nature biotechnology

Profiling of engineering hotspots identifies an allosteric CRISPR-Cas9 switch

Benjamin L Oakes[1], Dana C Nadler[1], Avi Flamholz[1], Christof Fellmann[1], Brett T Staahl[1], Jennifer A Doudna[1–5] & David F Savage[1,5,6]

The clustered, regularly interspaced, short palindromic repeats (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 disruption of 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 showed 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.

Domain insertions—the transfer of coding sequences for unique folds from one open reading frame (ORF) to another—are common occurrences throughout protein evolution[1,2]. These events facilitate the diversification and rapid functionalization of new types of protein modules that fill needed roles in both eukaryotic and prokaryotic genomes[3]. In this study we investigate how synthetic domain insertions can be used to functionalize the programmable *S. pyogenes* endonuclease Cas9 (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 noncoding genomic elements in a wide variety of organisms[4–6]. Recently, there have been numerous attempts to develop Cas9 variants with additional functions by fusing protein domains directly to its N or C terminus[7–9]. 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 such fusions[10,11]. This close spatial proximity can lead to steric incompatibility and may explain the relative lack of activity for many fusions, such as that with the viral transcriptional activator VP16 (ref. 12). Another approach would be to identify insertion points within Cas9 that are capable of accepting functional domains, as occurs naturally in evolution[1–3]. 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[5,6,13,14].

Here we profile the inherent plasticity of the Cas9 structure by examining its ability to tolerate a synthetic domain insertion while retaining RNA-guided DNA-binding activity. We created an unbiased Cas9 insertion library using randomized transposition (**Fig. 1a**). Briefly, an engineered Mu transposon possessing an antibacterial selection marker flanked by BsaI endonuclease sites was inserted randomly throughout a plasmid encoding a catalytically inactive Cas9 (dCas9) by *in vitro* transposition[15] (**Fig. 1a** and **Supplementary Fig. 1**). After selection and subcloning to isolate plasmids with single transposition events within the dCas9 ORF, we characterized the library by deep sequencing. This analysis revealed that the transposition library possessed good insertion coverage, with insertions at >70% of all possible amino acid 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. 1a** and **Supplementary Figs. 1** and **2**).

We used the PDZ domain of mouse α1-syntrophin (PDZ) as a proof-of-concept insertion domain owing to its small size (86 aa), well-folded nature and adjacent N and C termini[16,17] (**Supplementary Fig. 3a**). We hypothesized that this domain would be minimally perturbative and would act as a molecular 'potentiometer'—i.e., the capacity of Cas9 to accommodate a PDZ insertion would be indicative of the general insertion potential of a given amino acid site within the protein. Moreover, as PDZs are known protein–protein interaction domains, the PDZ-Cas9 constructs identified here may be further used as protein scaffolds to recruit other domains for editing, epigenetic modification and activation or repression purposes[17] (**Supplementary Fig. 3b**).

We cloned a PDZ domain with flanking amino acid linkers into the naive library (**Supplementary Figs. 1** and **2**) and subjected this library

[1]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, California, USA. [2]Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, California, USA. [3]Innovative Genomics Initiative, University of California, Berkeley, Berkeley, California, USA. [4]Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA. [5]Department of Chemistry, University of California, Berkeley, Berkeley, California, USA. [6]Energy Biosciences Institute, University of California, Berkeley, Berkeley, California, USA. Correspondence should be addressed to D.F.S. (savage@berkeley.edu).

Received 24 July 2015; accepted 7 March 2016; published online 2 May 2016; doi:10.1038/nbt.3528

