

Supplementary Information

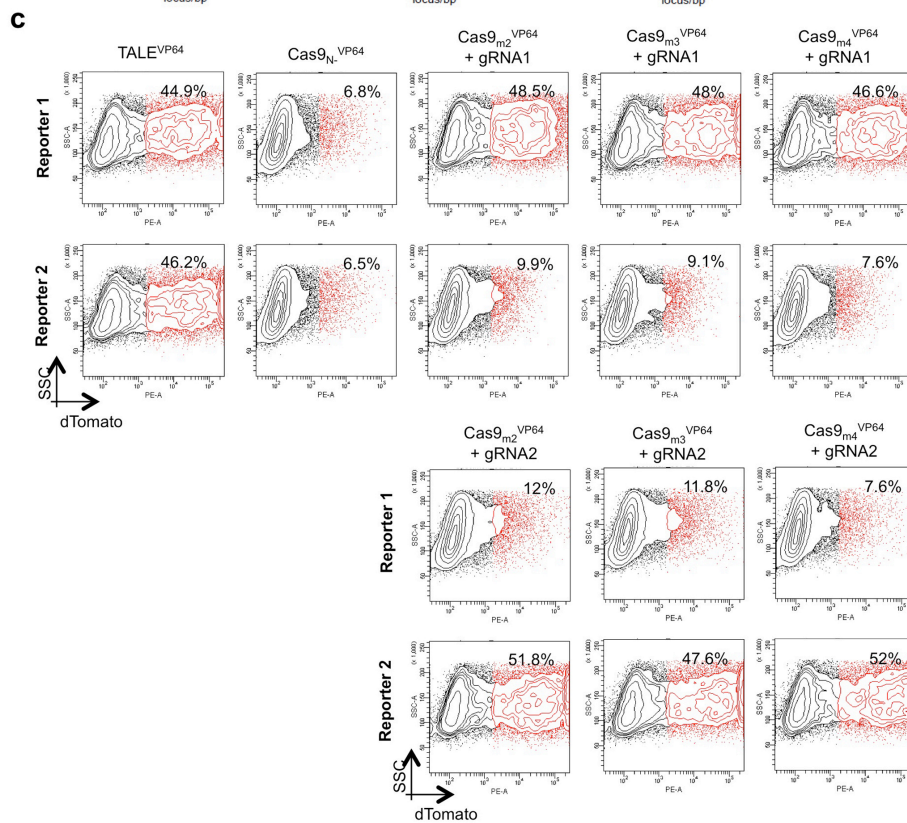
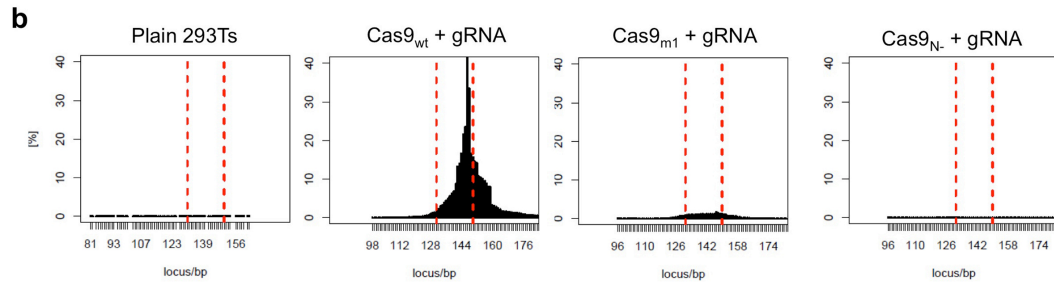
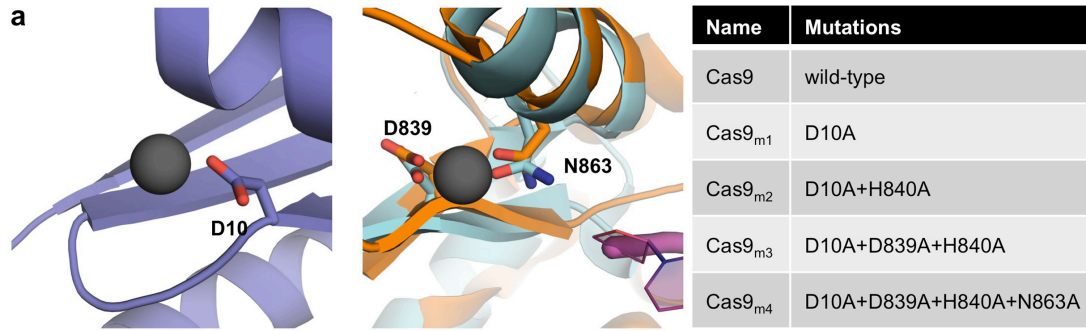
Use of adjacent sgRNA:Cas9 complexes for transcriptional activation and genome engineering.

