Title

Engineering CRISPR-Cas9 Systems to Expand Functionality

Author

Oakes, Benjamin L.

Publication Date

2017

Peer reviewed|Thesis/dissertation

Engineering CRISPR-Cas9 Systems to Expand Functionality

by
Benjamin L Oakes

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Molecular and Cell Biology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David F. Savage, Co-Chair
Professor Jennifer A. Doudna, Co-Chair
Professor John E. Dueber
Professor Jeremy W. Thorner

Spring 2017

Abstract

Engineering CRISPR-Cas9 Systems to Expand-Functionality

by

Benjamin L Oakes

Doctor of Philosophy in Molecular and Cell Biology

University of California, Berkeley

Professors David F. Savage and Jennifer A. Doudna, Co-Chairs

CRISPR-Cas9 is a RNA-protein complex from adaptive bacterial immune systems that has been repurposed to enable convenient and specific alteration of any genome via the introduction of a double-strand DNA break. By its complementarity to a corresponding DNA sequence, the guide RNA directs the Cas9 protein, a DNA endonuclease, to a target site where it binds to and cleaves the DNA. In this dissertation, the Cas9 ortholog of *Streptococcus pyogenes* was studied. First, detailed structural insights and knowledge about the biochemical requirements of Cas9-mediated cleavage were exploited to generate a method to efficiently bind a single-stranded RNA target rather than double-stranded DNA. Thus, Cas9 can be used as an RNA-guided RNA-binding protein. Second, methods were devised and implemented to improve and expand the function of Cas9 as an RNA-guided DNA endonuclease, especially means to control its activity spatially and temporally. Systematic unbiased and comprehensive exploration of sites within Cas9 that can tolerate insertions of various protein functionalities, yet retain overall Cas9 activity, had not been undertaken. To engineer more complex and dynamic control over Cas9 action, such as allosteric regulation and location-sensitive molecular scaffolding, I developed a ratiometric analysis protocol that is able to identify one functional Cas9 protein out of tens of millions, permitting the screening of large libraries of engineered Cas9 proteins. These methods were used to successfully modify Cas9 via the insertion of various protein-protein and protein-ligand binding domains that confer new functionalities. Three distinct types of protein domains were inserted into Cas9 with minimal disruption of its core binding and enzymatic functions. In particular, insertions of the ligand-binding domain of a mammalian estrogen receptor generated a Cas9 derivative whose activity is specifically activated upon addition of the synthetic estrogen 4-hydroxy tamoxifen. This particular Cas9 derivative has no detectable background activity in the absence of the hormone analog, thus providing a fine-tuned mechanism to control genome editing and modification with a small molecule. Overall, the methods described in this thesis can be applied to engineer any programmable DNA-binding protein and thereby confer the ability to be activated or deactivated in response to specific molecular signals.

Acknowledgements

It starts with my family, thank you Mom and Dad for always supporting me in every way; even when I veered away from being a doctor and decided that instead I wanted to create invisible contraptions that would modify invisible things. You have made me who I am and I can only hope to one day be able to give back all that you have given me in love, motivation, and opportunity. To my sister who, despite being two years younger, successfully beat me out of graduate school; you are amazing, you always make me laugh, I can't wait to see what you have in store for us all and I can't wait until you move to California! Finally to Danielle, thank you for being my rock and for enriching my life in so many ways. You challenge me to think outside of my own perspective and as a scientist and individual I do not think there could be any greater gift.

To my advisors past and present thank you:

Dave, I vividly remember when I realized that Berkeley was where I was going to attend grad school. We were sitting on the couch at Jennifer's, during recruitment weekend 4 years ago, chatting about using domain insertion for biosensor creation. Until then I thought of protein engineering through a lens of mutation and, that afternoon, you expanded my viewpoint. I cannot think of anything I have enjoyed more over these past 4 years than devising both the daft and the only slightly off kilter molecular engineering ideas with you: ABACas, CAMO, SIREs and MISER just to name a few. I can't wait to see so many more of these IDEAS come to fruition. Thank you so much for allowing me the freedom to chase after these crazy dreams while always finding a way to help me become a better scientist, writer and presenter.

Jennifer, when I first read the 2012 Cas9 paper I was floored. There I was, working in the Noyes lab, performing tens of thousands of selections on engineered zinc fingers to bind any given target and you had just solved the whole problem with a single RNA (RNA lab… go figure). Working with you since then has been an enlightening and gratifying experience that has changed the way I think about science, society and the necessary interaction between the two. You have taught me how to compose the best possible scientific story while also thinking about the consequences of our actions on the greater world. If I am able to make a lasting impact it will be in large part because you have showed me the way to do so; by mixing grace and humility in equal measure with strength and determination.

Marcus, I would be remiss if did not acknowledge how strongly you have guided my scientific and personal journey. When we first began to work together, suffice it to say, I was a bit green. It was only working side by side that I learned many of the lessons that I carry with me every day, such as: if it sounds like more work, it's probably the right thing to do and when you try to do everything, you do nothing. These and others; prepare for failure, you get only what you select for, and always run an no-insert control have fundamentally guided my approach to science at a much deeper level than one, at first glance, might think such 'scientific catechisms' could. Thank you for taking the time and energy to impart such sage advice while also laying out the intricacies of cloning, protein engineering, and the scientific technique I would not be where I am right now without your guidance.

To the many other professors, instructors and mentors who I have had the pleasure of working with over the years, thank you all for imparting you knowledge and instilling your love of learning. To Debra Laskin thank you for introducing me to laboratory science and allowing me to work with you at Rutgers during high school. To Andrew Gow and Alba Rossi-George thank you for being the first to familiarize me to the western blot and the rigors required for scientific publishing. To Alison Forgie thank you for helping me to learn about other side of research science, biotechnology. To Frank Fekete, and Megan McClean thank you for letting me spearhead my own research under your guidance. To John Dueber, thank you for acting as a sounding board and for all your help with many specific ideas on research projects and my thesis as a whole. To Jeremy Thorner, while your breadth and depth of knowledge, including seemingly every yeast gene, maybe every gene, ever, have been one of the more intimidating aspects of graduate school your input as both my quals chair and my thesis advisor has been an immensely valuable and enjoyable experience, thank you.

To my colleagues:

There are too many of you to properly thank everybody but be sure that I am immensely thankful for everyone I have interacted with here at Berkeley. Specifically, I must first thank Addison Von Wright and Mitchell O'Connell for always providing advice and guidance on many of my most important efforts and manuscripts. To Sam Sternberg and Charles Hesser, thank you for always listening and providing feedback on my many crazy ideas. To Sean Higgins and Dana Nadler, for holding down the Savage lab pro-eng fort with me, the cyano forces are strong, but we will maintain! Brett Staahl and Christof Fellmann, you help with eukaryotic experiments and assay development was invaluable and I look forward to many more productive collaborations in the future. Avi Falmhoz and Harneet Rishi, thank you for shepherding our work together through seas of data. To Matt Kim and Kian Tylor, working with you two through your honors thesis has been a wonderful experience and I can't wait to see what you accomplish in the future! To all of the other undergrads, summer and rotation students I have worked with, thank you for always being excited to come in and explore. Finally thank you everyone else that I have not been able to name specifically, for all the day to day support, discussions, coffee breaks, Indian buffet lunches, East Bay Spice trips, sushi burritos and just general cavorting, graduate school has been wonderful. I owe this first and foremost to all the people around me. I wish you all the best of luck!

Table of Contents

List of Figures & Tables

List of commonly used abbreviations

4-HT	4-Hydorxytamoxifen
5-FOA	5-Fluoroorotic Acid
A	Alanine
arC9	allosterically regulated Cas9
aTC	Anhydrotetracycline
BE	β-estradiol
bp	Base pair(s)
Cas	CRISPR associated
CFU	Colony forming units
CRISPRi	CRISPR interference
crRNA	CRISPR RNA
D	Aspartic Acid
darC9	dead allosterically regulated Cas9
DBD	DNA binding domain
dCas9	nuclease dead Cas9
DES	Diethylstilbestrol
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic acid
dsDNA	double stranded DNA
ER	Estrogen receptor
FACS	Fluorescence Activated Cell Sorting
G	Guanine
GFP	Green Fluorescent Protein
gRNA	guide RNA (same as sgRNA)
H	Histidine
HDR	Homology Directed Repair
HR	Homologous Recombination
IPTG	Isopropyl β-D-1 Thiogalactopyranoside
LBD	Ligand Binding Domain

NHEJ	Non Homologous End Joining
NLS	Nuclear localization signal
OD_{600}	Optical Density at 600 nm
PAM	Protospacer adjacent motif
PAMmer	PAM-presenting oligonucleotides
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
RCas9	RNA binding Cas9
RFP	Red Fluorescent Protein
RNA	Ribonucleic acid
sgRNA	single guide RNA
spyCas9	*Streptococcus pyogenes* Cas9
ssRNA	single stranded RNA
TE	Transposable element
tracrRNA	trans activating CRISPR RNA
wt	Wild Type

Introduction:

Engineering CRISPR-Cas9 systems to expand functionality

Since the identification of DNA as the heritable material of the cell, biologists have sought reliable, programmable methods to manipulate gene sequence and expression (Carroll, 2014). More recently research has demonstrated that site-specific double-strand breaks (DSBs) made by targetable DNA nucleases enable such genetic manipulation. Specifically, DSBs are known to stimulate either homologous recombination (HR), allowing the researcher to introduce exogenous DNA via homology, or non-homologous end joining (NHEJ) repair, which can mediate a gene knockout event via small insertions or deletions of sequence (Carroll, 2014; Porteus and Carroll, 2005). Initially, the design of a targeted DNA endonuclease to induce DSBs required that sequence-independent nuclease domains, such as Fok I, linked to DNA-binding proteins such as re-engineered zinc fingers (ZFs) (Boch et al., 2009; Christian et al., 2010; Kim and Chandrasegaran, 1994; Kim et al., 1996; Li et al., 2011; Pavletich and Pabo, 1991; Persikov et al., 2015). These ZF nucleases allowed researchers to generate and test precise hypotheses; causally linking specific genetic sequences to phenotypes, and are the basis for gene therapies currently in clinical development (Carroll, 2014; Cradick et al., 2013; Doudna and Charpentier, 2014; Hsu et al., 2014; Oakes et al., 2016). Nevertheless, such technologies required engineering of a new protein for every DNA target, limiting their use to specialized labs. Only recently has the RNA-guided endonuclease, Cas9, democratized access to the genome by imparting the ability to precisely alter endogenous DNA sequences rapidly, at a massive scale, in nearly any organism of choice(Barrangou and Doudna, 2016).

Cas9 is a RNA-guided double-stranded DNA endonuclease that can bind and cleave nearly any sequence (Carroll, 2014; Doudna and Charpentier, 2014; Hsu et al., 2014; Shalem et al., 2014). In its native context, Cas9 is a crucial component of prokaryotic type II Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) loci, which function as an adaptive immune system of many bacteria and archaea (Chylinski et al., 2013; Gasiunas et al., 2012; Jinek et al., 2012). Cas9 recognizes and degrades invading nucleic acids, especially infecting bacteriophage, via RNA-directed double strand DNA cleavage. Targeting of pathogenic nucleic acids is achieved via an endogenously encoded CRISPR RNA (crRNA) and trans-activating crRNA (tracrRNA). These RNAs hybridize to form a ‘guide RNA’ structure that is specifically recognized and bound by Cas9 (Gasiunas et al., 2012; Jinek et al., 2012). The holo complex, referred to as the Cas9 ribonucleotide protein complex or Cas9-RNP, is able to recognize a *bona fide* DNA target via 20 bp of complementary to the crRNA in the guide RNA. Upon full RNA:DNA hybridization, catalysis of two single-stranded cleavage events occurs within the targeted sequence (Jinek et al., 2012). The Cas9 protein of the Cas9-RNP also plays a crucial role in target specificity and recognition. Cas9 binds a short sequence on the DNA strand not engaged in the RNA target-DNA hybrid (R-loop) that is located 5' and adjacent to the R-loop; this sequence is called the protospacer adjacent motif (PAM) (Anders et al., 2014; Jinek et al., 2012; Sternberg et al., 2014). PAM recognition is critical for the initiation of DNA binding during target search and is vital to the correct positioning of the double-stranded DNA target for ATP-independent unwinding (Anders et al., 2014; Sternberg et al., 2014). Although there is some sequence variation in the PAM element recognized by various Cas9 orthologs, *Streptococcus pyogenes* Cas9 (SpyCas9), the most widely used protein, recognizes the PAM sequence 5'-NGG-3'.

Structure-function studies have on elucidated many of the biochemical mechanisms of Cas9 activity. The first major effort demonstrated that the two-piece cr/tracrRNA could be converted into a chimeric, "single guide" RNA (sgRNA) that was both necessary and sufficient for targeting the dsDNA cleavage ability of Cas9(Jinek et al., 2012). This study also established mutations that inactivate both nuclease domains of Cas9 without perturbing DNA binding ability, leading to the creation of systems that use such a nuclease "dead" Cas9 (dCas9), or a dCas9 to which other effector proteins have been fused. This has enabled transcriptional activation via fusions to domains such as herpes simplex viral protein 16 (VP16) or the core catalytic domain of the histone acetyltransferase p300, transcriptional repression via fusions to the Krüppel-associated box (KRAB) domain of Kox1, and even chromosomal imaging via fusions to fluorophore such as Green fluorescent protein (GFP) (Chen et al., 2013; Gilbert et al., 2013; Hilton et al., 2015). Further termini fusions to dCas9 have harnessed two dCas9 proteins and the requisite dimerization of the split FokI endonuclease to make a more specific, easily programmable endonuclease complex (Guilinger et al., 2014; Tsai et al., 2014a) and recent work has even shown how Cas9 can be adapted for use as base editor via a fusion to the cytidine deaminase APOBEC1 or AID (Kim et al., 2017; Komor et al., 2016; Ma et al., 2016; Nishida et al., 2016).

Nevertheless, the N- and C-termini are the only two points of access on Cas9 to which orthogonal domains can be linked. Moreover, they are within ~4 nm of each other, and are not ideally placed to gain access to the bound DNA, requiring long linkers for most fusions (Anders et al., 2014; Guilinger et al., 2014; Jiang et al., 2016; Komor et al., 2016; Nishida et al., 2016; Nishimasu et al., 2014; Tsai et al., 2014a). This dearth of options when attempting to build new Cas9 functionalities likely explains the relative lack of activity and undesired effects for many of these constructs. For example, dCas9 transcriptional activators require numerous activation domains (up to 24) (Tanenbaum et al., 2014) or combinations of gRNAs to achieve robust activation (Hilton et al., 2015). dCas9-FokI fusions are not nearly as efficient at indel induction as Cas9 itself, (Guilinger et al., 2014; Tsai et al., 2014a) and the base editing cytidine deaminase fusions which result in strong C to T editing within a 12 bp target window also cause significant amounts, 2-3%, of undesirable deamination up to 15bp outside of the Cas9 target sequence (Kim et al., 2017; Komor et al., 2016; Nishida et al., 2016). Furthermore, engineering more complex and dynamic functions, such as allosteric control of Cas9 activity will require the specific internal coupling of protein domains (Ostermeier, 2005; Reynolds et al., 2011). Taken together it is apparent that a more holistic approach to engineering novel Cas9 functionalities will be required in order to enable more proficient generation of Cas9 based tools.

The following thesis describes an in depth analysis and engineering of Cas9 to build enhanced functionality into this important genome modifying protein. The results address many pressing needs of the scientific community including: robust methods to identify functional Cas9 proteins from a large pool of variants, mechanisms to enable the binding of Cas9 to a target RNA rather than DNA, Cas9 proteins which can enable enhanced fusions, and Cas9 proteins that can be regulated in a spatiotemporal fashion (Jinek et al., 2012; O’Connell et al., 2014; Oakes et al., 2014, 2016). Enabling many of these new functionalities required extensive protein engineering. To this end, I also present a detailed set of methods and experiments that describe the first ever comprehensive engineering of Cas9 to identify useful and novel RNA guided DNA binding and cleavage proteins.

Chapter 1

Protein engineering of Cas9 for enhanced function

† The work presented in this chapter has previously been published in the following review article: Oakes, B.L., Nadler, D.C., and Savage, D.F. (2014). Protein Engineering of Cas9 for Enhanced Function. Methods Enzymol. 546C, 491–511.

‡ The work presented in this chapter precedes many of the discoveries reported later in this thesis and should serve as background information informing the following results. Benjamin Oakes designed the research and performed the experiments Dana Nadler & Dave Savage assisted in writing and Dave Savage supervised the project

Abstract

CRISPR/Cas systems act to protect the cell from invading nucleic acids in many bacteria and archaea. The bacterial immune protein Cas9 is a component of one of these CRISPR/Cas systems and has recently been adapted as a tool for genome editing. Cas9 is easily targeted to bind and cleave a DNA sequence via a complimentary RNA; this straightforward programmability has gained Cas9 rapid acceptance in the field of genetic engineering. While this technology has developed quickly, a number of challenges regarding Cas9 specificity, efficiency, fusion protein function, and spatiotemporal control within the cell remain. In this work, we develop a platform for constructing novel proteins to address these open questions. We demonstrate methods to either screen or select active Cas9 mutants and use the screening technique to isolate functional Cas9 variants with a heterologous PDZ domain inserted directly into the protein. As a proof of concept, these methods lay the groundwork for the future construction of diverse Cas9 proteins. Straightforward and accessible techniques for genetic editing are helping to elucidate biology in new and exciting ways; a platform to engineer new functionalities into Cas9 will help forge the next generation of genome modifying tools.

Introduction

The manipulation of gene sequence and expression is fundamental to unraveling the complexity of biological systems. However, our inability to make such manipulations across different organisms and cell types has limited the power of recombinant DNA technology to a handful of model systems. Consequently, numerous strategies for genome engineering – the ability to programmably disrupt or replace genomic loci - have emerged in recent years, yet there remains no universal solution to the problem(Carroll, 2014).

Investigation of bacterial adaptive immunity, known as the ‘clustered regularly interspaced short palindromic repeats’ (CRISPR) system, led to the discovery of the RNA-guided DNA nuclease Cas9, which has proven particularly potent for genome engineering (Barrangou et al., 2007; Deltcheva et al., 2011; Jinek et al., 2012; Wiedenheft et al., 2012). In its biological context, Cas9 is part of a Type II CRISPR interference system that functions to degrade pathogenic phage or plasmid DNA (Fig. 1.1A). Targeting of Cas9 is enabled by host-encoded CRISPR-RNAs (crRNAs), which recognize, through RNA:DNA hybridization, 20 bp of complementary target DNA sequence (referred to as a protospacer). The Cas9 protein itself also plays a role in target recognition by binding a short DNA sequence adjacent and opposite the protospacer, called the protospacer adjacent motif (PAM). Although there is significant variation in PAM specificity among Cas9 orthologs the commonly employed Cas9 from *Streptococcus pyogenes* (SpCas9) recognizes the PAM sequence 5'-NGG-3'. PAM binding is thought to prime Cas9 for target recognition by the crRNA sequence(Sternberg et al., 2014). Upon target recognition, two nuclease domains, termed the RuvC and HNH domains because of their sequence similarity to other endonucleases, engage and cleave the separated strands of DNA between 3 and 4 bp upstream of the PAM site (Jinek et al., 2012). A second *trans*-activating RNA (tracrRNA), with partial complementarity to crRNA, is also required for crRNA maturation and activity. Doudna and colleagues have shown that the crRNA and tracrRNA can be fused together with a tetraloop insertion to form a single guide RNA (sgRNA or ‘guide’) (Jinek et al., 2012). Expression of Cas9 and this sgRNA is both necessary and sufficient for targeting DNA. Therefore, the rapid success of Cas9-based engineering has been driven by programmability – Cas9 can be targeted to any DNA locus by simply changing the sgRNA sequence.

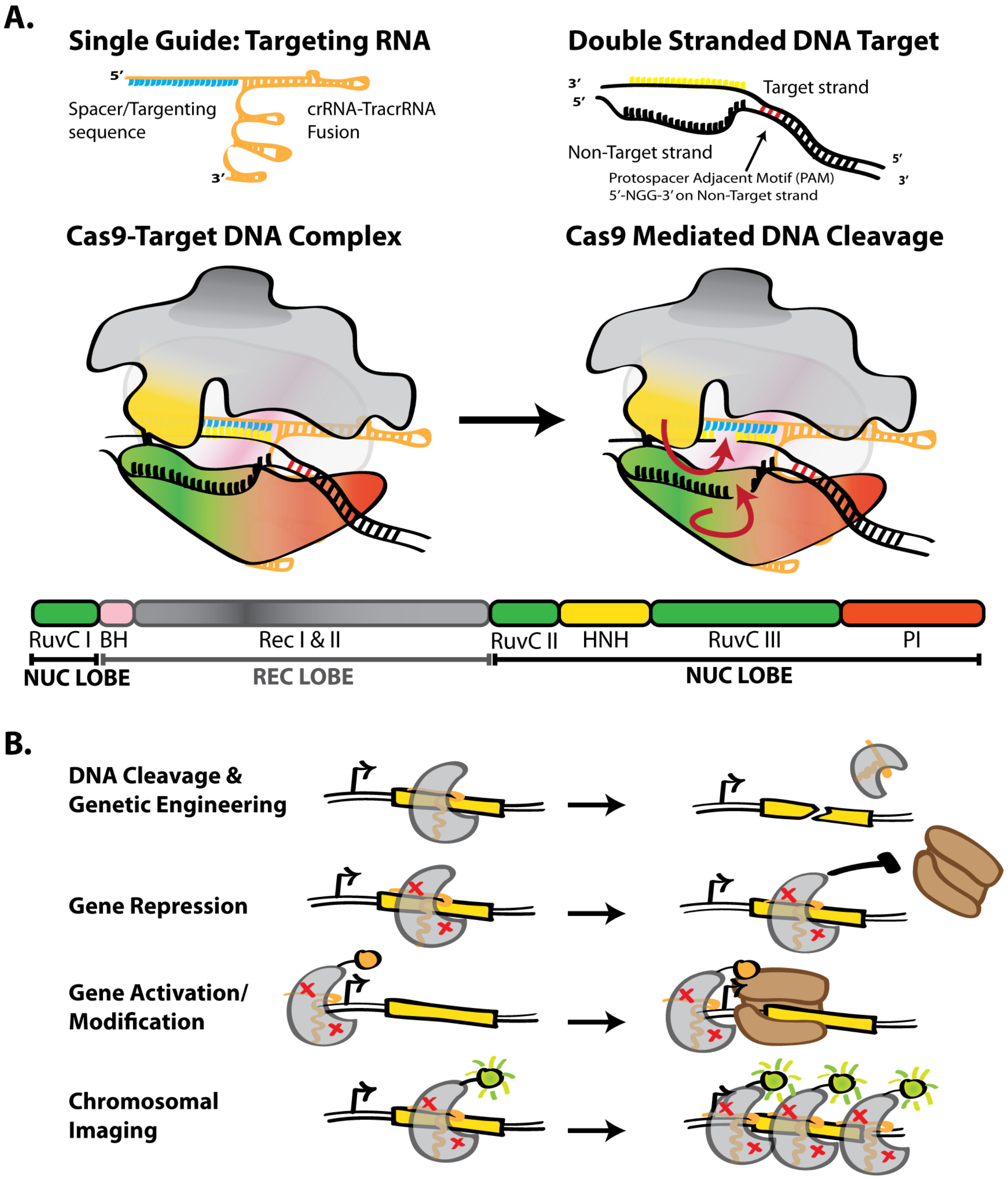


Figure 1.1 | Cas9 structure and its potential uses. (A) Single guide RNA, target dsDNA and Cas9 modeled. Domains of Cas9 are colored accordingly, RuvC-green, BH-pink, RecI-gray, RecII- dark grey, HNH-yellow, PI-red. (B) Common uses of Cas9 as a tool. Red x's in Cas9 represent a nuclease dead variant.

Intense interest in Cas9-based genomic engineering has already led to a number of directed alterations to change or improve Cas9 functionality (Fig. 1.1B). Based on sequence conservation of the RuvC and HNH nuclease domains, a number of point mutants were constructed to transform the normal endonuclease activity into either a nickase (for genome editing) or a catalytically dead mutant (dCas9) that can function as a transcription inhibitor (CRISPRi) (Cong et al., 2013; Jinek et al., 2012; Qi et al., 2013). As PAM recognition is critical to functionality but encoded by the protein, a number of efforts have identified Cas9 orthologs with minimal PAM requirements for use in conjunction with, or in place of, SpCas9 (Esvelt et al., 2013). Finally, a number of N- and C-terminal fusions to Cas9 have been used to recruit alternative factors to specific DNA loci, including RNA polymerase subunits to activate transcription and additional nuclease domains for improving the on-target specificity of genome editing (Bikard et al., 2013; Guilinger et al., 2014; Tsai et al., 2014a). These advances show that, from an engineering perspective, Cas9 can be thought of as a unifying factor able to recruit any protein, RNA, and DNA together in the cell (Mali et al., 2013a). However, even with these recent improvements, there are a number of additional desirable that could be engineered into Cas9.

Enabling complex functions requires more elaborate protein engineering efforts than previously attempted. Fortunately, high-resolution structures of apo and holo Cas9 were recently solved and inform this process greatly (Jinek et al., 2014; Nishimasu et al., 2014). Thus, we begin with a discussion of structure, which frames the protein engineering effort. This is followed by a discussion of potentially feasible and advantageous manipulations of Cas9. Ultimately, the isolation of an enhanced Cas9 with new function will require assaying the activity of large libraries (> 10^6) of variants. Fortuitously, the basic function of Cas9, disrupting gene sequence or gene expression, facilitates the construction of a genetic screen and means that directed evolution methods can be employed in the same conditions functionality is ultimately desired. To this end, we present a detailed set of methods, employing either screening or selection, for the directed evolution of novel Cas9 proteins.

1.1 The structure of Cas9

Recently published high-resolution structures of Cas9 serve as a useful starting point to inform the protein engineering process. Doudna and colleagues reported the high-resolution structure of apoSpCas9 using x-ray crystallography and a holo complex using electron microscopy (EM) (Jinek et al., 2014). Concurrently, Zhang and colleagues solved the x-ray structure of SpCas9 bound to sgRNA and single-stranded target DNA (ssDNA) (Nishimasu et al., 2014). Here, we summarize these structures in the context of engineering novel Cas9 variants.

Cas9 possesses a hand-shaped structure of size 100 Å x 100 Å x 50 Å and is composed of two major lobes, an N-terminal recognition (REC) lobe and a C-terminal lobe (NUC) possessing two endonuclease domains (Fig. 1. 2). The REC lobe is composed of three segments: a small portion of the RuvC domain, an arginine-rich bridge helix (BH) which links the REC to the NUC lobe and an α-helical recognition segment with two subdomains (RecI and II). The NUC lobe possesses three domains: the RuvC endonuclease, the HNH endonuclease, and a PAM-interacting (PI) C-terminal domain. The sgRNA-DNA complex lies at the interface of the two lobes, with the BH and RecI domains making many of the primary interactions. It should be noted that the sgRNA-target DNA heteroduplex generally makes non-specific, sequence independent interactions with the protein, while the sgRNA repeat:anti-repeat region makes sequence-dependent interactions; this is consistent with the precise sgRNA recognition yet broad target flexibility of Cas9. Interestingly, although the sgRNA has extensive structural features – there are three stem loops in the tracrRNA sequence – stem loops 2 and 3 exit out the 'back' of the structure and make few sequence-specific contacts with Cas9 (inset Fig. 1. 2).

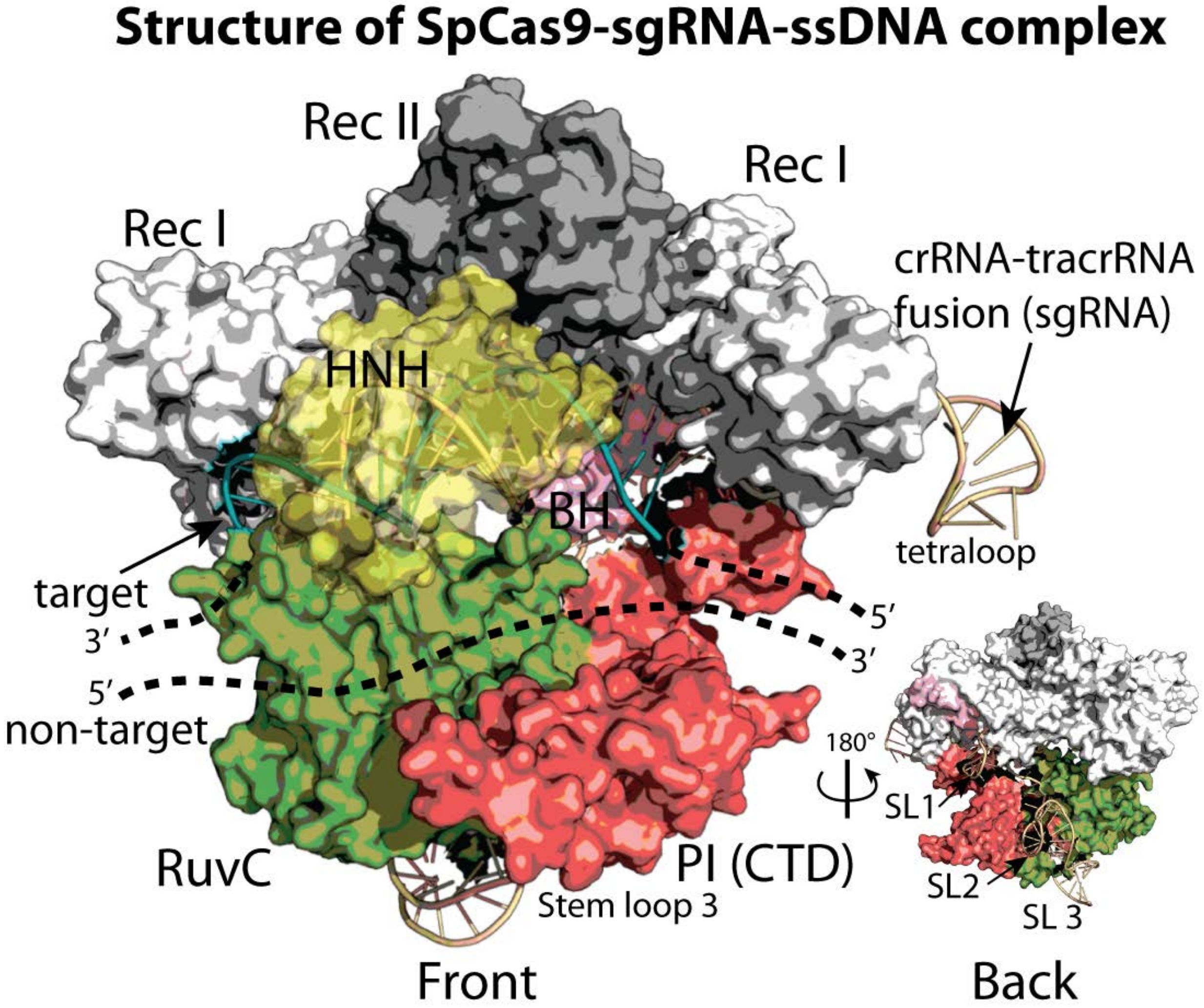


Figure 1.2 | The structure of Cas9 in complex with sgRNA and a single-stranded target DNA (ssDNA) adapted from Nishimasu et al. 2014 (PDB code 4OO8). The structure is shown in a surface representation and colored as in Fig. 1. 1. Note, the inset is rotated 180° and displays the location of the sgRNA stem loops (SLs).

The structures inform a number of rational protein engineering design strategies. Many publications have employed N and C-terminal fusions to Cas9 as a means of recruiting specific factors to a genomic locus, such as RNA polymerase for transcriptional activation (Bikard et al., 2013; Mali et al., 2013b). The structures suggest the protein N-terminus is adjacent to the 3' end of the target DNA exiting Cas9 while the C-terminus is adjacent to the 5'-end of the same DNA. The close proximity of DNA to protein termini explains probably explains why many simple fusions have been successful. Intriguingly, the N and C termini are in the same lobe and roughly 50 Å apart, suggesting it might be possible to make circularly permuted forms of Cas9 for more elaborate protein engineering uses. Moreover, the bi-lobed nature of the apo complex indicates it could be possible to construct split variants of Cas9 that are only active under conditions when both halves are recruited together. Such variants would be a useful tool in the construction of two-hybrid systems and for the engineering of allosterically controlled Cas9 derivatives, such as with optogenetic domains.

The sequence conservation of Cas9 orthologs mapped onto the structure also suggests domains that may prove malleable for engineering. Phylogenetic variation among Cas9 orthologs is significantly higher in the RecII domain of the REC lobe and the PI domain in the NUC lobe (Chylinski et al., 2014; Nishimasu et al., 2014). The RecII domain makes very few contacts with the sgRNA or target DNA in the holo complex structure and Nishimasu et al. demonstrated that the domain (Δ175-307) can be completely eliminated yet still retain roughly 50% of editing activity. Conversely, two deletions (Δ97-150 and Δ312-409) in the highly conserved RecI domain were catalytically dead in the same assay. This is intriguing because Cas9 is very large; all known Cas9 proteins range from 984-1629 amino acids (Chylinski et al., 2013). Thus, a current goal in the community is to find or engineer smaller, but equally effective, Cas9 proteins and this may be possible through multiple domain deletions. Sequence diversity between PI domains also suggests a means to alter PAM specificity. Cross-linking experiments and structural evidence conclude that the C-terminal PI domain interacts directly with the PAM (Jinek et al., 2014). However, this specificity may in fact be modular. Domain swapping of the SpCas9 PI domain for that of the orthologous sequence from *S. thermophilus* imbues a preference for the *S. thermophilus* PAM sequence (TGGCG) (Nishimasu et al., 2014). Expanded domain swapping experiments, coupled with directed evolution, could therefore provide a means for creating orthogonal Cas9 variants from a single, well-characterized SpCas9 scaffold.

1.2 Current uses

Cas9 has rapidly established itself as a promising genome engineering technology in widely used model organisms (Friedland et al., 2013; Gratz et al., 2013; Guilinger et al., 2014; Hsu et al., 2013; Hwang et al., 2013; Materials et al., 2013; Nishimasu et al., 2014; Niu et al., 2014; Shan et al., 2013; Tsai et al., 2014b; Wang et al., 2013a). In these systems, Cas9 has been used to create both small genomic insertions and deletions (indels) via non-homologus end joining (NHEJ) and to facilitate larger sequence manipulation with homologous recombination (HR). Cas9 also allows for multiplexed genome engineering, and has been used to create large knock-out libraries in human cells, a feat both surprising in its simplicity and impressive in its efficacy (Shalem et al., 2014; Zhou et al., 2014). Decoupling the DNA binding activity of Cas9 proteins from cleavage activity has lead to a broader set of uses such as repression and activation of transcription (Gilbert et al., 2013). Finally, recent evidence suggests that Cas9 may even be able to manipulate RNA (Sampson et al., 2013). Although still nascent, the simple programmability and effectiveness of Cas9-based technology promises to democratize access to genome manipulation.

1.3 Initial engineering questions

As previously mentioned, there are a number of clear initial questions pertaining to Cas9 that are addressable using existing protein engineering tools. Namely, we believe that designing novel Cas9s with domain insertions and deletions will lead to the creation of a new family of synthetic orthologs whose outputs are manifold. For example, domain insertions could act to recruit a additional protein partners with desired activity onto Cas9-associated nucleic acids; domain deletions will reduce Cas9's size and increase its versatility.

Alternatively, improving N or C-terminal fusions with engineered linkers or creating Cas9s with new N and C-termini altogether, may greatly increase the efficacy of fusions. For example, to address issues of Cas9 targeting specificity dCas9 has been fused to FokI, an obligate dimeric sequence-independent nuclease (Guilinger et al., 2014; Tsai et al., 2014a). This system requires the mutual on-target activity of two different FokI-dCas9 fusions to adjacent sites, a combined 40 bp of targeting, to catalyze a DNA cleavage event. Unfortunately, these FokI-dCas9 fusions are substantially less active than either WT Cas9 or the dual nickase strategy at inducing indels, rendering them a less attractive tool. Nevertheless, it is known that FokI protein fusions to other DNA binding domains can achieve cleavage efficiencies similar to that of WT Cas9 (Hwang et al., 2013; Mali et al., 2013b) Therefore, lower activity of the current FokI-dCas9 is likely due to imperfect positioning of the FokI nuclease domain and further engineering the dCas9-FokI interface should yield an increase in activity.

Lastly, split proteins are known to function as switches or response elements in many different systems (Olson and Tabor, 2014). Splitting Cas9, as mentioned above, would be a simple method for engineering allosteric control and open the door to a number of uses including optogenetics, small-molecule dependence, or linking function to a cellular signal, such as a phosphorylation-dependent signal transduction cascade. Nevertheless, all of the previous engineered scenarios require that Cas9 is active despite the modifications introduced. Therefore, it is imperative that any engineering attempt start with an assay allowing for the separation of active mutants from the inactive.

Methods

To advance some of the aforementioned goals related to Cas9 protein engineering, we have developed a suite of protocols allowing for the rapid isolation of functional Cas9 proteins. These techniques rely on the ability to screen or select active Cas9 from a large pool of variants. As such, the methods may be applied equally to either rational or library-based approaches for engineering Cas9.

2.1 A note on applications

Although we present here a general method for directed evolution of Cas9, it is impossible to cover the myriad of nuances inherent to different applications. Instead, as a proof of concept, we demonstrate that Cas9 can be manipulated by domain insertion, a common event found in eukaryotic proteomes (Lander et al., 2001). Specifically, we created libraries of Cas9 in which the α-syntrophin PDZ domain was randomly inserted throughout the SpCas9 gene once per variant using *in vitro* transposition-based methods (Edwards et al., 2008). PDZ domains are small (~ 100 amino acids) proteins with adjacent N- and C-termini that mediate protein-protein interactions by specifically binding the C-terminal peptide of a cognate partner in the cell (Nourry et al., 2003). These domains have been used extensively in synthetic biology applications as a tool for scaffolding proteins (Dueber et al., 2003, 2007). In the context of Cas9, PDZ insertion could recruit additional factors to a DNA-bound Cas9 in the cell, such as florescent tags, chromatin remodeling machinery and nuclease domains.