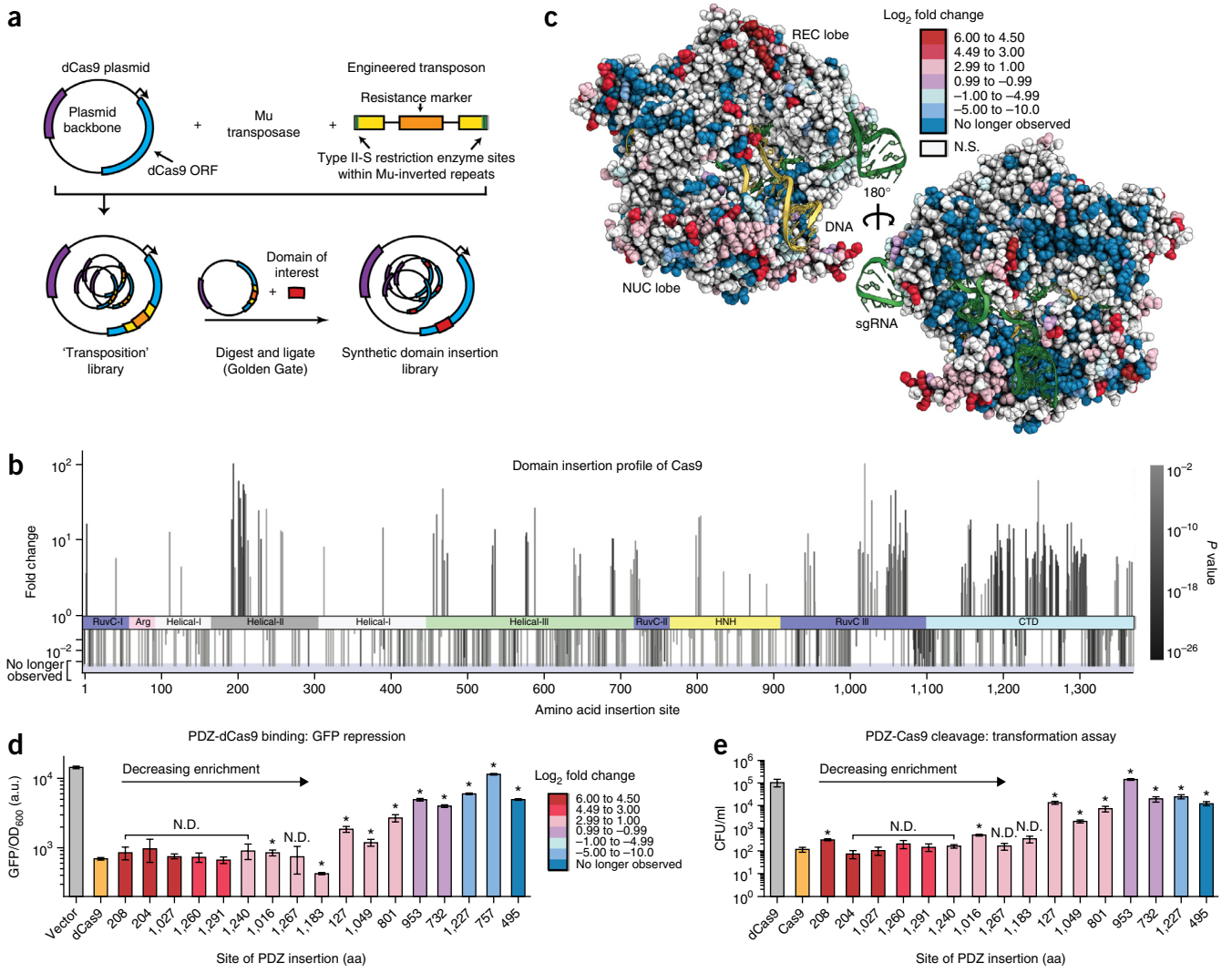

Figure 1 Mapping the insertion potential of Cas9 with the PDZ domain. (**a**) Generation of the transposon-based domain insertion library. (**b**) Fold-change values for insertions at specific amino acid sites derived from sequencing data over two rounds of screening. A positive value indicates the preference 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. Bars that reach 10^2 represent sites that were not sequenced before screening. Bars that extend into the shaded region represent clones that were cleared from the library (i.e., were not observed after screening). *P* values were determined in DESeq with multiple-hypothesis-testing correction. (**c**) Log_2 fold-change values from **b** mapped onto the structure of Cas9 (PDB 4UN3)[11]. N.S., not significant. (**d**) GFP repression activity of individual PDZ insertion sites. Data represent mean ± s.d. ($n = 3$ biological replicates from one experiment); constructs are arranged in descending order of fold change from sequencing. dCas9, positive control; vector only, negative control. N.D., no difference detected from dCas9 ($P > 0.01$); $^*P < 0.01$ (*t*-test). (**e**) Cleavage activity of clones via an *E. coli*–based transformation assay; data represent biological triplicates and s.d. ($n = 3$ assayed for cfu/mL in technical triplicate plating assays in one experiment). Cas9, positive control; dCas9, negative control. N.D., no difference detected from WT Cas9 ($P > 0.01$); $^*P < 0.01$ (*t*-test). CFU, colony-forming units; a.u., arbitrary units.

to two rounds of a CRISPR interference (CRISPRi) screen[18,19]. Briefly, cells expressing RFP and GFP were assayed by fluorescence-activated cell sorting (FACS) to identify Cas9 variants capable of repressing RFP in a single-guide RNA (sgRNA)-dependent fashion[18,19] (**Supplementary Fig. 4**). After sorting, the PDZ-dCas9 libraries were subjected to deep sequencing to identify insertion sites. Comparison of the transposition and PDZ libraries revealed that whereas the unscreened PDZ library was enriched in frameshift and reverse mutations, 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 sites in Cas9 (**Fig. 1b**,**c** and **Supplementary Table 1**). Domain insertions at the majority of residues within Cas9 were strongly selected against, with the bulk of clones falling out of the pool by the second round of screening (**Fig. 1b**,**c** and **Supplementary Table 1**). Sites with negative fold changes were highly overrepresented in crucial 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. 1c** and **Supplementary Fig. 6**). Nevertheless, we identified small local clusters of amino acids that were tolerant to insertions and recovered a total of 175 statistically significant ($P < 0.1$)

sites enriched ≥2-fold (**Fig. 1b**,**c**). When mapped onto the holo Cas9–DNA–RNA crystal structure[11], 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 at many locations within distinct Cas9 domains: at six clusters within the helical recognition (REC) lobe, in the linker between the REC and nuclease lobes, in the HNH domain, at three extended sites in the RuvCIII region and throughout the PAM-interacting (C-terminal) domain (**Fig. 1b**,**c** and **Supplementary Fig. 7**). These insertion sites provide access to both 5′ and 3′ ends of the bound DNA (~10 Å) as well as to the groove hypothesized to hold the nontargeted DNA strand[10,11]. Consequently, these insertions might allow engineering of specific and functional interactions with bound DNA. Although insertions were enriched in undefined regions from three Cas9 X-ray crystal structures, they also occurred in nearly every secondary structure element of Cas9 (**Supplementary Fig. 8**). This poses the possibility that the often-used rational design strategy of inserting domains into flexible loops underestimates the protein space available for manipulation. Moreover, as the insertions into an α-helix or β-sheet presumably disrupt the secondary structure in these areas, such insertions may be informative in future efforts to dissect Cas9 structure and function.

