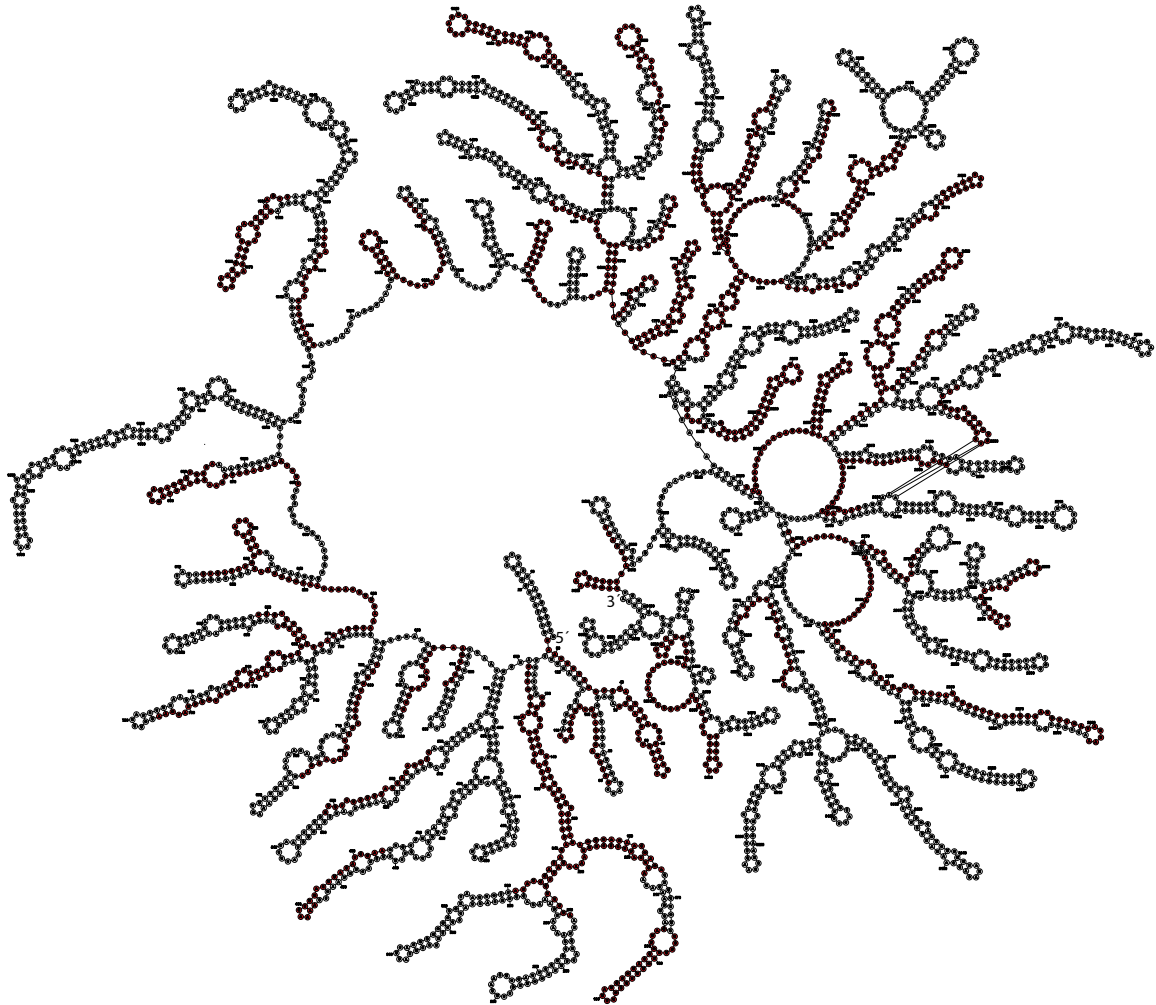
**Figure 3 – Figure Supplement 2: Enriched MS2 targeting guides mapped to MS2 genome structure.**