2.2 Electrocompetent *E. coli* preparation for library construction.

Construction of libraries containing 10^6-10^9 diverse members is a basic step for engineering new functions into Cas9. A simple method for the creation of electrocompetent cells that consistently yields *E. coli* with transformation efficiencies of $\geq 10^8$ per µg of plasmid is described. If necessary, cells can be pre-transformed with additional screening/selection plasmids, as required, to maximize library transformation efficiency.

1. Begin with the desired strain grown as single colonies on appropriate plates. For example, we use plasmid 44251: pgRNA-bacteria (addgene) with the RFP guide from Qi et al. (2013) transformed and grown on carbenicillin plates (100µg/mL) overnight at 37°C to single colonies.
2. Pick a single colony and inoculate into 5mL SOC (BD Difco 244310: Super Optimal Broth + 20 mL 20 % glucose per L) plus carbenicillin. Grow overnight at 37°C.
3. Inoculate 1 L of SOC + carbenicillin with the 5 mL overnight culture. Grow at 37°C for 2-4 hours to an OD_{600} between 0.55 and 0.65.
4. Rapidly cool the culture in an ice bath by swirling. Keep cells at during all subsequent steps.
5. Centrifuge cells at 4000g for 10 min. Wash the cell culture with 500 mL ice-cold water by gentle resuspension. Centrifuge at 4000g for 10 min. Repeat wash step twice.
6. Wash cells with 500mL of ice cold 10% glycerol. Repeat twice.
7. Perform a final spin and discard the supernatant. Resuspend the pellet in 2.75 mL of 10% glycerol and aliquot into cold microcentrifuge tubes, 80 µL each. Flash freeze.
8. Transforming cells: Thaw on ice, add library plasmid to 75 µL cells and vortex. Electroporate at 1800 V, 200ohms, 25uF in 0.1cm cuvettes, resuspending in warm SOC immediately afterward. Recover for one hour at 37°C, before adding antibiotics, as required.

2.3 Discovery of functional, engineered, variants of Cas9 proteins:

After generating any library of Cas9 variants, it is necessary to have an platform which can separate active variants from the library with minimal effort. Two assays to probe this functionality are the coupling of dCas9 activity to either RFP expression or media-dependent cell growth. When devising these systems, it is important to keep the First Law of Directed Evolution (Schmidt-Dannert and Arnold, 1999) in mind: ‘you get what you screen for.’

2.4 Screening for Cas9

The catalytically dead version of Cas9, has a functional output that can be tied directly to transcription in *E.coli;* namely, it can repress transcription of a desired gene (Qi et al., 2013). Qi et al. previously demonstrated that dCas9 with a guide sequence of 5'-AACTTTCAGTTTAGCGGTCT-3' can target and repress a genome-encoded RFP while avoiding repression of a genome-encoded upstream GFP (Qi et al., 2013). In a screening context, this provides a simple output for assaying dCas9 functionality (i.e. RFP knock-down) while correcting for extrinsic noise in the population by monitoring GFP (Elowitz, 2002). The basic method of screening is schematized in Figure 3A. Briefly, cells containing functional dCas9s will repress RFP and express GFP while those with non-functional dCas9s will express both fluorescent proteins. This signal is easily distinguished using flow cytometry and florescence imaging (Fig. 1. 3B and 3C).

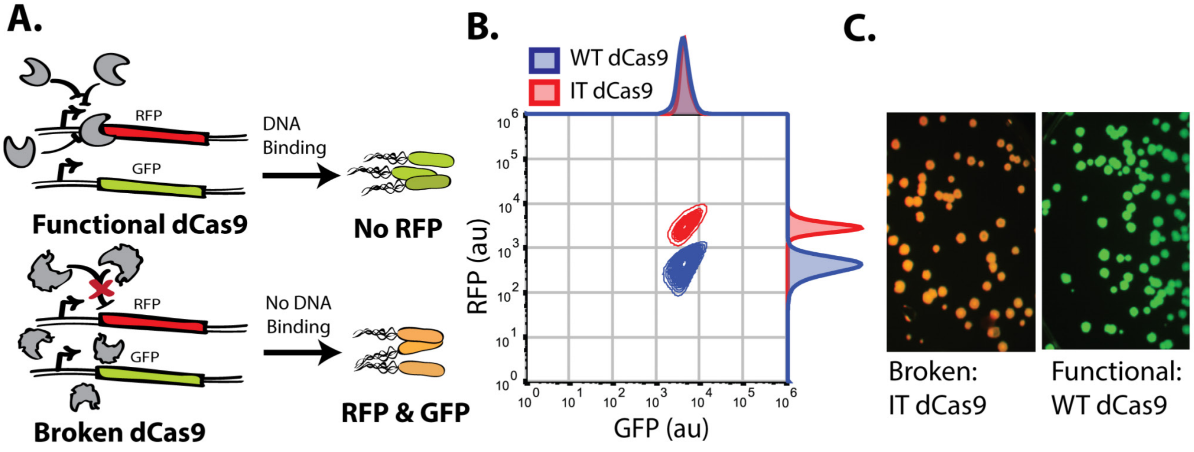


Figure 1.3 | Screen for functional Cas9s. (A) Schematic representation of the screen. (B) Flow cytometry data of the functional positive (WT dCas9) control in blue and negative 'Inactive Truncation' Cas9 (IT dCas9) control in Red, IT dCas9 contains only the C-terminal 250 amino acids. Both controls contain the sgRNA plasmid targeting RFP for repression. Samples were grown overnight in rich induction media. (C) Colony fluorescence of the functional (WT dCas9) and 'Broken' negative (IT dCas9) controls.

2.5 Selecting Cas9

To complement the screening method and for use with larger libraries of Cas9 mutants, we have also developed a technique to select functional dCas9s using cellular growth. We fashioned a derivative of the classic yeast counter-selection method, which takes advantage of the toxicity of 5-fluoroorotic acid (5-FOA) in cells with the URA3 gene (Fig. 1. 4A)(Boeke et al., 1984). In yeast, URA3 encodes orotidine-5'phosphate decarboxylase, which catalyzes 5-FOA into a highly toxic compound(Boeke et al., 1984). The *E.coli* homolog of URA3, PyrF, is thought to act in a similar manner, and PyrF and the upstream gene PyrE are known to function as a selectable marker in other gram-negative bacteria (Galvao and de Lorenzo, 2005; Yano et al., 2005). Nevertheless, it was unclear whether dCas9 based repression would mimic the effects of these full gene knockouts in an *E. coli* system. To this end, we tested whether repression of either of PyrF and PyrE by dCas9 was sufficient to rescue a slow growth phenotype on 5-FOA. After creating a number of different sgRNA's to the start of PyrE and PyrF, we determined that the guides 5'-accttcttgatgatgggcac-3' for PyrF and 5'-taagcgcaaattcaataaac-3' for PyrE rescued growth in 5mM of 5-FOA (Fig. 1. 4B).

Ultimately, it is important to decide which approach, screening or selection, will be used to enrich for functional engineered Cas9 mutants. A primary determining factor is the theoretical library size. Screening systems can effectively cover libraries of sizes up to ~10^6, which is roughly the amount of *E. coli* that can conveniently be sorted 10x by flow cytometer in an hour. On the other hand, selection systems, which rely on repression/activation of a toxic/essential gene for growth, can screen libraries of random protein variants of up to ~10^9 in size (Persikov et al., 2013).

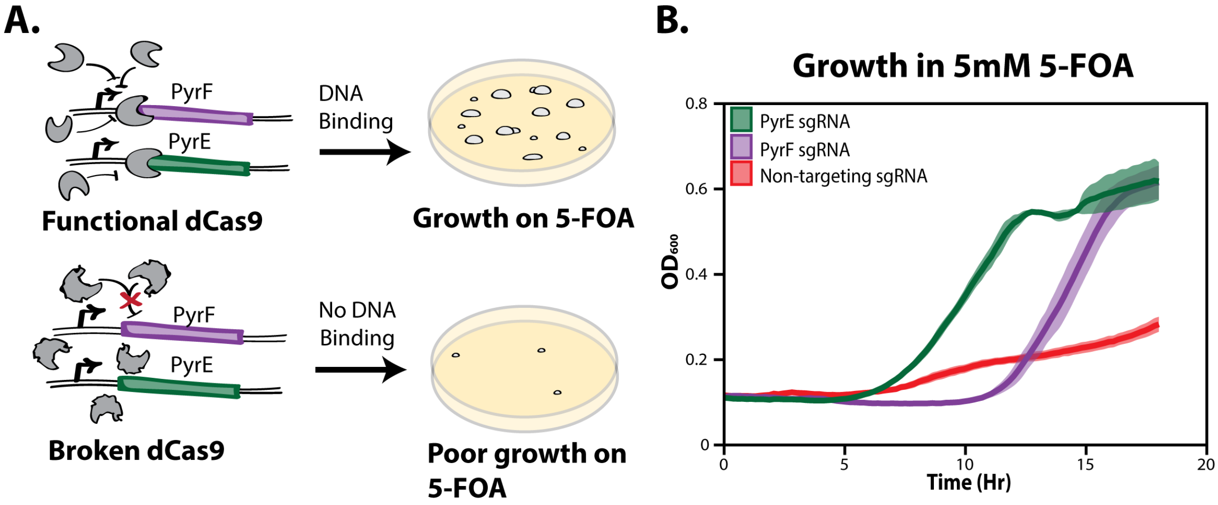


Figure 1.4. | Functional Cas9 selection overview. (A) Schematic representation of the selection system. (B) Growth rate of a functional dCas9 + sgRNAs repressing the PyrF gene (purple), PyrE gene (green), and a no guide sequence control (red). Samples were grown in rich induction media +5mM 5-Fluoroorotic Acid. All measurements represent the average (line) and standard deviation (shading) of three biological replicates.

2.6 Screening for functional Cas9 variants

We have found that Fluorescence-activated cell sorting (FACS) is a convenient method for isolating functional Cas9 variants. This approach is somewhat more flexible than a selection – a gating strategy in FACS is easily manipulated while the growth constraints of a selection are not – yet still provides reasonable throughput. As a proof of concept, we demonstrate the FACS-based screening of functional dCas9 variants possessing the insertion of an α-syntrophin PDZ domain.

1. Obtain a dCas9 library containing ≤ 10^6 variants on an expression plasmid of choice. Here we used a tetracycline inducible expression plasmid, plasmid 44249: pdCas9-bacteria from addgene (Qi et al., 2013) to create a library which has a SNTA1 PDZ domain inserted across the whole dCas9 protein (Dueber et al., 2003). Based on the possible insertion sites and linkers, the size of this library is roughly equal to 10^6.
2. Transform electrocompetent *E. coli* expressing GFP and RFP with 1 µg of the library plasmid and 1 µg/µL of a sgRNA to RNA, if necessary. Here, the *E. coli* strain and guide RNA plasmid come from Qi et al. 2013 (plasmid is 44251: pgRNA-bacteria; addgene).
3. To ensure adequate coverage of the library the transformation efficiency should be at least 5-10x greater than the theoretical library size. To determine this plate 5 µl aliquots of serially-diluted transformants, and grow to colonies (overnight, 37°C) on double selection-media (chloramphenicol (50 µg/mL) to maintain the engineered dCas9 plasmids and carbenicillin (100 µg/mL) for maintenance of the guide RNA plasmid). Store the remaining transformed cells at 4 °C overnight.
4. Determine the volume of the transformation mixture needed to cover the theoretical library size 5-10x and inoculate into 5 mL of rich induction media: SOC, chloramphenicol, carbenicillin and 2 µM anhydrotetracycline (aTC). Concurrently, inoculate tubes of rich induction media with controls WT dCas9 and IT dCas9 with the RFP sgRNA. Grow at 37°C; we have found shaking ≥ 250 rpm is helpful for maximum RFP and GFP fluorescence.

5. After 8-12 hours of growth, centrifuge 500 µL of each sample, wash 2x with 1 mL 1x PBS and re-suspend 1:20 in 1x PBS for flow cytometry.
6. Run the controls on a FACS instrument to establish correct positive and negative gating (Fig 5.).
7. Screen the library dCas9's using FACS and collect the cells falling into the previously determined positive gate in rich media without antibiotic (Fig. 1. 5). Screen at least 10x the library size to, as cellular viability post-FACS is often substantially less than 100%.
8. Recover sorted cells for 2 hours at 37 °C.
9. Depending on the library enrichment after a single round of sorting, repeating steps 4-8 may be necessary to further enrich for functional, engineered dCas9 clones.

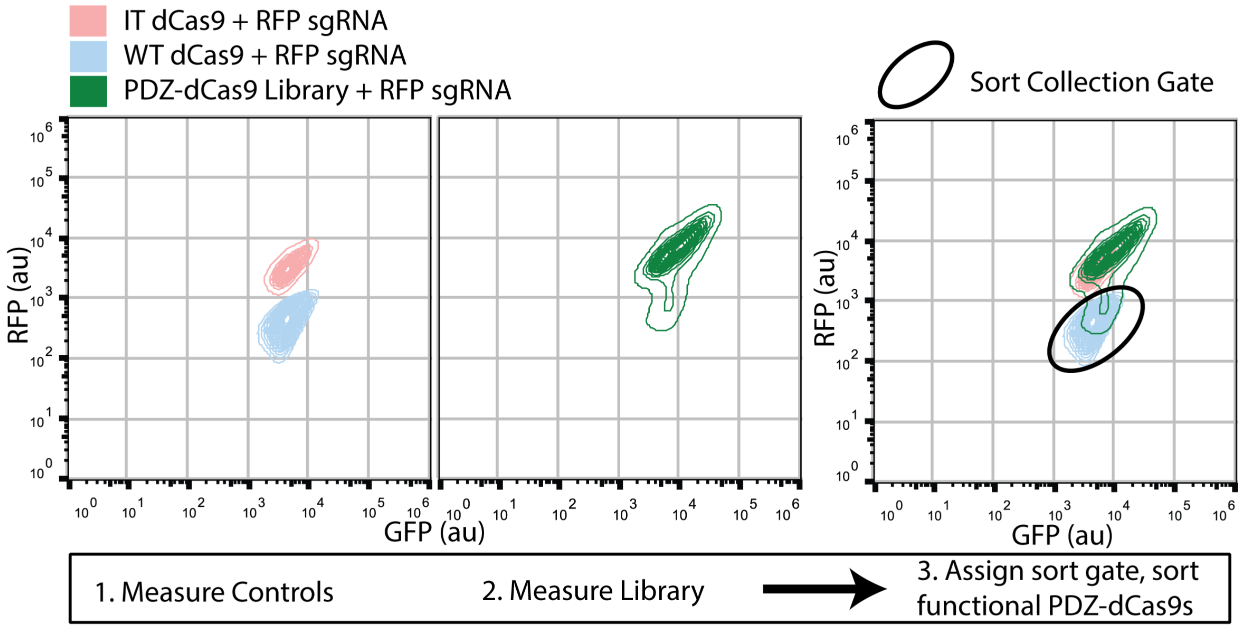


Figure 1.5 | Cell sorting data from the GFP-RFP screen. The first panel depicts the RFP vs GFP measurements of WT dCas9 (blue) and IT dCas9 (pink) as run on a Sony SH800 cell sorter. The separation of these two controls into distinct populations is readily apparent. The second panel portrays the spread of the PDZ-Cas9 intercalation library (green). The last panel shows the overlay of all three FACS plots. It is evident that there are populations of functional and non-functional proteins within the single PDZ-dCas9 library and the third panel also provides a demonstration of a gate for isolation of functional PDZ- dCas9 intercalations.

2.7 Determining screening enrichment of PDZ-dCas9 domain insertions

A successful round of screening with the PDZ-dCas9 library should enrich for functional PDZ-dCas9 insertion mutants (‘intercalations’). A straightforward method to check for enrichment is to PCR with a primer specific to the inserted domain and one external to the engineered Cas9 (Fig 6A). An amplified smear indicates a relatively ‘naïve’ library while specific bands indicate enriched library members. The following is a representative protocol for checking screening success.

1. Plate approximately 1000-10,000 of the sorted and recovered cells on rich induction plates with antibiotics and inducer. Add the remaining cells in liquid media with appropriate antibiotics.
2. Grow the induction plate(s) overnight at 37°C and allow 12 hrs at room temperature for RFP to fully mature. *As described in section 2.8 below, this plate will be used to pick colonies with functionally intercalated PDZ-dCas9s.*
3. Grow the liquid culture at 37°C overnight and prepare a glycerol stock of the sorted cells for future use by mixing 800 µL of culture with 400 µL of 50% glycerol in lysogeny broth (LB).
4. Centrifuge the remaining liquid culture and miniprep to recover plasmid DNA (Qiagen).
5. Perform a PCR on the original and the screened libraries with the primers described above (Fig. 1. 6A). The screened library PCR should show enrichment bands (Fig. 1. 6B). If bands are not evident, the library may require further rounds of screening (section 2.6, step 4). Alternatively, deep sequencing may be used to rigorously characterize the library.

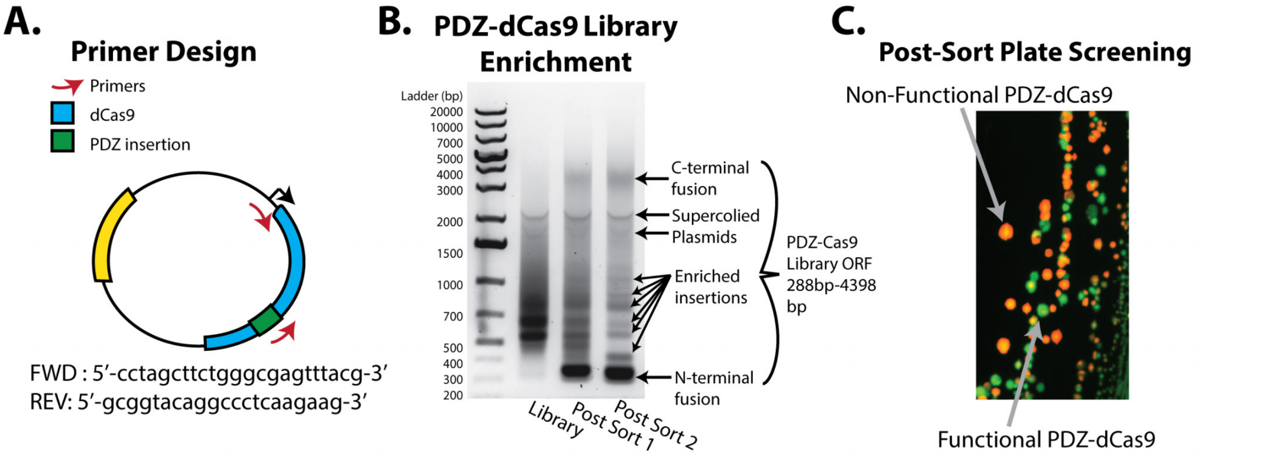


Figure 1.6 | Checking success of a screen and picking final clones. (A) An overview of primer design for the PDZ-dCas9 library. (B) Gel electrophoresis of PCRs run on the original PDZ-dCas9 library and the first and second round of screening. The banding patterns that appear after the first and second sorts are indicative of library enrichment, representing the insertion sites of a PDZ domain. It is also evident that the N and C termini fusions to PDZ are also enriched. Since these fusions are expected to be functional this serves as an internal control. (C) Fluorescent image of the on-plate 'finishing' screen. Colonies that express only GFP are expected to have a functional PDZ-dCas9.

2.8 Identifying and testing PDZ-Cas9 clones from a screened library.

Next, it is next necessary to isolate functional dCas9 clones with intercalated PDZ domains. This is done via a final plate-based screen. Once identified and isolated, it is then possible to collect, test, and verify unique PDZ-dCas9 clones in a secondary, discrete, screening method, such as repression of an alternative gene.

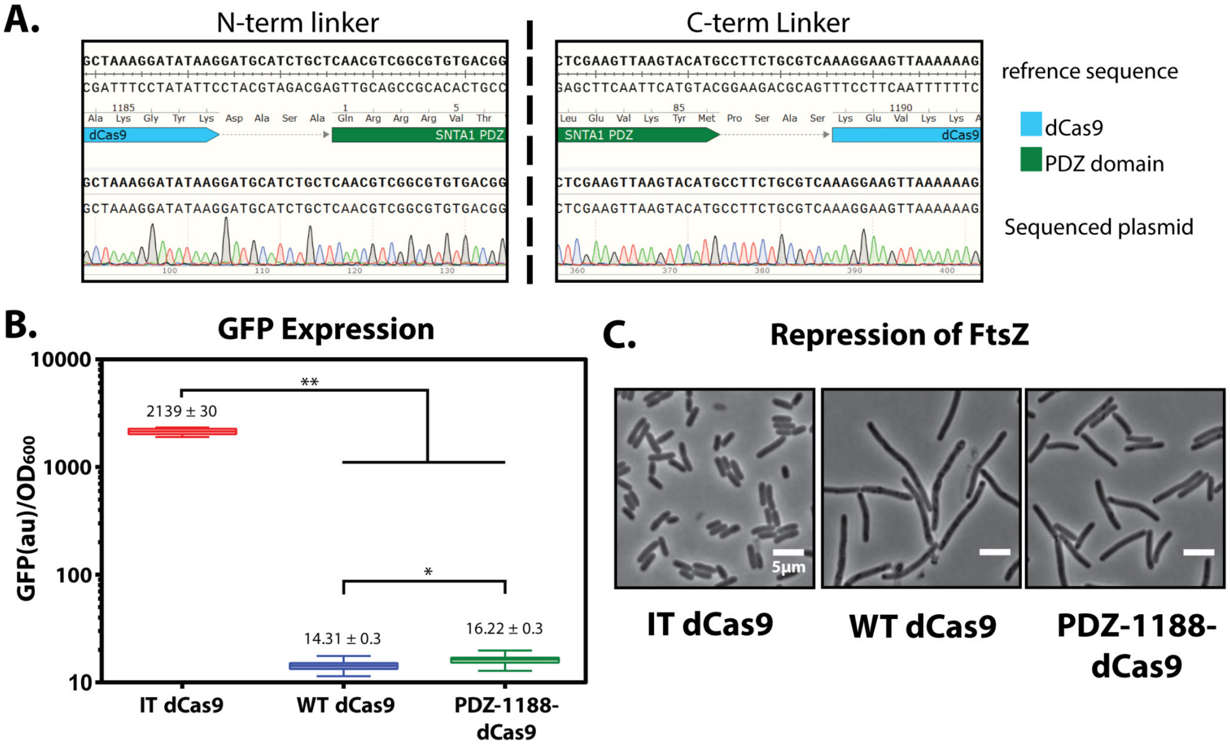


Figure 1.7 | Validating functionality of engineered dCas9. (A) Sequence validation of the site 1188 PDZ intercalation. Sequence alignment via SnapGene. (B) Quantitative repression of GFP by the PDZ-1188 -dCas9 intercalation clone. Bulk florescence measurements of GFP expression levels over five hrs. Double asterisks represents a p value of < 0.0001 in a 1 way ANOVA. Single asterisk represents p values of < 0.0001 in an unpaired Student's T-test. (C) Qualitative repression of ftsZ gene by PDZ-dCas9 and controls, scale bar is 5 μm.

1. Set up a 96-well plate of 50 μL PCR reactions using the primers from section 2.7, step 5. In parallel, fill a 96-well plate with 100 μL of rich media per well.
2. From the induction plates grown in section 2.7, step 2, pick colonies expressing only GFP (Fig. 1. 6C), as assayed via fluorescence imaging (Bio-Rad Chemidoc MP). Spot each colony into a well of the PCR plate by pipetting up and down 5x and then resuspend the colony in the corresponding well of the media plate.
3. Run the aforementioned PCR and sequence the amplicons. Store the resuspended colony plate at 4 °C.
4. Align sequences to the original plasmid map using appropriate software. Ensure variants are in-frame and determine unique clones.
5. Take 50 μL of the corresponding unique clones from the colony resuspension plate and grow overnight in 5 mL of rich media with antibiotics. Miniprep DNA to obtain a mix of both the engineered PDZ-dCas9 plasmid and the RFP guide plasmid for each isolate.
6. Using a primer upstream of the dCas9 insertion site, sequence the plasmid to determine the insertion site and linkers. (Fig 7. A)
7. Digest 5 μg of plasmid mixture with BsaI to remove the guide plasmid and clean up DNA (Qiagen) to remove restriction enzyme. (Digestion rxn occurs as follows 1. 37°C-60 min, 2. 50°C-60 min, 3. 80°C-10 min)

8. Transform the digested plasmid mixture with 200ng of new guide plasmids to examine the functionally of the intercalated PDZ-dCas9 on other genes and endogenous loci. Specifically, we transformed one of our PDZ-dCas9 intercalations with guides for GFP and FtsZ, an essential cell division protein (sequences 5'-atctaattcaacaagaatt-3', 5'-tcggcgtcggcggcggcgg-3' respectively).
9. Culture the bacteria with the PDZ-dCas9 intercalation isolates and the new guides. Grow the original dCas9 controls with these new guides, induce with 2µM aTC and measure the phenotypes accordingly.
10. Validate that the qualitative and quantitative phenotypes are within range of the WT Cas9 (Fig. 1. 1.7 B-C).

Expanding Horizons

While the democratization of simple and multiplex genome engineering is central to the story of Cas9, the question of reliable specificity is paramount to its future use. Ideally, Cas9 would target and cleave one site in a complex genome yet leave other, similar, sites un-marred. Scarring the genome obscures the genotype-phenotype relationship, limiting basic science utility, and Cas9 cannot translate into the therapeutic arena if it is known to induce spurious mutations. Thus, how and when PAM and guide interactions integrate to provide specificity and activate Cas9-mediated cleavage is essential. Studies have shown that while the spCas9 PAM 5'-NGG-3' requirement is strict, only poorly tolerating one other target site (5'-NAG-3'), sgRNA:target-DNA hybridization may accept a number of mismatches, especially towards the 5' end of the sgRNA (Hsu et al., 2013; Mali et al., 2013c; Pattanayak et al., 2013). Accordingly, detrimental off-target binding and cleavage activity when using Cas9 is a pressing issue.

A number of reports have addressed this concern. Truncating the guide sequence appears to lessen the accepted number of mismatches in a given guide (Fu et al., 2014). Alternatively, Cas9 nickases, which cleave only one strand, can be multiplexed to require mutual on-target activity of two Cas9's in order for editing to occur (Mali et al., 2013b; Ran et al., 2013). Finally, it has been shown that lowering the expression of Cas9 can lessen off-target effects (Hsu et al., 2013). Nevertheless, rigorous engineering may lead to superior solutions.

While the systems for isolating active Cas9 variants presented here are not designed to directly address specificity concerns, we envisage that with small changes our selection and screening platforms could separate mutant Cas9s with less specificity from those with more. For example, it should be possible to introduce high affinity off-target binding-sites in front of the fluorescent reporter not actively targeted, such that any binding to these mock sites would act as an internal counterscreen. We are motivated by the likelihood that, in the future, screens and selections of this vein may be used to engineer synthetic Cas9 proteins with can tolerate few, if any mismatches, in the guide and/or PAM sequence.

Conclusion

Cas9 has fundamentally altered the genome engineering landscape due to its simple programmability and overall effectiveness. Here, for the first time, we have delineated protein engineering-based methods for the directed evolution of Cas9 proteins with novel functions. We believe that such techniques will be critical for answering unresolved biochemical questions of protein structure and function. Moreover, directed evolution of Cas9 will allow for more refined improvement of this singular protein and the construction of next-generation tools for both interrogating the genome and biomedical therapies.

Chapter 2

Programmable RNA recognition and cleavage by CRISPR/Cas9

† The work presented in this chapter has previously been published in the following research article: O'Connell, M.R., Oakes, B.L., Sternberg, S.H., East-Seletsky, A., Kaplan, M., Doudna, J.A. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature, doi:10.1038/nature13769.

‡ Mitchell O'Connell and Benjamin Oakes designed experiments, performed Cas9 binding and cleavage assays, and conducted pull-down experiments. Samuel Sternberg and Alexandra East-Seletsky assisted with experimental design and experiments. Mitchell O'Connell, Benjamin Oakes and Jennifer Doudna wrote the manuscript. Samuel Sternberg Matias Kaplan generated biotin-Cas9 with the assistance of Samuel Sternberg. Jennifer Doudna supervised the project.

Abstract

The CRISPR-associated protein Cas9 is an RNA-guided DNA endonuclease that uses RNA:DNA complementarity to identify target sites for sequence-specific double-stranded DNA (dsDNA) cleavage. In its native context, Cas9 acts on DNA substrates exclusively because both binding and catalysis require recognition of a short DNA sequence, the protospacer adjacent motif (PAM), next to and on the strand opposite the 20-nucleotide target site in dsDNA. Cas9 has proven to be a versatile tool for genome engineering and gene regulation in many cell types and organisms, but it has been thought to be incapable of targeting RNA. Here we show that Cas9 binds with high affinity to single-stranded RNA (ssRNA) targets matching the Cas9-associated guide RNA sequence when the PAM is presented *in trans* as a separate DNA oligonucleotide. Furthermore, PAM-presenting oligonucleotides (PAMmers) stimulate site-specific endonucleolytic cleavage of ssRNA targets, similar to PAM-mediated stimulation of Cas9-catalyzed DNA cleavage. Using specially designed PAMmers, Cas9 can be specifically directed to bind or cut RNA targets while avoiding corresponding DNA sequences, and we demonstrate that this strategy enables the isolation of a specific endogenous mRNA from cells. These results reveal a fundamental connection between PAM binding and substrate selection by Cas9, and highlight the utility of Cas9 for programmable and tagless transcript recognition.

Introduction

CRISPR–Cas immune systems must discriminate between self and non-self to avoid an autoimmune response (Marraffini and Sontheimer, 2010). In type I and II systems, foreign DNA targets which contain adjacent PAM sequences are targeted for degradation, whereas potential targets in CRISPR loci of the host do not contain PAMs and are avoided by RNA-guided interference complexes (Garneau et al., 2010; Gasiunas et al., 2012; Mojica et al., 2009; Sashital et al., 2012). Single-molecule and bulk biochemical experiments showed that PAMs act both to recruit Cas9–guide RNA complexes (Cas9–gRNA) to potential target sites and to trigger nuclease domain activation (Sternberg et al., 2014). Cas9 from *Streptococcus pyogenes* recognizes a 5’-NGG-3’ PAM on the non-target (displaced) DNA strand (Jinek et al., 2012; Mojica et al., 2009), suggesting that PAM recognition may stimulate catalysis through allosteric regulation. Moreover, the HNH nuclease domain of Cas9, which is responsible for target strand cleavage (Gasiunas et al., 2012; Jinek et al., 2012), is homologous to other HNH domains shown previously to cleave RNA substrates (Hsia et al., 2004; Pommer et al., 2001). Based on the observation that single-stranded DNA (ssDNA) targets can be activated for cleavage by a separate PAMmer oligonucleotide (Sternberg et al., 2014), and that similar HNH domains can cleave RNA, we wondered whether a similar strategy would enable Cas9 to bind and cleave ssRNA targets in a programmable fashion (Fig. 2.1a).

Results:

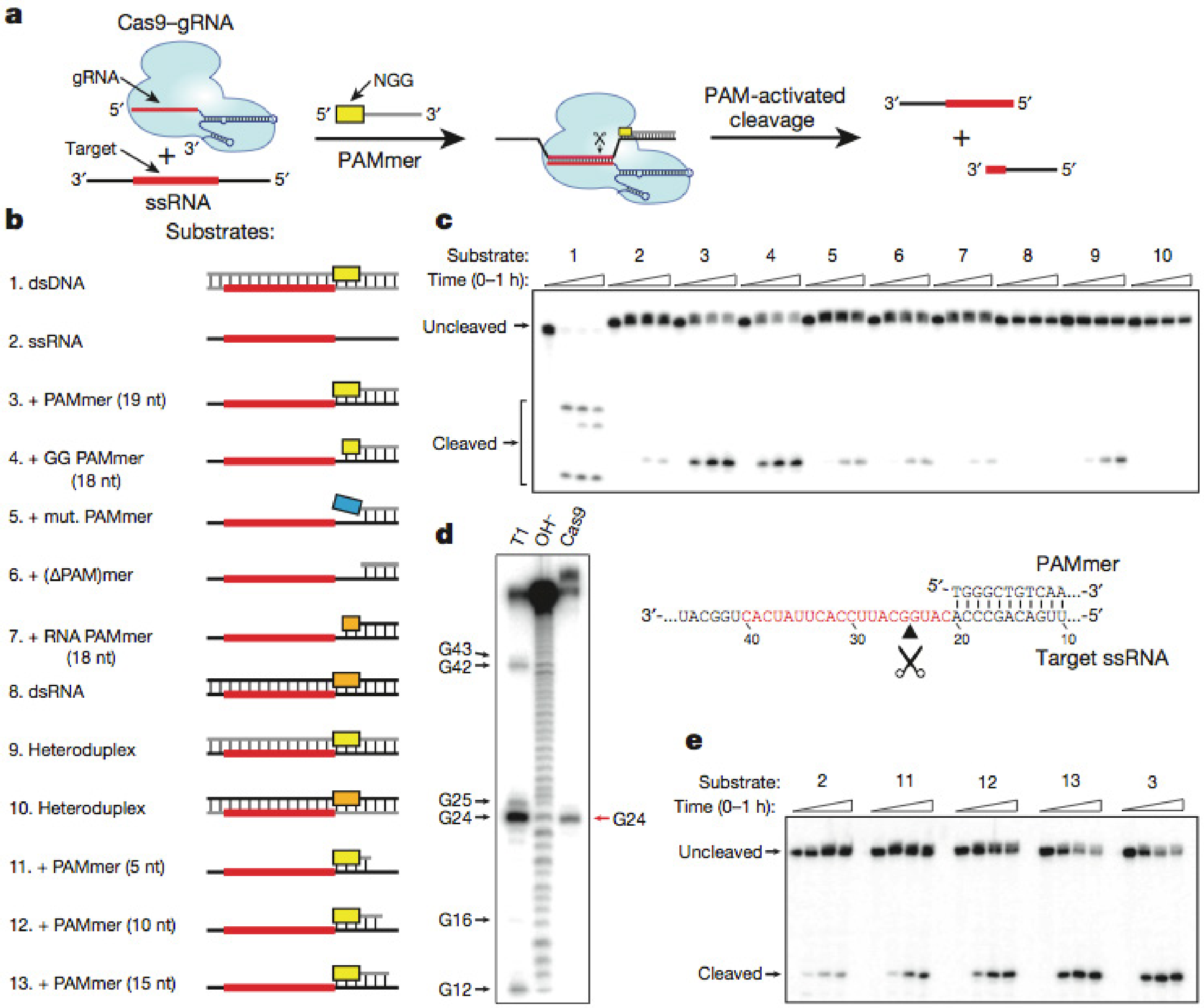


Figure 2.1 | RNA-guided Cas9 cleaves ssRNA targets in the presence of a short PAM-presenting DNA oligonucleotide (PAMmer). (a) Schematic depicting the approach used to target ssRNA for programmable, sequence-specific cleavage. (b) The panel of nucleic acid substrates examined in this study. Substrate elements are colored as follows: DNA (grey), RNA (black), guide RNA target sequence (red), DNA PAM (yellow), mutated DNA PAM (blue), RNA PAM (orange). (c) Representative cleavage assay for 5'-radiolabeled nucleic acid substrates using Cas9–gRNA, numbered as in (b). (d) Cas9–gRNA cleavage site mapping assay for substrate 3. T1 and OH^- denote RNase T1 and hydrolysis ladders, respectively; the sequence of the target ssRNA is shown at right. (e) Representative ssRNA cleavage assay in the presence of PAMmers of increasing length, numbered as in (b).

Using *S. pyogenes* Cas9 and dual-guide RNAs (Methods), we performed *in vitro* cleavage experiments using a panel of different RNA and DNA targets (Fig. 2.1b). Deoxyribonucleotide-comprised PAMmers specifically activated Cas9 to cleave ssRNA (Fig. 2.1c), an effect that required a 5'-NGG-3' or 5'-GG-3' PAM. RNA cleavage was not observed using ribonucleotide-based PAMmers, suggesting that Cas9 may recognize the local helical geometry and/or deoxyribose moieties within the PAM. Consistent with this idea, dsRNA targets were not cleavable, and RNA–DNA heteroduplexes could only be cleaved when the non-target strand was composed of deoxyribonucleotides. Interestingly, we found that Cas9 cleaved the ssRNA target strand between positions 4 and 5 of the base-paired guide RNA-target RNA hybrid (Fig. 2.1d), in contrast to the cleavage between positions 3 and 4 observed for dsDNA (Garneau et al., 2010; Gasiunas et al., 2012; Jinek et al., 2012) likely due to subtle differences in substrate positioning. However, we did observe a significant reduction in the pseudo-first order cleavage rate constant of PAMmer-activated ssRNA as compared to ssDNA (Sternberg et al., 2014) (Fig. 2.2).

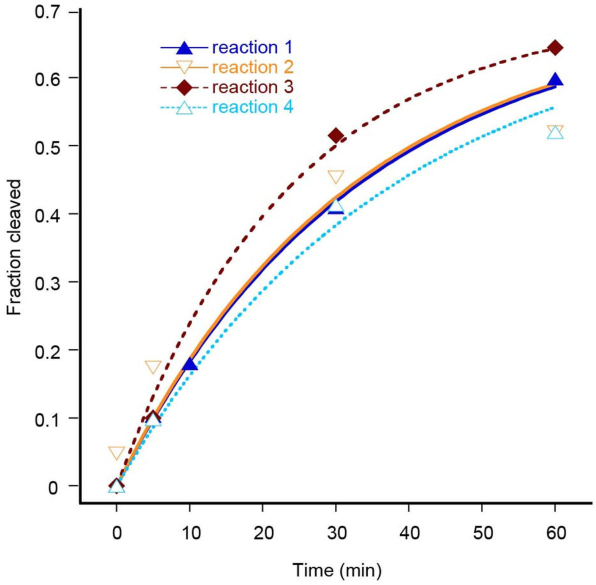


Figure 2.2 | Quantified data for cleavage of ssRNA by Cas9–gRNA in the presence of a 19-nt PAMmer. Cleavage assays were conducted as described in the Materials and Methods, and the quantified data were fit with single-exponential decays. Results from four independent experiments yielded an average apparent pseudo-first order cleavage rate constant of 0.032 ± 0.007 min^{-1}. This is slower than the rate constant determined previously for ssDNA in the presence of the same 19-nt PAMmer (7.3 ± 3.2 min^{-1}) (Sternberg et al., 2014).