To confirm the binding activity of dCas9 proteins with PDZ domain insertions, we isolated plasmid DNA for individual clones at random at all stages of selection and used them to cotransform *E. coli* with a GFP-targeting sgRNA expression plasmid. We found that effective GFP repression corresponded well with the calculated fold change of the insertion site; highly enriched clones performed at levels comparable to that of wild-type (WT) dCas9 (ref. 18) (**Fig. 1d** and **Supplementary Fig. 9**). To determine whether clones with WT Cas9 and PDZ insertions (PDZ-Cas9) also possessed nuclease activity, the catalytic residues (D10 and H840) were reintroduced and tested in an *E. coli*–based transformation assay[4,20]. In this assay, nuclease activity leads to genomic cleavage and cell death. Cleavage activity correlated well with fold change; the most highly enriched PDZ-Cas9 clones, except for one containing an insertion at residue 208, maintained levels of cleavage-induced cell death similar to that of WT Cas9 (**Fig. 1e** and **Supplementary Fig. 9**). Thus, it is possible to insert an entire exogenous domain at numerous sites within the Cas9 primary structure while maintaining near-native levels of activity.

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 adaptor protein Crk (SH3)[17]. The SH3 domain was inserted into highly enriched (>10-fold) sites chosen to represent a diverse sampling of the Cas9 primary structure. We found that all clones were functional for GFP repression, and seven of the eight obtained near-native activity levels (**Supplementary Fig. 10**). Finally, 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 previously validated PDZ-dCas9 and SH3-dCas9 insertions as well as three PDZ insertions, a PDZ–C terminus fusion and two SH3 insertions and created stacked constructs with up to three separate domain insertions and a terminal fusion. Although higher numbers of insertions and fusions perturbed binding and repression activity, many of the stacked constructs were capable of repressing GFP to a degree comparable to that of dCas9 (**Supplementary Fig. 10**).

Because the design of synthetic protein switches and sensors is limited by the difficulty of predicting insertion sites that have the potential to confer allostery[21], we next explored whether domain insertion profiling could be used to reveal sites in Cas9 amenable to allosteric coupling[14,21]. For a proof-of-concept experiment, we used the well-studied ligand-binding domain (residues 302–552) of human estrogen receptor-α (ER-LBD)[22–24]. Crystallographic data show that this domain adopts 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 the apo or agonist-bound form (**Supplementary Fig. 11**). This conformational switch can serve as a mechanism by which an allosteric signal is transduced[25].

In order to create an allosterically regulated Cas9 (arC9), we introduced ER-LBD into the naive dCas9 transposition library as described

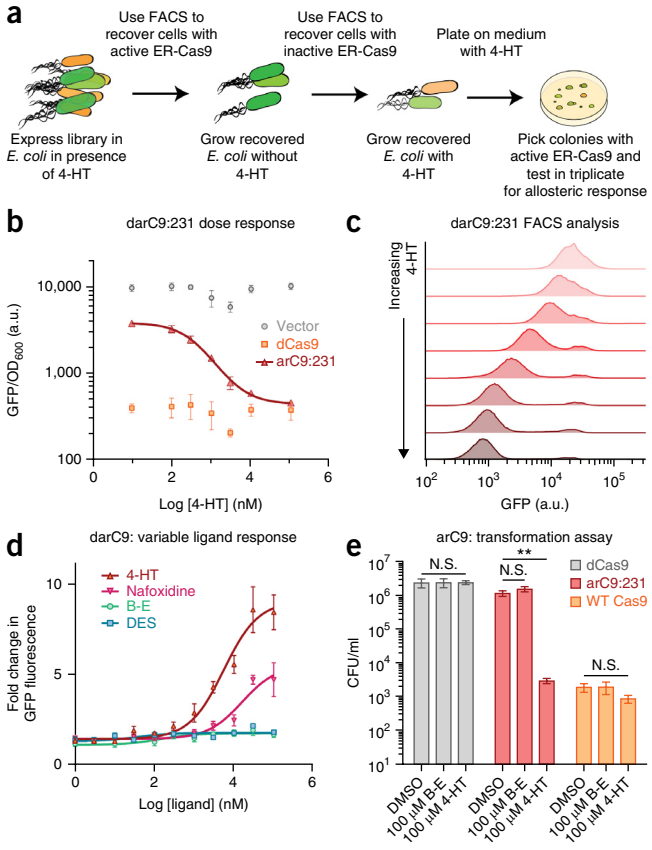

Figure 2 Creation of a switch-like Cas9 though insertion of the ER-LBD in *E. coli*. (**a**) Schematic of the screen–counter-screen procedure to select for ligand responsive ER-LBD (ER-Cas9) insertions. (**b**) Dose-response curve of darC9:231 to 4-HT. darC9:231 has a half-maximal effective concentration (EC_{50}) of 440 ± 70 nM (s.d.) and a Hill coefficient of 1.04, as expected for noncooperative binding of 4-HT to the ER-LBD. Data represent the mean ± s.d. ($n = 3$) from one experiment using biological replicates containing the respective constructs. (**c**) Single-cell analysis of darC9:231 binding in response to increasing concentrations of 4-HT. Flow cytometry ensemble data from one experiment (n >10,000 events). (**d**) Dose response of darC9:231 binding to various ligands. Response is normalized to fluorescence of vector control under the same conditions. B-E, β-estradiol; DES, diethylstilbestrol. Data represent the mean ± s.d. ($n = 3$) from one experiment of biological replicates. (**e**) Switching of arC9:231 cleavage activity. Transformation assays show that ligand-dependent arC9 switching also extends to cleavage activity. $^{**}P < 0.01$, t-test. Data represent the mean ± s.d. ($n = 3$) from one experiment of biological replicates assayed for cfu/mL in technical triplicate plating assays containing the respective constructs. CFU, colony-forming units; a.u., arbitrary units; N.S., not significant.

above and passaged it 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 sgRNA targeting GFP and assayed for repression in *E. coli*. We identified an insertion site at residue 231 that demonstrated a 4-HT-dependent decrease in GFP fluorescence, indicating switch-like behavior (**Supplementary Fig. 12**). Notably, this site was also enriched in the PDZ profile of Cas9 (**Supplementary Table 1**).