Structure of the MS2 genome inside the viral particle was obtained from a recently published EM structure (Dai et al. 2017) and guides (red) significantly enriched in the MOI-100 treatment (FDR-adjusted p-value < 0.05) were mapped to the MS2 genome and subsequently visualized in Forna (Kerpedjiev et al. 2015).
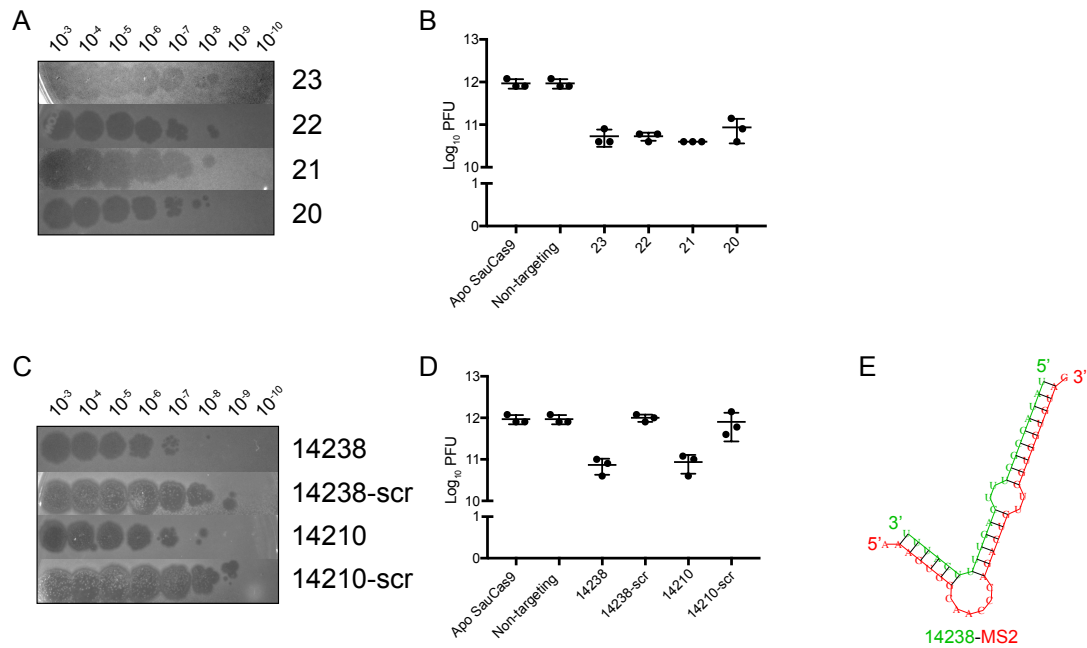
**Figure 3 – Figure Supplement 3: Confirmation that enriched guides from the MS2 screen confer protection against MS2 infection.**