We hypothesized that PAMmer nuclease activation would depend on the stability of the hybridized PAMmer–ssRNA duplex and tested this by varying the PAMmer length. As expected, ssRNA cleavage was lost when the predicted melting temperature for the duplex decreased below the temperature used in our experiments (Fig. 2.1e). In addition, large molar excesses of di- or tri-deoxyribonucleotides in solution were poor activators of Cas9 cleavage (Fig. 2.3). Collectively, these data demonstrate that hybrid substrate structures composed of ssRNA and deoxyribonucleotide-based PAMmers that anneal upstream of the RNA target sequence can be efficiently cleaved by RNA-guided Cas9.

We next investigated the binding affinity of catalytically inactive (dCas9; D10A/H840A) dCas9–gRNA for ssRNA targets with and without PAMmers using native gel mobility shift experiments. Intriguingly, while our previous results showed that ssDNA and PAMmer-activated ssDNA targets are bound with indistinguishable affinity (Sternberg et al., 2014), PAMmer-activated ssRNA targets were bound >500-fold tighter than ssRNA alone (Fig. 2.4a,b). A recent crystal structure of Cas9 bound to a ssDNA target revealed deoxyribose -specific van der Waals interactions between the protein and the DNA backbone (Nishimasu et al., 2014), suggesting that energetic penalties associated with ssRNA binding must be attenuated by favorable compensatory binding interactions with the provided PAM. The equilibrium dissociation constant measured for a PAMmer–ssRNA substrate was within 5-fold of that for dsDNA (Fig. 2.4b), and this high-affinity interaction again required a cognate deoxyribonucleotide-comprised 5’-GG-3’ PAM (Fig. 2.4a). Tight binding also scaled with the PAMmer length (Fig. 2.4c), consistent with the cleavage data presented above.

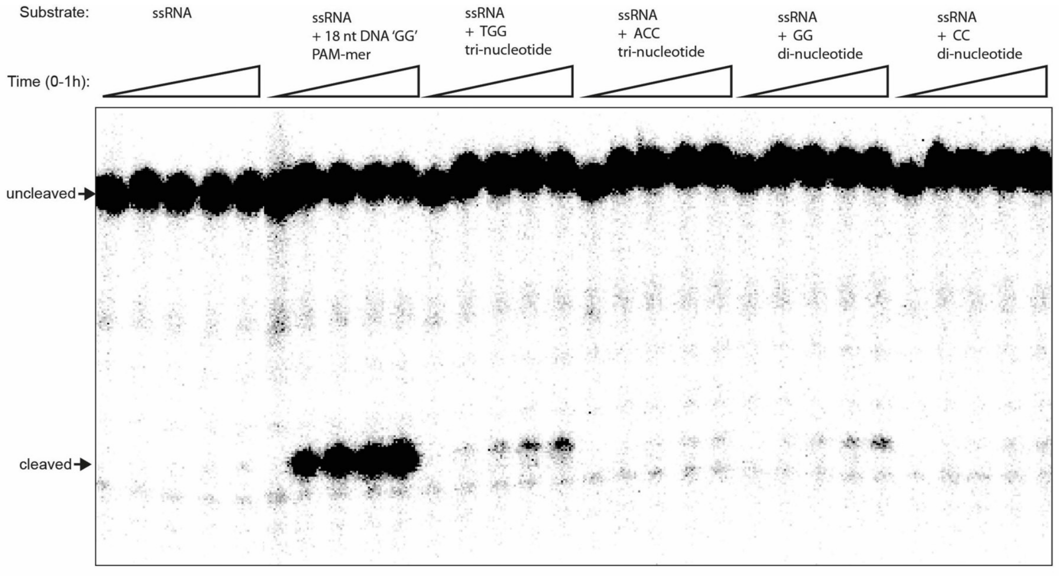


Figure 2.3 | RNA cleavage is marginally stimulated by di- and tri-deoxyribonucleotide PAMmers. Cleavage reactions contained ~1 nM 5′- radiolabelled target ssRNA and no PAMmer (left), 100 nM 18-nt PAMmer (second from left), or 1 mM of the indicated di- or tri-nucleotide (remaining lanes). Reaction products were resolved by 12% denaturing polyacrylamide gel electrophoresis (PAGE) and visualized by phosphorimaging.

To verify the programmable nature of PAMmer-mediated ssRNA cleavage by Cas9–gRNA, we prepared three distinct guide RNAs (λ2, λ3, and λ4) and showed that their corresponding ssRNA targets could be efficiently cleaved using complementary PAMmers without any detectable cross-reactivity (Fig. 2.5a). This result indicates that complementary RNA–RNA base-pairing is critical in these reactions. Surprisingly, though, dCas9 programmed with the λ2 guide RNA bound all three PAMmer–ssRNA substrates with similar affinity (Fig. 2.5b). This observation suggests that high-affinity binding in this case may not require correct base-pairing between the guide RNA and the ssRNA target, particularly given the compensatory role of the PAMmer.

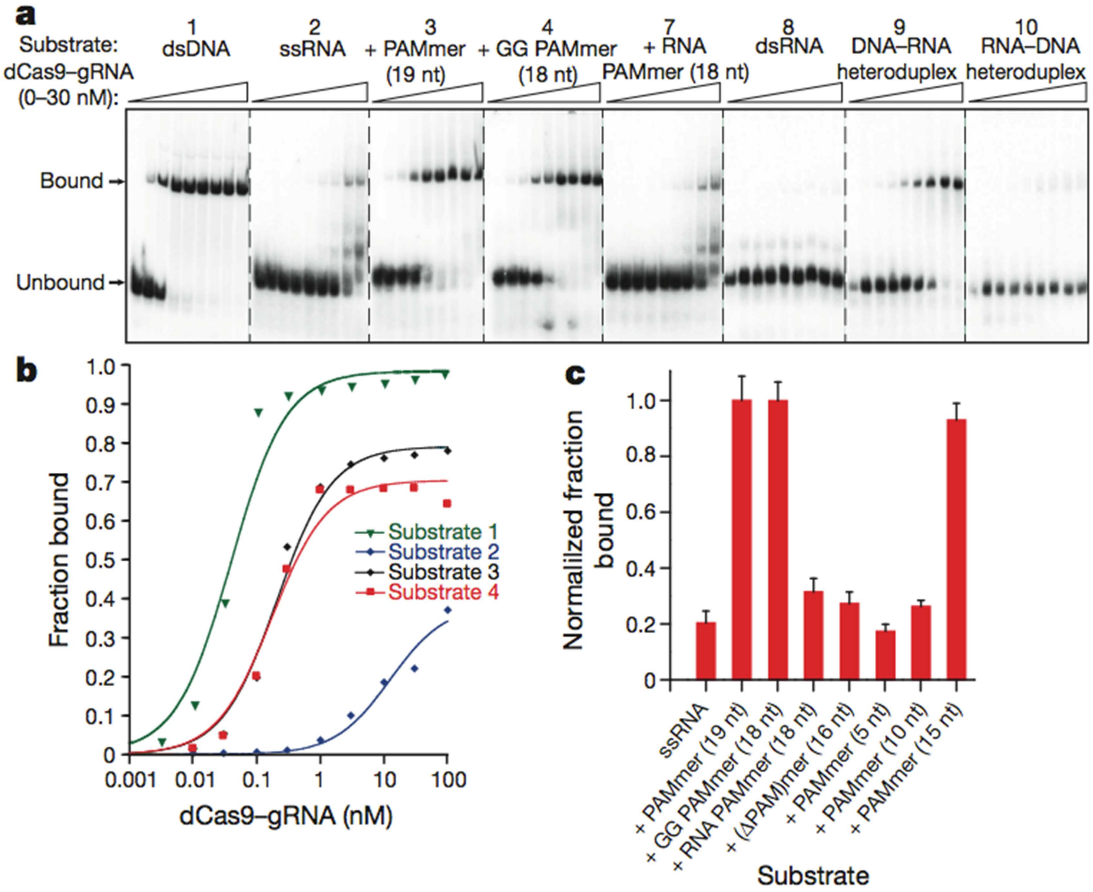


Figure 2.4 | dCas9–gRNA binds ssRNA targets with high affinity in the presence of PAMmers. (a) Representative electrophoretic mobility shift assay for binding reactions with dCas9–gRNA and a panel of 5'-radiolabeled nucleic acid substrates, numbered as in Fig. 2.1b. (b) Quantified binding data for substrates 1–4 from (a) fit with standard binding isotherms. Measured dissociation constants from three independent experiments (mean ± s.d.) were 0.036 ± 0.003 nM (1), >100 nM (2), 0.20 ± 0.09 nM (3), and 0.18 ± 0.07 nM (4). (c) Relative binding data for 1nM dCas9–gRNA and 5'-radiolabeled ssRNA with a panel of different PAMmers. The data are normalized to the amount of binding observed at 1 nM dCas9–gRNA with a 19 nt PAMmer; error bars represent the standard deviation from three independent experiments.

During dsDNA targeting by Cas9–gRNA, duplex melting proceeds directionally from the PAM and strictly requires formation of complementary RNA–DNA base-pairs to offset the energetic costs associated with dsDNA unwinding (Sternberg et al., 2014). We therefore wondered whether binding specificity for ssRNA substrates would be recovered using PAMmers containing 5'-extensions that create a partially double-stranded target region requiring unwinding (Fig. 2.5c). Indeed, we found that use of a 5'-extended PAMmer enabled dCas9 bearing the λ2 guide sequence to bind sequence-selectively to the λ2 PAMmer–ssRNA target. The λ3 and λ4 PAMmer–ssRNA targets were not recognized under these conditions (Fig. 2.5d & 2.6), although we did observe a 10-fold reduction in overall ssRNA substrate binding affinity. By systematically varying the length of the 5' extension, we found that PAMmers containing 2–8 additional nucleotides upstream of the 5'-NGG-3' offer an optimal compromise between gains in binding specificity and concomitant losses in binding affinity and cleavage efficiency (Fig. 2.7).

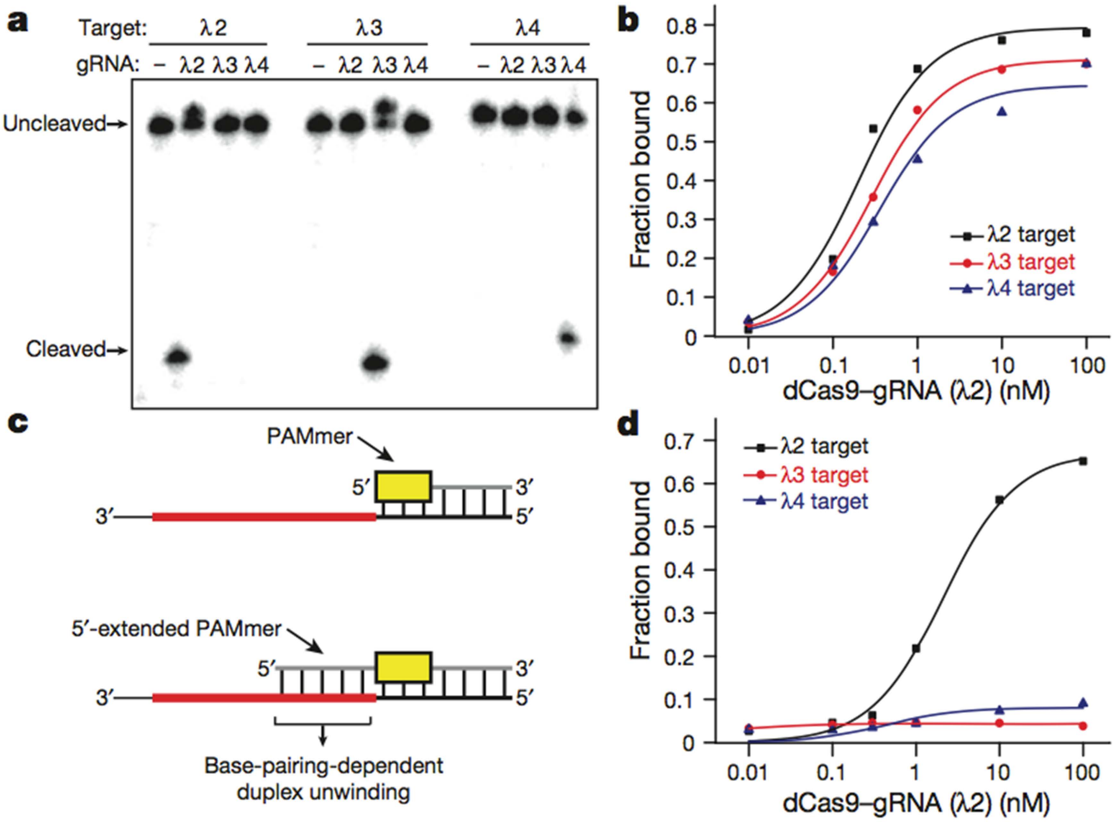


Figure 2.5 | 5'-extended PAMmers are required for specific target ssRNA binding. (a) Cas9 programmed with either λ2, λ3, or λ4-targeting gRNAs exhibits sequence-specific cleavage of 5'-radiolabeled λ2, λ3, and λ4 target ssRNAs, respectively, in the presence of cognate PAMmers. (b) dCas9 programmed with a λ2-targeting gRNA exhibits similar binding affinity to λ2, λ3, and λ4 target ssRNAs in the presence of cognate PAMmers. Dissociation constants from three independent experiments (mean ± s.d.) were 0.20 ± 0.09 nM (λ2), 0.33 ± 0.14 nM (λ3), and 0.53 ± 0.21 nM (λ4). (c) Schematic depicting the approach used to restore guide RNA-mediated ssRNA binding specificity, which involves 5'-extensions to the PAMmer that cover part of the target sequence. (d) dCas9 programmed with a λ2-targeting gRNA specifically binds the λ2 ssRNA but not λ3 and λ4 ssRNAs in the presence of 5'-extended PAMmers. Dissociation constants from three independent experiments (mean ± s.d.) were 3.3 ± 1.2 nM (λ2) and >100 nM (λ3 and λ4).

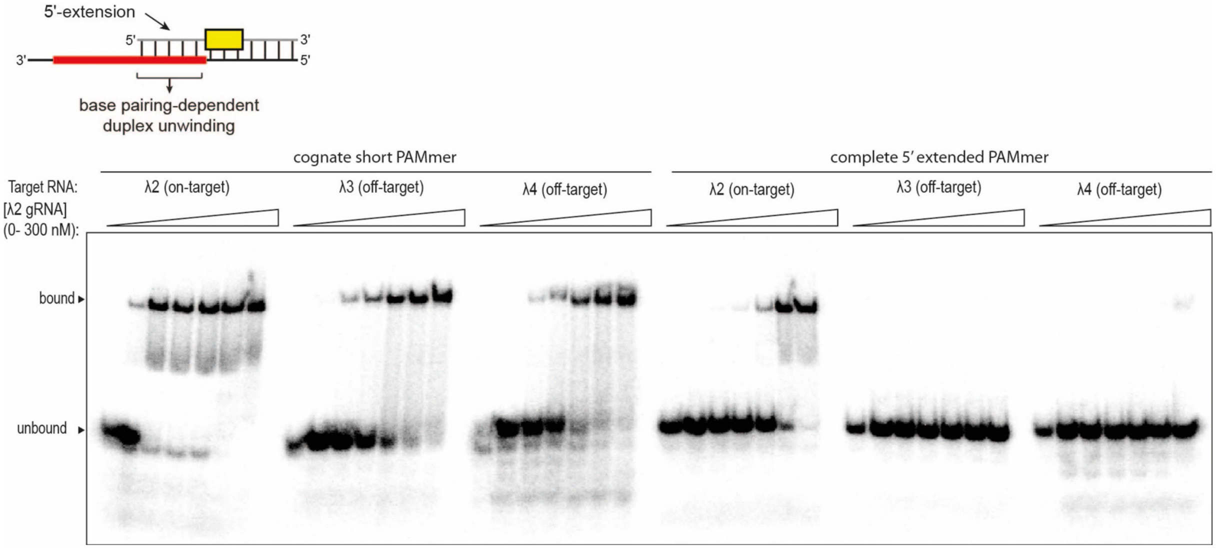


Figure 2.6 | Representative binding experiment demonstrating guide-specific ssRNA binding with 5’-extended PAMmers. Gel shift assays were conducted as described in the Materials and Methods. Binding reactions contained Cas9 programmed with λ2-gRNA and either λ2 (on-target), λ3 (off-target) or λ4 (off-target) ssRNA in the presence of short cognate PAMmers or cognate PAMmers with complete 5’-extensions, as indicated. The presence of a cognate 5’-extended PAM- mer abrogates off-target binding. Three independent experiments were conducted to produce the data shown in Fig. 2.5b,d.

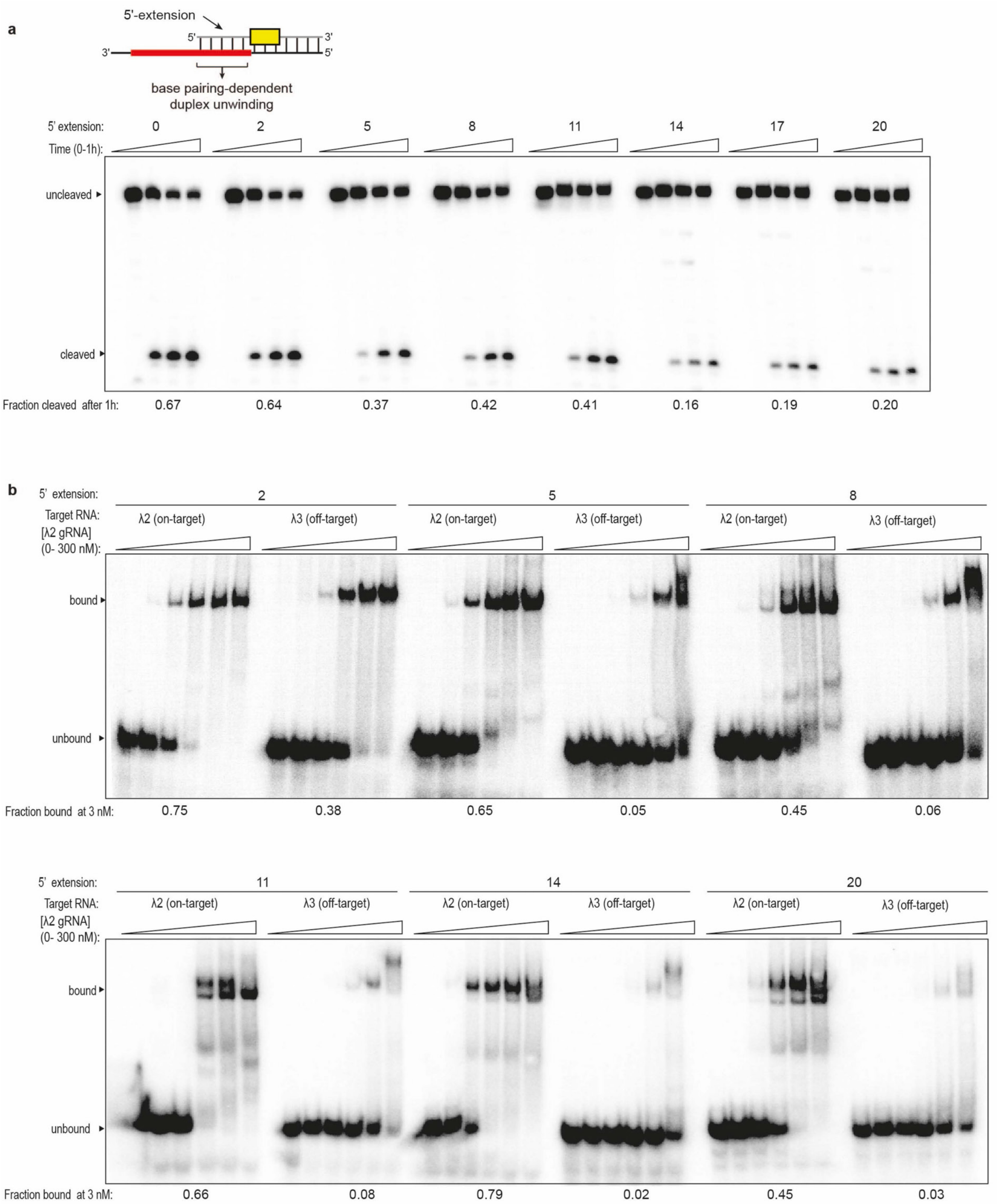


Figure 2.7 | Exploration of RNA cleavage efficiencies and binding specificity using PAMmers with variable 5’-extensions. (a) Cleavage assays were conducted as described in the Materials and Methods. Reactions contained Cas9 programmed with λ2-gRNA and λ2 ssRNA target in the presence of PAMmers with 5’-extensions of variable length. The ssRNA cleavage efficiency decreases as the PAMmer extends further into the target region, as indicated by the fraction RNA cleaved after 1 h. (b) Binding assays were conducted as described in the Materials and Methods, using mostly the same panel of 5’-extended PAMmers as in (a). Binding reactions contained Cas9 programmed with λ2-gRNA and either λ2 (on-target) or λ3 (off-target) ssRNA in the presence of cognate PAMmers with 5’-extensions of variable length. The binding specificity increases as the PAMmer extends further into the target region, as indicated by the fraction of λ3 (off-target) ssRNA bound at 3 nM Cas9-gRNA. PAMmers with 5’ extensions also cause a slight reduction in the relative binding affinity of λ2 (on-target) ssRNA.

Next we investigated whether nuclease activation by PAMmers requires base-pairing between the 5'-NGG-3' and corresponding nucleotides on the ssRNA. Prior studies showed that DNA substrates containing a cognate PAM that is mismatched with the corresponding nucleotides on the target strand are cleaved as efficiently, under some conditions, as a fully base-paired PAM (Jinek et al., 2012). Importantly, this could enable targeting of RNA while precluding binding or cleavage of corresponding genomic DNA sites lacking PAMs (Fig. 2.8a). To test this possibility, we first demonstrated that Cas9–gRNA cleaves PAMmer–ssRNA substrates regardless of whether or not the PAM is base-paired (Fig. 2.8b,c). When Cas9–RNA was incubated with both a PAMmer–ssRNA substrate and the corresponding dsDNA template containing a cognate PAM, both targets were cleaved. In contrast, when a dsDNA target lacking a PAM was incubated together with a PAMmer-ssRNA substrate bearing a mismatched 5'-NGG-3' PAM, Cas9–gRNA selectively targeted the ssRNA for cleavage (Fig. 2.8c). The same result was obtained using a mismatched PAMmer with a 5' extension (Fig. 2.8c), demonstrating that this general strategy enables the specific targeting of RNA transcripts while effectively eliminating any targeting of their corresponding dsDNA template loci.

We next explored whether Cas9-mediated RNA targeting could be applied for tagless transcript isolation from HeLa cells (Fig. 2.8d). To immobilize Cas9 on a solid-phase resin, we mutagenized Cas9 to remove both wild-type cysteine residues and introduced a unique cysteine at the N-terminus distal from any nucleic acid binding surfaces. Chemical labeling of purified Cas9 at this position with a biotin moiety was specific and robust, and the resulting biotin-Cas9 protein was fully active and could be quantitatively retained by magnetic streptavidin beads (Fig. 2.9).

As a proof of concept, we first isolated *GAPDH* mRNA from HeLa total RNA using biotinylated dCas9, gRNAs and PAMmers that target four non-PAM-adjacent sequences within exons 5–7 (Fig. 2.8e). We observed a substantial enrichment of *GAPDH* mRNA relative to a control *β-actin* mRNA by Northern blot analysis, but saw no enrichment using a non-targeting gRNA or dCas9 alone (Fig. 2.8f).

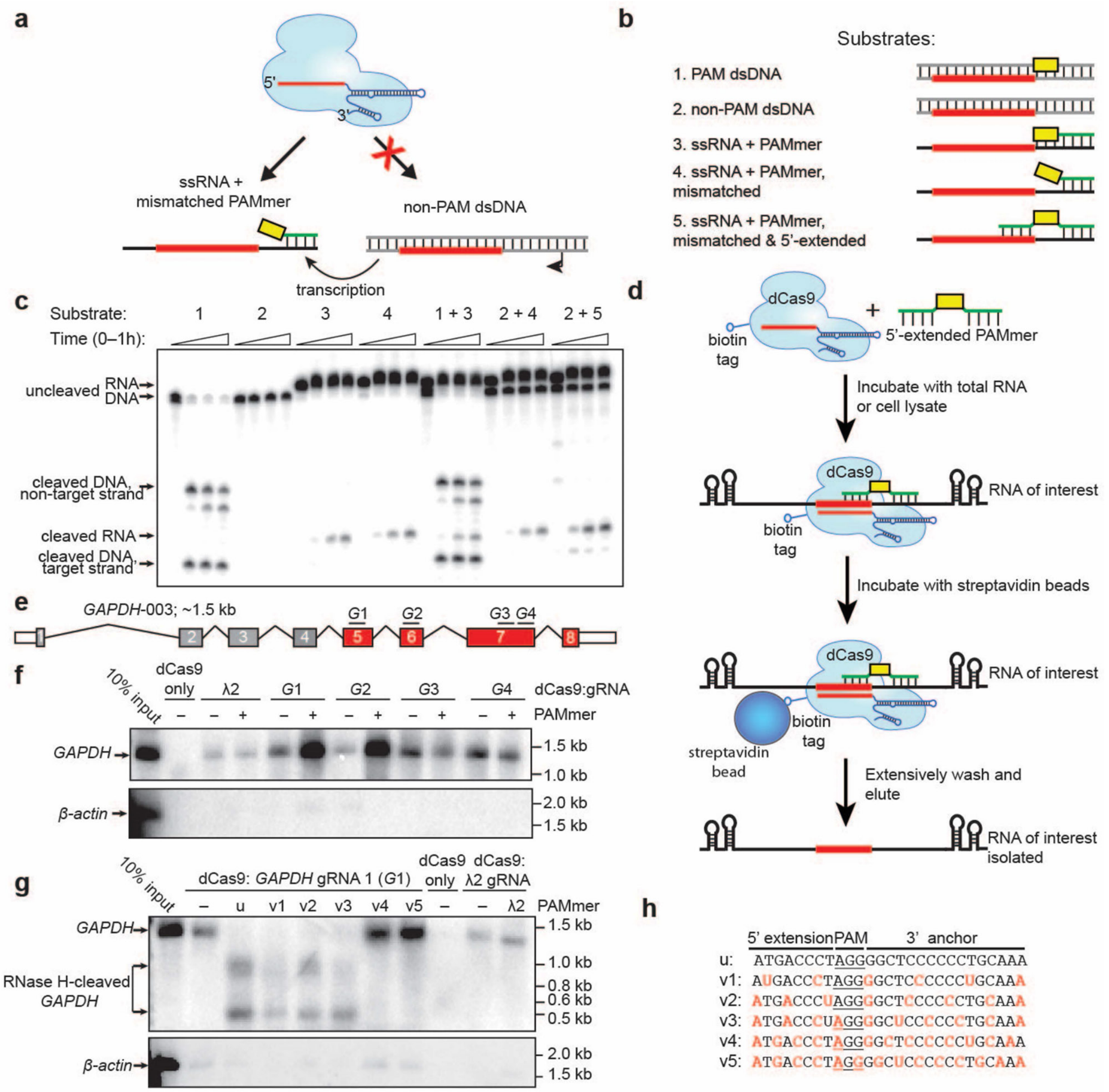
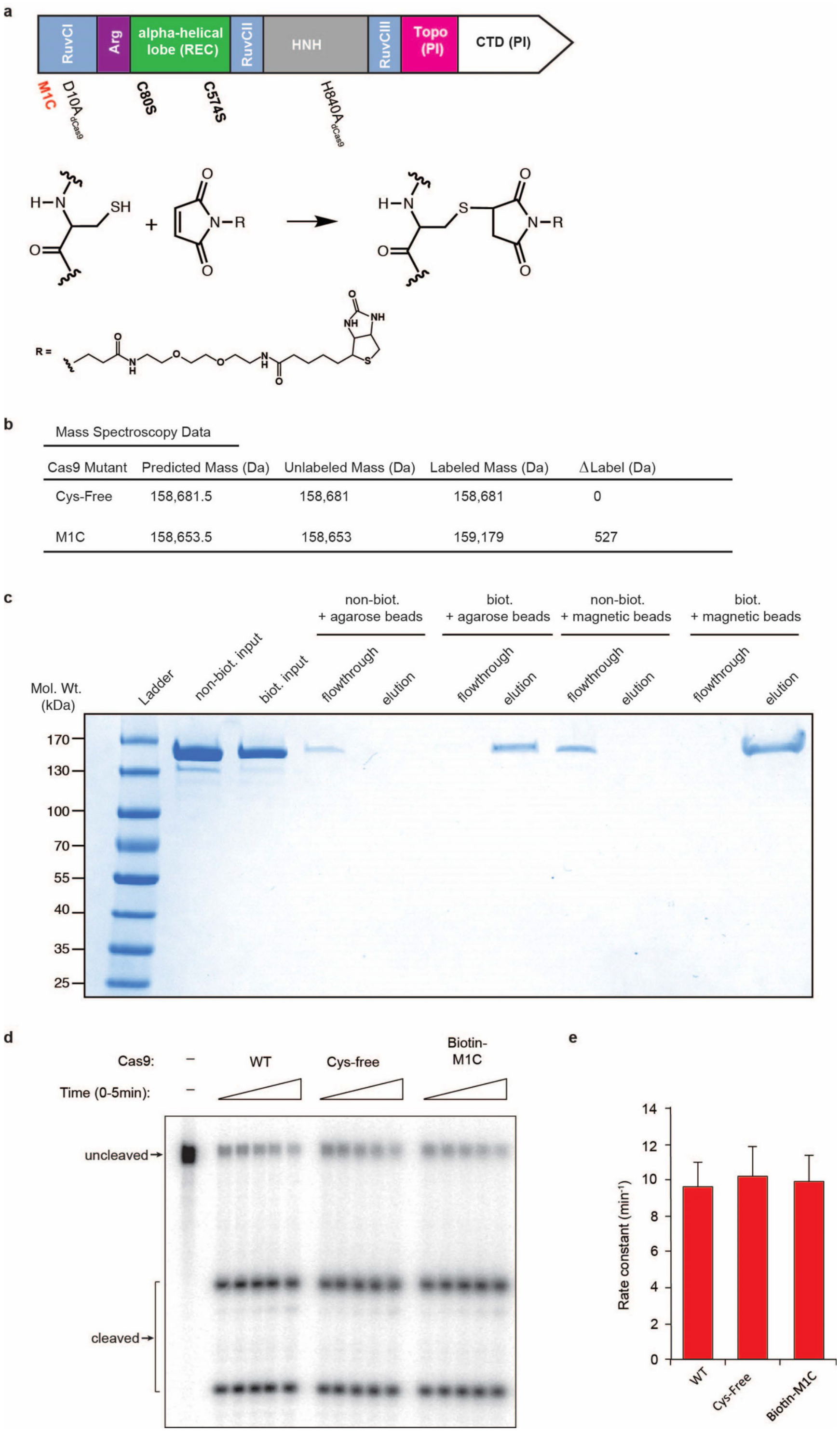


Figure 2.8 | RNA-guided Cas9 can target non-PAM sites on ssRNA and isolate *GAPDH* mRNA from HeLa cells in a tagless manner. (a) Schematic of the approach designed to avoid cleavage of template DNA by targeting non-PAM sites in the ssRNA target. (b) The panel of nucleic acid substrates tested in (c). (c) Cas9–gRNA cleaves ssRNA targets with equal efficiency when the 5'-NGG-3' of the PAMmer is mismatched with the ssRNA. This strategy enables selective cleavage of ssRNA in the presence of non-PAM target dsDNA; at cognate PAM sites, Cas9–gRNA cleaves both ssRNA and dsDNA. (d) Schematic of the dCas9 RNA pull-down expriment. (e), *GAPDH* mRNA transcript isoform 3 shown schematically, with exons common to all *GAPDH* protein-coding transcripts in red and gRNA/PAMmer targets 1-4 indicated. (f) Northern blot showing that gRNAs and 5'-extended PAMmers enable tagless isolation of *GAPDH* mRNA from HeLa total RNA; *β-actin* mRNA is shown as a control. (g) Northern blot showing tagless isolation of *GAPDH* mRNA from HeLa cell lysate with varying 2'-OMe-modified PAMmers. RNase H cleavage is abrogated with v4 and v5 PAMmers; *β-actin* mRNA is shown as a control. (h) Sequences of unmodified and modified *GAPDH* PAMmers used in (g); 2'-OMe-modified nucleotides are shown in red.



Mass Spectroscopy Data

Cas9 Mutant	Predicted Mass (Da)	Unlabeled Mass (Da)	Labeled Mass (Da)	ΔLabel (Da)
Cys-Free	158,681.5	158,681	158,681	0
M1C	158,653.5	158,653	159,179	527

Figure 2.9 | Site-specific biotin labeling of Cas9. (a) In order to introduce a single biotin moiety on Cas9, the solvent accessible, non-conserved N- terminal methionine was mutated to a cysteine (M1C; red text) and the naturally occurring cysteine residues were mutated to serine (C80S and C57S; bold text). This enabled cysteine-specific labeling with EZ-link Maleimide-PEG2-biotin through an irreversible reaction between the reduced sulfhydryl group of the cysteine and the maleimide group present on the biotin label. dCas9 mutations are also indicated in the domain schematic. (b) Mass spectrometry analysis of the Cas9 biotin labeling reaction confirmed

that successful biotin labeling only occurs when the M1C mutation is present in the Cys-Free background (C80S,C574S). The mass of the Maleimide- PEG2-biotin reagent is 525.6 Da. (c) Streptavidin bead binding assay with biotinylated (biot.) or non-biotinylated (non-biot.) Cas9 and streptavidin agarose or streptavidin magnetic beads. Cas9 only remains specifically bound to the beads after biotin labeling. (d) Cleavage assays were conducted as described in the Materials and Methods and resolved by denaturing PAGE. Reactions contained 100 nM Cas9 programmed with λ2-gRNA and ~1 nM 5′-radiolabelled λ2 dsDNA target. (e) Quantified cleavage data from triplicate experiments were fit with single-exponential decays to calculate the apparent pseudo-first order cleavage rate constants (average ± standard deviation). Both Cys-Free and Biotin-M1C Cas9 retain WT activity.

We then used this approach to isolate endogenous *GAPDH* transcripts from HeLa cell lysate under physiological conditions. In initial experiments, we found that Cas9–gRNA captured two *GAPDH*-specific RNA fragments rather than the full-length mRNA (Fig. 2.8g). Based on the sizes of these bands, we hypothesized that RNA:DNA heteroduplexes formed between the mRNA and PAMmer were cleaved by cellular RNase H. Previous studies have shown that modified DNA oligonucleotides can abrogate RNase H activity (Wu et al., 1999), and therefore we investigated whether Cas9 would tolerate chemical modifications to the PAMmer. We found that a wide range of modifications still enabled PAMmer-mediated nuclease activation, including locked nucleic acids, 2’-OMe and 2’-F ribose moieties (Fig. 2.10). Importantly, by varying the pattern of 2’-OMe modifications in the PAMmer, we could completely eliminate RNase H-mediated cleavage during the pull-down experiment and successfully isolate intact *GAPDH* mRNA (Fig. 2.8g,h). Interestingly, we consistently observed specific isolation of *GAPDH* mRNA in the absence of any PAMmer, albeit with lower efficiency, suggesting that Cas9–gRNA can bind to *GAPDH* mRNA through direct RNA:RNA hybridization (Fig. 2.8f,g & 2.11). Taken together, these experiments demonstrate that RNA-guided Cas9 can be used to purify endogenous untagged RNA transcripts. In contrast to current oligonucleotide-mediated RNA-capture methods, this approach works well under physiological salt conditions and doesn't require crosslinking or large sets of biotinylated probes (Chu et al., 2011; Engreitz et al., 2013; Simon et al., 2011).

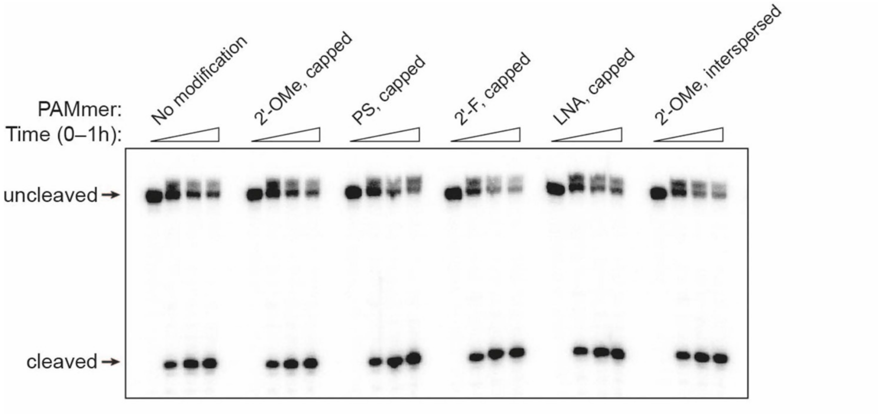


Figure 2.10 | RNA-guided Cas9 can utilize chemically modified PAMmers. 19-nt PAMmer derivatives containing various chemical modifications on the 5’ and 3’ ends (capped) or interspersed still activate Cas9 for cleavage of ssRNA targets. These types of modification are often used to increase the in vivo half-life of short oligonucleotides by preventing exo- and endonuclease-mediated degradation. Cleavage assays were conducted as described in the Methods. PS, phosphorothioate bonds; LNA, locked nucleic acid.