Insertion of ER-LBD at residue 231 (arC9:231) was first characterized in *E. coli*. A catalytically dead arC9:231 (darC9) showed 4-HT-dose-dependent repression of GFP with a ~10-fold change in activity in pooled and single-cell experiments (**Fig. 2b**,**c**). darC9 therefore represents a tunable CRISPRi effector. darC9 also showed clear ligand discrimination. CRISPRi-based repression of GFP was higher with ligands 4-HT and nafoxidine, which encourage the ER-LBD to enter an antagonist conformation, than with β-estradiol and diethylstilbestrol, which promote the agonist conformation (**Fig. 2d**). This further supports the argument that the ER-LBD insertion at residue 231 is able to transduce ligand-specific binding of 4-HT into Cas9 activity through induction of the antagonist ER-LBD conformation. To determine whether arC9:231 also exhibited allosteric control over cleavage activity, we reintroduced the catalytic residues (D10 and H840) and tested arC9:231 in the *E. coli* transformation assay described above (**Fig. 2e**). We found that arC9 increased chromosomal cleavage and death by at least 100-fold in the presence of 4-HT as compared with a DMSO vehicle control ($P < 0.01$), indicating that allosteric modulation of arC9 also extends to cleavage activity.

W 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, EGFP fused to the degradation domain of mouse ornithine decarboxylase, and assayed for EGFP disruption[26] (**Fig. 3a**). After 72 h, WT Cas9 was found to disrupt 91% of the EGFP signal regardless of treatment condition (**Fig. 3b** and **Supplementary Fig. 12**). A sixfold induction of arC9-mediated EGFP disruption upon addition of 4-HT (10.9 ± 0.5% to 66 ± 1%, s.d.) (**Fig. 3b**) was observed over the same time period. We also found that higher expression conditions led to an increase of EGFP disruption in the absence of 4-HT (**Supplementary Fig. 13**).

To improve arC9 control, we sought to take advantage of the known ligand-dependent nuclear localization activity of ER-LBD[27]. We removed the dual NLS from the arC9 construct and found that its EGFP-disruption activity was reduced to background levels in the absence of 4-HT. In the presence of 300 nM 4-HT, arC9's disruption