(A) Representative plaque assay for lawns of *E. coli* expressing wtSauCas9 and sgRNA of different length spotted with phage dilutions indicated. Here, the sgRNA with the highest fold-change in both MOI-10 and -100 samples was chosen for each length. The 23-mer sgRNA produces hazy plaques for an unknown reason. All other guides tested, including a different 23-mer sgRNA, produced clear plaques. (B) Quantification of relative plaque forming units (PFU, mean ± S.D., n=3) from data in (A). (C) Same as in (A). Guides were two 'control' guides that were significantly enriched in both MOI-10 and -100 treatments during phage selection. Scrambled (scr) indicates random shuffling of the target sequence to serve as a non-targeting control. Scrambled sequences were verified against the MS2 genome and its reverse-complement to ensure no partial matches. (D) Quantification of relative PFU as in (C). Guides 14238 and 14210 confer ~10-fold protection over their scrambled counterparts. The level of protection is similar to perfectly complementary guides (B, and see Figure 3E and 3F). (E) Predicted binding of guide 14238 (green) to a fragment of the MS2 genome (red, nts: 1533-1563) using RNAhybrid (Rehmsmeier et al. 2004).
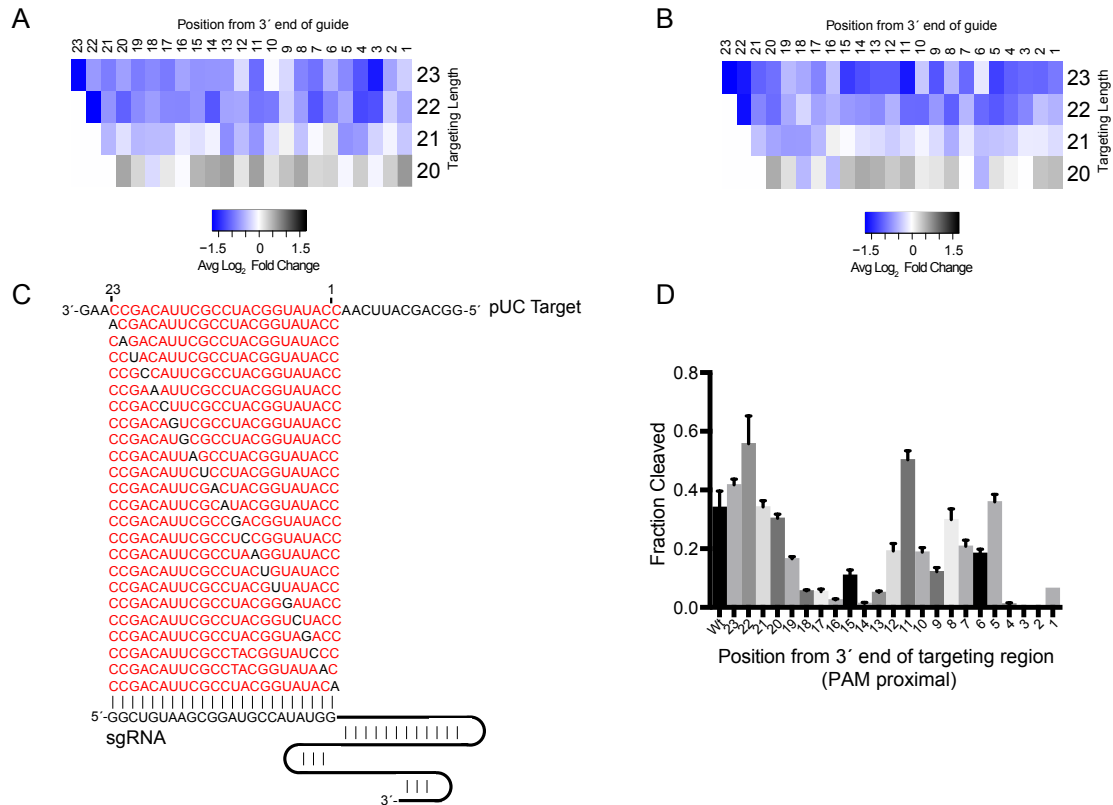
**Figure 3 – Figure Supplement 4: Effect of single-nucleotide mismatches on ssRNA targeting**