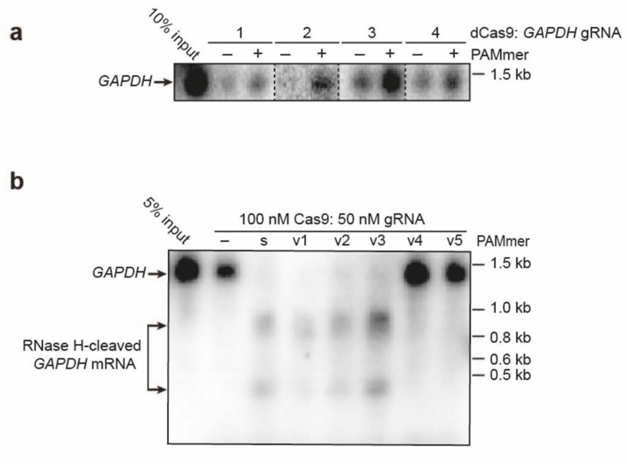


Figure 2.11 | Cas9 programmed with GAPDH-specific gRNAs can pull-down GAPDH mRNA in the absence of PAMmer. (a) Northern blot showing that, in some cases, Cas9-gRNA is able to pull down detectable amounts of GAPDH mRNA from total RNA without requiring a PAMmer. (b) Northern blot showing that Cas9-gRNA 1 is also able to pull-down quantitative amounts of GAPDH mRNA from HeLa cell lysate without requiring a PAMmer. s: standard; v: 2'-OMe-modified PAMmers.

Discussion

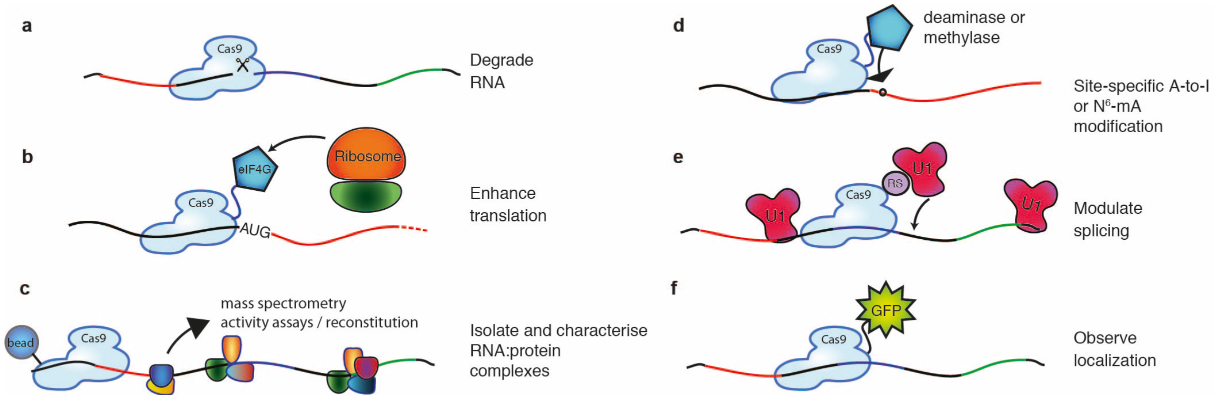


Figure 2.12 | Potential applications of RCas9 for untagged transcript analysis, detection, and manipulation. (a) Catalytically-active RCas9 could be used to target and cleave RNA, particularly those for which RNAi-mediated repression/degradation is not possible. (b) Tethering the eukaryotic initiation factor eIF4G to a catalytically inactive dRCas9 targeted to the 5' untranslated region of an mRNA could drive translation. (c) dRCas9 tethered to beads could be used to specificially isolate RNA or native RNA:protein complexes of interest from cells for downstream analysis or assays including identification of bound protein complexes, probing of RNA structure under native protein-bound conditions, and enrichment of rare transcripts for sequencing analysis. (d) dRCas9 tethered to RNA deaminase or N^6-mA methylase domains could direct site-specific A-to-I editing or methylation or RNA, respectively. (e) dRCas9 fused to a U1 recruitment domain (arginine- and serine- rich (RS) domain) could be programmed to recognize a splicing enhancer site and thereby promote the inclusion of a targeted exon. (f) dRCas9 tethered to a fluorescent protein such as GFP could be used to observe RNA localization and transport in living cells.

Here we have demonstrated the ability to re-direct the dsDNA targeting capability of CRISPR/Cas9 for RNA-guided ssRNA binding and/or cleavage (RCas9). Programmable RNA recognition and cleavage has the potential to transform the study of RNA function much as site-specific DNA targeting is changing the landscape of genetic and genomic research (Mali et al., 2013a) (Fig. 2.12). Although certain engineered proteins such as PPR proteins and Pumilio/FBF

(PUF) repeats show promise as platforms for sequence-specific RNA targeting (Filipovska and Rackham, 2011; Mackay et al., 2011; Wang et al., 2013b; Yagi et al., 2013; Yin et al., 2013), these strategies suffer from the need to re-design the protein for every new RNA sequence of interest. While RNA interference has proven useful for manipulating gene regulation in certain organisms (Kim and Rossi, 2008), there has been a strong motivation to develop orthogonal nucleic acid-based RNA recognition systems, such as the CRISPR/Cas Type III-B Cmr complex (Hale et al., 2009, 2012; Spilman et al., 2013; Staals et al., 2013; Terns and Terns, 2014) and the atypical Cas9 from *Francisella novicida* (Sampson and Weiss, 2014; Sampson et al., 2013). In contrast to these systems, the molecular basis for RNA recognition by RCas9 is now clear and requires only the design and synthesis of a matching gRNA and complementary PAMmer. The ability to recognize endogenous RNAs within complex mixtures with high affinity and in a programmable manner paves the way for direct transcript detection, analysis and manipulation without the need for genetically encoded affinity tags.

Materials and Methods

Cas9 and nucleic acid preparation

Wild-type Cas9 and catalytically inactive dCas9 (D10A/H840A) from *S. pyogenes* were purified as previously described (Jinek et al., 2012). crRNAs (42 nt) were either ordered synthetically (Integrated DNA Technologies) or transcribed *in vitro* with T7 polymerase using single-stranded DNA templates, as described (Sternberg et al., 2012). tracrRNA was transcribed *in vitro* and contained nucleotides 15–87 following the numbering scheme used previously (Jinek et al., 2012). λ-targeting sgRNAs were *in vitro* transcribed from linearized plasmids and contain full-length crRNA and tracrRNA connected via a GAAA tetraloop insertion. *GAPDH* mRNA-targeting sgRNAs were *in vitro* transcribed from dsDNA PCR products based on an optimized sgRNA design (Chen et al., 2013). Target ssRNAs (55–56 nt) were *in vitro* transcribed using single-stranded DNA templates. Sequences of all nucleic acid substrates used in this study can be found in Table 2.1.

All RNAs were purified using 10–15% denaturing polyacrylamide gel electrophoresis (PAGE). crRNA–tracrRNA duplexes were prepared by mixing equimolar concentrations of each RNA in hybridization buffer (20 mM Tris-HCl pH 7.5, 100 mM KCl, 5 mM $MgCl_2$), heating to 95 °C for 30 s and slow cooling. Fully double-stranded DNA/RNA substrates were prepared by mixing equimolar concentrations of each nucleic acid strand in hybridization buffer, heating to 95 °C for 30 s, and slow cooling. RNA, DNA, and chemically modified PAMmers were synthesized commercially (Intergrated DNA Technologies). DNA and RNA substrates were 5’-radiolabeled using [γ-^{32}P]-ATP (PerkinElmer) and T4 polynucleotide kinase (New England Biolabs). dsDNA and dsRNA substrates were 5’-radiolabeled on both strands, whereas only the target ssRNA was 5’-radiolabeled in other experiments.

Cleavage assays

Cas9–gRNA complexes were reconstituted before cleavage experiments by incubating Cas9 and the crRNA–tracrRNA duplex for 10 min at 37 °C in reaction buffer (20 mM Tris-HCl pH 7.5, 75 mM KCl, 5 mM $MgCl_2$, 1 mM dithiothreitol (DTT), 5% glycerol). Cleavage reactions were conducted at 37 °C and contained ~1 nM 5′-radiolabeled target substrate, 100 nM Cas9–RNA, and 100 nM PAMmer, where indicated. Aliquots were removed at each time point and quenched by the addition of RNA gel loading buffer (95% deionized formamide, 0.025% (w/v) bromophenol blue, 0.025% (w/v) xylene cyanol, 50 mM EDTA (pH 8.0), 0.025% (w/v) SDS). Samples were boiled for 10 min at 95 °C prior to being resolved by 12% denaturing PAGE. Reaction products were visualized by phosphorimaging and quantified with ImageQuant (GE Healthcare).

RNA cleavage site mapping

A hydrolysis ladder (OH^-) was obtained by incubating ~25 nM 5′-radiolabeled λ2 target ssRNA in hydrolysis buffer (25 mM CAPS (*N*-cyclohexyl-3-aminopropanesulfonic acid), pH 10.0, 0.25 mM EDTA) at 95 °C for 10 min, before quenching on ice. An RNase T1 ladder was obtained by incubating ~25 nM 5′-radiolabeled λ2 target ssRNA with 1 Unit RNase T1 (NEB) for 5 min at 37 °C in RNase T1 buffer (20 mM sodium citrate, pH 5.0, 1 mM EDTA, 2 M urea, 0.1 mg mL^{-1} yeast tRNA). The reaction was quenched by phenol/chloroform extraction before adding RNA gel loading buffer. All products were resolved by 15% denaturing PAGE.

Electrophoretic mobility shift assays

In order to avoid dissociation of the Cas9–gRNA complex at low concentrations during target ssRNA binding experiments, binding reactions contained a constant excess of dCas9 (300 nM), increasing concentrations of sgRNA, and 0.1–1 nM of target ssRNA. The reaction buffer was supplemented with 10 µg ml^{-1} heparin in order to avoid non-specific association of apo-dCas9 with target substrates (Sternberg et al., 2014). Reactions were incubated at 37 °C for 45 min before being resolved by 8% native PAGE at 4 °C (0.5× TBE buffer with 5 mM $MgCl_2$). RNA and DNA were visualized by phosphorimaging, quantified with ImageQuant (GE Healthcare), and analyzed with Kaleidagraph (Synergy Software).

Cas9 biotin labeling

To ensure specific labeling at a single residue on Cas9, two naturally occurring cysteine residues were mutated to serine (C80S and C574S) and a cysteine point mutant was introduced at residue M1. To attach the biotin moiety, 10 µM WT Cas9 or dCas9 was reacted with a 50-fold molar excess of EZ-Link Maleimide-PEG2-Biotin (Thermo Scientific) at 25 °C for 2 h. The reaction was quenched by the addition of 10 mM DTT, and unreacted Maleimide-PEG2-Biotin was removed using a Bio-Gel P-6 column (Bio-Rad). Labeling was verified using a streptavidin bead binding assay, where 8.5 pmol of biotinylated Cas9 or non-biotinylated Cas9 was mixed with either 25 µL streptavidin-agarose (Pierce Avidin Agarose; Thermo Scientific) or 25 µL streptavidin magnetic beads (Dynabeads MyOne Streptavidin C1; Life Technologies). Samples were incubated in Cas9 reaction buffer at RT for 30 minutes, followed by three washes with Cas9 reaction buffer and elution in boiling SDS-PAGE loading buffer. Elutions were analysed using SDS-PAGE. Cas9 M1C biotinylation was also confirmed using mass spectroscopy performed in the QB3/Chemistry Mass Spectrometry Facility at UC Berkeley. Samples of intact Cas9 proteins were analyzed using an Agilent 1200 liquid chromatograph equipped with a Viva C8 (100 mm × 1.0 mm, 5 µm particles, Restek) analytical column and connected in-line with an LTQ Orbitrap XL mass spectrometer (Thermo Fisher Scientific). Mass spectra were recorded in the positive ion mode. Mass spectral deconvolution was performed using ProMass software (Novatia).

GAPDH mRNA pull-down

HeLa-S3 cell lysates were prepared as previously described (Lee et al., 2013). Total RNA was isolated from HeLa-S3 cells using Trizol reagent according to the manufacturer’s instructions (Life Technologies). Cas9–sgRNA complexes were reconstituted before pull-down experiments by incubating a two-fold molar excess of Cas9 with sgRNA for 10 min at 37 °C in reaction buffer. HeLa total RNA (40 µg) or HeLa lysate (~5×10^6 cells) was added to reaction buffer with 40U RNasin (Promega), PAMmer (5 µM) and the biotin-dCas9 (50 nM):sgRNA (25 nM) in a total volume of 100 µL and incubated at 37 °C for 1 h. This mixture was then added to 25 µL magnetic streptavidin beads (Dynabeads MyOne Streptavidin C1; Life Technologies) pre-equilibrated in reaction buffer and agitated at 4 °C for 2 h. Beads were then washed six times with 300 µL wash buffer (20 mM Tris-HCl pH 7.5, 150 mM NaCl, 5mM $MgCl_{2,}$ 0.1% Triton X-100, 5% glycerol, 1mM DTT, 10 µg ml^{-1} heparin). Immobilized RNA was eluted by heating beads at 70 °C in the presence of DEPC-treated water and a phenol/chloroform mixture. Eluates were then treated with an equal volume of glyoxal loading dye (Life Technologies) and heated at 50 °C for 1 h before separation via 1% BPTE agarose gel (30 mM Bis-Tris, 10 mM PIPES, 10 mM EDTA, pH 6.5). Northern blot transfers were carried out according to Chomczynski *et al.* (Chomczynski, 1992). Following transfer, membranes were crosslinked using UV radiation and incubated in pre-hybridization buffer (UltraHYB Ultrasensitive Hybridization Buffer; Life Technologies) for 1 h at 46 °C prior to hybridization. Radioactive Northern probes were synthesized using random priming of *GAPDH* and *β-actin* partial cDNAs (for cDNA primers, see Table 6.1) in the presence of [α-^{32}P]-dATP (PerkinElmer), using a Prime-It II Random Primer Labeling kit (Agilent Technologies). Hybridization was carried out for 3 h in pre-hybridization buffer at 46 °C followed by two washes with 2×SSC (300mM NaCl, 30 mM trisodium citrate, pH 7, 0.5% (w/v) SDS) for 15 min at 46 °C. Membranes were imaged using a phosphorscreen.

Table 2.1 | RNA and DNA substrates used in this study

Description	Sequence[a]
Oligo for preparing dsDNA T7 promoter, in vitro transcription	5’-TAATACGACTCACTATA-3’
λ2-targeting crRNA	5’-GUGAUAAGUGGAAUGCCAUGGUUUUAGAGCUAUGCUGUUUUG-3’
λ3-targeting crRNA	5’-CUGGUGAACUUCCGAUAGUGGUUUUAGAGCUAUGCUGUUUUG-3’
λ4-targeting crRNA	5’-CAGATATAGCCTGGTGGTTCGUUUUAGAGCUAUGCUGUUUUG-3’
ssDNA T7 template[b]: tracrRNA	5’- AAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTT ATTTTAACTTGCTATGCTG TCCTATAGTGAGTCGTATTA
tracrRNA (nt 15-87)	GGACAGCAUAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAA AGUGGCACCGAGUCGGUGCUUUUU
λ2-targeting sgRNA T7 template[c]	5’TAATACGACTCACTATAGGTGATAAGTGGAATGCCATGGTTTTAGAGCTA TGCTGTTTTGGAAACAAAACAGCATAGCAAGTTAAAATAAGGCTAGTCCGT TATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT-3’
λ2-targeting sgRNA	5’- GGUGAUAAGUGGAAUGCCAUGGUUUUAGAGCUAUGCUGUUUUGGAAACA AAACAGCAUAGCAAGUUAAAA
λ2 target dsDNA duplex	5'- GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAATT GAGG-3'

λ2 ssDNA target strand (used to make heteroduplex DNA:RNA)	3'-CTCACCTTCCTACGGTCACTATTCACCTTACGGTACACCCGACAGTTTTAACTCG-5'
λ2 ssDNA non-target strand (used to make heteroduplex DNA:RNA)	5'-GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGC-3'
λ2 ssRNA target strand T7 template	5’-GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGCCTATAGTGAGTCGTA TTA-3’
λ2 ssRNA target strand	3'-CUCACCUUCCUACGGUCACUAUUCACCUUACGGUACACCCGACAGUUUUAACUCGG-5'
λ2 ssRNA non-target strand T7 template	5’-GCTCAATTTTGACAGCCCACATGGCATTCCACTTATCACTGGCATCCTTCCACTCCTATAGTGAGTCGTA TTA-3’
λ2 ssRNA non-target strand (used to make dsRNA)	5’-GGAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGC-3’
19 nt λ2 DNA PAMmer	5'-TGGGCTGTCAAAATTGAGC-3'
18 nt λ2 “GG” PAMmer	5'-GGGCTGTCAAAATTGAGC-3'
19 nt λ2 DNA mutated PAMmer	5'-ACCGCTGTCAAAATTGAGC-3'
16 nt λ2 DNA “PAM-less” PAMmer	5'-GCTGTCAAAATTGAGC-3'
18 nt λ2 RNA PAMmer	5'-GGGCUGUCAAAAUUGAGC-3'
5 nt λ2 DNA PAMmer	5'-TGGGC-3'
10 nt λ2 DNA PAMmer	5'-TGGGCTGTCA-3'
15 nt λ2 DNA PAMmer	5'-TGGGCTGTCAAAATT-3'
λ3 ssRNA target strand T7 template	5’-AACGTGCTGCGGCTGGCTGGTGAACTTCCGATAGTGCGGGTGTTGAATGATTTCCTATAGTGAGTCGTAT TA-3’
λ3 ssRNA target strand	3'-UUGCACGACGCCGACCGACCACUUGAAGGCUAUCACGCCCACAACUUACUAAAGG-5'
λ4 ssRNA target strand T7 template	5’-TCACAACAATGAGTGGCAGATATAGCCTGGTGGTTCAGGCGGCGCATTTTTATTGCCTATAGTGAGTCGT ATTA-3’
λ4 ssRNA target strand	3'-AGUGUUGUUACUCACCGUCUAUAUCGGACCACCAAGUCCGCCGCGUAAAAAUAACGG-5'
λ3 ssDNA non-target strand	5'-AACGTGCTGCGGCTGGCTGGTGAACTTCCGATAGTGCGGGTGTTGAATGAT

	TTCC-3'
λ4 ssDNA non-target strand	5'-TCACAACAATGAGTGGCAGATATAGCCTGGTGGTTCAGGCGGCGCATTTTTATTG-3'
19 nt λ3 DNA PAMmer	5’-CGGGTGTTGAATGATTTCC-3’
19 nt λ4 DNA PAMmer	5’-AGGCGGCGCATTTTTATTG-3’
21 nt λ2 5’-extended DNA PAMmer	5’-TGTGGGCTGTCAAAATTGAGC-3’
21 nt λ3 5’-extended DNA PAMmer	5’-TGCGGGTGTTGAATGATTTCC-3’
24 nt λ2 5’-extended DNA PAMmer	5’-CCATGTGGGCTGTCAAAATTGAGC-3’
24 nt λ3 5’-extended DNA PAMmer	5’-TAGTGCGGGTGTTGAATGATTTCC-3’
27 nt λ2 5’-extended DNA PAMmer	5’-ATGCCATGTGGGCTGTCAAAATTGAGC-3’
27 nt λ3 5’-extended DNA PAMmer	5’-CGATAGTGCGGGTGTTGAATGATTTCC-3’
30 nt λ2 5’-extended DNA PAMmer	5’-GGAATGCCATGTGGGCTGTCAAAATTGAGC-3’
30 nt λ3 5’-extended DNA PAMmer	5’-TTCCGATAGTGCGGGTGTTGAATGATTTCC-3’
33 nt λ2 5’-extended DNA PAMmer	5’-AGTGGAATGCCATGTGGGCTGTCAAAATTGAGC-3’
33 nt λ3 5’-extended DNA PAMmer	5’-AACTTCCGATAGTGCGGGTGTTGAATGATTTCC-3’
36 nt λ2 5’-extended DNA PAMmer	5’-ATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGC-3’
39 nt λ2 5’-extended DNA PAMmer	5’-GTGATAAGTGGAATGCCATGTGGGCTGTCAAAATTGAGC-3’
39 nt λ3 5’-extended DNA PAMmer	5’-CTGGTGAACTTCCGATAGTGCGGGTGTTGAATGATTTCC-3’
non-PAM λ2 dsDNA	5'-GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGACCGCTGTCAAAATTGAGC-3' 3'-CTCACCTTCCTACGGTCACTATTCACCTTACGGTACTGGCGACAGTTTTAACTCG-5'
non-PAM λ2 ssRNA target strand T7 template	5’-GAGTGGAAGGATGCCAGTGATAAGTGGAATGCCATGACCGCTGTCAAAATTGAGCCTATAGTGAGTCGTA TTA-3’
non-PAM λ2 ssRNA target strand	3'-CUCACCUUCCUACGGUCACUAUUCACCUUACGGUACTGGCGACAGUUUUAACUCGG-5'
λ2 2’OMe capped PAMmer^{d}	5’-*UGGGCTGTCAAAATTGAG*C-3’
λ2 PS capped PAMmer^{d}	5’-T*GGGCTGTCAAAATTGAG*C-3’
λ2 2’F capped PAMmer^{d}	5’-*UGGGCTGTCAAAATTGAG*C-3’
λ2 LNA capped PAMmer^{d}	5’-*TGGGCTGTCAAAATTGAG*C-3’

λ2 19 nt 2’OMe interspersed PAMmer[d]	5’-*UGGGC*UGTCA*AAATT*GAG*C-3’
GAPDH-targeting sgRNA 1 T7 template[e]	5’-TAATACGACTCACTATAGGGGCAGAGATGATGACCCTGTTTAAGAGCTATGCTGGAAACAGCATAGCAAG TTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT-3’
GAPDH-targeting sgRNA 1	5’-GGGGCAGAGAUGAUGACCCUGUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAGUC CGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU-3’
GAPDH-targeting sgRNA 2 T7 template[e]	5’-TAATACGACTCACTATAGGCCAAAGTTGTCATGGATGACGTTTAAGAGCTATGCTGGAAACAGCATAGCA AGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT-3’
GAPDH-targeting sgRNA 2	5’-GGCCAAAGUUGUCAUGGAUGACGUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU-3’
GAPDH-targeting sgRNA 3 T7 template[e]	5’-TAATACGACTCACTATAGGCCAAAGTTGTCATGGATGACGTTTAAGAGCTATGCTGGAAACAGCATAGCA AGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT-3’
GAPDH-targeting sgRNA 3	5’-GGAUGUCAUCAUAUUUGGCAGGGUUUAAGAGCUAUGCUGGAAACAGCAUAGCAAGUUUAAAUAAGGCUAG UCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUUU-3’
GAPDH-targeting sgRNA 4 T7 template[e]	5’-TAATACGACTCACTATAGGATGTCATCATATTTGGCAGGGTTTAAGAGCTATGCTGGAAACAGCATAGCA AGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT-3’
GAPDH-targeting sgRNA 4	5’-GGATGTCATCATATTTGGCAGGGTTTAAGAGCTATGCTGGAAACAGCATAGCAAGTTTAAATAAGGCTAG TCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTT-3’
GAPDH PAMmer 1	5’-ATGACCCTTGGGGCTCCCCCTGCAAA-3’
GAPDH PAMmer 2	5’-TGGATGACCGGGGCCAGGGGTGCTAAG-3’
GAPDH PAMmer 3	5’-TTGGCAGGTGGTTCTAGACGGCAGGTC-3’
GAPDH PAMmer 4	5’-CCCCAGCGTGGAAGGTGGAGGAGTGGG-3’
GAPDH PAMmer 1 2’OMe v1	5’-A*UGACC*CTAGG*GGCTC*CCCCC*UGCAA*A-3’
GAPDH PAMmer 1 2’OMe v2	5’-*ATG*ACCC*UAGG*GGCT*CCCC*CCTG*CAA*A-3’
GAPDH PAMmer 1 2’OMe v3	5’-*ATG*ACC*CU*AGG*GGC*UCC*CCC*CTG*CAA*A-3’
GAPDH PAMmer 1 2’OMe v4	5’-*AT*GA*CC*CT*AGG*GG*CT*CC*CC*CC*UG*CA*AA-3’
GAPDH PAMmer 1 2’OMe v5	5’-*AT*GA*CC*CT*AG*GG*GC*TC*CC*CC*CU*GC*AA*A-3’
GAPDH cDNA primer Fwd	5’-CTCACTGTTCTCTCCCTCCGC-3’

GAPDH cDNA primer Rev	5’-AGGGGTCTACATGGCAACTG-3’
β-actin cDNA primer Fwd	5’-AGAAAATCTGGCACCACACC-3’
β-actin cDNA primer Rev	5’-GGAGTACTTGCGCTCAGGAG-3’

[a] Guide crRNA sequences and complementary DNA target strand sequences are shown in red. PAM sites (5'-NGG-3') are highlighted in yellow on the non-target strand when adjacent to the target sequence or in the PAMmer oligonucleotides.

[b] The T7 promoter is indicated in bold (or reverse complement of), as well as 5’ G or GG included in the ssRNA product by T7 polymerase.

NA, not applicable.

[c] sgRNA template obtained in pIDT, subsequently linearised by AflII for run-off transcription.

[d] Positions of modifications depicted with asterisks preceding each modified nucleotide in each case (except for PS linkages which are depicted between bases)

PS: phosphorothioate bond

LNA: locked nucleic acid

[e] sgRNAs for *GAPDH* were designed according to (Chen et al., 2013).

Chapter 3

Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch

† The work presented in this chapter has previously been published in the following research article: Oakes, B.L., Nadler, D.C., Flamholz, A., Fellmann, C., Staahl, B.T., Doudna, J.A., and Savage, D.F. (2016). Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch. Nat. Biotechnol. vol: 34, 646–651.

‡ Benjamin Oakes & David Savage conceived this project. Benjamin Oakes, Dana Nadler, Christof Fellmann, Brett Staahl and David Savage designed the research. Benjamin Oakes, Dana Nadler & Christof Fellmann, performed the experiments. Abraham Flamholz performed the computational analysis. Benjamin Oakes, Abraham Flamholz, Christof Fellmann, and David Savage analyzed the data. Benjamin Oakes, Christof Fellmann, Jennifer Doudna and David Savage wrote the paper. Jennifer Doudna and David Savage provided funding support.

Abstract:

The CRISPR-associated protein Cas9 from *Streptococcus pyogenes* is an RNA-guided DNA endonuclease with widespread utility for genome modification. However, the structural constraints limiting the engineering of Cas9 have not been determined. Here we experimentally profile Cas9 using randomized insertional mutagenesis and delineate hotspots in the structure capable of tolerating insertions of a PDZ domain without disrupting the enzyme’s binding and cleavage functions. Orthogonal domains or combinations of domains can be inserted into the identified sites with minimal functional consequence. To illustrate the utility of the identified sites, we construct an allosterically regulated Cas9 by insertion of the Estrogen Receptor α Ligand Binding Domain. This protein displayed robust, ligand-dependent_activation in prokaryotic and eukaryotic cells, establishing a versatile one-component system for inducible and reversible Cas9 activation. Thus, domain insertion profiling facilitates the rapid generation of new Cas9 functionalities and provides useful data for future engineering of Cas9.

Introduction:

Domain insertions—the transfer of coding sequences for unique folds from one open reading frame (ORF) to another—are common occurrences throughout protein evolution(Consortium et al., 2001; Peisajovich et al., 2010). These events are known to facilitate the diversification and rapid functionalization of new types of protein modules that fill needed roles in both eukaryotic and prokaryotic genomes (Chothia, 2003). In this study we investigated how synthetic domain insertions can be used to functionalize the programmable endonuclease Cas9 from *S. pyogenes* (hereafter Cas9).

Cas9 is an RNA-guided, DNA-binding and cleaving protein that has been adapted to enable the facile modification or perturbation of genes and regulatory and non-coding genomic elements in a wide variety of organisms (Doudna and Charpentier, 2014; Hsu et al., 2014; Jinek et al., 2012). Recently, there have been numerous attempts to develop Cas9 variants with additional functions by fusing protein domains directly to its N- or C-terminus(Chen et al., 2013; Gilbert et al., 2013; Guilinger et al., 2014). However, the N- and C-termini of Cas9 are within ~40 Å of each other, leaving a large fraction of the protein structure unexplored by termini fusions(Anders et al., 2014; Nishimasu et al., 2014). This close spatial proximity can lead to steric incompatibility and may explain a relative lack of activity for many fusions, such as with VP64(Tanenbaum et al., 2014). Another solution would be to identify insertion points within Cas9 that are capable of accepting functional domains, as occurs naturally in evolution(Chothia, 2003; Consortium et al., 2001; Peisajovich et al., 2010). This strategy would allow the engineering of complex functionalities. For example, an allosterically regulated Cas9 would permit conditional control of activity and allow precise interrogation of development, disease progression and differentiation(Davis et al., 2015; Doudna and Charpentier, 2014; Hsu et al., 2014; Reynolds et al., 2011).

Results:

Here, we profiled the inherent plasticity of the Cas9 structure by examining its ability to tolerate a synthetic domain insertion while retaining RNA-guided DNA binding activity. An unbiased Cas9 insertion library was created using randomized transposition (Fig. 3.1A). Briefly, an engineered Mu transposon possessing an antibacterial selection marker flanked by BsaI endonuclease sites was inserted randomly throughout a catalytically inactive Cas9 (dCas9)-containing plasmid by *in vitro* transposition (Fig. 3.1A , Supplementary Fig. 1) (Edwards et al., 2008). After selection and sub-cloning to isolate plasmids with single transposition events within the dCas9 ORF, the library was characterized by deep sequencing. This analysis revealed that the transposition library possessed good insertion coverage, with insertions at >70% of all possible amino acid (AA) sites observed at least once (Supplementary Fig. 2). Once isolated, this library was used to construct specific domain insertion libraries by cleavage with BsaI and re-ligation with DNA fragments containing an ORF of interest. (Fig. 3.1A, Supplementary Fig. 1, Supplementary Fig. 2).

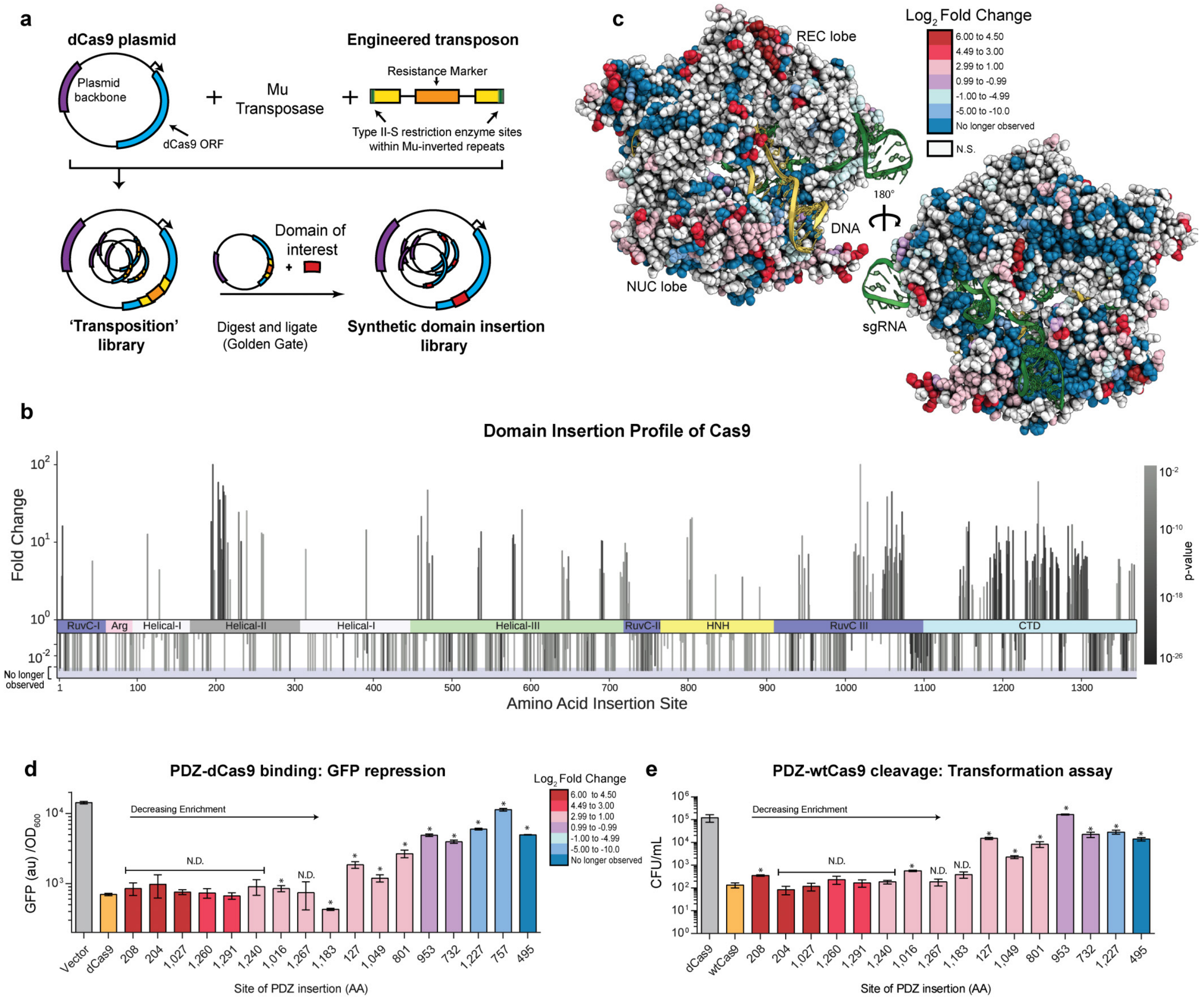


Figure 3.1 | Mapping the insertion potential of Cas9 with the Alpha-Syntrophin PDZ protein interaction domain. (a) Generation of the transposon-based domain insertion library. (b) Fold change values for insertions at specific amino acid sites derived from sequencing data over two rounds of screening. A positive value indicates the preference of the domain insertion at a site to remain in the library after screening for function. A negative value indicate a loss of the clone with an insertion at the site. More significant P values (DESeq, multiple hypothesis testing corrected) are represented as darker color bars. Positive values which attain 102 represent sites that were not sequenced before screening. Negative bars that extend into the shaded region represent clones which have been cleared from the library (i.e. not observed after screening). (c) Log2 fold change values from (b) mapped onto the structure of Cas9 (PDB ID:4UN3). (d) GFP repression activity of individual PDZ insertion sites. Values represent biological replicates with standard deviation (n=3), constructs are in order of decreasing fold change from sequencing. Positive and negative controls dCas9 and vector only are colored orange and grey, respectively. N.D. stands for no difference detected from dCas9, t-test ($p > 0.01$), (* = $p < 0.01$) (e) Cleavage activity of clones via an E. coli based transformation assay. Cas9 activity results in genomic cleavage and lower CFU/mL, values represent biological triplicates with standard deviation

(n=3). Positive and negative controls Cas9 and dCas9 are colored orange and grey, respectively. N.D. stands for no difference detected from WT Cas9, t-test (p > 0.01), (* = p < 0.01)

S1.