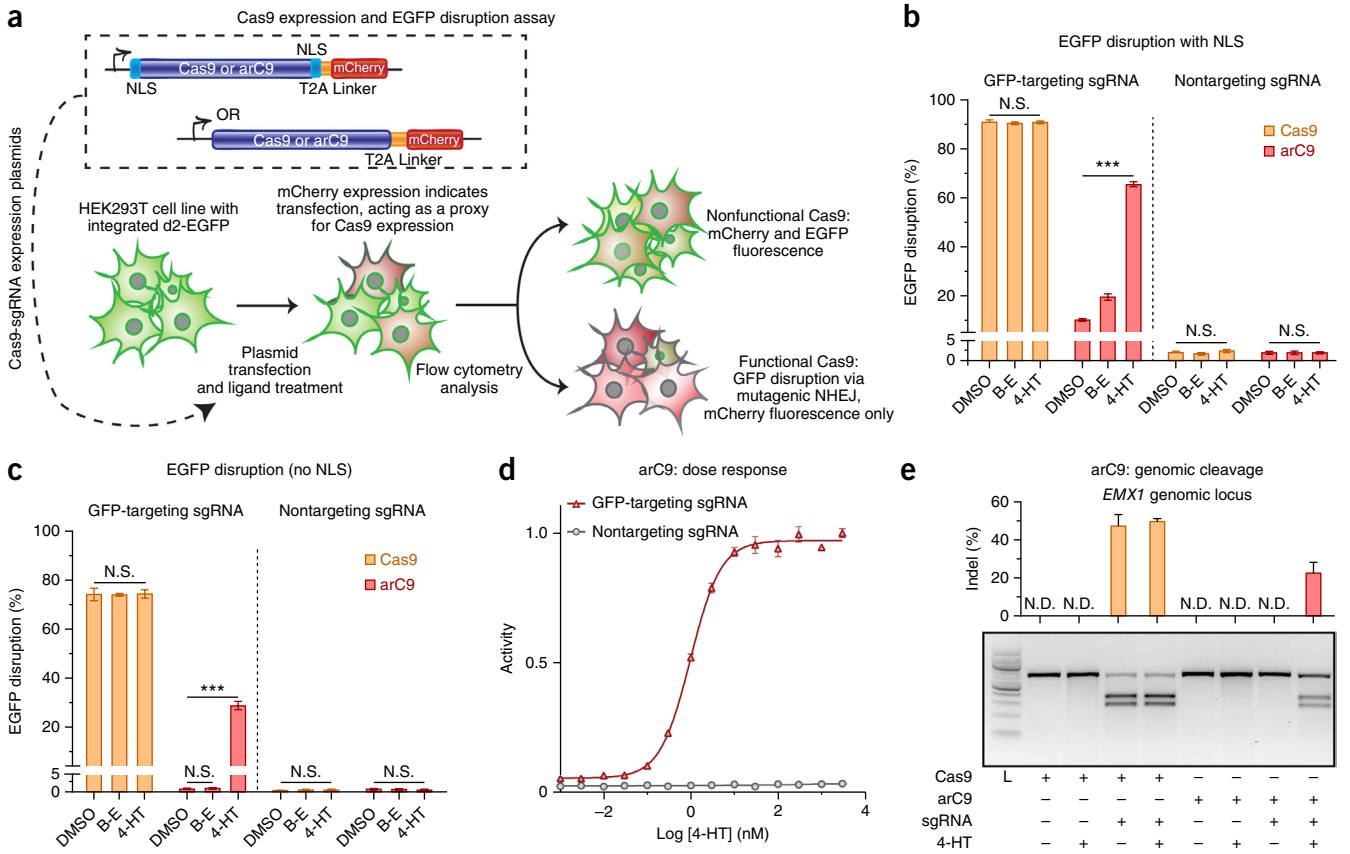

Figure 3 Validation of arC9 in eukaryotic cells. (**a**) Schematic of the arC9:231 expression constructs and EGFP disruption assay. (**b**) Quantification of EGFP disruption at 72 h for Cas9 and arC9 with an N- and C-terminal NLS ($n = 3$). Error bars represent mean ± s.d. of biological replicates from one experiment. (**c**) Quantification EGFP disruption at 72 h for Cas9 and arC9 without an NLS ($n = 3$, biological replicates from one experiment). Background activity of arC9 was not significantly different from a nontargeting negative control; $^{***}P < 0.001$ (*t*-test). (**d**) Dose response of arC9 without NLS, normalized to maximum activity. (**e**) T7EI assay of Cas9- and arC9-mediated indel formation at the *EMX1* locus at 72 h. Indel modification efficiency estimated via 2% TAE agarose gel electrophoresis analysis (below); quantification of the T7EI cleavage intensities are shown above. The 2% TAE agarose gel is a representative image and the quantification data represent the mean + s.d. of biological replicates ($n = 3$) from one experiment. N.D., not detected; N.S., not significant.

activity increased to 30 ± 2% (s.d.) that of Cas9, representing at least a 24-fold increase (**Fig. 3c**). Using the EGFP disruption assay, we observed arC9 dose-dependent response, with a half-maximal effective concentration (EC_{50}) of 1.0 ± 0.2 nM (s.d.) and full induction at 100–1,000 nM 4-HT (**Fig. 3d**). The lack of background disruption suggests that arC9 can be adapted for tight, reversible control of Cas9 cleavage activity.

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[27,28]. 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 controlled Cas9 EGFP disruption, providing a modest (0.21-fold) increase in repression upon the addition of 4-HT, with more than 50% of cells disrupted without activating ligand (**Supplementary Fig. 14**). This result is consistent with data demonstrating that WT Cas9 does not require an NLS to function in dividing human cells (**Fig. 3c**). To further validate arC9, we tested its nuclease activity on two endogenous human loci, *EMX1* and *DYRK1* (*DYRK1A*). T7 endonuclease I (T7EI) analysis showed efficient insertion or deletion (indel) formation and gene editing in the presence of 4-HT, whereas there was no detectable arC9 activity without ligand (**Fig. 3e**, **Supplementary Fig. 15** and **Supplementary Table 2**). Thus, arC9 acts as a 4-HT-inducible nuclease in human cells, demonstrating that unbiased domain insertion can be used to engineer robust Cas9 functionality not achievable with terminal fusions.

The discovery and application of CRISPR-Cas9 technology 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. Tetracyline-based systems can enable control of Cas9 expression[29–33] but are cumbersome owing to the need for additional components. We therefore tested whether arC9 could be stably integrated into the genome from a single-vector lentiviral system to 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 mouse BNL CL.2 cells with a retroviral vector expressing EGFP (**Supplementary Fig. 16a**). Upon low-copy infection with the Cas9 and arC9 lentiviral constructs and hygromycin B selection, we measured the reduction in the number of EGFP-expressing cells for arC9 over 24 d. Whereas WT Cas9 showed up to ~80% EGFP disruption with EGFP-targeting sgRNAs, arC9 with GFP-targeting sgRNAs in DMSO demonstrated levels of activity comparable to that of the non-targeting guides (**Supplementary Fig. 16c**). After 12 d, we treated a subpopulation of arC9-expressing cells with 4-HT (1 μM), β-estradiol (1 μM) or vehicle control (DMSO). In the presence of 4-HT, arC9 disrupted EGFP 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.

To test whether the activation of arC9 is reversible we recovered the 4-HT treated cells at the end of the treatment, cultured them for 2 d in regular culture medium, and then infected them with a secondary lentiviral vector expressing sgRNA targeting Pcsk9. T7EI analysis of the endogenous *Pcsk9* locus after 6 d of treatment with DMSO or 4-HT showed that arC9 can be turned off in the absence of ligand or used for controlled serial genome editing by repeated addition of 4-HT (**Supplementary Fig. 16e**). After 4-HT was removed from the medium, a small amount of residual arC9 activity remained. This may be due to the high affinity of arC9 for 4-HT (EC_{50} 1 nM) and the likelihood of slow dissociation of the complex (**Fig. 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.