(A-B) Heatmap of average $\log_2$ fold-change for all single-nucleotide mismatch (SNP) guides in MOI-10 (A) and MOI-100 (B) treatment. Deeper blue represents greater negative selection of guides indicating greater sensitivity to mismatches at that position. While deeper black represents greater positive selection indicating that mismatches at that position are more tolerated. Positions are given as distance from 3′ end of the targeting region of the sgRNA. (C) Diagram of target ssRNAs with SNPs for *in vitro* cleavage assays. Red highlights the region complementary to the guide while black nucleotides indicate the mismatched base in the targeting region. Numbering of nucleotides is labeled from 1 to 23 to reflect positions in (A) and (B). (D) Quantification of *in vitro* cleavage assays with mismatched targets in (C). Bars represent the mean ± S.D. (n=3). 'Wt' indicates 23nt of perfect complementarity between the sgRNA and the target.
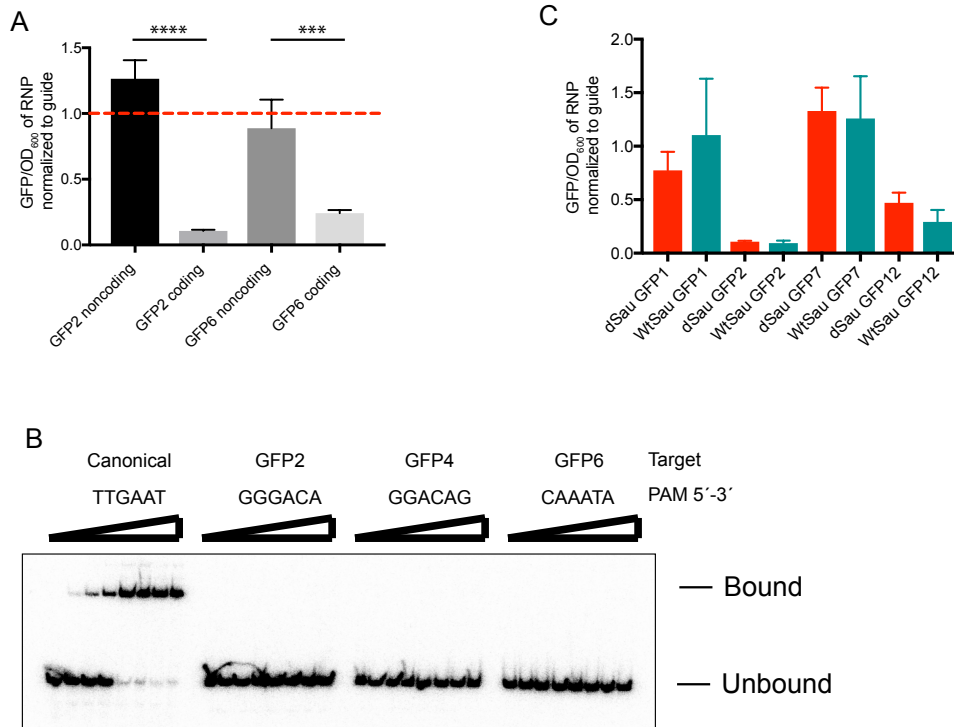
**Figure 4 – Figure Supplement 1: Repression of GFP mRNA**

(A) dSauCas9-sgRNA directed against a GFP mRNA (coding) or antisense (noncoding) sequence. GFP2 and GFP6 refer to guides diagrammed in Figure 4C. Dashed red line indicates that the sgRNA alone is as efficient as the SauCas9 RNP for GFP repression. Bars represent mean ± S.D. (n=3). ***$p < 0.001$, ****$p < 0.0001$, by one-way ANOVA. (B) Electrophoretic mobility shift assay (EMSA) confirming that dSauCas9 does not bind dsDNA adjacent to non-canonical PAMs. Targeting sequence is identical for all substrates but with varied PAM sequences as indicated for the guides in Figure 4C. Final concentrations of dSauCas9 from left to right: 0, 0.1, 0.5, 1, 5, 10, 50, 100nM. (C) Comparison of ability of dSau and wtSauCas9 to repress GFP expression *in vivo*. Bars represent mean ± S.D. (n=3).

**Supplemental Files**

Supplemental File 1: List of sequences used in this study

**Source Data Files**

Figure 2 Source Data 1: Quantification of cleavage and binding of structured RNA substrates by SauCas9

Figure 3 Source Data 1: List of guides, normalized read counts, enriched guides, fold change distribution and plaque enumeration

Figure 4 Source Data 1: Raw data for PAM dependency, length efficiency, and GFP mRNA tiling for GFP repression assays

Figure 1-figure supplement 1 Source Data 1: ssRNA cleavage time course for Cas9 homologs

Figure 1-figure supplement 2 Source Data 1: Quantification of multi-turnover cleavage

Figure 1-figure supplement 3 Source Data 1: Quantification of SauCas9 cleavage of nucleic acid substrates

Figure 1-figure supplement 4 Source Data 1: Cleavage and binding data for different length guides

Figure 2-figure supplement 1 Source Data 1: Quantification of ssRNA after additional protein and target spike-in

Figure 3-figure supplement 1 Source Data 1: Enriched guide length distribution, sequences, and targeting location on MS2 genome

Figure 3-figure supplement 2 Source Data 1: Location of enriched guides from MOI-100 condition mapped to MS2 genome

Figure 3-figure supplement 3 Source Data 1: Plaque enumeration for SauCas9-mediated MS2 protection

Figure 3-figure supplement 4 Source Data 1: Heatmaps of single-nucleotide mismatches from MS2 screen and *in vitro* mismatch cleavage

Figure 4-figure supplement 1 Source Data 1 s: Raw data for GFP repression assays