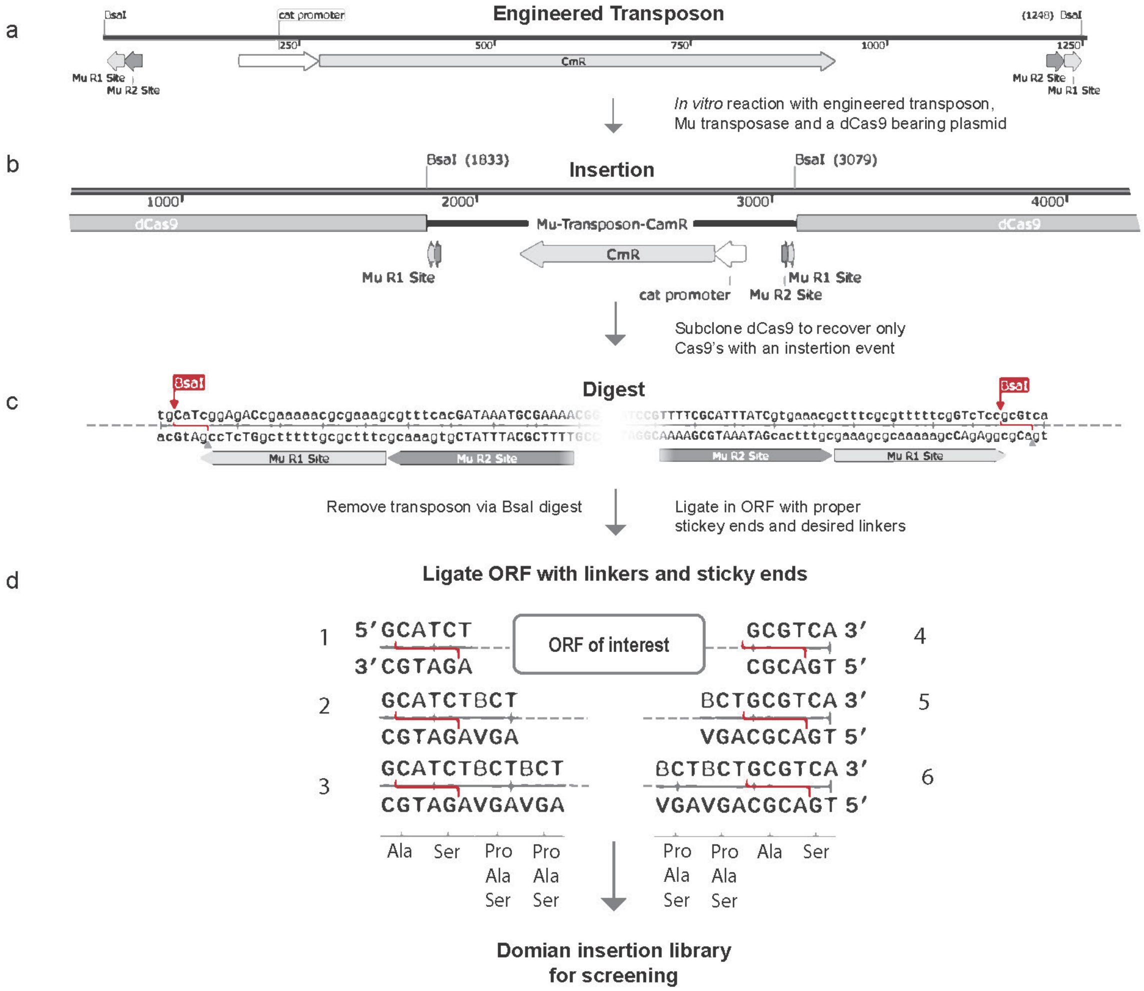

a
BsaI
cat promoter
Engineered Transposon
(1248) BsaI
250
500
750
1000
1250
CmR
Mu R1 Site
Mu R2 Site
Mu R2 Site
Mu R1 Site
In vitro reaction with engineered transposon, Mu transposase and a dCas9 bearing plasmid
b
BsaI (1833)
Insertion
BsaI (3079)
1000
2000
3000
4000
dCas9
Mu-Transposon-CamR
dCas9
CmR
Mu R1 Site
cat promoter
Mu R1 Site
Mu R2 Site
Subclone dCas9 to recover only Cas9's with an instertion event
c
BsaI
Digest
BsaI
tgCaTcggAgACcgaaaaacgcgaaagcgtttcacGATAAATGCGAAAACGG
TCCGTTTTCGCATTTATCgtgaaacgctttcgcgtttttcgGTcTccgcGtca
acGtAgccTcTGgctttttgcgctttcgcaaagtgCTATTTACGCTTTTGCC
AGGCAAAAGCGTAAATAGcactttgcgaaagcgcaaaaagcCAgAggcgCagt
Mu R1 Site
Mu R2 Site
Mu R2 Site
Mu R1 Site
Remove transposon via BsaI digest
Ligate in ORF with proper stickey ends and desired linkers
d
Ligate ORF with linkers and sticky ends
1
5'GCATCT
3'CGTAGA
ORF of interest
GCGTCA 3'
CGCAGT 5'
4
2
GCATCTBCT
CGTAGAVGA
BCTGCGTCA 3'
VGACGCAGT 5'
5
3
GCATCTBCTBCT
CGTAGAVGAVGA
BCTBCTGCGTCA 3'
VGAVGACGCAGT 5'
6
Ala
Ser
Pro Ala Ser
Pro Ala Ser
Pro Ala Ser
Pro Ala Ser
Ala
Ser
Domian insertion library for screening

Figure 3.S1 | Construction of a transposition library (a) Schematic of the engineered Mu transposon used for transposition. (b) Schematic showing the inserted transposon. (c) ORF excision of an intra-Cas9 transposon using Type II-S restriction enzyme BsaI. Any ORF can be subsequently ligated into these sticky ends. (d) Description of the sticky ends and linkers used to ligate in the PDZ and ER-LBD domains. Ala and Ser are hardcoded on each side providing the correct sticky ends for a Golden Gate ligation and additional diversity is provided by BCT codons (encoding Ala, Ser, or Pro). In total, there are 13 possible amino acid variations per terminus

S2

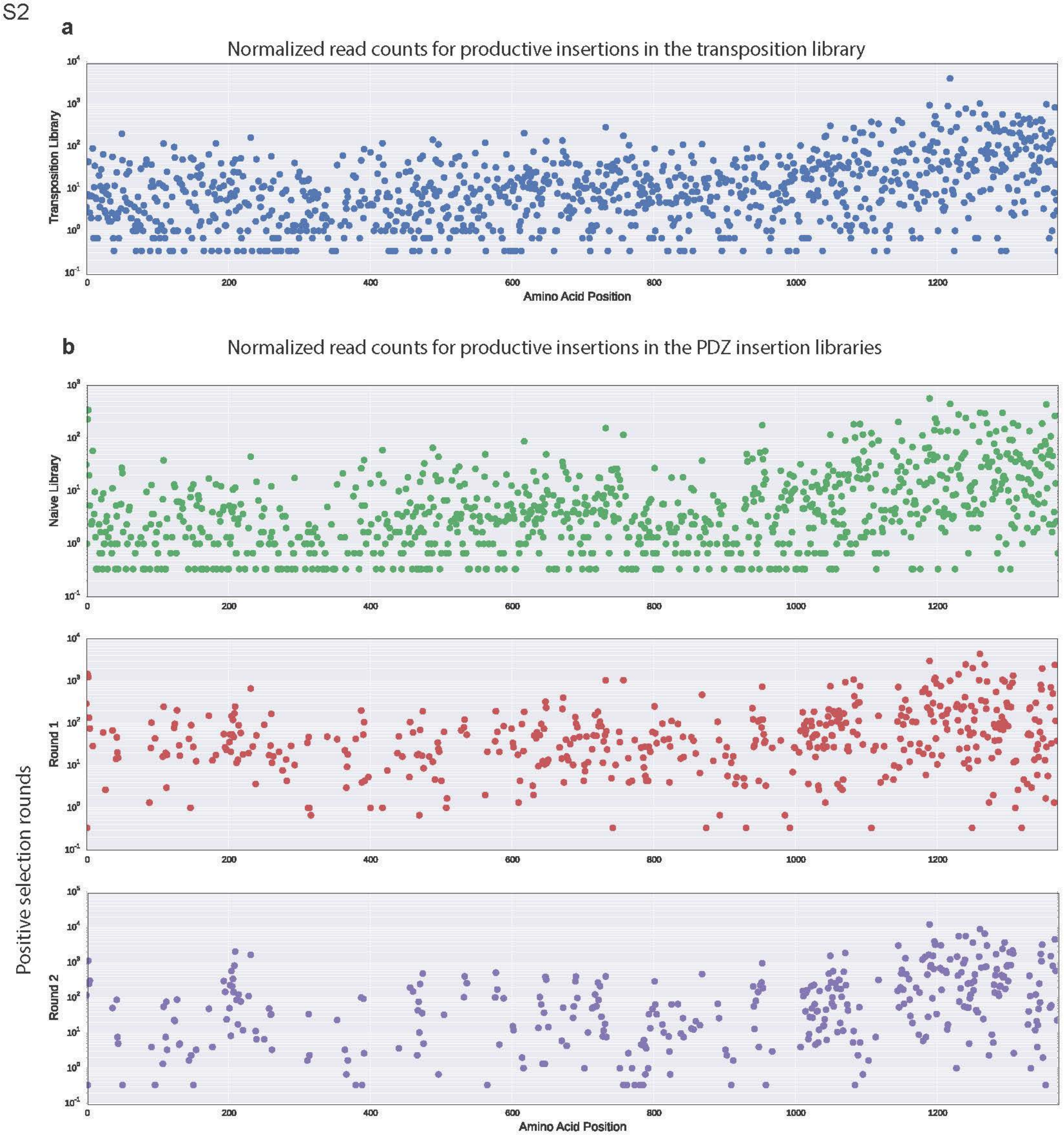


Figure 3.S2 | Deep sequencing of the transposition, naïve PDZ and selected PDZ libraries. (a) Sequencing and alignment of the transposon insertion library indicates coverage across the Cas9 coding sequence with a bias for insertion towards the C-terminus. We observe 973 productive insertions covering 71% of all possible sites. (b) Deep sequencing of the PDZ libraries, naïve (pre-screened) and post one and two rounds of CRISPRi screening. Sequencing of the naïve PDZ library indicates that good coverage of the protein is maintained upon cloning in of the PDZ domain with 953 productive insertions observed. Upon screening many clones are cleared out of the library while a smaller number are enriched. All read counts represent the corrected replicated averages generated from DESeq.

S3

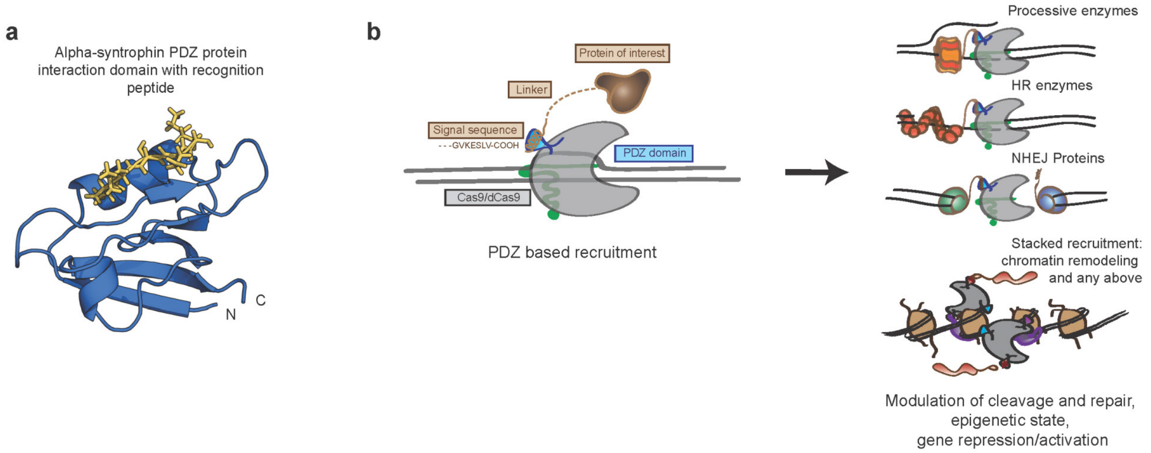

a
Alpha-syntrophin PDZ protein interaction domain with recognition peptide
N
C
b
Protein of interest
Linker
Signal sequence
- - -GVKESLV-COOH
PDZ domain
Cas9/dCas9
PDZ based recruitment
Processive enzymes
HR enzymes
NHEJ Proteins
Stacked recruitment: chromatin remodeling and any above
Modulation of cleavage and repair, epigenetic state, gene repression/activation

Figure 3.S3 | The PDZ domain as a potential scaffolding element (a) The Alpha-Syntrophin PDZ protein interaction domain. The adjacent N- and C-termini and its peptide ligand are depicted (PDB ID: 2PDZ) (b) PDZ based recruitment. The PDZ domain is a protein interaction domain that specifically recognizes a seven amino acid C-terminal motif that can be modularly attached to any protein of interest. This provides a mechanism by which it is possible to recruit one or many different proteins to a cleavage site or binding site to increase the local concentration. This may allow for the recruitment of proteins that may be processive (such as helicases), DNA repair enzymes (such as HR and NHEJ machinery, Rad51 and Ku70/80), activators/repressor or epigenetic modifying machinery, and even libraries of multiple protein domains fused to the PDZ recruitment amino acid sequence.

S4

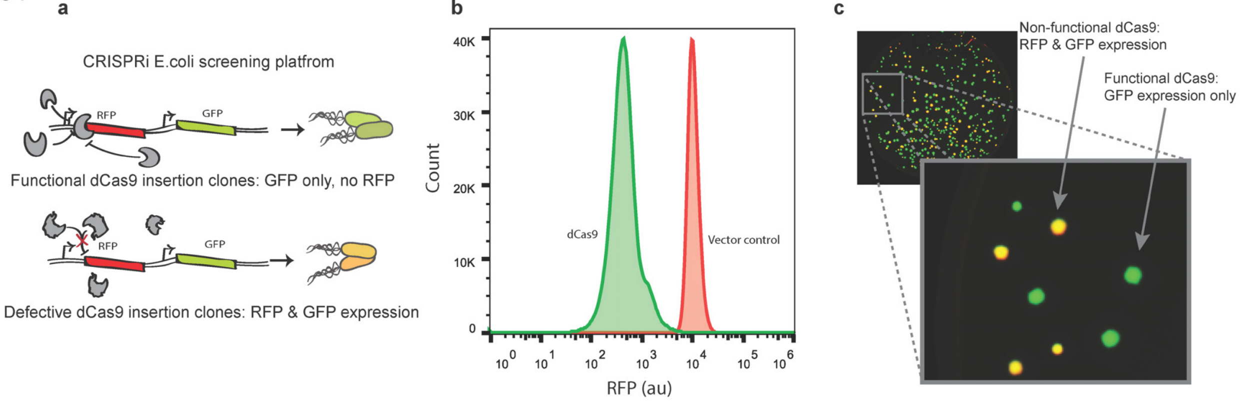


Figure 3.S4 | CRISPRi screening protocol controls (a) Schematic of the E. coli screening platform for determining DNA-binding competent dCas9 insertion mutants. (b) Flow cytometry of the RFP repression by dCas9. A 20 fold change in RFP fluorescence when dCas9 is present for 6+ hours similar to the results reported in Qi et al. 20132, this provides a clear signal by which to select functional Cas9 insertion mutants (c) On-plate screening of RFP repression by dCas9. The lack of RFP signal is visible by eye when screening colonies after overnight growth on plates.

The mouse alpha-1-syntrophin, Snta1, PDZ domain (PDZ) was chosen as a proof of concept insertion domain due to its small size (86 amino acids), well-folded nature and adjacent N- and C-termini (Supplementary Fig 3a)(Dueber et al., 2009; Schultz et al., 1998). We hypothesized that this domain would be minimally perturbative and act as a molecular 'potentiometer'— where the capacity to accommodate a PDZ insertion is indicative of the general insertion potential of a given amino acid site within Cas9. Moreover, as the PDZs are known protein-protein interaction domains, PDZ-Cas9s identified here may be further used as protein scaffolds to recruit other domains for editing, epigenetic modification and activation or repression purposes(Dueber et al., 2009) (Supplementary Fig 3b).

A PDZ domain with flanking amino acid linkers was cloned into the naïve library (Supplementary Fig 1, Supplementary Fig. 2) and passaged through two rounds of a CRISPRi screen (Oakes et al., 2014; Qi et al., 2013). Briefly, cells expressing Red Fluorescent Protein (RFP) and Green Fluorescent Protein (GFP) were assayed using fluorescence activated cell-sorting (FACS) to identify Cas9 variants capable of repressing RFP in a single guide RNA (sgRNA)-dependent fashion (Supplementary Fig 4)(Oakes et al., 2014; Qi et al., 2013). After sorting, the dCas9-PDZ libraries were subjected to deep sequencing to identify insertion sites. Comparing the transposition and PDZ libraries revealed that while the unscreened PDZ library was enriched in out-of-frame and reverse insertions, screening for dCas9-mediated gene repression increased the fraction of in-frame insertions by ~20 fold (Supplementary Fig 5). We refer to such insertions as 'productive' because they produce a full-length insertion protein. Thus, the screen enriched for productive PDZ-dCas9 insertion constructs.

Calculation of the $\log_2$-fold enrichment of insertion sites between the unscreened and final PDZ insertion library revealed statistically significant ($p < 0.1$)changes for roughly half of the amino acid (AA) sites in Cas9 (Fig. 3.1B, C Supplementary Table 1). Domain insertions at the majority of residues within Cas9 are strongly selected against, with the bulk of clones falling out of the pool by the second round of screening (Fig 3.1B, C Supplementary Table 1). Sites with negative fold changes were highly overrepresented in critical motifs such as the globular core of Cas9, sgRNA-binding grooves, the bridge helix, the PAM-binding pocket and the DNA:RNA heteroduplex annealing channel (Fig. 3.1C, Supplementary Fig. 6). Nevertheless, we identified small local clusters of amino acids that are tolerant to insertions and recovered a total of 175 statistically significant ($p < 0.1$) sites enriched ≥ 2-fold (Fig. 3.1B, C). When mapped onto the holo Cas9-DNA-RNA crystal structure(Anders et al., 2014), the enriched insertion sites tend to cluster in discrete regions, often around flexible loops, the ends of helices, and at solvent-exposed residues. Specifically, hotspots are found in an abundance of Cas9's domains: at six clusters within the helical recognition (REC) lobe, the linker between the REC and nuclease lobes, the HNH domain, three extended sites in the RuvCIII region and throughout the PAM interacting (C-terminal) domain (Fig. 3. 1B,C, Supplementary Fig. 7). These insertion sites provide access to both 5' and 3' ends of the bound DNA (~10 Å) as well as the groove hypothesized to hold the non-targeted DNA strand(Anders et al., 2014; Nishimasu et al., 2014). Consequently, these insertions might allow engineering of specific and functional interactions with bound DNA. Although insertions are enriched in undefined regions from three different Cas9 x-ray crystal structures, they also occur in nearly every secondary structure element of Cas9 (Supplementary Fig. 8). This poses the possibility that the oft-used rational design strategy of inserting domains into flexible loops underestimates the protein space available for manipulation. Moreover, as the insertions into an alpha-helix or beta sheet presumably disrupt the secondary structure in these areas, such insertions may be informative in future efforts to dissect Cas9 structure and function.

S5

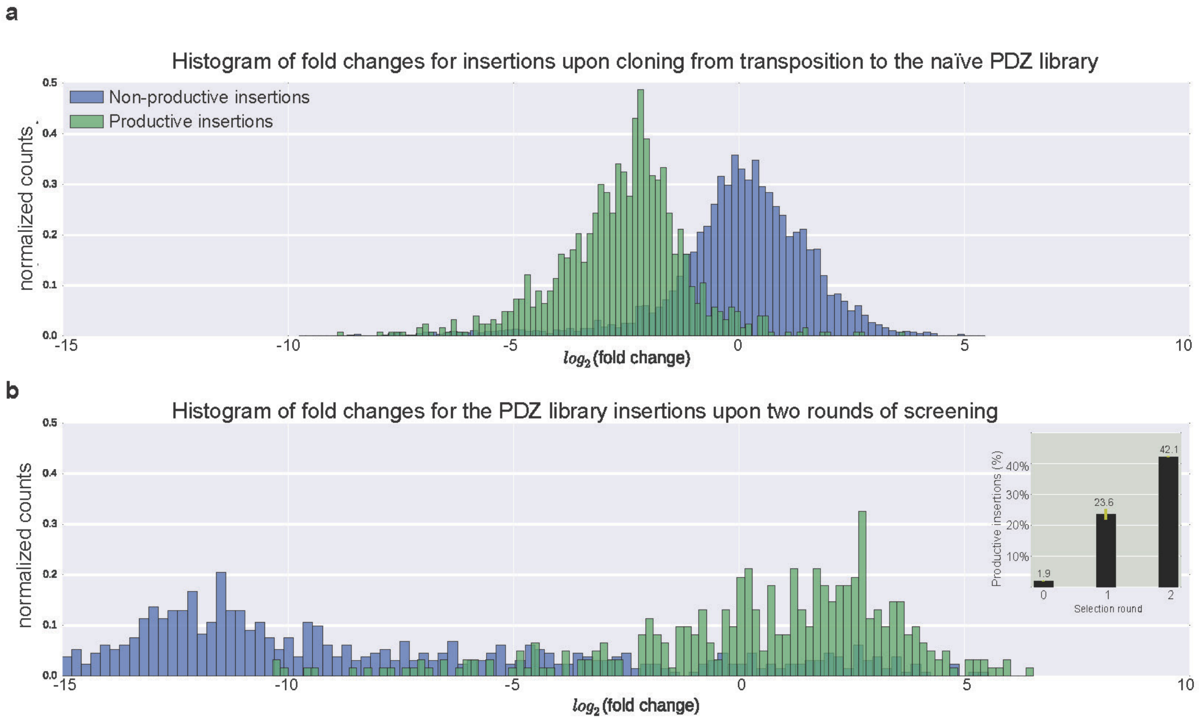


Figure 3.S5 | Enrichment for productive clones during domain insertion profiling of the PDZ-insertion library (a) Histogram of fold changes from the transposition to naïve PDZ libraries. The transposition library technique will theoretically create out of frame or reverse insertions that do not code for full length proteins ~5/6 of the time. During library outgrowth and cloning of the PDZ library we observe depletion of clones with in-frame, forward ('productive') insertions. This is presumably due to the cost of producing full length, >1400 amino acid, DNA binding proteins. (b) Passage of the PDZ insertion library through rounds of screening enriches productive insertions and substantially depletes non-productive insertions. Thus the CRISPRi based screen selects for full length coding sequences that can translate into full Cas9 proteins. Inset: percentage of productive insertions in the library during each round. All counts represent the normalized and corrected replicated averages generated from DESeq, error bars represent standard deviation of four technical replicates.

S6

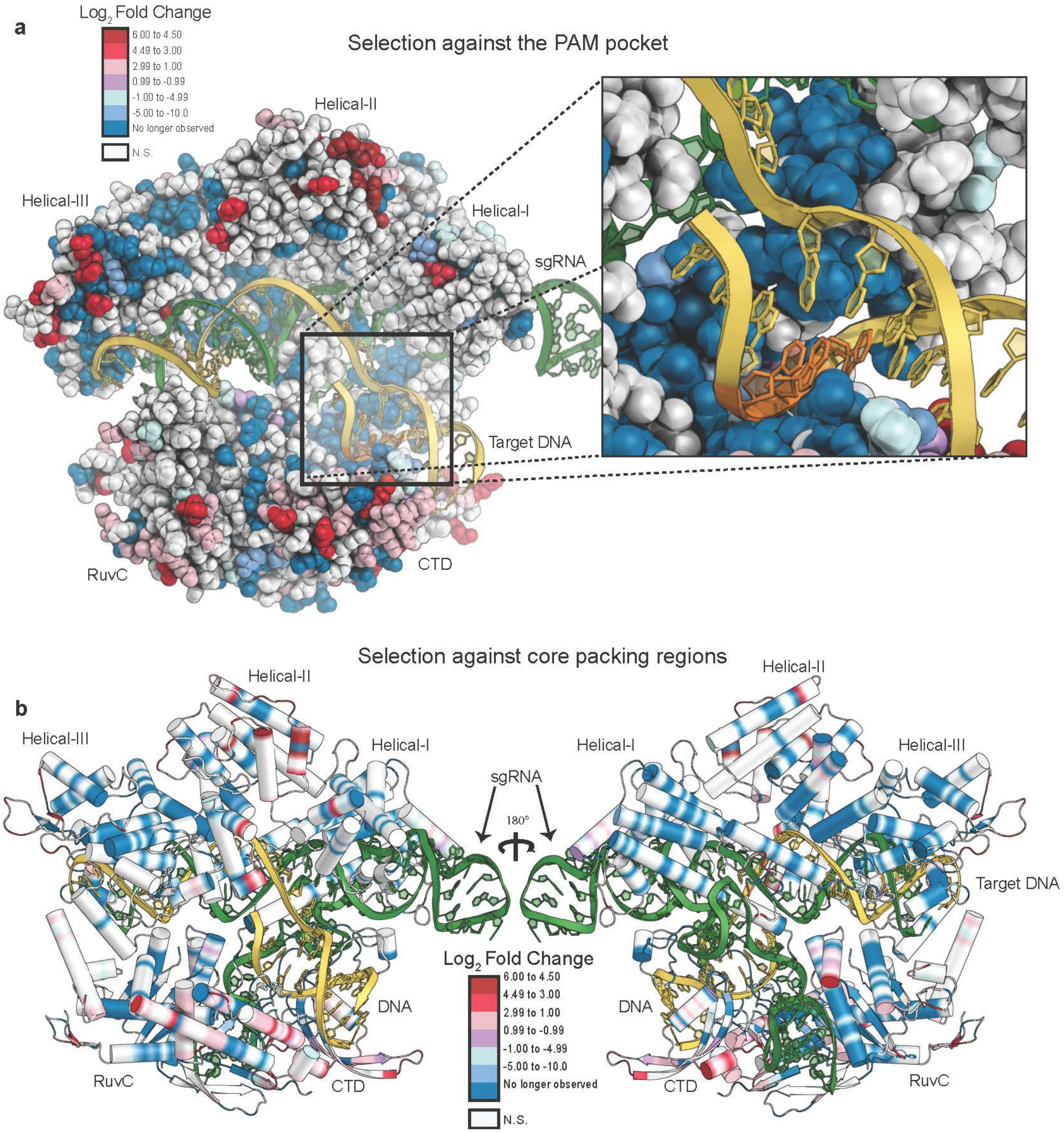


Figure 3.S6 | PDZ-insertion sites avoid critical structural motifs. (a) Mapping the log2 fold change of the PDZ-insertions onto the RNA:DNA holo Cas9 crystal structure (PDB ID: 4UN3) demonstrates how the unbiased insertion technique can experimentally delineate regions of critical structure and function. The PAM binding pocket (PAM residues in orange) and the RNA:DNA channel are selected against (blue). (b) PDZ insertions are not readily observed in core packing regions of the RuvC and Helical-III domains and are also selected against in the sgRNA binding grooves (Fig 1C) and Arg helix.

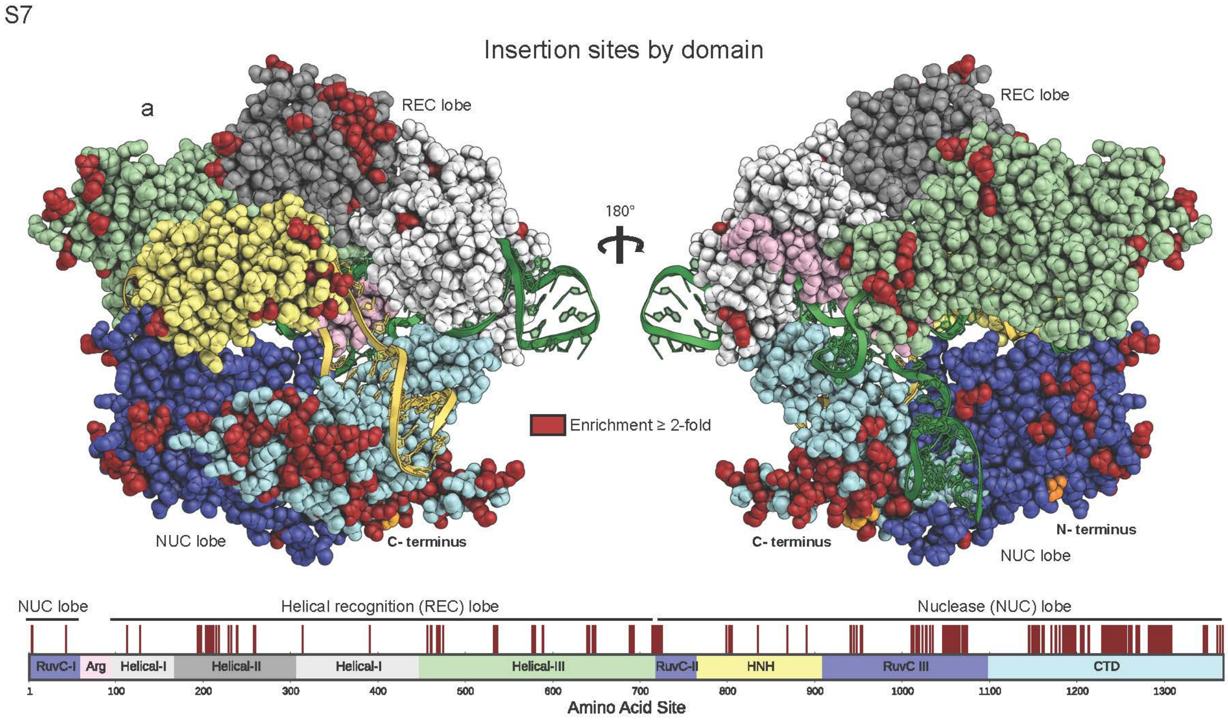


Figure 3.S7 | Map of enriched PDZ-insertion sites by Cas9 domain (a) Mapping the enriched insertions sites onto a RNA:DNA bound Cas9 crystal structure (PDB ID: 4UN3). Red denotes statistically enriched sites (figure 1b) greater than two-fold. Domains are colored in according to the primary sequence bar. Many sites with high fold-change mapped to amino acids that are unresolved in the crystal structures and are presumably in unstructured loops.

S8

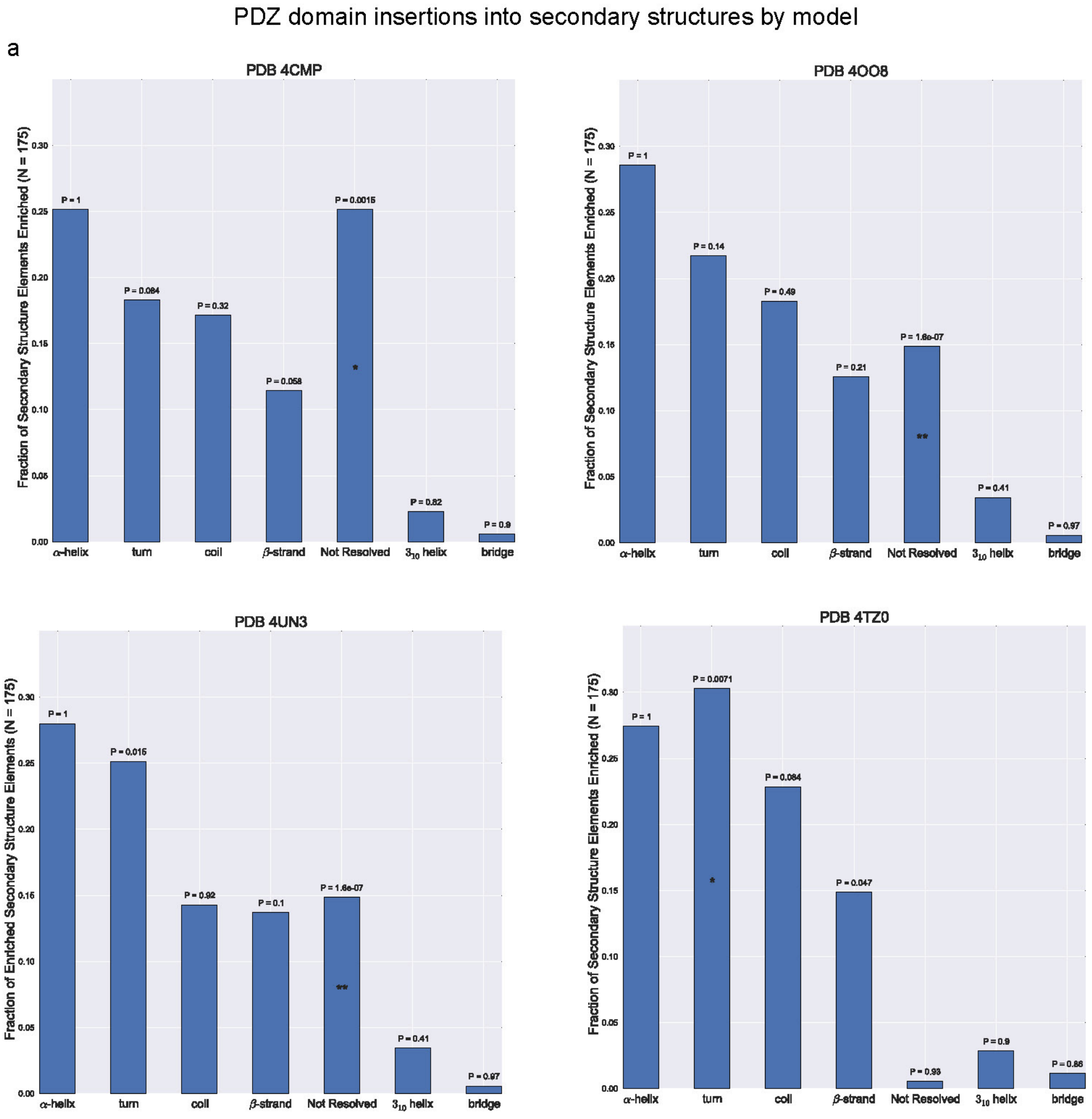


Figure 3.S8| Secondary structure of enriched PDZ insertions (a) Secondary structure annotation for each atomic model (PDB ID: 4CMP, 4OO8, 4UN3 & 4TZ0) was determined using STRIDE. In each of the structures insertions occur in all annotated types of structural elements. The fraction of insertions into each element was also compared to the overall prevalence of that element in the model and P-values were calculated using a null model where insertions were picked at random from the structure. For every structure except for 4TZ0 insertions into regions which could not be modeled (not resolved) were statistically overrepresented. This presents an interesting finding which could inform future rational efforts. In 4ZT0 where many more residues were resolved, insertions into turn elements were found to be over represented.

In order to confirm the binding activity of dCas9 proteins with PDZ domain insertions, plasmid DNA for individual clones were isolated, at random, from all stages of the selection and used to co-transform *E. coli* with a GFP-targeting sgRNA-expression plasmid. We found that effective GFP repression corresponds well with the calculated fold change of the insertion site; highly enriched clones perform at near wild-type (WT) dCas9 levels (Fig. 3.1D) (Supplementary Fig. 9) (Qi et al., 2013). To determine whether the PDZ-Cas9 clones also possessed nuclease activity, the catalytic residues (D10 & H840) were reintroduced and tested in an *E. coli* based transformation assay(Briner et al., 2014; Jinek et al., 2012). In this assay, nuclease activity leads to genomic cleavage and cell death. Once again, cleavage activity correlated well with fold change, the most highly enriched PDZ-Cas9 clones, except for the insertion at amino acid 208, maintained levels of cleavage-induced cell death similar to WT Cas9 (Fig. 3.1E) (Supplementary Fig. 9). Thus, it is possible to insert an entire exogenous domain at numerous sites within Cas9's primary sequence while maintaining near-native levels of activity.

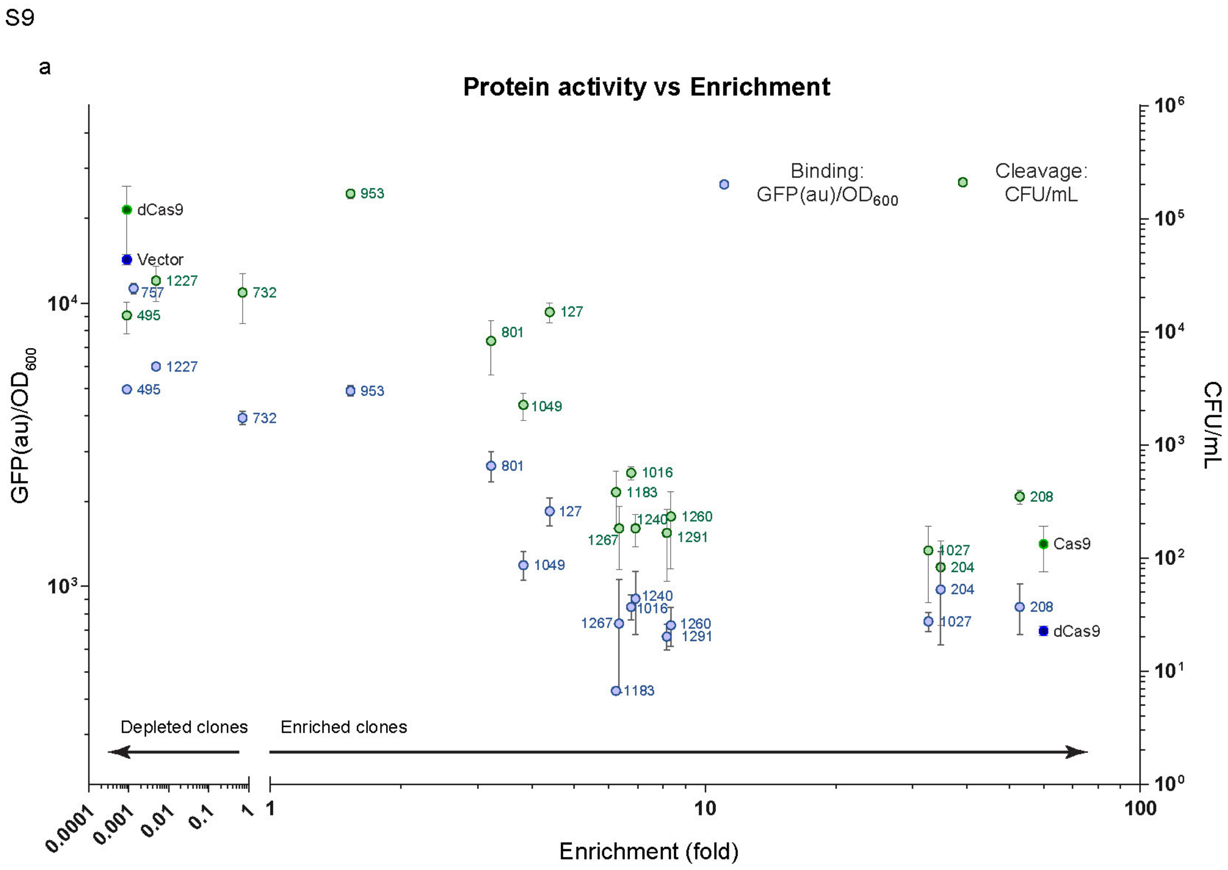


Figure 3.S9 | PDZ insertion clone activity vs fold change (a) The data from figures 1d & e has been graphed according to the enrichment (fold change) score for each construct, binding data based on repression of genomic GFP is graphed on the left y axis in blue. Positive and negative controls dCas9 and vector are dark blue. Cleavage data for each construct is mapped onto the right Y axis and colored in green with positive and negative controls Cas9 and dCas9 respectively in dark green. A monotonic relationship between fold change and binding/cleavage ability is readily apparent. PDZ-Cas9 activity levels plateau (at > ~4 fold enrichment) near wt Cas9 activity levels for each assay despite further increased enrichment. The spearman correlation for the cleavage data to be r= -.85, p value of <0.0001, and the binding data to be r= -.76, p value of 0.0006, indicating a strong monotonic correlation between enrichment and increased ability to both bind and cleave the genome of E. coli.