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 activity of 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[10]. A recent crystal structure of sgRNA-bound Cas9 (ref. 34) helps explain the ligand dependence of arC9, suggesting that ER-LBD may disrupt the required conformational changes of the REC2 domain. This disruption would sterically occlude the RNA–DNA binding channel of Cas9 and prevent functional DNA unwinding and RNA–DNA hybridization[34] (**Supplementary Fig. 17**).

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[13,35]. Previous efforts to engineer Cas9 as a molecular scaffold have constructed systems in which the sgRNA can recruit effector proteins[36,37]. Others have built inducible Cas9 nucleases using intein splicing or the splitting of the protein itself[13,38,39]. All of these efforts have used iterative, rational design to isolate functional molecules. By contrast, we demonstrate unbiased profiling of domain insertions across the 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 to (≤2 amino acids) a small fraction of our defined insertional hotspots[13,38–40] (**Supplementary Fig. 18**). 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. arC9 functions as a conditional DNA-binding protein and nuclease in both prokaryotes and eukaryotes and displays substantial ligand-dependent activity with very low background.

METHODS

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

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

ACKNOWLEDGMENTS

We thank S. Qi (Stanford) and J. Dueber (UC Berkeley) for providing the *E. coli* strain and the PDZ and SH3 domains, respectively. We would like to thank M. O'Connell, S. Sternberg, A. Wright and S. Higgins for productive discussions and readings of the manuscript. This work was supported by a NIH New Innovator Award (1DP2EB018658-01) and a Basil O'Connor Starter Scholar Research Award from the March of Dimes (D.F.S.) and by the National Science Foundation (IQJEDMS001 to J.A.D.); A.F. is funded by a National Science Foundation Graduate Research Fellow, and B.T.S. is funded by a Roche Postdoctoral Fellowship (RPF 311).

AUTHOR CONTRIBUTIONS

B.L.O., D.C.N., C.F., J.A.D. and D.F.S. designed the research. B.L.O., D.C.N., C.F., A.F. and B.T.S. performed the experiments. A.F. performed the computational analysis. B.L.O., A.F., C.F. and D.F.S. analyzed the data. B.L.O., C.F., J.A.D. and

D.F.S. wrote the paper. Reagents described in this work are available on Addgene (https://www.addgene.org/David_Savage/).

COMPETING FINANCIAL INTERESTS

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

1. Lander, E.S. *et al.* International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Peisajovich, S.G., Garbarino, J.E., Wei, P. & Lim, W.A. Rapid diversification of cell signaling phenotypes by modular domain recombination. *Science* **328**, 368–372 (2010).
3. Chothia, C., Gough, J., Vogel, C. & Teichmann, S.A. Evolution of the protein repertoire. *Science* **300**, 1701–1703 (2003).
4. Jinek, M. *et al.* A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
5. Doudna, J.A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science* **346**, 1258096 (2014).
6. Hsu, P.D., Lander, E.S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).
7. Gilbert, L.A. *et al.* CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
8. Guilinger, J.P., Thompson, D.B. & Liu, D.R. Fusion of catalytically inactive Cas9 to FokI nuclease improves the specificity of genome modification. *Nat. Biotechnol.* **32**, 577–582 (2014).
9. Chen, B. *et al.* Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**, 1479–1491 (2013).
10. Nishimasu, H. *et al.* Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
11. Anders, C., Niewoehner, O., Duerst, A. & Jinek, M. Structural basis of PAM-dependent target DNA recognition by the Cas9 endonuclease. *Nature* **513**, 569–573 (2014).
12. Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S. & Vale, R.D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).
13. Davis, K.M., Pattanayak, V., Thompson, D.B., Zuris, J.A. & Liu, D.R. Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat. Chem. Biol.* **11**, 316–318 (2015).
14. Reynolds, K.A., McLaughlin, R.N. & Ranganathan, R. Hot spots for allosteric regulation on protein surfaces. *Cell* **147**, 1564–1575 (2011).
15. Edwards, W.R., Busse, K., Allemann, R.K. & Jones, D.D. Linking the functions of unrelated proteins using a novel directed evolution domain insertion method. *Nucleic Acids Res.* **36**, e78 (2008).
16. Schultz, J. *et al.* Specific interactions between the syntrophin PDZ domain and voltage-gated sodium channels. *Nat. Struct. Biol.* **5**, 19–24 (1998).
17. Dueber, J.E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753–759 (2009).
18. Qi, L.S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
19. Oakes, B.L., Nadler, D.C. & Savage, D.F. Protein engineering of Cas9 for enhanced function. *Methods Enzymol.* **546**, 491–511 (2014).
20. Briner, A.E.E.E. *et al.* Guide RNA functional modules direct Cas9 activity and orthogonality. *Mol. Cell* **56**, 333–339 (2014).
21. Stein, V. & Alexandrov, K. Synthetic protein switches: design principles and applications. *Trends Biotechnol.* **33**, 101–110 (2015).
22. Shiau, A.K. *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927–937 (1998).
23. Tanenbaum, D.M., Wang, Y., Williams, S.P. & Sigler, P.B. Crystallographic comparison of the estrogen and progesterone receptor's ligand binding domains. *Proc. Natl. Acad. Sci. USA* **95**, 5998–6003 (1998).
24. Wärnmark, A. *et al.* Interaction of transcriptional intermediary factor 2 nuclear receptor box peptides with the coactivator binding site of estrogen receptor-α. *J. Biol. Chem.* **277**, 21862–21868 (2002).
25. Tucker, C.L. & Fields, S. A yeast sensor of ligand binding. *Nat. Biotechnol.* **19**, 1042–1046 (2001).
26. Tsai, S.Q. *et al.* Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing. *Nat. Biotechnol.* **32**, 569–576 (2014).
27. McIsaac, R.S. *et al.* Synthetic gene expression perturbation systems with rapid, tunable, single-gene specificity in yeast. *Nucleic Acids Res.* **41**, e57 (2013).
28. Feil, R. *et al.* Ligand-activated site-specific recombination in mice. *Proc. Natl. Acad. Sci. USA* **93**, 10887–10890 (1996).
29. Dow, L.E. *et al.* Inducible *in vivo* genome editing with CRISPR-Cas9. *Nat. Biotechnol.* **33**, 390–394 (2015).
30. González, F. *et al.* An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215–226 (2014).
31. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. USA* **89**, 5547–5551 (1992).
32. Gossen, M. *et al.* Transcriptional activation by tetracyclines in mammalian cells. *Science* **268**, 1766–1769 (1995).
33. Kearns, N.A. *et al.* Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. *Development* **141**, 219–223 (2014).
34. Jiang, F., Zhou, K., Ma, L., Gressel, S. & Doudna, J.A. A Cas9-guide RNA complex preorganized for target DNA recognition. *Science* **348**, 1477–1481 (2015).
35. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
36. Zalatan, J.G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).
37. Shechner, D.M., Hacisuleyman, E., Younger, S.T. & Rinn, J.L. Multiplexable, locus-specific targeting of long RNAs with CRISPR-Display. *Nat. Methods* **12**, 664–670 (2015).
38. Nihongaki, Y., Kawano, F., Nakajima, T. & Sato, M. Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat. Biotechnol.* **33**, 755–760 (2015).
39. Zetsche, B., Volz, S.E. & Zhang, F. A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat. Biotechnol.* **33**, 139–142 (2015).
40. Truong, D.-J.J. *et al.* Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res.* **43**, 6450–6458 (2015).

ONLINE METHODS

Strains and media. *E. coli* MG1655 (ref. 18), which has chromosomally integrated constitutive GFP and RFP expression was used for *in vivo* screening, fluorescence measurements and transformation assays. Cell transformation, plasmid maintenance, and verification of transformation were done as described using AmpR and CmR as selectable markers[19]. EZ-rich defined growth medium (EZ-RDM, Teknoka) was used for *in vivo* fluorescence assays unless otherwise noted. SOB medium was used for library outgrowth, screening and for experiments shown in **Figure 2d**. arC9 ligands 4-HT, nafoxidine, β-estradiol, and diethylstilbestrol were from Sigma-Aldrich and resuspended in dimethyl sulfoxide. No blinding or randomization was done for any of the experiments reported.

Transposon library construction. A modified transposon containing an antibiotic resistance marker (**Supplementary Table 2** and **Supplementary Fig. 1**) was inserted into pUC19 plasmid carrying dCas9 in an overnight *in vitro* reaction (0.5 molar ratio transposon to 100 ng dCas9 plasmid) using 1 µl of MuA Transposase (F-750, Thermo Fisher). Three of these reactions were carried out in parallel. The DNA was electroporated into cells, which were selected for the transposon antibiotic resistance to achieve $>10^7$ CFU or 100,000-fold over the possible library size of 8,262. This was done 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)[18]. A BsaI Golden Gate reaction was then used to remove the modified transposon and insert the domain of interest (PDZ, ER-LBD) in its place. No selection was used for this cloning step, but efficiencies of reaction were >99%. Completed libraries were transformed into *E. coli* cells[18] using electroporation.

Library sequencing and analysis. The ORFs encoding 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 were analyzed with a custom Python pipeline (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 amino acid 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 1 million synthetic reads for each library. It should be noted that Mu transposase 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 amino acid of our synthetic insertion (e.g., insertion at site 208 indicates the inserted domain begins as amino acid 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[41]. All *P* values reported were corrected for multiple hypothesis testing using the Benjamini–Hochberg procedure implemented by DESeq. Domain-insertion-profile sequencing data are available at https://github.com/SavageLab/arC9_data.

CRISPRi screen and FACS selection. Screening of the PDZ libraries was accomplished by transforming and screening >10-fold more *E. coli* than the theoretical library size and repeating two rounds of positive screening and FACS selection as described previously[19]. Screening for the ER library followed the same methods; however, the primary screen and FACS selection were for function with 4-HT, the secondary screen and FACS selection were against function without 4-HT, and the final screen was for function with 4-HT on plates. From these plates colonies were chosen by eye (**Supplementary Fig. 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 or without 10 µM 4-HT in SOB medium.