S10

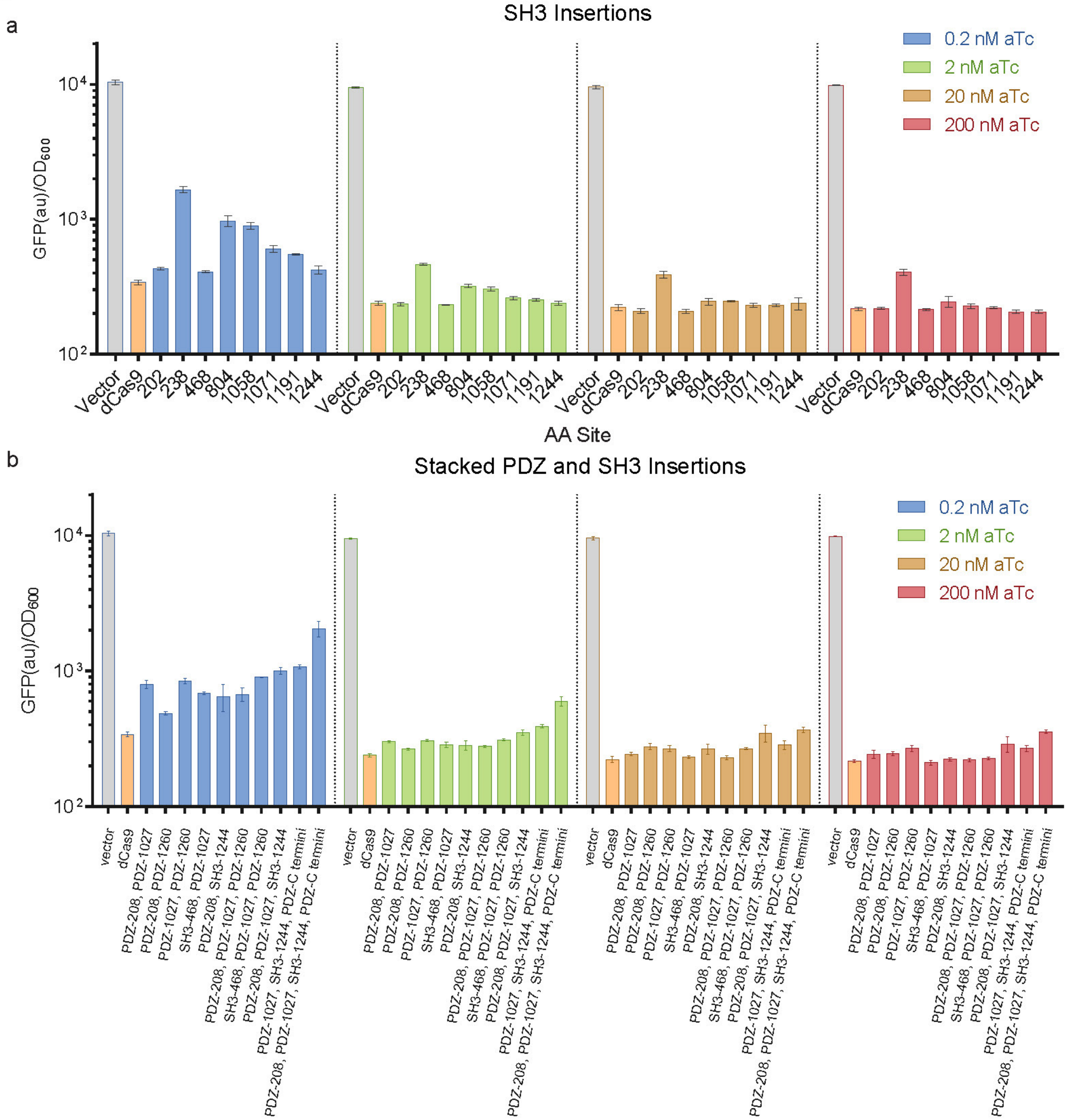


Figure 3.S10 | Testing of the SH3 and stacked domain insertions. (a) The crk SH3 domain was cloned into 8 enriched sites in dCas9 with linkers that mimic those used for the PDZ domain insertions (Supplementary Table 2). These clones were then tested for GFP repression ability at increasing levels of Tet promoter induction. While the SH3 insertions are less functional than expected at very low levels of induction, all but two, at sites 238 and 804, are able attain GFP repression levels at or near (±5%)WT dCas9 at higher levels of induction (based on mean repression, 238 maintains 51% and 804, 88% of dCas9 activity). (b) Previously validated PDZ and SH3 domain insertions were cloned into a single dCas9 and tested for GFP repression ability. These stacked domain insertions are also able to maintain DNA binding and gene repression ability with 3 at or near (±5%) WT dCas9, 3 at >80% of dCas9 and 4 between 58 and 79% of d Cas9 activity. Error bars represent one standard deviation of biological triplicates. Vector and dCas9 controls at each induction level is shown next to the group in grey and orange respectively.

To determine the effect of inserting an orthogonal domain at the sites recovered with the PDZ screen, we created eight synthetic domain insertions using the mouse SRC Homology 3 domain of adapter protein Crk (SH3)(Dueber et al., 2009). The SH3 domain was inserted into highly-enriched sites (> 10 fold) chosen to represent a diverse sampling of the Cas9 primary sequence. We found that all clones were functional for GFP repression and seven out of eight were able to obtain near-native activity levels (Supplementary Fig. 10). Finally, in order to examine the potential restrictions for domain insertion into Cas9 we surveyed the effect of multiple domain insertions on Cas9 function. We selected a subset of the PDZ-dCas9 and SH3-dCas9 insertions previously validated, 3 PDZ insertions, a PDZ C-termini fusion and 2 of the SH3 insertions, and created stacked constructs with up to three separate domain insertions and a terminal fusion. Although greater numbers of insertions and fusions have a perturbative effect on binding and repression activity, many of the stacked constructs are capable of repressing GFP to a degree comparable with dCas9 (Supplementary Fig. 10).

The design of synthetic protein switches and sensors is limited by the difficulty of predicting insertion sites that have the potential to confer allostery(Stein and Alexandrov, 2015). Therefore we next explored whether domain insertion profiling could be used to reveal sites in Cas9 amenable to allosteric coupling(Reynolds et al., 2011; Stein and Alexandrov, 2015). As a proof of concept, we chose to use the well-studied Human Estrogen Receptor α Ligand Binding Domain (ER-LBD; residues 302-552 of ESR1)(Shiau et al., 1998; Tanenbaum et al., 1998; Warnmark, 2002). Based on crystallographic data, this domain is known to adopt distinct conformations dependent on ligand binding; the antagonist-bound conformation places the N- and C-termini of ER-LBD substantially closer together (~20 Å) than either the apo or agonist bound forms (Supplementary Fig. 11). This conformational switch can serve as a mechanism by which an allosteric signal is transduced(Tucker and Fields, 2001).

In order to create an allosterically-regulated Cas9 (arC9), the ER-LBD was introduced into the naïve dCas9 transposition library in the manner previously described and passaged through a modified version of the CRISPRi-based screen. Briefly, a positive screen in the presence of ligand 4-hydroxytamoxifen (4-HT; antagonist) was carried out, followed by a negative screen for loss of activity in the absence of ligand (Fig 2A). Clones were recovered by plating, re-transformed with a sgRNA targeting GFP, and assayed for repression in *E. coli*. We identified an insertion site at AA 231 that demonstrated a 4-HT-dependent decrease in GFP fluorescence, indicating switch-like behavior (Supplementary Fig. 12). Notably, this site was also enriched during the PDZ profile of Cas9 (Supplementary Table 1).

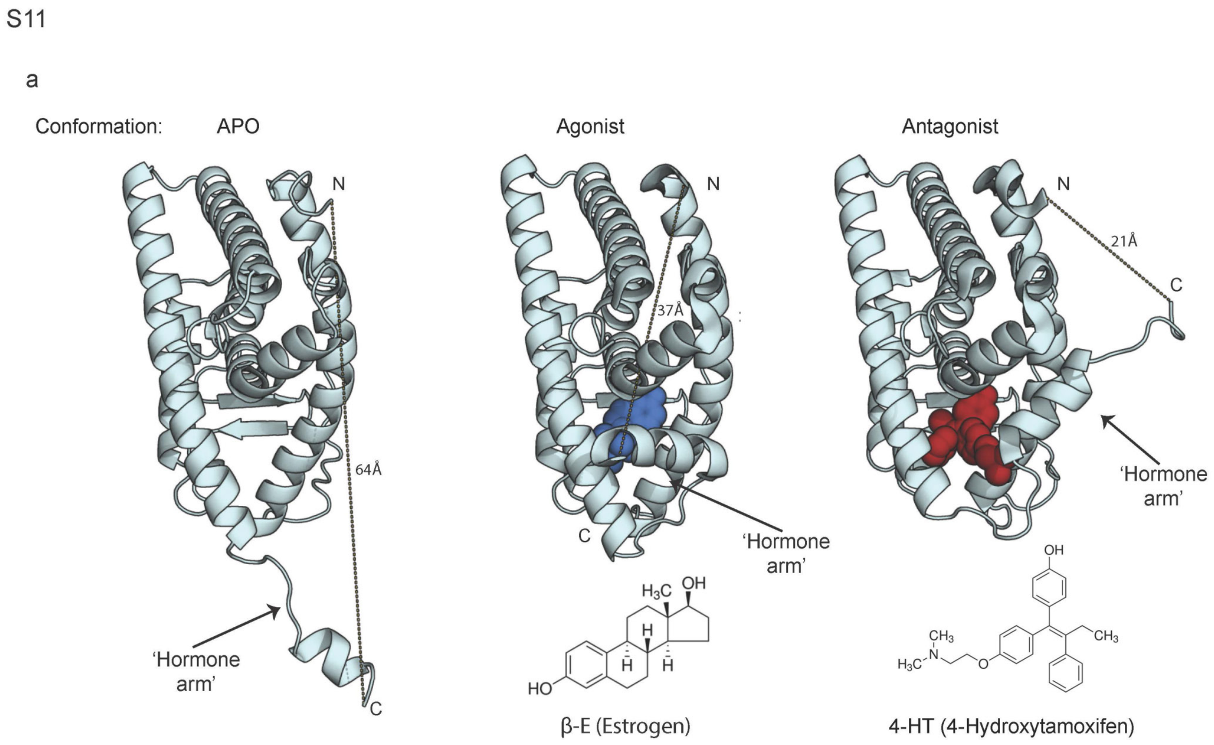


Figure 3.S11 | Conformations of the estrogen receptor alpha ligand binding domain (ER-LBD). (a) Conformations of the ER-LBD. Crystal structures of the apo,17-beta-estradiol and 4-hydoxytamoxifen-bound ER-LBD (PDB ID's:1A52, 1GWR, 3ERT respectively) clearly demonstrate the range of conformational change this domain can undergo. The 'hormone arm,' or helix 12, places the modeled N- and C-termini up to 63 Å apart in the apo form, 37 Å in the agonist bound form (β-E, blue), and 21 Å in the antagonist bound structure (4-HT, red).

Insertion of ER-LBD at AA 231 (arC9:231) was first characterized in *E. coli*. A catalytically dead arC9:231 (darC9) exhibited 4-HT dose-dependent repression of GFP with a ~10 fold change in activity in pooled and single cell experiments (Fig. 3.2B, C). Therefore darC9 represents a tunable CRISPRi effector. darC9 also exhibited clear ligand discrimination. CRISPRi-based GFP repression increased with ligands that encourage the ER-LBD to enter an antagonist conformation, 4-HT and nafoxidine, as opposed to ligands that promote the agonist conformation, beta-estradiol and diethylstilbestrol (Fig. 3.2D). This further supports the argument that the ER-LBD insertion at AA 231 is able to transduce ligand specific binding of 4-HT, through induction of the antagonist ER-LBD conformation, into Cas9 activity. To determine if arC9:231 also exhibited allosteric control over cleavage activity, the catalytic residues were reintroduced (D10 & H840) and tested in the *E. coli* transformation assay described above (Fig. 3.2E). We found that arC9 increased chromosomal cleavage and death at least 100-fold more in the presence of 4-HT than with a DMSO vehicle control (p value < 0.01), indicating that allosteric modulation of arC9 also extends to cleavage activity.

S12

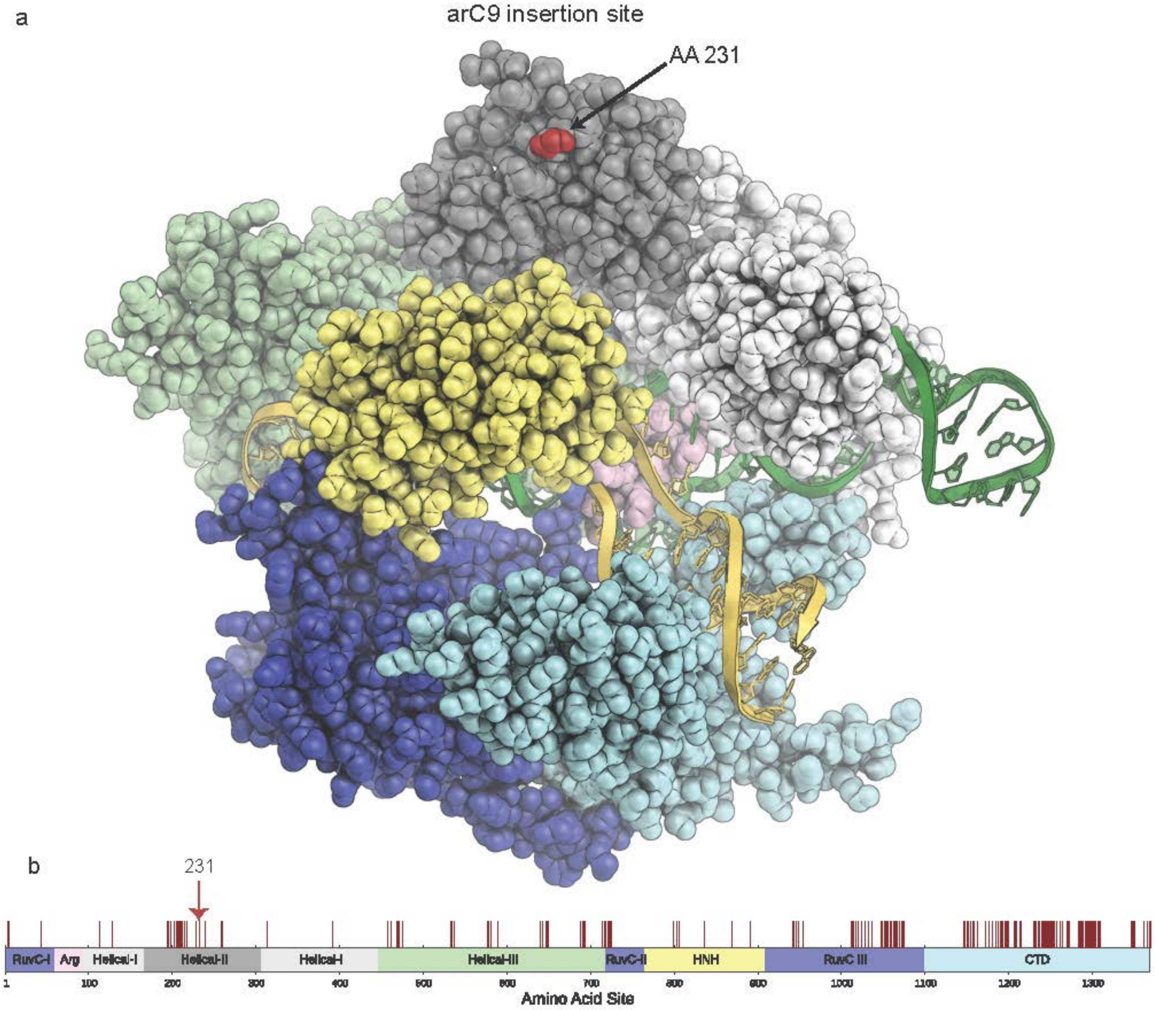


Figure 3.S12 | arC9 hit (a) After rounds of screening and counter-screening (Figure 2A) clones were picked from plates and tested in a 96 well assay for switch like behavior. One clone with the ER-LBD insertion site at amino acid Gly231 (indicated in red) demonstrated a change in activity upon addition of 4-HT and was further validated. (PDB ID: 4UN3)

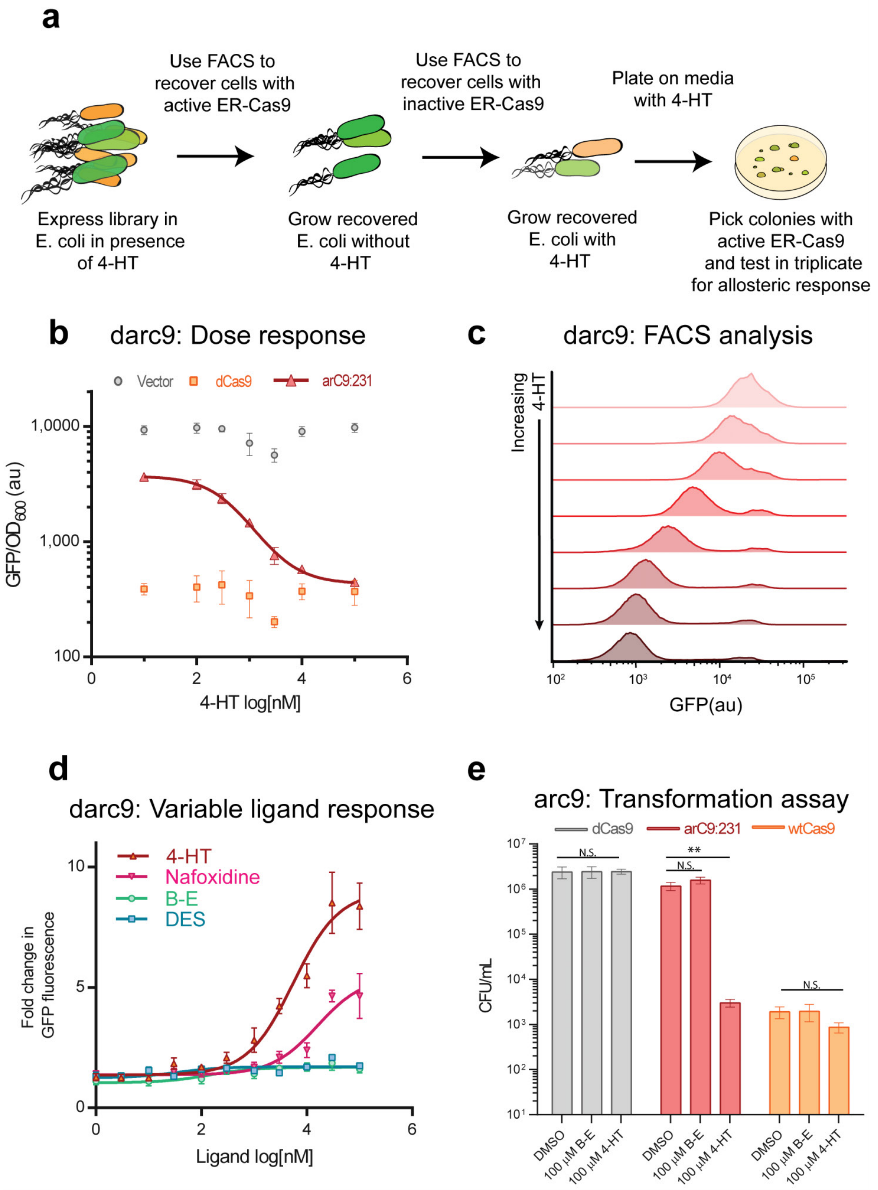


Figure 3.2 | Creation of a switch-like Cas9 though insertion of the Estrogen Receptor ligand binding domain. (a) Schematic of the screen / counter-screen procedure to select for ligand responsive Estrogen Receptor ligand binding domain (ER-Cas9) insertions. (b) Dose-response curve to 4-HT. darC9:231 has an IC50 of 440 ± 70 nM (S.D) and a Hill coefficient of 1.04 as expected for non-cooperative binding of 4-HT to ER-LBD. (c) Single cell analysis of darC9:231 binding in response to increasing concentrations of 4-HT. Flow cytometry data tracks ensemble data demonstrating a > 9 fold switch between darC9 based GFP repression plus and minus 4-HT (darC9 GFP signal without 4-HT mean: 24,310 (au) and with 100 µM 4-HT: 2,631 (au) (d) Dose response of darC9:231 binding to various ligands (B-E and DES are beta-estradiol & diethylstilbestrol). Response is normalized to vector control fluorescence under the same conditions. (e) Switching of arC9:231 cleavage activity. Transformation assays demonstrate that ligand dependent arC9 switching also extends to cleavage activity, t-test (** p-value <0.01). All experiments performed in E. coli.

We next tested arC9:231 function in eukaryotic cells. Cas9 and arC9:231 constructs flanked by nuclear localization sequences (NLS) were transfected into a Human Embryonic Kidney (HEK293T) cell line expressing a stably integrated EGFP-PEST, and assayed for GFP disruption (Fig. 3.3A)(Tsai et al., 2014a). After 72 hours, WT Cas9 was found to disrupt 91% of the EGFP signal regardless of treatment condition (Fig. 3.3B, Supplementary Fig 12). A 6-fold induction of arC9-mediated GFP disruption upon the addition of 4-HT (10.9 ± 0.5 % to 66 ± 1%) (Fig. 3.3B) was observed over the same time period. We also found that higher expression conditions led to an increase of GFP disruption even in the absence of 4-HT (Supplemental Fig. 13).

To improve arC9 control we sought to take advantage of the known ligand-dependent nuclear localization activity of ER-LBD (McIsaac et al., 2013). We removed the dual NLS from the arC9 construct and found that this reduced activity without 4-HT to background levels while maintaining roughly 30% of

Cas9 disruption activity in the presence of 4-HT (Fig. 3.3C). Using the GFP disruption assay, we observed arC9's dose-dependent response with an EC_{50} of 10 nM and full induction at 100-1000 nM 4-HT (Fig. 3.3D). The lack of background disruption suggests that arC9 can be adapted for very tight, reversible control of Cas9 cleavage activity.

S13

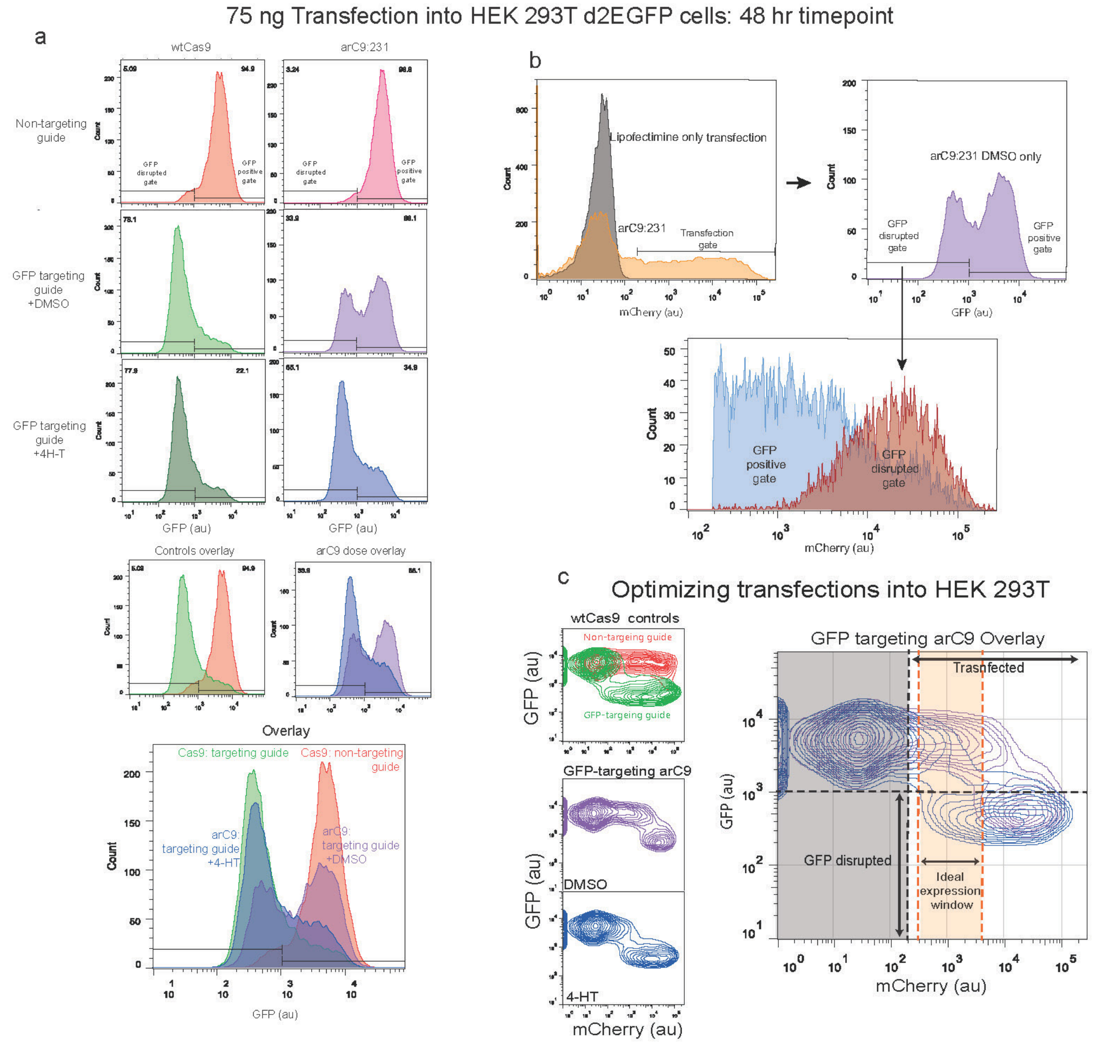


Figure 3.S13 | Optimization of arC9+NLS expression levels (a) Upon transfection of 75 ng of plasmid into the HEK293T cell line and measurement EGFP fluorescence via Flow cytometry at 48 hours we observed that 2xNLS-Cas9 disrupts ~70-80% of EGFP signal regardless of treatment condition. 2xNLS-arC9 also disrupted GFP signal with a 2 fold increase upon the addition of 4-HT (b). We detected that 4-HT-induced activation is seen at low levels of arC9-mCherry expression, but higher levels of arC9-mCherry expression cause GFP disruption regardless of the presence of ligand. Specifically, the mCherry signal for the arC9 transfected cells - in which GFP was disrupted regardless of 4-HT treatment - was 8.5x greater than that of the non-disrupted cells. Therefore we optimized transfection conditions and were able to reduce the expression levels similar to the ideal gates posed in (c) by lowering the plasmid transfection to 5 ng of DNA. This resulted in significantly less background activity while maintaining 4-HT activation (main Fig. 3B).

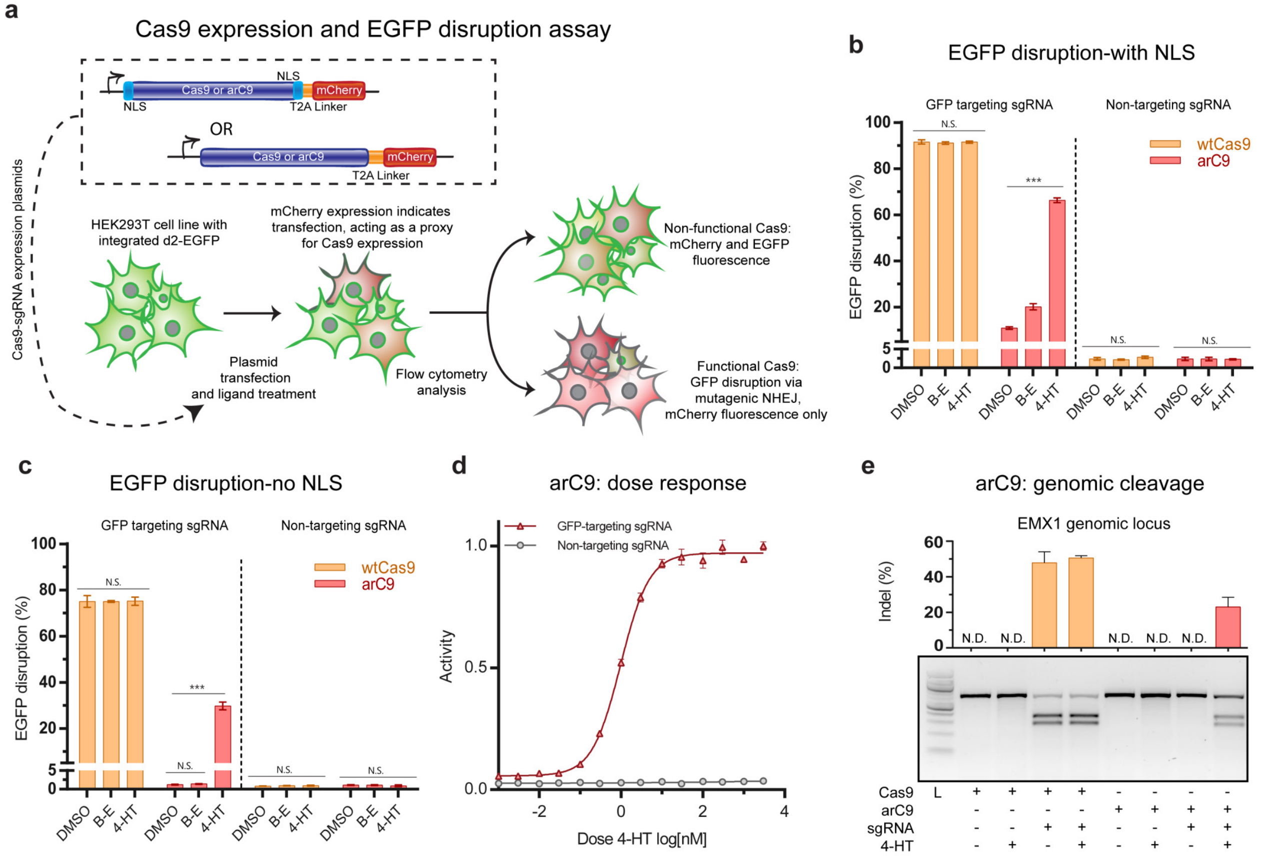


Figure 3.3 | Validating arC9 in eukaryotic cells. (a) Schematic of the arC9:231 expression constructs and GFP disruption assay. (b) Quantification of EGFP disruption at 72 hours for Cas9 and arC9 with a N- and C-terminal nuclear localization signal (NLS) (n=3). These data demonstrate a ~6 fold increase in EGFP disruption. Background activity of arC9 is 10.9 ± 0.5% (S.D.) while EGFP disruption in the presence of 300 nM 4-HT increases to 66 ± 1% (S.D.) Error bars represent one standard deviation of biological replicates. (c) Quantification EGFP disruption at 72 hours for Cas9 and arC9 without an NLS (n=3). Background activity of arC9 is not significantly different from a non-targeting negative control, t-test. EGFP disruption in the presence of 4-HT increases to 30 ± 2% (S.D.) this represents at least a 24-fold increase in arC9 activity in the presence of 300 nM 4-HT, t-test (*** p values < 0.001). (d) Dose response of arC9 w/o NLS normalized to maximum activity. IC50 is 1.0 ± 0.2 nM (S.D). (e) T7EI assay of Cas9 and arC9 mediated indel formation at the EMXI locus at 72 hours. Cas9 with the targeting guide causes indels regardless of treatment condition, arC9 only cleaves a genomic locus in the presence of 4-HT. Quantification and error bars represent the standard deviation of biological replicates (n=3). N.D. signifies not detected, below the detection limit of the assay.

In previous work, ER-LBD fusions to the termini of various DNA-binding proteins have been shown to regulate the ability of such proteins to enter the nucleus(Feil et al., 1996; McIsaac et al., 2013) To explicitly demonstrate the necessity and utility of internal domain insertion, we directly compared the activity of our arC9 construct with that of a C-terminal ER-LBD fusion. In contrast to the switch-like response of arC9, a C-terminal fusion of ER-LBD minimally controls Cas9 EGFP disruption, providing a modest 0.21 fold increase in repression upon the addition of 4-HT, with greater than 50% of cells disrupted without activating ligand (Supplementary Fig. 14). This result is consistent with our previous data demonstrating that WT Cas9 does not require an NLS to function in dividing human cells (Fig. 3.3C). To further validate arC9 we tested its nuclease activity on two endogenous human loci, EMX1 and DRKY1. T7 endonuclease I (T7EI) analysis showed efficient indel formation and gene editing in presence of 4-HT, while there was no detectable arC9 activity without ligand (Fig. 3.3E, Supplementary Fig. 15). Thus, arC9 acts as a 4-HT-inducible nuclease in human cells demonstrating that unbiased domain insertion can be used to engineer robust Cas9 functionalities not achievable using terminal fusions.

S14

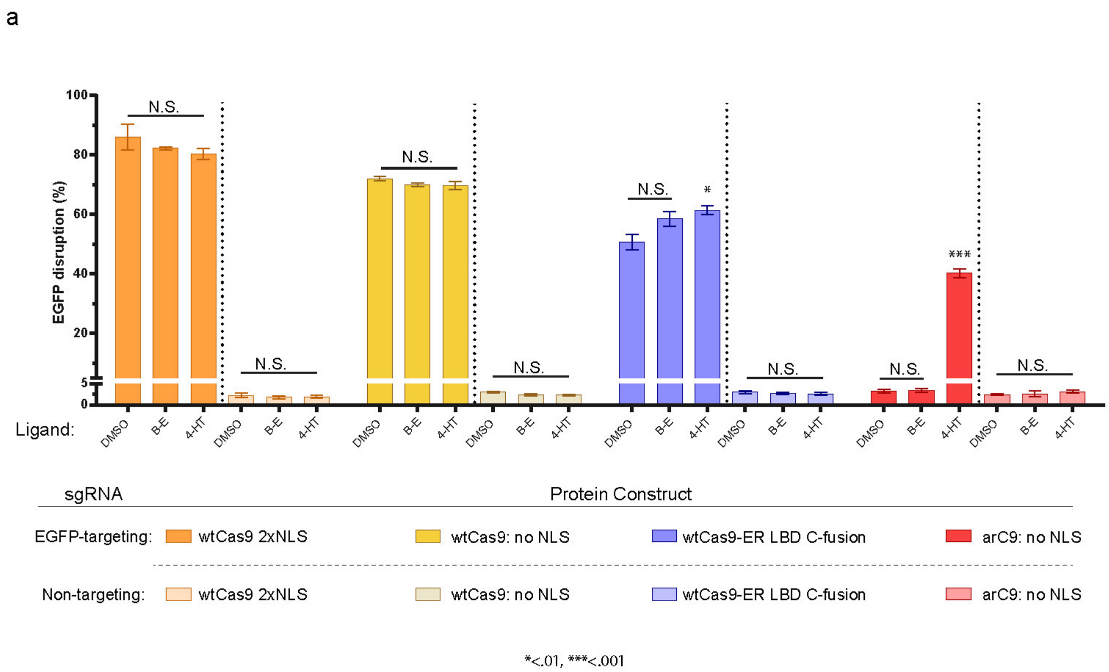


Figure 3.S14 | Comparison between termini and domain insertion fusions. (a) In order to examine the capabilities of termini fusions in comparison to the arC9 domain insertion the ER-LBD was cloned as a C-termini fusion to wtCas9 w/o a NLS. Examination of the ability to control Cas9 activity in the presence of distinct ligands reveals that the ER-LBD C-termini fusion can increase Cas9 based EGFP disruption from 50.7 ±2.5% in DMSO to 61.3±1.5% in 4-HT(p=0.003). This induction is compared to that of arC9 which induces Cas9 EGFP disruption from background levels (equivalent to non-targeting guide) in DMSO and B-E to 40.3 ±1.5% in 4-HT (p <0.001). All data is represented as a mean from biological triplicates, error is one standard deviation.

The discovery and application of CRISPR-Cas9 has revolutionized functional genomics. However, large-scale screens and more focused *in-vivo* applications that require precise timing, such as the study of development, differentiation and late onset disease, have been limited by the constitutive activity of Cas9. Tet-systems can enable Cas9 expression control(Dow et al., 2015; González et al., 2014; Gossen and Bujard, 1992; Gossen et al., 1995; Kearns et al., 2014), but are cumbersome due to the need for additional components. We therefore tested whether arC9 could be stably integrated into the genome from a single-vector lentiviral system and enable conditional gene editing. To assess long-term leakiness, induction, and reversibility of the arC9 protein, we also generated a sensitive monoclonal reporter cell line by transducing murine BNL CL.2 cells with a retroviral vector expressing EGFP (Supplementary Fig.16a). Upon low-copy infection with the Cas9/arC9 lentiviral constructs and hygromycin B selection, we measured the reduction in the number off GFP expressing cells for arC9 over 24 days. Whereas WT Cas9 showed up to ~80%

GFP disruption with GFP-targeting sgRNAs, no leakiness was observed with arC9 in DMSO (Supplementary Fig. 16c). After 12 days, we treated a subpopulation of arC9-expressing cells with 4-HT (1 µM), beta-estradiol (1 µM) or vehicle control (DMSO). In the presence of 4-HT, arC9 disrupted GFP in up to ~16% of cells, whereas no measurable editing was observed in the absence of ligand (Supplementary Fig. 16d). Some background GFP-negative cells were present in all samples.

S15

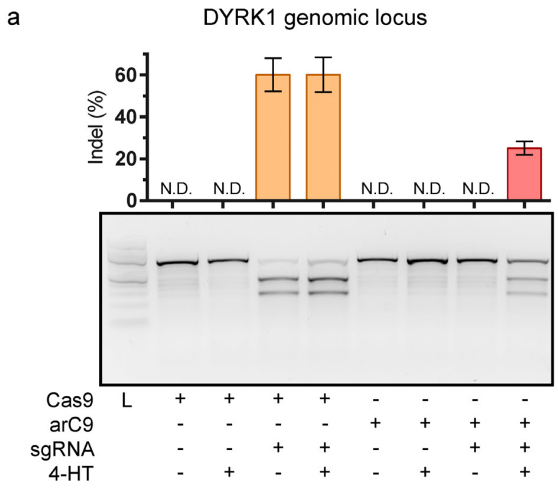


Figure 3.S15 | T7EI assay of the Cas9 and arC9 cleaved DYRK1 genomic locus. (a) Cas9 and arC9 targeting of a second genomic locus was carried out for 72 hrs in triplicate. T7EI assays were then performed on the recovered genomic DNA. Cas9 with targeting guides causes indels regardless of treatment condition, while arC9 cleaves a genomic locus only in the presence of 4-HT. Quantification and error bars represent the standard deviation of biological replicates. N.D. signifies not detected, below the detection limit of the assay. L is a lane with a 100bp Ladder.