CRISPRi GFP repression assays. For individual PDZ-construct testing, colonies were cultured from plates, then PDZ-dCas9 plasmid DNA was recovered, separated from the RFP guide plasmid (pgRNA-bacteria, Addgene ID: 44251) via restriction digestion with BsaI and cotransformed with a guide plasmid to repress chromosomal GFP (**Supplementary Table 2**). 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. Optical density at 600 nm (OD_{600}) was measured for each well. GFP and OD_{600} signals were measured and blanked and the GFP (a.u.) normalized to OD_{600}. The cultures were compared when approaching saturation (80% of the maximum OD_{600}, **Fig. 2**) or after 12 h of growth (**Fig. 1** and **Supplementary Fig. 10**). The darC9 construct was treated and assayed as above with induction using 2 µM aTc. Fold changes for the effect of different ligands were calculated by normalizing the GFP and OD_{600} of the darC9 construct to that of a vector control and dividing the fluorescence values without ligand by the fluorescence values with ligand treatment. For single-cell analysis of darC9, cells were grown for ~6 h 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 that targets the genome (either chromosomal RFP or GFP) were made electrocompetent as described previously[19] and, similarly to previous work, all tests were done with the same batch of electrocompetent cells to minimize transformation variability[20]. *E. coli* with these self-targeting guides were electroporated with 9 fmol of either WT Cas9, 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 SOC medium after electroporation for 1 h. CFU/ml was calculated by spotting two technical replicates of tenfold 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 the EF1-a promoter driving the gene of interest and a second PGK promoter driving production of a gene that confers resistance to hygromycin. The EGFP was destabilized by fusion to residues 422–461 of mouse ornithine decarboxylase, giving an *in vivo* half-life of ~2 h. HEK293T cells (obtained from the UC Berkeley Tissue Culture Facility and 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 cotransfected with packaging vectors pspax2 and pMD2.G essentially as described by Tiscornia *et al.*[42]. Cell lines were authenticated and confirmed as mycoplasma free.

HEK293T GFP disruption assays. GFP disruption assays were based on those previously described[26]. Briefly, HEK293T cells were cultured in 10-cm dishes using DMEM with 4.5 g/L glucose, L-glutamine, sodium pyruvate (Corning Cellgro) plus 10% FBS, 1× MEM Non-Essential Amino Acids Solution (Gibco) and penicillin and streptomycin (Gibco). One day before transfection, ~3×10^4 cells per well were plated into a 96-well plate with DMEM plus hygromycin and allowed to settle. One hour before transfection the medium was removed and replaced with medium containing ligand or vehicle control. Cells were transfected according to the manufacturer's protocol with Lipofectamine 2000 (Life Technologies) and 5 ng (**Fig. 3b**) or 50 ng (**Fig. 3c**,**d**) plasmid DNA[43] containing a sgRNA and Cas9, Cas9 ER-LBD terminal fusion or arC9 and a T2A-mCherry tag, as described (**Supplementary Table 2**). Cells were analyzed for EGFP and mCherry expression at 48 or 72 h after transfection using a BD

doi:10.1038/nbt.3528

LSR Fortessa high-throughput sequencer. Transfected cells were gated positive on the basis of mCherry florescence, and the percentage of EGFP disruption for three independent biological replicates was calculated from this gate (**Supplementary Fig. 13**).

HEK293T T7EI assays. HEK293T cells were cultured as described above. One day before transfection, ~2×10^4 cells per well were plated into a 96-well plate with DMEM. One hour before transfection, the medium was removed and replaced with medium containing ligand or vehicle control. Cells were transfected with Lipofectamine 2000 (Life Technologies) and plasmid DNA containing sgRNA and Cas9 or arC9 and a P2A-puromycin fusion (**Supplementary Table 2**) according to manufacturer's protocol. Cells were selected for transfection 24 h later with 1.5 μg/ml puromycin. Genomic DNA was collected and analyzed at 72 h via T7EI assays as previously described[43].

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[44] and grown in 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, was chosen for further experiments. Cells were grown at 37 °C with 5% CO_2.

Lentiviral vector construction. The arC9 lentiviral vector and all variants were constructed according to the sequences provided (**Supplementary Table 2**), using custom oligonucleotides (IDT), standard cloning and Gibson assembly techniques[45]. The pBC2101, pBC2102 and pBC2103 lentiviral vectors (U6-sgRNA-EFS-Cas9-T2A-mCherry-P2A-Hygro or U6-sgRNA-EFS-arC9-T2A-mCherry-P2A-Hygro) (**Supplementary Fig. 16a**) were constructed in the pRRL backbone[46] based on derivatives of SGEP[44]. The U6 promoter and core promoter for the human elongation factor EF-1α (EFS) were based on the lenti-CRISPR-v2 plasmid[47]. For sgRNA expression, an enhanced *Streptococcus pyogenes* Cas9 scaffold was used[9]. All sgRNAs (**Supplementary 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 lentiviral vector expressing a sgRNA for targeting Pcsk9 (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 cells were selected 24 h after infection with 400 μg/ml hygromycin B (Sigma-Aldrich). Treatment regimens for arC9-expressing cells included 1 μM 4-hydroxytamoxifen (4-HT, Sigma-Aldrich), 1 μM β-estradiol (B-E, Sigma-Aldrich), or 0.1% DMSO as control (Sigma-Aldrich).

Data, source code, and reagent availability. Domain-insertion-profile sequencing data are available at https://github.com/SavageLab/arC9_data or via NCBI BioProject accession code PRJNA314234 Sequencing data were analyzed with a custom Python pipeline available at http://github.com/SavageLab/dipseq. Reagents described in this work are available on Addgene (https://www.addgene.org/David_Savage/).

41. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
42. Tiscornia, G., Singer, O. & Verma, I.M. Production and purification of lentiviral vectors. *Nat. Protoc.* **1**, 241–245 (2006).
43. Lin, S., Staahl, B.T., Alla, R.K. & Doudna, J.A. Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife* **3**, e04766 (2014).
44. Fellmann, C. *et al.* An optimized microRNA backbone for effective single-copy RNAi. *Cell Rep.* **5**, 1704–1713 (2013).
45. Gibson, D.G. *et al.* Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).
46. Dull, T. *et al.* A third-generation lentivirus vector with a conditional packaging system. *J. Virol.* **72**, 8463–8471 (1998).
47. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).