To test whether the activation of arCas9 is reversible_we recovered the 4-HT treated cells at the end of the treatment, cultured them for two days in regular culture medium, and then infected them with a secondary lentiviral vector expressing a sgRNA targeting Pcsk9. T7EI analysis of the endogenous Pcsk9 locus after six days of treatment with DMSO or 4-HT showed that arC9 can be turned off in the absence of ligand the or used for controlled serial genome editing by repeated addition of 4-HT (Supplementary Fig. 16e). After removal of 4-HT from the media a small amount of residual arC9 activity remained. This may be due to the high affinity of arC9 for 4-HT (EC_{50} 10 nM) and likely slow dissociation of the complex (Figure 3D). Nevertheless, arC9 can serve as a single-component system for conditional genome editing and/or be combined with other sgRNA expression platforms to render gene editing inducible and reversible.

S16

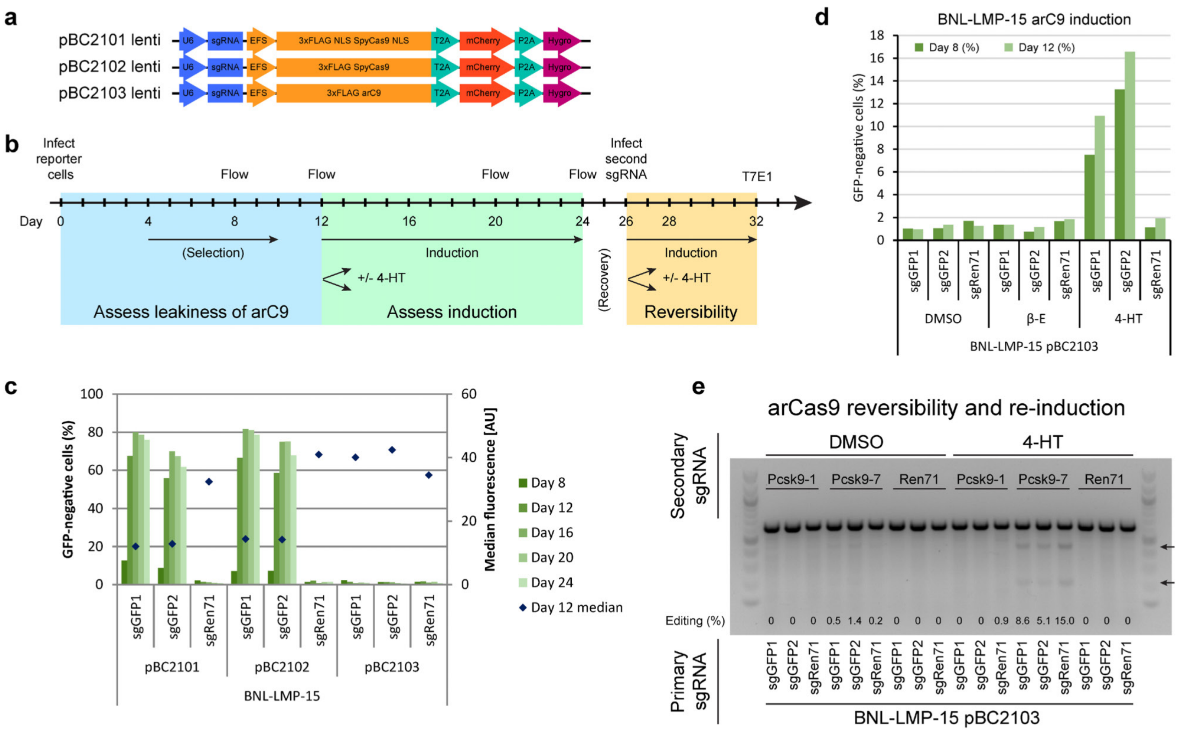
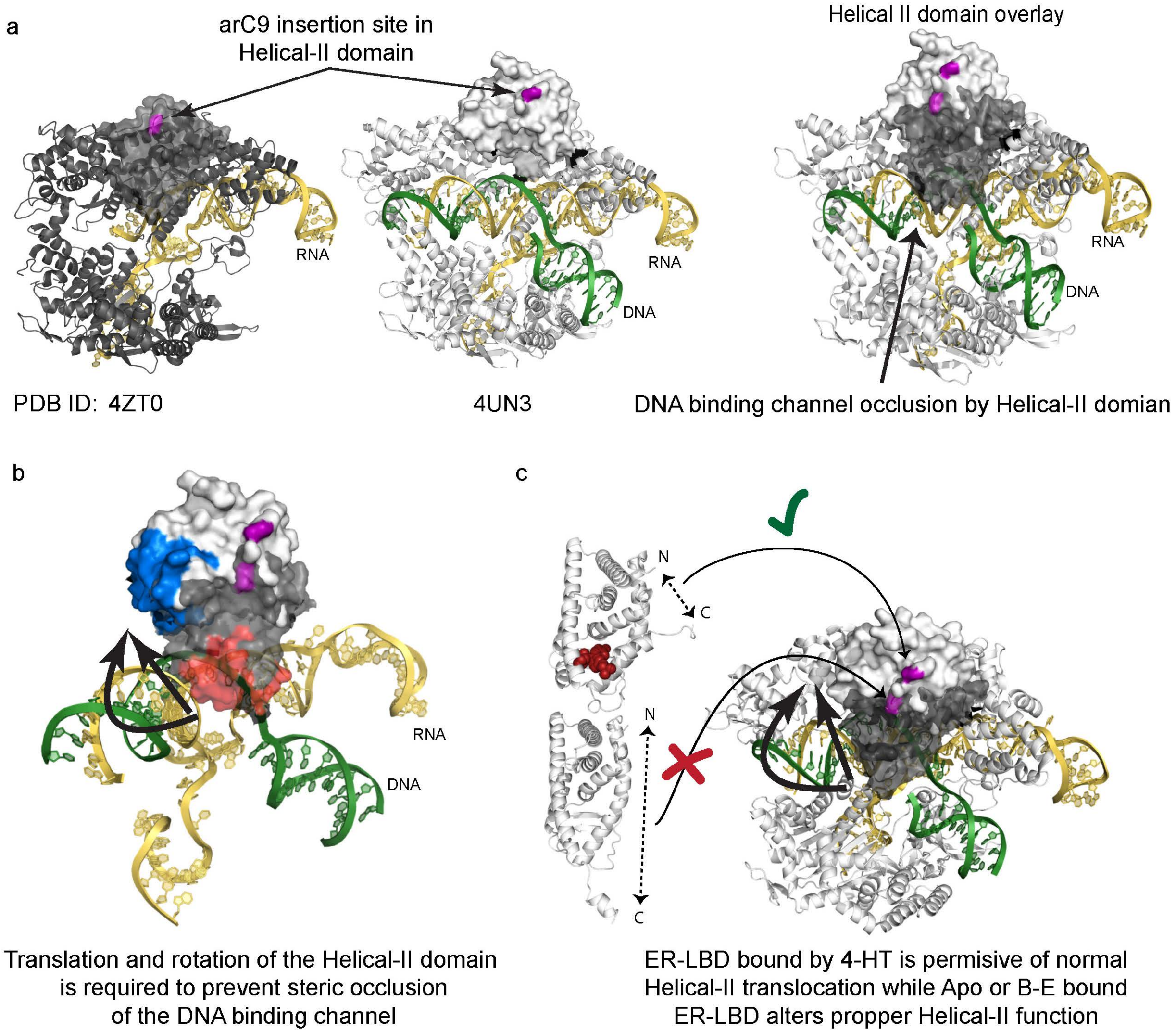


Figure 3.S16 | 4-hydroxy-tamoxifen inducible arC9 lentiviral vector analysis (a) Lentiviral vector maps. (b) Timeline of arC9 assessment with regards to leakiness, inducibility and reversibility of arCas9 activity. (c) Control quantification of the fraction of GFP negative BNL-LMP-15 reporter cells at the indicated time points and arC9 without ligand, no leakiness is observed. sgRen71 is a non-targeting, negative control guide (n >10,000 events for each measurement). (d) Quantification of the fraction of GFP negative BNL-LMP-15 pBC2103 cells treated with DMSO, beta-estradiol (β-E) or 4-hydroxy tamoxifen (4-HT) for the indicated amount of time, sgRen71 is a non-targeting, negative control guide (n >10,000 events for each measurement) (e) T7E1 quantification of editing at the Pcsk9 locus with secondary sgRNAs, 6 days after infection and start of secondary arC9 re-induction (see timeline for details). Regardless of the three previous guide conditions cleavage with sgPcsk9-7 is observed only in the presence of 4-HT, no signal above background is detected in the DMSO control. sgPcsk9-1 does not show activity. Arrows indicate the size of the canonical cleavage products for sgPcsk9-7 (414bp, 180bp)

Although arC9 functions robustly as an inducible gene repressor and endonuclease in prokaryotic and eukaryotic cells, it is unclear how arC9 is activated in the presence of 4-HT. The arC9:231 insertion is especially counterintuitive as the Helical-II (REC2) domain into which the ER-LBD is inserted has previously been shown to be dispensable for cleavage activity(Nishimasu et al., 2014). A recent crystal structure of sgRNA bound Cas9(Jiang et al., 2015) helps rationalize the ligand-dependence of arC9, suggesting that ER-LBD may disrupt the required conformational changes of the Helical-II domain. This disruption would sterically occlude the RNA:DNA binding channel of Cas9 and prevent functional DNA unwinding and RNA:DNA hybridization (Supplementary Fig. 17)(Jiang et al., 2015).


a
arC9 insertion site in Helical-II domain
Helical II domain overlay
RNA
RNA
DNA
RNA
DNA
PDB ID: 4ZT0
4UN3
DNA binding channel occlusion by Helical-II domian
b
c
N
C
N
C
RNA
DNA

Translation and rotation of the Helical-II domain is required to prevent steric occlusion of the DNA binding channel

ER-LBD bound by 4-HT is permisive of normal Helical-II translocation while Apo or B-E bound ER-LBD alters propper Helical-II function

Figure 3.S17 | Model for arC9 based regulation. (a) PDB models 4ZT0 (black) and 4UN3 (white) are Cas9 structures which are respectively sgRNA only and sgRNA and DNA bound. When aligned it is possible to observe that in the RNA only bound structure the helical-II domain, highlighted using surface representation, is sterically occluding the channel in which the DNA will be unwound and bind to the sgRNA. It is therefore assumed that this domain must rearrange and vacate this channel in order to unwind DNA. This provides a possible mechanism by which arC9 ER-LBD control may be accomplished. (b) Within the helical-II domain there is an alpha helix and adjacent residues (residues 255-283) that must translate~20 -30 Å and rotate from the position highlighted in red to that in blue in order to vacate the DNA binding channel. (c) The insertion of the apo or B-E bound ER-LBD at site 231(highlighted in purple) could affect the ability of Cas9 undergo this conformational change, thus rendering it unable to unwind, bind and cleave DNA targets. Upon the addition of 4-HT the insertion is no longer perturbative to the un-shielding of the DNA channel by the Helical-II domain and thus DNA binding and cleavage are restored.

Discussion:

In this work we identify a range of potential insertion sites for Cas9 engineering and outline a methodology for the development of new Cas9 constructs with improved control over genome editing and modification(Davis et al., 2015; Hsu et al., 2013). Previous efforts to engineer Cas9 as a molecular scaffold have constructed systems in which the sgRNA can recruit effector proteins(Shechner et al., 2015; Zalatan et al., 2014). Others have built inducible Cas9 nucleases using intein splicing or the splitting of the protein itself (Davis et al., 2015; Nihongaki et al., 2015; Zetsche et al., 2015). All of these efforts have used iterative, rational design to isolate functional molecules. By contrast, we demonstrate an unbiased profiling of domain insertions across Cas9 structure. Coupled with high-throughput screening and sequencing, this profiling rapidly queries structure and informs the protein engineering process. Specifically, we have generated a host of Cas9 scaffolds, containing PDZ or SH3 interaction domains that are functional and may prove useful for the recruitment of accessory proteins in future work. Notably, we also find that all previously reported and validated split Cas9 constructs fall within or adjacent (≤2 amino acids) to a small number of the insertional hotspots identified here (Supplementary Fig.18)(Davis et al., 2015; Nihongaki et al., 2015; Truong et al., 2015; Zetsche et al., 2015). To assess the generality of domain insertion profiling, the process was repeated with ER-LBD and identified a variant, arC9, whose activity is allosterically coupled to ER ligand binding. It functions as a conditional DNA binding protein and nuclease in both prokaryotes and eukaryotes and displays substantial ligand-dependent activity with very low background.

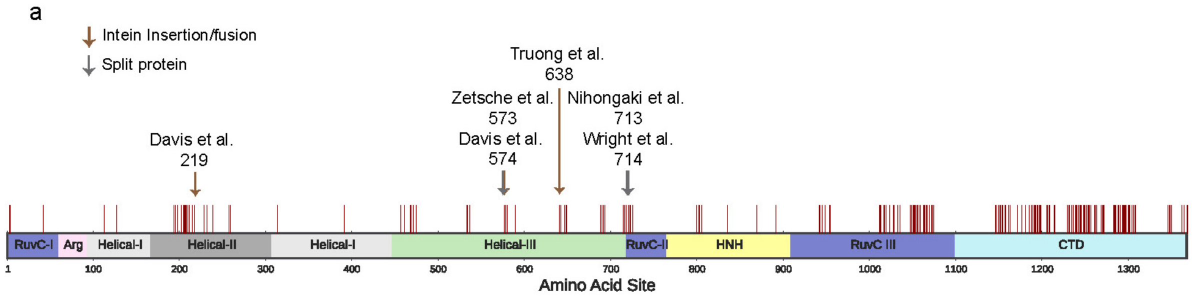


Figure 3. S17 | Comparison of previously identified sites with hotspots identified in this study. (a) Amino acid sites rationally identified and utilized in previous studies which have split Cas9, introduced intiens or split intiens into the protein have been mapped onto the hotspots identified with the PDZ insertion.

AA	Fold Change	P-value	AA	Fold Change	P-value	AA	Fold Change	P-value
0	0	8.88E-02	158	0	3.85E-07	388	0.034376063	8.51E-04
1	0.000826557	5.83E-28	159	0	3.90E-02	390	14.32567319	1.01E-10
3	3.601137388	2.30E-08	160	0	5.35E-03	391	0.017859286	9.80E-28
4	16.11436743	1.27E-27	162	0	3.30E-05	392	0	5.87E-05
6	0	1.36E-02	163	0	1.19E-02	397	0	8.79E-02
7	0	9.37E-03	171	0	7.81E-02	400	0	6.09E-13
8	0	1.79E-53	177	0.326662683	7.95E-02	401	0	1.83E-02
10	0	2.81E-03	182	0	2.10E-13	409	0	6.55E-02
11	0	4.99E-10	187	0	2.00E-04	411	0	7.84E-03
12	0	1.85E-02	188	0	3.48E-03	412	0	1.22E-24
17	0	1.71E-02	189	0	4.13E-14	417	0	6.73E-56
19	0	1.16E-02	192	0	1.22E-05	418	0	1.74E-13
20	0	5.34E-06	193	18.32575335	4.19E-32	420	0	1.49E-03
23	0	6.79E-10	195	inf	4.49E-25	423	0	3.88E-02
26	0	7.98E-03	197	4.301197109	1.36E-02	424	0	1.54E-03
27	0	4.91E-04	202	58.92282099	2.73E-29	427	0	7.11E-04
28	0	7.00E-04	204	34.79664821	6.15E-56	428	0	5.99E-02
39	0	3.95E-02	205	10.98458786	6.71E-19	435	0	3.86E-02
41	0	1.18E-02	206	7.947451778	6.25E-22	436	0	8.80E-02
42	5.675401141	4.75E-07	207	5.820770735	1.18E-11	438	0	1.04E-02
48	0	8.29E-02	208	52.92924613	4.01E-76	439	0	1.41E-04
49	0	2.98E-28	209	44.09401567	3.59E-118	440	0.110612574	5.10E-05
50	0.001512907	8.02E-06	210	0	8.80E-02	441	0	5.25E-02
51	0	8.93E-03	211	39.78901975	6.33E-11	445	0	1.62E-03
56	0	2.56E-02	214	9.754677145	5.19E-12	450	0	1.02E-10
59	0	1.44E-04	217	3.247607307	1.03E-03	451	0	4.16E-02
60	0	8.30E-06	228	24.15240414	1.44E-13	453	0	1.02E-14
72	0	5.83E-02	231	10.1408837	6.02E-58	454	0	2.78E-02
76	0	5.24E-03	234	0	1.22E-02	455	0	3.32E-02
80	0	8.06E-03	238	25.2213505	5.66E-02	456	12.05532582	2.03E-17
83	0	1.59E-03	239	0.120304697	2.25E-07	460	21.36444536	3.31E-16
84	0	1.07E-09	242	0	5.59E-09	463	0	5.23E-02
90	0	1.07E-03	243	0	3.90E-02	464	0	2.43E-03
91	0.101072022	1.02E-05	246	0	2.58E-02	466	0.229151682	1.02E-02
92	0	6.96E-04	247	0	7.92E-05	467	9.936097257	1.29E-03
94	0	1.06E-03	249	0	8.68E-02	468	46.7244051	1.18E-10
98	0	8.68E-02	257	13.0416803	1.00E-05	470	5.27934249	1.69E-13
102	0	6.76E-03	259	12.54667912	2.30E-03	474	6.627732797	2.37E-23
107	0.04431243	1.14E-04	272	0	5.36E-06	476	0	1.26E-04
108	0.421552818	9.69E-04	275	0	3.93E-02	477	0	2.87E-04
112	12.6091231	8.39E-09	276	0	2.45E-03	483	0	7.93E-06
118	0	6.60E-02	280	0	6.38E-05	485	0	2.63E-08
120	0	8.72E-02	281	0	4.00E-02	488	0	2.84E-62
123	0.449394768	9.21E-02	293	0	1.84E-19	489	0	1.03E-05
124	0.513023252	7.10E-02	313	8.040270093	1.24E-02	492	0	1.43E-19
127	4.385505124	4.14E-05	316	0	3.86E-02	495	0	1.72E-16
130	0.261106757	1.18E-02	319	0	8.42E-02	496	0.003154003	1.23E-42
134	0	6.67E-06	322	0	1.84E-02	497	0	1.22E-05
135	0	1.34E-04	325	0	8.72E-02	499	0	8.58E-18
136	0	2.34E-03	336	0	8.88E-02	507	0	2.66E-02
139	0	2.75E-02	339	0	1.62E-04	508	0	2.64E-07
146	0	3.46E-05	345	0	4.34E-03	509	0	8.84E-02
147	0.051883754	8.84E-08	353	0.428703399	2.19E-02	512	0	1.35E-02
150	0.00602536	1.61E-12	361	0	2.22E-24	514	0	3.13E-04
153	0	7.85E-03	376	0	7.77E-05	515	0	2.85E-17
155	0	3.98E-05	379	0.007720254	6.24E-11	519	0	7.94E-03
157	0	1.31E-07	386	0	2.31E-03	520	0	2.28E-26

521	0	3.88E-02	632	0	1.41E-03	740	0	3.41E-19
522	0	2.75E-03	633	0	5.72E-03	741	0	6.75E-11
523	0	1.83E-02	634	0	8.07E-15	742	0	1.01E-24
525	0	4.53E-30	635	0	1.26E-02	743	0	1.83E-03
526	0	4.81E-11	639	7.678755348	4.92E-08	744	0.035763142	1.03E-02
527	0	1.28E-02	640	0	4.03E-02	745	0	1.53E-24
528	0	1.05E-04	641	4.672334773	1.14E-04	746	0	2.07E-08
530	0	8.84E-02	642	0	5.63E-06	748	0	1.53E-24
532	4.655987592	2.13E-06	643	0.037760487	8.14E-06	749	0	3.85E-10
533	8.304643471	1.84E-23	644	0	2.79E-02	750	0	1.06E-03
536	13.55966485	5.25E-23	645	3.308409271	1.87E-02	751	0	1.85E-05
542	0	3.96E-05	646	0.045492087	1.19E-05	752	0.013379116	9.48E-13
543	0	3.89E-02	647	6.458051453	1.08E-20	753	0	3.30E-10
545	0	1.20E-03	648	1.724848068	4.25E-04	757	0.00132474	5.96E-93
547	0	5.65E-06	649	0	4.41E-10	760	0	3.84E-09
548	0	7.05E-25	658	0	1.83E-09	761	0	1.15E-02
549	0	2.81E-02	660	0	4.82E-03	765	0	4.21E-02
550	0	6.74E-08	664	0	3.63E-06	769	0.083074548	2.77E-02
554	0	1.00E-03	667	0	1.27E-08	773	0.046159765	2.24E-02
555	0	3.09E-04	668	0	1.13E-02	779	0	8.82E-02
558	0	4.28E-02	669	0	4.06E-02	789	0.234577623	8.25E-02
560	0	8.88E-02	670	0.223243968	3.48E-02	790	0	5.88E-02
561	0	7.23E-23	671	0	2.28E-38	792	0.1641852	4.96E-02
562	0	1.51E-50	672	0.550265888	2.56E-02	794	0	5.78E-02
565	0.008438882	2.09E-09	673	0	1.41E-03	796	0	7.36E-07
566	0	4.47E-02	676	0	2.47E-20	798	11.3647931	2.55E-05
570	0	1.18E-02	678	0	4.48E-04	801	3.22221779	4.76E-08
576	11.127212	5.67E-11	679	0	3.22E-19	802	19.47542725	2.16E-08
577	12.32261079	1.17E-33	681	0	2.61E-02	804	20.47634985	2.64E-04
578	0	4.20E-02	687	4.471065472	1.04E-03	805	0	7.81E-03
579	9.345388331	2.66E-14	689	10.39300096	1.16E-21	807	0	5.74E-03
588	26.02963007	1.58E-12	690	10.24437596	1.26E-29	823	0.11340013	9.15E-02
590	0	1.06E-03	692	3.941549008	7.34E-07	824	0	3.91E-02
592	0	3.38E-19	694	0	2.41E-03	826	0	4.03E-02
593	0	1.62E-04	695	0	3.48E-13	829	0	4.24E-02
594	0	1.50E-03	696	0	1.17E-02	830	0	8.88E-02
595	0	2.19E-04	697	0	8.73E-03	832	0	1.07E-02
598	0	1.85E-06	698	0	5.86E-02	834	3.778993968	1.22E-02
599	0	8.64E-15	701	0	1.74E-02	839	0	1.20E-02
603	0	1.62E-04	702	0.072223072	3.18E-03	841	0.515775793	2.00E-02
608	0	3.42E-08	703	0	1.85E-08	852	0	5.01E-04
611	0	2.65E-02	704	0	7.03E-04	853	0.328112805	7.94E-02
614	0.110087143	1.21E-02	705	0	1.31E-04	854	0	4.38E-03
615	0	2.32E-03	706	0	1.64E-06	856	0.269717721	1.49E-02
616	0.041354084	5.36E-03	713	4.046282285	3.41E-03	857	0	2.90E-02
617	0	3.02E-31	715	3.503798722	1.18E-02	868	3.525511326	4.79E-14
618	0	1.37E-11	717	5.485686225	4.83E-08	871	0	1.19E-02
619	0	3.95E-02	719	9.63896339	1.56E-06	872	0	1.80E-03
620	0	5.13E-10	721	7.273774143	1.75E-18	873	0	8.79E-02
621	0	1.79E-02	722	1.597571458	9.96E-02	874	0	2.31E-03
622	0	2.29E-03	724	3.029404622	1.17E-06	875	0	4.41E-03
623	0	5.88E-02	725	0	2.12E-04	877	0	1.61E-03
624	0	2.71E-07	728	0.26931352	1.19E-02	879	0	3.90E-02
625	0	3.13E-04	729	0.553877951	9.27E-02	880	0	5.25E-03
626	0	6.00E-06	730	0	3.32E-04	884	0	1.99E-08
627	0	3.29E-04	732	0.709301771	4.96E-02	890	2.61179512	6.59E-03
629	0	5.78E-02	734	0	5.02E-14	893	0	9.21E-06
630	0.006860526	9.62E-22	736	0	8.84E-02	895	0	1.05E-03
631	0	5.46E-07	739	0	1.49E-07	898	0	1.57E-05

902	0	2.75E-02	1047	4.332441954	1.12E-17	1136	0	1.06E-03
915	0	5.74E-03	1049	3.819572268	1.56E-27	1138	0	1.26E-49
923	0	8.88E-02	1050	8.656112275	4.63E-21	1139	0.010198583	3.00E-04
928	0	3.96E-05	1051	3.276316853	2.63E-03	1140	0	2.91E-04
929	0	4.85E-08	1052	38.21050925	4.79E-12	1141	0	1.14E-05
930	0	5.82E-05	1053	6.524411351	1.92E-19	1143	0	4.48E-42
931	0	2.12E-48	1054	4.222170095	8.48E-05	1144	1.977626023	1.28E-03
932	0	4.51E-36	1055	0	1.39E-03	1145	4.375146951	2.81E-32
934	0	3.88E-02	1056	8.134795976	5.37E-15	1147	0	5.23E-03
936	0	1.04E-03	1058	44.53199587	1.41E-23	1148	2.699557228	4.69E-07
939	0	2.52E-39	1061	5.400493687	1.02E-12	1149	4.8280186	1.93E-17
940	6.747007103	1.90E-15	1062	6.309346696	4.62E-16	1150	1.911171949	4.63E-05
941	2.791295114	6.35E-02	1063	8.517313886	5.94E-02	1152	3.104924825	1.52E-06
942	0.358595484	2.07E-02	1064	16.53274783	8.14E-17	1153	15.5112011	5.35E-08
944	11.9946334	4.71E-08	1065	3.911922199	1.81E-05	1155	13.56147615	5.65E-66
946	0	6.57E-02	1066	0	2.44E-04	1156	18.18159821	2.54E-42
947	4.537005519	8.85E-09	1068	7.130283291	1.99E-28	1160	2.771843675	6.12E-09
952	6.941977334	3.29E-19	1069	0	5.86E-03	1161	2.398800805	1.19E-03
953	1.528060108	1.28E-03	1070	5.733374729	3.31E-40	1163	0	8.84E-02
955	0	2.55E-09	1071	24.43447647	5.08E-29	1165	0.073012672	8.90E-30
956	0	1.24E-23	1073	12.25977086	3.78E-08	1167	0	1.52E-05
957	0	1.42E-58	1075	0	2.95E-14	1168	0	4.09E-08
958	0.004124532	6.29E-16	1081	0.129167334	9.88E-05	1169	0	5.14E-06
961	0	3.32E-04	1082	0.016724244	1.18E-98	1170	1.893108564	2.87E-03
968	0	8.68E-02	1083	0	5.12E-39	1171	0	2.43E-04
971	0	1.07E-03	1084	0.00076798	4.28E-85	1175	2.63360621	4.80E-04
973	0	6.98E-04	1085	0	1.84E-02	1179	3.026350288	3.61E-18
976	0	1.29E-04	1087	0	5.15E-06	1182	0	2.07E-08
979	0	7.63E-22	1088	0.021248097	1.59E-27	1183	6.242541179	1.31E-40
980	0	3.59E-08	1089	0	9.51E-30	1185	0	5.25E-11
984	0	1.76E-05	1091	0.007516618	8.89E-114	1186	3.421370543	2.09E-14
985	0	6.38E-07	1092	0	8.02E-03	1188	7.000670333	4.36E-49
986	0	2.94E-05	1093	0	2.69E-16	1189	5.980166164	1.39E-57
988	0	7.28E-15	1095	0.002567576	1.70E-91	1190	4.973090192	2.65E-22
990	0	1.23E-06	1096	0	2.03E-57	1191	16.04907315	1.65E-52
992	0	1.63E-20	1097	0	4.29E-02	1193	2.423138831	1.54E-02
993	0	3.53E-07	1098	0	1.83E-03	1194	9.790614353	2.45E-53
994	0	2.55E-08	1099	0	4.60E-29	1195	5.391714835	5.11E-23
995	0	5.38E-03	1100	0	3.57E-03	1196	7.812708321	4.59E-64
996	0	1.16E-10	1101	0	1.07E-14	1197	3.884207283	1.84E-29
997	0	2.66E-19	1102	0	2.27E-18	1198	0	1.31E-04
998	0	8.61E-10	1103	0.010327755	1.49E-28	1199	0.053408144	1.68E-62
999	0	6.47E-13	1104	0	2.42E-03	1200	0	5.40E-23
1007	0.031909063	4.81E-19	1106	0	8.73E-13	1202	0	2.16E-04
1010	7.442208058	5.57E-11	1107	0	1.22E-103	1204	10.27058721	1.30E-29
1011	2.242670601	4.34E-03	1108	0	2.09E-72	1205	6.913063393	5.16E-38
1012	9.426578756	1.22E-12	1109	0	6.86E-03	1206	15.68434306	2.99E-17
1015	6.605513356	1.72E-09	1110	0	2.37E-06	1207	10.0616133	4.90E-35
1016	6.764228074	4.74E-12	1113	0.136112798	4.91E-08	1208	0	1.19E-03
1018	inf	8.68E-02	1118	0	8.94E-02	1209	0	1.55E-03
1022	6.428385309	1.44E-13	1120	0	1.27E-31	1210	0	5.94E-05
1025	0	7.17E-04	1121	0	7.89E-06	1212	14.02178353	3.20E-36
1026	2.74377414	8.57E-07	1122	0	1.37E-02	1213	3.208603561	1.10E-12
1027	32.64272023	9.99E-09	1123	0	3.78E-03	1214	0.212097696	4.45E-04
1031	2.236344352	2.68E-02	1125	0	3.41E-03	1215	0.033819006	7.35E-69
1033	0.244790689	5.20E-03	1126	0	1.04E-03	1216	0	9.47E-05
1034	3.85035514	7.69E-08	1127	0	1.48E-12	1218	0.823177111	6.85E-02
1044	0.121994264	2.49E-06	1128	0	3.97E-02	1219	0	1.07E-23
1046	2.47964675	5.55E-06	1135	0	2.92E-04	1221	0	5.27E-03

1222	0	7.11E-08	1305	0	1.40E-03
1223	0	1.17E-03	1306	6.367614469	1.66E-48
1226	0	2.28E-10	1307	2.443511564	1.91E-12
1227	0.004842303	5.55E-65	1308	0.154571426	1.39E-02
1228	13.74029131	4.74E-67	1309	0	1.17E-02
1229	4.09336741	9.04E-14	1310	0	1.24E-35
1230	5.666300567	1.16E-49	1311	0	5.97E-52
1233	2.967469593	1.34E-06	1312	0	7.20E-08
1234	3.186642835	4.17E-13	1313	0	1.51E-03
1237	5.626266939	7.56E-31	1314	0	1.39E-87
1238	3.407872689	1.46E-02	1315	0	1.64E-19
1239	6.413839465	2.01E-22	1316	0	2.11E-46
1240	6.919992568	6.03E-34	1317	0	3.17E-41
1242	3.408032387	4.87E-03	1319	0	2.59E-19
1243	17.02798231	6.61E-25	1320	0	5.91E-02
1244	60.06944896	6.21E-08	1322	0	6.55E-32
1246	2.711548525	3.51E-03	1323	0	1.15E-13
1247	3.496791914	1.63E-05	1325	0	1.49E-10
1248	5.186555665	1.95E-14	1327	0	8.90E-03
1249	0	1.69E-02	1328	0.266832735	1.83E-04
1250	6.103062278	1.26E-48	1329	0.393814156	3.37E-04
1251	15.57297383	1.87E-17	1332	0.015267312	9.49E-13
1252	12.05909022	9.00E-24	1333	0	4.24E-08
1253	8.333359425	2.03E-28	1334	0	2.64E-05
1255	2.33036367	8.60E-03	1336	0	1.85E-02
1259	4.636907917	1.13E-31	1337	0	3.05E-07
1260	8.335168028	4.54E-75	1338	0.090253724	3.12E-10
1261	0	9.13E-05	1339	0.556182527	2.09E-03
1262	7.315900199	1.37E-32	1340	0.270962015	1.32E-05
1267	6.343411344	1.17E-58	1344	2.898378397	1.82E-02
1268	5.315169321	6.25E-25	1346	5.952182203	5.68E-04
1269	5.06131415	4.92E-25	1347	5.965234375	4.44E-11
1270	4.5218108	8.12E-35	1348	4.754963947	2.91E-06
1271	0	6.63E-31	1349	0.015045233	7.59E-15
1272	0	1.40E-36	1350	0	5.27E-03
1273	0.021265262	2.38E-23	1351	0	2.58E-05
1274	0	7.02E-84	1352	0	5.87E-63
1277	0.072729947	2.01E-17	1353	0.000956176	4.46E-25
1278	0	2.64E-09	1354	0	2.32E-110
1279	0.14608086	1.13E-05	1355	0	1.25E-08
1281	2.207627839	2.45E-03	1356	0	3.79E-30
1282	15.91686323	8.71E-37	1357	0	4.41E-44
1283	3.572132708	1.04E-23	1358	0	1.82E-02
1284	3.347780958	4.13E-14	1359	0.245693154	2.03E-08
1286	1.557362596	9.72E-02	1360	4.197321376	1.54E-13
1287	16.69084795	1.27E-15	1362	0	4.28E-02
1289	4.971356962	2.33E-12	1363	5.214957636	7.02E-12
1291	8.159166899	1.75E-30	1364	0	5.14E-02
1292	2.293648137	4.35E-14	1365	0	1.99E-04
1293	6.413337332	2.92E-21	1366	5.044946073	3.49E-18
1294	6.743134303	1.32E-20	1367	10.51794047	1.20E-22
1295	7.026587363	1.52E-56	1368	0	7.07E-03
1296	3.773500401	4.41E-07			
1298	4.316811596	8.74E-22			
1299	4.858251998	5.17E-23			
1300	10.16299549	6.44E-21			
1301	0	1.02E-07			
1302	5.928394706	3.78E-42			
1304	6.533686876	7.56E-31			

Supplementary Table 3.1 | Significant PDZ-insertion data. The full data for enrichment or depletion of each amino acid site for the PDZ domain insertions after 2 rounds of CRISPRi with the calculated p-values. Fold change values of 0 indicate that the clone was no longer observed when the second round of screening was sequenced. Inf indicates the clone was not sequenced in the pre-screened library.

Materials and Methods

Strains and Media

E. coli MG1655 from Qi et al. 2013, which have chromosomally integrated constitutive GFP and RFP expression, were used for *in vivo* screening, fluorescence measurements, and transformation assays(Qi et al., 2013).Cell transformation, plasmid maintenance, and verification of transformation were done as previously described using AmpR and CmR as selectable markers(Oakes et al., 2014) EZ-rich defined growth medium (EZ-RDM, Teknoka) was used for *in vivo* fluorescence assays unless otherwise noted. S.O.B medium was used for library outgrowth and screening and Figure 2D. The arC9 ligands used; 4-hydroxytamoxifen, nafoxidine, beta-estradiol, and diethylstilbestrol, were ordered from sigma and resuspended in dimethyl sulfoxide.

Transposon library construction

A modified transposon containing an antibiotic resistance marker (Supplementary figure 1) was inserted into pUC19 plasmid carrying dCas9 in an overnight *in vitro* reaction (.5 molar ratio transposon to 100ng dCas9 plasmid) using 1uL of MuA Transposase (#F-750, Thermo Fisher). Three of these reactions were carried out in parallel, the DNA was electroporated into cells and selected for the transposon antibiotic resistance to achieve > 10^7 CFU or 100,000 fold over the possible library size of 8,262. This is done in order to help ensure proper library diversity. The coding sequence of dCas9 was subsequently excised via restriction digest with BglII and XhoI, size-selected for a single successful transposon insertion and cloned, using standard molecular biology procedures, into the expression plasmid pdCas9-bacteria (Addgene ID: 44249)(Qi et al., 2013). A BsaI golden gate reaction was then used to remove the modified transposon and insert the domain of interest (PDZ, ER) in its place. No selection was used for this cloning step, but efficiencies of reaction were >99%. Completed libraries were transformed into *E. coli* from Qi et al. 2013 using electroporation.

Library sequencing and analysis

The ORF coding for the Cas9 insertion constructs were excised from plasmids via restriction digestion and then sheared to ~300 bp fragments for sequencing. All libraries were prepped with a NEBnext DNA Library Prep Kit (New England Biolabs) and sequenced on Illumina platform sequencers (MySeq and HiSeq). Sequencing data was analyzed with a custom Python pipeline that is available online at http://github.com/SavageLab/dipseq. In brief, reads were filtered to remove those that did not contain both Cas9 and sequence from the inserted domain (PDZ or ER-LBD). The inserted domain was then trimmed from the read and the remaining sequence was aligned to dCas9 to calculate the insertion site in nucleotides from the start codon. Linker sequences were extracted from the original sequence. The insertion site, linker length, and insert sequence were then used to calculate whether the insertion was in-frame and forward-oriented relative to dCas9. For such productive insertions, the AA insertion site was calculated as the C-terminal most amino acid after which the N-terminus of the insert was detected. In this manner, we ensured that reads catching the N- or C-terminal end of the insertion would result in the same calculated insertion site. This scheme was tested for correctness and recall by processing one million synthetic reads for each library. It should be noted that Mu transposases duplicates 5 bp or ~2 AA of native sequence on either side of an insertion. Therefore the C-terminal most amino acid after which the N-terminus of the insert was detected is considered to be the first AA of our synthetic insertion (e.g insertion at site 208 indicates the inserted domain begins as AA 208 in the synthetic protein chimera). Each library was sequenced twice and reads identifying N- and C-terminal insertions for the same site were used as internal technical replicates, giving four technical replicates with which to calculate fold changes and associated p-values. Fold changes were calculated using the DESeq package, which uses a negative binomial model so as not to underestimate the dispersion of read counts at each site(Anders and Huber, 2010). All p-values reported were corrected for multiple hypothesis testing using the Benjamini-Hochberg procedure implemented by DESeq. Domain Insertion Profile-sequencing data is available at https://github.com/SavageLab/arC9_data.

CRISPRi Screen and FACS selection

Screening of the PDZ libraries was accomplished by transforming and screening at least >10 fold more *E. coli* than the theoretical library size and repeating two rounds of positive screening and FACS selection as described previously(Oakes et al., 2014). Screening for the ER library followed these same methods however the primary screen and FACS selection was for function with 4-HT, the secondary screen and FACS selection was against function without 4-HT, and the final screen was for function with 4-HT on plates. From these plates colonies were picked by eye (Supplementary Figure 4) for dCas9 based repression of RFP and tested in triplicate for RFP based repression after overnight growth in 2 µM anhydrotetracycline (aTc) and with and without 10 µM 4-HT in SOB media.

CRISPRi GFP repression assays

For individual PDZ-construct testing, colonies were cultured from plates, PDZ-dCas9 plasmid DNA recovered, separated from the RFP guide plasmid (pgRNA-bacteria Addgene ID: 44251) via restriction digestion with BsaI and co-transformed with a guide plasmid to repress chromosomal GFP. GFP repression for each construct was then tested in triplicate in a 96-well microplate reader (Tecan M1000) at 37°C. PDZ clones were grown with appropriate antibiotics and 0.2 nM aTc unless otherwise noted. OD600 nm absorbance was measured for each well. GFP and OD signals were measured and blanked and the GFP (au) normalized to OD. The cultures were compared when approaching saturation (80% of the maximum OD600 nm, Figure 2) or after 12 hours of growth (Figure 1, S10). The arC9 construct was treated and assayed in the same fashion as above with induction using 2 µM aTc. Fold changes for the effect of different ligands were calculated by normalizing the GFP/OD600 of the arC9 construct to that of a vector control and dividing the fluorescent values without ligand by the fluorescent values with ligand treatment. For single cell analysis of arC9, cells were grown for ~6 hours with antibiotics and inducer, washed with PBS, and assayed for GFP fluorescence using a Sony SH800 cell sorter.

Transformation assay

E. coli containing a sgRNA plasmid which targets the genome (either the chromosomal RFP or GFP) were made electrocompetent as described previously(Oakes et al., 2014) and similar to previous work all tests were done with the same batch of electrocompetent cells to minimize transformation variability(Briner et al., 2014). *E. coli* with these self-targeting guides were electroporated with 9 fmol of either wtCas9, dCas9 plasmid, or a test construct plasmid (active PDZ-Cas9s, or arC9) in triplicate using a BTX Harvard apparatus ECM 630 High Throughput Electroporation System. Cells were recovered in 1 mL S.O.C. medium post-electroporation for 1 hr. CFU/mL was calculated by spotting 2 technical replicates of 10-fold serial dilutions onto plates containing antibiotics for both plasmids.

HEK293T-EGFP-PEST cell line creation

The d2EGFP reporter construct was created in a modified lentivirus backbone with EF1-a promoter driving the gene of interest and a second PGK promoter driving production of a gene which confers resistance to hygromycin. The EGFP is destabilized by fusion to residues 422-461 of mouse ornithine decarboxylase, giving an *in vivo* half-life of ~2 hours. 293T cells (obtained from the UC Berkeley Tissue Culture Facility, not authenticated) were transduced and selected with hygromycin (250 µg/ml). d2EGFP clones were isolated by sorting single cells into 96-well plates and characterized by intensity of d2EGFP. Lentivirus was produced by PEI (Polysciences Inc., 24765) transfection of 293T cells with gene delivery vector co-transfected with packaging vectors pspax2 and pMD2.G essentially as described by Tiscornia et al., 2006(Tiscornia et al., 2006). Cells lines were confirmed and tested for mycoplasma contamination.

HEK 293T GFP disruption assays.

GFP disruption assays were based on those previously described(Tsai et al., 2014a). Briefly, HEK cells were cultured in 10 cm dishes using Dulbecco's Modification of Eagle's Medium (DMEM) with 4.5 g/L glucose, L-glutamine, sodium pyruvate (Corning cellgro) plus 10% fetal bovine serum, 1x MEM Non-Essential Amino Acids Solution (Gibco) and Pen-Strep (gibco). One day before transfection, ~3×10^4 cells were plated into each well of a 96-well plate with the DMEM medium plus hygromycin and allowed to settle. One hour before transfection the media was removed and replaced with media with ligand treatment or vehicle control. Cells were transfected according to the manufacturer's protocol with Lipofectamine 2000 (Life Technologies) and(5 figure 3B or 50ng figure 3C, D) plasmid DNA based on (Lin et al., 2014) containing a sgRNA, either Cas9, Cas9 ER-LBD termini fusion or arC9 and a T2A-mCherry tag as described. Cells were analyzed for EGFP and mCherry expression at 48 or 72 hours post transfection using a BD LSR Fortessa high-throughput sequencer. Transfected cells were gated positive based on mCherry florescence and the percent EGFP disruption for three independent biological replicates was calculated from this gate (Supplementary Figure 13).

HEK 293T T7EI assays

HEK cells were cultured as described above. One day before transfection, ~2×10^4 cells were plated into wells of a 96-well plate with the DMEM medium. One hour before transfection the media was removed and replaced with media containing a ligand treatment or vehicle control. Cells were transfected with Lipofectamine 2000 (Life Technologies) and plasmid DNA containing sgRNA, Cas9 or arC9 and a P2A-Puromycin fusion as described according to manufacturer's protocol. Cells were selected for transfection 24 hrs later with 1.5ug/ml of Puromycin. Genomic DNA was collected and analyzed at 72 hrs via T7EI assays as previously described.(Lin et al., 2014)

BNL CL.2 GFP reporter cell line (BNL-LMP-15)

A monoclonal BNL CL.2 reporter cell line stably expressing GFP (EGFP) was derived from a single-cell clone of murine BNL CL.2 cells (ATCC) transduced with a MSCV-PGK-Puro-IRES-GFP (LMP Pten.1524) retrovirus(Fellmann et al., 2013), and grown in Dulbecco's Modified Eagle Medium (DMEM, Corning #10-013) supplemented with 10% FBS, 100 U/ml penicillin, and 100 µg/ml streptomycin. Several clones were tested for their GFP fluorescence and growth properties, and clone 15 – termed "BNL-LMP-15" – chosen for all further experiments. Cells were grown at 37°C with 5% CO2.

Lentiviral vector construction

The arC9 lentiviral vector and all variants were constructed according to the sequences provided (Supplemental Table 2), using custom oligonucleotides (IDT), standard cloning and Gibson assembly techniques(Gibson et al., 2009). The pBC2101, pBC2102 and pBC2103 lentiviral vectors (U6-sgRNA-EFS-Cas9/arC9-T2A-mCherry-P2A-Hygro, Supplemental Figure S16a) were constructed in the pRRL backbone (Dull et al., 1998) based on derivatives of SGEP (Fellmann et al., 2013). The U6 promoter and core promoter for the human elongation factor EF-1α (EFS) were based on the lenti-CRISPR-v2 plasmid(Sanjana et al., 2014). For sgRNA expression, an enhanced Streptococcus pyogenes Cas9 scaffold was used(Chen et al., 2013). All sgRNAs (Supplemental Table 2) were designed with a G preceding the 20 nt guide for better expression, and cloned into the lentiviral vectors using the BsmBI restriction sites. The pBC2201 lentiviral vector (U6-sgRNA-EFS-mCherry-P2A-Hygro) was built by removing Cas9 from pBC2101 using standard cloning techniques.

BNL-LMP-15 cell culture and editing assessment

Percentages of GFP-negative cells were assessed by flow cytometry (guava easyCyte, Millipore) on at least 10,000 acquired events. T7 endonuclease I (T7EI, NEB) assays were carried out according to manufacturer's procedures and visualized using SYBR Gold (Thermo Fisher Scientific). Transduced BNL-LMP-15 were selected 24 h post infection with 400 µg/ml hygromycin B (Sigma). Treatment regimens for arC9 expressing cells included 1 µM 4-hydroxytamoxifen (4-HT, Sigma), 1 µM beta-estradiol (B-E, Sigma), or 0.1% dimethyl sulfoxide as control (DMSO, Sigma).

Please see: http://www.nature.com/nbt/journal/v34/n6/abs/nbt.3528.html for important sequences in Supplementary table 2 not included here due to space constrictions.

Supplementary figure citations: (Schultz et al., 1998) (Qi et al., 2013),(Anders et al., 2014)(Jinek et al., 2014)(Nishimasu et al., 2014) (Jiang et al., 2015) (Heinig and Frishman, 2004)
(Dueber et al., 2009) ER(Shiau et al., 1998; Tanenbaum et al., 1998; Warnmark, 2002)E(Anders et al., 2014)(Jiang et al., 2015)(Davis et al., 2015) (Truong et al., 2015) (Zetsche et al., 2015)(Nihongaki et al., 2015)(Wright et al., 2015)

Accession codes:
The Domain Insertion Profile-sequencing data is available: PRJNA314234
Sequencing data was analyzed with a custom Python pipeline that is available online at http://github.com/SavageLab/dipseq

Conclusion:

Engineering CRISPR-Cas9 systems to expand functionality

In the early 1990s, researchers demonstrated that cleaving genomic DNA at a specific target site could induce mutations at the cut site (Rouet et al., 1994). This discovery triggered an intense and focused effort to develop enzymes that would cut and therefore mutate any specific DNA target. Twenty years of research initially yielded promising but cumbersome results; new proteins had to be engineered for each desired cleavage site (Christian et al., 2010; Kim and Chandrasegaran, 1994; Kim et al., 1996; Porteus and Baltimore, 2003; Porteus and Carroll, 2005; Urnov et al., 2005). Nevertheless, the recent introduction of RNA-guided endonucleases such as CRISPR-Cas9, that can bind and cleave specific DNA sequences based on easily programmable RNA-DNA base pairing, has made the targeting and cleavage of the genome quite straightforward (Cho et al., 2013; Cong et al., 2013; Friedland et al., 2013; Gratz et al., 2013; Hwang et al., 2013; Jinek et al., 2012; Li et al., 2015; Materials et al., 2013; Nekrasov et al., 2013; Shen et al., 2013). While CRISPR has resolved the problem of simple and programmable DNA cleavage, the possibilities of how we will harness and apply this technology beyond straightforward genome cleavage and error prone repair are just beginning to be explored (Chavez et al., 2015; Chen et al., 2013; Hilton et al., 2015; Komor et al., 2016; O'Connell et al., 2014; Qi et al., 2013; Tanenbaum et al., 2014; Zalatan et al., 2014). **The question is no longer, "How can we cut the genome where we want to?" but rather, "What else can be accomplished with this versatile technology?"**

The future of CRISPR-Cas9 lies in the ability to engineer this protein complex to suit our needs, accomplishing goals never intended when Cas9 evolved as a protective nucleic acid degradation mechanism for prokaryotes. For example, a great deal of biochemical work has enabled the rational design of Cas9 mutants that have higher specificity DNA targeting (Kleinstiver et al., 2016; Slaymaker et al., 2015). Moreover, structural understanding has facilitated the rational splitting of the Cas9 enzyme into two fragments that reconstitute activity upon co-localization. This has allowed for partially functional light, chemical or genetic regulation (Nihongaki et al., 2015; Truong et al., 2015; Wright et al., 2015; Zetsche et al., 2015). Finally, recent research has demonstrated that fusing enzymes including deaminases to Cas9 or dCas9 allows for the mutation of specific nucleotides, even without cutting both strands of the genome (Komor et al., 2016; Ma et al., 2016; Zong et al., 2017). Other Cas9 fusions to transcriptional activators, repressors and epigenetic modifiers have been created to enable the simple alteration of expression levels for any given gene up or down (Chavez et al., 2015; Gilbert et al., 2013; Hilton et al., 2015; Mali et al., 2013c; Qi et al., 2013; Zalatan et al., 2014). Nevertheless, all of these Cas9 modifications are in their infancy, are designed rationally, have minimal validation outside of their initial reports, and hence, they cannot be assumed to be ideal technologies. Here we have established a new paradigm for the generation of robust novel Cas9 functionalities by combining deep biochemical understanding with comprehensive querying of protein structure. Ultimately these techniques have led to the creation of an unbiased methodology that enabled the creation of an entirely new type of RNA guided DNA binding protein which can sense its own environment and allosterically activate in response to a small molecule. The knowledge that we 'know only that we do not know' what will work best when engineering Cas9 has enabled a suite of unbiased protein engineering technologies and protocols, that are universal and therefore can be easily adapted to any of the Class II single molecule CRISPR effector proteins such as Cpf1 or C2C2 and even used to engineer never before seen classes of RNA guided DNA binding proteins.

The prototype Cas9-biosensor described in Oakes et al. 2016 represents the first step in engineering pragmatic functions into CRISPR proteins and harnessing these tools for biotechnological purposes. In the future I foresee suites of engineered RNA guided DNA binding proteins that can sense and respond to numerous signals including those important within an organism such as immune signals and post translational modification. By sensing and responding to such cellular signals these RNA guided DNA binding proteins will be able to synthetically program prokaryotic and eukaryotic cells to undertake complex transcriptional programs upon specific stimulation, ushering a new era of programmable synthetic biology sensors and circuits.

Beyond facilitating complex, synthetic, sense-and-respond schemes, the comprehensive protein engineering approaches laid out here will enable the further development of genome editing technologies. Already it is likely that the allosteric Cas9 we have created will significantly reduce non-desirable off-target effects by limiting active nuclease dose and exposure (Davis et al., 2015; Ran et al., 2013). Moreover, while querying spyCas9 protein sequence space for mutability this work has led to a much deeper understanding of Cas9 structure by yielding insights into the necessary components for Cas9 RNA guided DNA binding. This research has significantly enhanced our knowledge and ability to engineer a protein that is quickly proving to be one of the most important tools for molecular biology and genome engineering.

BIBLIOGRAPHY

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome Biol *11*, R106.

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

Barrangou, R., and Doudna, J.A. (2016). Applications of CRISPR technologies in research and beyond. Nat. Biotechnol. *34*, 933–941.

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., and Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. Science *315*, 1709–1712.

Bikard, D., Jiang, W., Samai, P., Hochschild, A., Zhang, F., and Marraffini, L.A. (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. Nucleic Acids Res. *41*, 7429–7437.

Boch, J., Scholze, H., Schornack, S., Landgraf, a., Hahn, S., Kay, S., Lahaye, T., Nickstadt, a., and Bonas, U. (2009). Breaking the Code of DNA Binding Specificity of TAL-Type III Effectors. Science (80-.). *326*, 1509–1512.

Boeke, J.D., Lacroute, F., and Fink, G.R. (1984). A positive selection for mutants lacking orotidine-5 â€^{TM} -phosphate decarboxylase activity in yeast : 5-fluoro-orotic acid resistance. Mol. Gen. Genet. *197*, 345–346.

Briner, A.E.E.E., Donohoue, P.D.D.D., Gomaa, A.A. a. a, Selle, K., Slorach, E.M.M.M., Nye, C.H.H.H., Haurwitz, R.E.E.E., Beisel, C.L.L.L., May, A.P.P.P., and Barrangou, R. (2014). Guide RNA Functional Modules Direct Cas9 Activity and Orthogonality. Mol. Cell *56*, 333–339.

Carroll, D. (2014). Genome engineering with targetable nucleases. Annu. Rev. Biochem. *83*, 409–439.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. Nat. Methods *12*, 326–328.

Chen, B., Gilbert, L. a, Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. Cell *155*, 1479–1491.

Cho, S.W., Kim, S., Kim, J.-S.J.M.J.-S., and Kim, J.-S.J.M.J.-S. (2013). Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat. Biotechnol. *31*, 230–232.

Chomczynski, P. (1992). One-Hour Downward Alkaline Capillary Transfer for Blotting of DNA and Rna. Anal. Biochem. *201*, 134–139.

Chothia, C. (2003). Evolution of the Protein Repertoire. Science (80-.). *300*, 1701–1703.

Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, a., Bogdanove, a. J., and Voytas, D.F. (2010). Targeting DNA Double-Strand Breaks with TAL Effector Nucleases. Genetics *186*, 757–761.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic Maps of Long

Noncoding RNA Occupancy Reveal Principles of RNA-Chromatin Interactions. Mol. Cell *44*, 667–678.

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

Chylinski, K., Makarova, K.S., Charpentier, E., and Koonin, E. V. (2014). Classification and evolution of type II CRISPR-Cas systems. Nucleic Acids Res. *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 System. Science (80-.). *339*, 819–823.

Consortium, I.H.G.S., Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., et al. (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860–921.

Cradick, T.J., Fine, E.J., Antico, C.J., and Bao, G. (2013). CRISPR/Cas9 systems targeting -globin and CCR5 genes have substantial off-target activity. Nucleic Acids Res. *41*, 9584–9592.

Davis, K.M., Pattanayak, V., Thompson, D.B., Zuris, J. a, and Liu, D.R. (2015). Small molecule–triggered Cas9 protein with improved genome-editing specificity. Nat. Chem. Biol. *11*, 316–318.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature *471*, 602–607.

Doudna, J.A., and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. Science (80-.). *346*, 1258096–1258096.

Dow, L.E., Fisher, J., O'Rourke, K.P., Muley, A., Kastenhuber, E.R., Livshits, G., Tschaharganeh, D.F., Socci, N.D., and Lowe, S.W. (2015). Inducible in vivo genome editing with CRISPR-Cas9. Nat. Biotechnol. *33*, 390–394.

Dueber, J.E., Yeh, B.J., Chak, K., and Lim, W.A. (2003). Reprogramming control of an allosteric signaling switch through modular recombination. Science *301*, 1904–1908.

Dueber, J.E., Mirsky, E.A., and Lim, W.A. (2007). Engineering synthetic signaling proteins with ultrasensitive input/output control. Nat. Biotechnol. *25*, 660–662.

Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A. V, Prather, K.L.J., and Keasling, J.D. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. Nat. Biotechnol. *27*, 753–759.

Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., and Naldini, L. (1998). A third-generation lentivirus vector with a conditional packaging system. J. Virol. *72*, 8463–8471.

Edwards, W.R., Busse, K., Allemann, R.K., and Jones, D.D. (2008). Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. Nucleic Acids Res. *36*, e78.

Elowitz, M.B. (2002). Stochastic Gene Expression in a Single Cell. Science (80-.). *297*, 1183–1186.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA Exploits Three-Dimensional Genome Architecture to Spread Across the X Chromosome. Science (80-.). *341*.

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. Nat. Methods *10*, 1116–1121.

Feil, R., Brocard, J., Mascrez, B., LeMeur, M., Metzger, D., and Chambon, P. (1996). Ligand-activated site-specific recombination in mice. Proc. Natl. Acad. Sci. U. S. A. *93*, 10887–10890.

Fellmann, C., Hoffmann, T., Sridhar, V., Hopfgartner, B., Muhar, M., Roth, M., Lai, D.Y., Barbosa, I.A.M., Kwon, J.S., Guan, Y., et al. (2013). An optimized microRNA backbone for effective single-copy RNAi. Cell Rep. *5*, 1704–1713.

Filipovska, A., and Rackham, O. (2011). Designer RNA-binding proteins: New tools for manipulating the transcriptome. RNA Biol *8*, 978–983.

Friedland, A.E., Tzur, Y.B., Esvelt, K.M., Colaiácovo, M.P., Church, G.M., and Calarco, J.A. (2013). Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat. Methods *10*, 741–743.

Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M., and Joung, J.K. (2014). Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat. Biotechnol. *32*, 279–284.

Galvao, T.C., and de Lorenzo, V. (2005). Adaptation of the Yeast URA3 Selection System to Gram-Negative Bacteria and Generation of a betCDE Pseudomonas putida Strain. Appl. Environ. Microbiol. *71*, 883–892.

Garneau, J.E., Dupuis, M.E., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadan, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature *468*, 67–+.

Gasiunas, G., Barrangou, R., Horvath, P., and Siksnys, V. (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. U. S. A. *109*, E2579-86.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

González, F., Zhu, Z., Shi, Z.-D., Lelli, K., Verma, N., Li, Q. V, and Huangfu, D. (2014). An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. Cell Stem Cell *15*, 215–226.

Gossen, M., and Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. Proc. Natl. Acad. Sci. U. S. A. *89*, 5547–5551.

Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science *268*, 1766–1769.

Gratz, S.J., Cummings, A.M., Nguyen, J.N., Hamm, D.C., Donohue, L.K., Harrison, M.M., Wildonger, J., and O'connor-Giles, K.M. (2013). Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics *194*, 1029–1035.

Guilinger, J.P., Thompson, D.B., and Liu, D.R. (2014). Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. *32*, 577–582.

Hale, C.R., Zhao, P., Olson, S., Duff, M.O., Graveley, B.R., Wells, L., Terns, R.M., and Terns, M.P. (2009). RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. Cell *139*, 945–956.

Hale, C.R., Majumdar, S., Elmore, J., Pfister, N., Compton, M., Olson, S., Resch, A.M., Glover, C.V.C., Graveley, B.R., Terns, R.M., et al. (2012). Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. Mol. Cell *45*, 292–302.

Heinig, M., and Frishman, D. (2004). STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. Nucleic Acids Res. *32*, W500-2.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Hsia, K.C., Chak, K.F., Liang, P.H., Cheng, Y.S., Ku, W.Y., and Yuan, H.S. (2004). DNA binding and degradation by the HNH protein ColE7. Structure *12*, 205–214.

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. Nat. Biotechnol. *31*, 827–832.

Hsu, P.D., Lander, E.S., and Zhang, F. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell *157*, 1262–1278.

Hwang, W.Y., Fu, Y., Reyon, D., Maeder, M.L., Tsai, S.Q., Sander, J.D., Peterson, R.T., Yeh, J.-R.J., and Joung, J.K. (2013). Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat. Biotechnol. *31*, 227–229.

Jiang, F., Zhou, K., Ma, L., Gressel, S., and Doudna, J. a. (2015). A Cas9-guide RNA complex preorganized for target DNA recognition. Science (80-.). *348*, 1477–1481.

Jiang, F., Taylor, D.W., Chen, J.S., Kornfeld, J.E., Zhou, K., Thompson, A.J., Nogales, E., and Doudna, J.A. (2016). Structures of a CRISPR-Cas9 R-loop complex primed for DNA cleavage. Science (80-.). *351*, 867–871.

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 (80-.). *337*, 816–822.

Jinek, M., Jiang, F., Taylor, D.W., Sternberg, S.H., Kaya, E., Ma, E., Anders, C., Hauer, M., Zhou, K., Lin, S., et al. (2014). Structures of Cas9 Endonucleases Reveal RNA-Mediated Conformational Activation. Science (80-.). *343*, 1247997–1247997.

Kearns, N.A., Genga, R.M.J., Enuameh, M.S., Garber, M., Wolfe, S.A., and Maehr, R. (2014). Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development *141*, 219–223.

Kim, D., and Rossi, J. (2008). RNAi mechanisms and applications. Biotechniques 613–616.

Kim, Y.G., and Chandrasegaran, S. (1994). Chimeric restriction endonuclease. Proc. Natl. Acad. Sci. U. S. A. *91*, 883–887.

Kim, Y.B., Komor, A.C., Levy, J.M., Packer, M.S., Zhao, K.T., and Liu, D.R. (2017). Increasing the genome-targeting scope and precision of base editing with engineered Cas9-cytidine deaminase fusions. Nat. Biotechnol. *35*, 371–376.

Kim, Y.G., Cha, J., and Chandrasegaran, S. (1996). Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. Proc. Natl. Acad. Sci. U. S. A. *93*, 1156–1160.

Kleinstiver, B.P., Pattanayak, V., Prew, M.S., Tsai, S.Q., Nguyen, N.T., Zheng, Z., and Joung, J.K. (2016). High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects. Nature *529*, 490–495.

Komor, A.C., Kim, Y.B., Packer, M.S., Zuris, J.A., and Liu, D.R. (2016). Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature *advance on*.

Lee, H.Y., Haurwitz, R.E., Apffel, a., Zhou, K., Smart, B., Wenger, C.D., Laderman, S., Bruhn, L., and Doudna, J. a. (2013). RNA-protein analysis using a conditional CRISPR nuclease. Proc. Natl. Acad. Sci. *110*, 5416–5421.

Li, T., Huang, S., Jiang, W.Z., Wright, D., Spalding, M.H., Weeks, D.P., and Yang, B. (2011). TAL nucleases (TALNs): hybrid proteins composed of TAL effectors and FokI DNA-cleavage domain. Nucleic Acids Res. *39*, 359–372.

Li, Z., Liu, Z.-B., Xing, A., Moon, B.P., Koellhoffer, J.P., Huang, L., Ward, R.T., Clifton, E., Falco, S.C., and Cigan, A.M. (2015). Cas9-Guide RNA Directed Genome Editing in Soybean. Plant Physiol. *169*, 960–970.

Lin, S., Staahl, B., Alla, R.K., and Doudna, J.A. (2014). Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. Elife *3*, e04766.

Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat. Methods.

Mackay, J.P., Font, J., and Segal, D.J. (2011). The prospects for designer single-stranded RNA-binding proteins. Nat Struct Mol Biol *18*, 256–261.

Mali, P., Esvelt, K.M., and Church, G.M. (2013a). Cas9 as a versatile tool for engineering biology. Nat. Methods *10*, 957–963.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013b). RNA-Guided Human Genome Engineering via Cas9 Prashant. Science (80-.). *339*, 823–826.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013c). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. *31*, 833–838.

Marraffini, L.A., and Sontheimer, E.J. (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. Nature *463*, 568–571.

Materials, S.I., Hou, Z., Zhang, Y., Propson, N.E., Howden, S.E., Chu, L.-F., Sontheimer, E.J., and Thomson, J. a (2013). Efficient genome engineering in human pluripotent stem cells using

Cas9 from Neisseria meningitidis. Proc. Natl. Acad. Sci. U. S. A. *110*, 2–3.

McIsaac, R.S., Oakes, B.L., Wang, X., Dummit, K. a., Botstein, D., and Noyes, M.B. (2013). Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. Nucleic Acids Res. *41*, e57–e57.

Mojica, F.J.M., Diez-Villasenor, C., Garcia-Martinez, J., and Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology-Sgm *155*, 733–740.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J.D.G., and Kamoun, S. (2013). Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. Nat. Biotechnol. *31*, 691–693.

Nihongaki, Y., Kawano, F., Nakajima, T., and Sato, M. (2015). Photoactivatable CRISPR-Cas9 for optogenetic genome editing. Nat. Biotechnol. *9*, 1–8.

Nishida, K., Arazoe, T., Yachie, N., Banno, S., Kakimoto, M., Tabata, M., Mochizuki, M., Miyabe, A., Araki, M., Hara, K.Y., et al. (2016). Targeted nucleotide editing using hybrid prokaryotic and vertebrate adaptive immune systems. Science (80-.).

Nishimasu, H., Ran, F.A., Hsu, P.D., Konermann, S., Shehata, S.I., Dohmae, N., Ishitani, R., Zhang, F., and Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. Cell *156*, 935–949.

Niu, Y., Shen, B., Cui, Y., Chen, Y., Wang, J., Wang, L., Kang, Y., Zhao, X., Si, W., Li, W., et al. (2014). Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-MediatedGene Targeting in One-Cell Embryos. Cell *156*, 836–843.

Nourry, C., Grant, S.G.N., and Borg, J.-P. (2003). PDZ Domain Proteins: Plug and Play! Sci. Signal. *2003*, re7-re7.

O'Connell, M.R., L. Oakes, B., Sternberg, S.H., East-Seletsky, A., Kaplan, M., and Doudna, J. a. (2014). Programmable RNA recognition and cleavage by CRISPR/Cas9. Nature *516*, 263–266.

Oakes, B.L., Nadler, D.C., and Savage, D.F. (2014). Protein Engineering of Cas9 for Enhanced Function. Methods Enzymol. *546C*, 491–511.

Oakes, B.L., Xia, D.F., Rowland, E.F., Xu, D.J., Ankoudinova, I., Borchardt, J.S., Zhang, L., Li, P., Miller, J.C., Rebar, E.J., et al. (2016). Multi-reporter selection for the design of active and more specific zinc-finger nucleases for genome editing. Nat. Commun. *7*, 10194.

Olson, E.J., and Tabor, J.J. (2014). Optogenetic characterization methods overcome key challenges in synthetic and systems biology. Nat. Chem. Biol. *10*, 502–511.

Ostermeier, M. (2005). Engineering allosteric protein switches by domain insertion. Protein Eng. Des. Sel. *18*, 359–364.

Pattanayak, V., Lin, S., Guilinger, J.P., Ma, E., Doudna, J.A., and Liu, D.R. (2013). High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nat. Biotechnol. *31*, 839–843.

Pavletich, N.P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science *252*, 809–817.

Peisajovich, S.G., Garbarino, J.E., Wei, P., and Lim, W. a. (2010). Rapid Diversification of Cell

Signaling Phenotypes by Modular Domain Recombination. Science (80-.). *328*, 368–372.

Persikov, A. V, Rowland, E.F., Oakes, B.L., Singh, M., and Noyes, M.B. (2013). Deep sequencing of large library selections allows computational discovery of diverse sets of zinc fingers that bind common targets. Nucleic Acids Res. *42*, 1–12.

Persikov, a. V., Wetzel, J.L., Rowland, E.F., Oakes, B.L., Xu, D.J., Singh, M., and Noyes, M.B. (2015). A systematic survey of the Cys2His2 zinc finger DNA-binding landscape. Nucleic Acids Res. *43*, 1965–1984.

Pommer, a J., Cal, S., Keeble, A.H., Walker, D., Evans, S.J., Kuhlmann, U.C., Cooper, a, Connolly, B. a, Hemmings, A.M., Moore, G.R., et al. (2001). Mechanism and cleavage specificity of the H-N-H endonuclease colicin E9. J. Mol. Biol. *314*, 735–749.

Porteus, M.H., and Baltimore, D. (2003). Chimeric Nucleases Stimulate Gene Targeting in Human Cells. Science (80-.). *300*, 763–763.

Porteus, M.H., and Carroll, D. (2005). Gene targeting using zinc finger nucleases. Nat. Biotechnol. *23*, 967–973.

Qi, L.S., Larson, M.H., Gilbert, L. a, Doudna, J. a, Weissman, J.S., Arkin, A.P., and Lim, W. a (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Ran, F.A., Hsu, P.D., Lin, C.-Y.Y., Gootenberg, J.S., Konermann, S., Trevino, A.E., Scott, D. a., Inoue, A., Matoba, S., Zhang, Y., et al. (2013). Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. Cell *154*, 1380–1389.

Reynolds, K.A., McLaughlin, R.N., and Ranganathan, R. (2011). Hot Spots for Allosteric Regulation on Protein Surfaces. Cell *147*, 1564–1575.

Rouet, P., Smih, F., and Jasin, M. (1994). Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell. Biol. *14*, 8096–8106.

Sampson, T.R., and Weiss, D.S. (2014). Exploiting CRISPR/Cas systems for biotechnology. Bioessays *36*, 34–38.

Sampson, T.R., Saroj, S.D., Llewellyn, A.C., Tzeng, Y.-L., and Weiss, D.S. (2013). A CRISPR/Cas system mediates bacterial innate immune evasion and virulence. Nature *497*, 254–257.

Sanjana, N.E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nat. Methods *11*, 783–784.

Sashital, D.G., Wiedenheft, B., and Doudna, J.A. (2012). Mechanism of foreign DNA selection in a bacterial adaptive immune system. Mol. Cell *46*, 606–615.

Schmidt-Dannert, C., and Arnold, F.H. (1999). Directed evolution of industrial enzymes. Trends Biotechnol. *17*, 135–136.

Schultz, J., Hoffmuüller, U., Krause, G., Ashurst, J., Macias, M.J., Schmieder, P., Schneider-Mergener, J., and Oschkinat, H. (1998). Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. Nat. Struct. Biol. *5*, 19–24.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D. a, Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science *343*, 84–87.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., Zhang, K., Liu, J., Xi, J.J., Qiu, J.-L., et al. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. Nat. Biotechnol. *31*, 686–688.

Shechner, D.M., Hacisuleyman, E., Younger, S.T., and Rinn, J.L. (2015). Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. Nat. Methods *12*, 664–670.

Shen, B., Zhang, J., Wu, H., Wang, J., Ma, K., Li, Z., Zhang, X., Zhang, P., and Huang, X. (2013). Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Nat. Publ. Gr. *23*, 720–723.

Shiau, A.K., Barstad, D., Loria, P.M., Cheng, L., Kushner, P.J., Agard, D. a., and Greene, G.L. (1998). The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. Cell *95*, 927–937.

Simon, M.D., Wang, C.I., Kharchenko, P. V, West, J.A., Chapman, B.A., Alekseyenko, A.A., Borowsky, M.L., Kuroda, M.I., and Kingston, R.E. (2011). The genomic binding sites of a noncoding RNA. Proc. Natl. Acad. Sci. U. S. A. *108*, 20497–20502.

Slaymaker, I.M., Gao, L., Zetsche, B., Scott, D.A., Yan, W.X., and Zhang, F. (2015). Rationally engineered Cas9 nucleases with improved specificity. Science (80-.). *351*, 84–88.

Spilman, M., Cocozaki, A., Hale, C., Shao, Y., Ramia, N., Terns, R., Terns, M., Li, H., and Stagg, S. (2013). Structure of an RNA Silencing Complex of the CRISPR-Cas Immune System. Mol. Cell *52*, 146–152.

Staals, R.H.J., Agari, Y., Maki-Yonekura, S., Zhu, Y., Taylor, D.W., van Duijn, E., Barendregt, A., Vlot, M., Koehorst, J.J., Sakamoto, K., et al. (2013). Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. Mol. Cell *52*, 135–145.

Stein, V., and Alexandrov, K. (2015). Synthetic protein switches: design principles and applications. Trends Biotechnol. *33*, 101–110.

Sternberg, S.H., Haurwitz, R.E., and Doudna, J. a. (2012). Mechanism of substrate selection by a highly specific CRISPR endoribonuclease. Rna *18*, 661–672.

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*, 62–67.

Tanenbaum, D.M., Wang, Y., Williams, S.P., and Sigler, P.B. (1998). Crystallographic comparison of the estrogen and progesterone receptor’s ligand binding domains. Proc Natl Acad Sci U S A *95*, 5998–6003.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. Cell *159*, 635–646.

Terns, R.M., and Terns, M.P. (2014). CRISPR-based technologies: prokaryotic defense weapons repurposed. Trends Genet *30*, 111–118.

Tiscornia, G., Singer, O., and Verma, I.M. (2006). Production and purification of lentiviral vectors. Nat. Protoc. *1*, 241–245.

Truong, D.-J.J., Kühner, K., Kühn, R., Werfel, S., Engelhardt, S., Wurst, W., and Ortiz, O.

(2015). Development of an intein-mediated split-Cas9 system for gene therapy. Nucleic Acids Res.

Tsai, S.Q., Wyvekens, N., Khayter, C., Foden, J.A., Thapar, V., Reyon, D., Goodwin, M.J., Aryee, M.J., and Joung, J.K. (2014a). Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. Nat. Biotechnol. *32*, 569–576.

Tsai, S.Q., Zheng, Z., Nguyen, N.T., Liebers, M., Topkar, V. V, Thapar, V., Wyvekens, N., Khayter, C., Iafrate, a J., Le, L.P., et al. (2014b). GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. Nat. Biotechnol. *33*, 187–197.

Tucker, C.L., and Fields, S. (2001). A yeast sensor of ligand binding. Nat. Biotechnol. *19*, 1042–1046.

Urnov, F.D., Miller, J.C., Lee, Y.-L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D., and Holmes, M.C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. Nature *435*, 646–651.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., and Jaenisch, R. (2013a). One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Cell *153*, 910–918.

Wang, Y., Wang, Z., and Tanaka Hall, T.M. (2013b). Engineered proteins with Pumilio/fem-3 mRNA binding factor scaffold to manipulate RNA metabolism. FEBS J *280*, 3755–3767.

Warnmark, a. (2002). Interaction of Transcriptional Intermediary Factor 2 Nuclear Receptor Box Peptides with the Coactivator Binding Site of Estrogen Receptor alpha. J. Biol. Chem. *277*, 21862–21868.

Wiedenheft, B., Sternberg, S.H., and Doudna, J. a. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331–338.

Wright, A. V, Sternberg, S.H., Taylor, D.W., Staahl, B.T., Bardales, J.A., Kornfeld, J.E., and Doudna, J.A. (2015). Rational design of a split-Cas9 enzyme complex. Proc. Natl. Acad. Sci. U. S. A. *112*, 2984–2989.

Wu, H., Lima, W.F., and Crooke, S.T. (1999). Properties of cloned and expressed human RNase H1. J. Biol. Chem. *274*, 28270–28278.

Yagi, Y., Nakamura, T., and Small, I. (2013). The potential for manipulating RNA with pentatricopeptide repeat proteins. Plant J.

Yano, T., Sanders, C., Catalano, J., and Daldal, F. (2005). Bidirectional Selection for Integration of Unmarked Alleles into the Chromosome of Rhodobacter capsulatus sacB – 5-Fluoroorotic Acid – pyrE -Based Bidirectional Selection for Integration of Unmarked Alleles into the Chromosome of Rhodobacter capsulatus.

Yin, P., Li, Q., Yan, C., Liu, Y., Liu, J., Yu, F., Wang, Z., Long, J., He, J., Wang, H.-W., et al. (2013). Structural basis for the modular recognition of single-stranded RNA by PPR proteins. Nature *504*, 168–171.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., et al. (2014). Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. Cell *160*, 339–350.

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

Zhou, Y., Zhu, S., Cai, C., Yuan, P., Li, C., Huang, Y., and Wei, W. (2014). High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. Nature *509*, 487–491.

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.-L., Wang, D., and Gao, C. (2017). Precise base editing in rice, wheat and maize with a Cas9- cytidine deaminase fusion. Nat. Biotechnol.