Prashant Mali^{1,4}, John Aach^{1,4}, P. Benjamin Stranges¹, Kevin M. Esvelt², Mark Moosburner¹,
Sriram Kosuri², Luhan Yang³, George M. Church^{1,2,5}

¹*Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA.* ²*Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, Massachusetts, USA.*

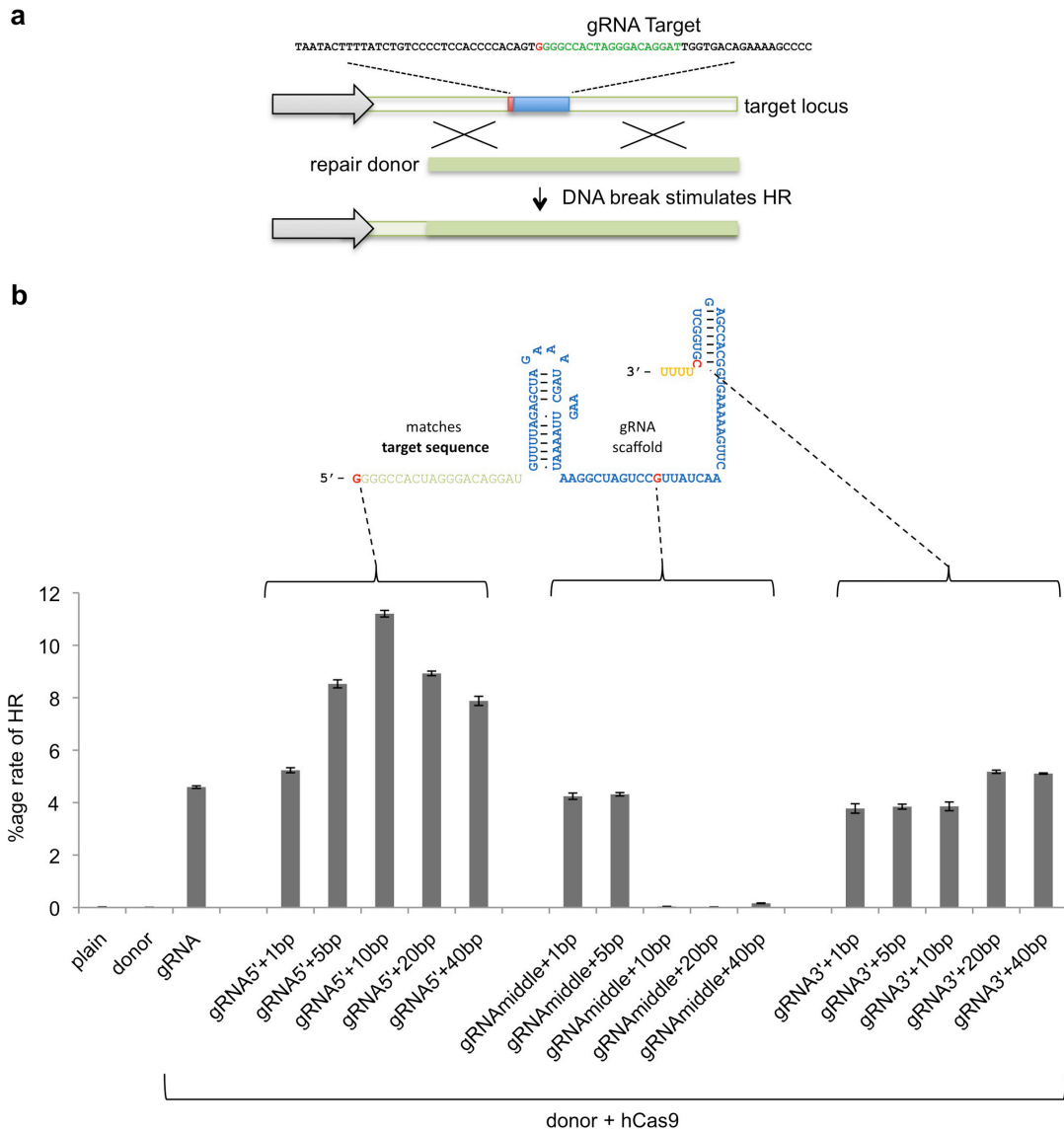
³*Biological and Biomedical Sciences Program, Harvard Medical School, Boston, MA, USA.* ⁴*These authors contributed equally to this work,* ⁵*Correspondence should be addressed to gchurch@genetics.med.harvard.edu.*

Section	Page
Supplementary Fig. 1 Generation of nuclease deficient Cas9.	2
Supplementary Fig. 2 gRNA flexibility to random base insertions	4
Supplementary Fig. 3 RNA-guided REX1 reporter regulation.	5
Supplementary Fig. 4 RNA-guided OCT4 regulation using Cas9 _N .VP64.	6
Supplementary Fig. 5 RNA-guided SOX2 and NANOG regulation using Cas9 _N .VP64.	7
Supplementary Fig. 6 High level specificity analysis processing flow for calculation of normalized expression levels illustrated with examples from experimental data.	8
Supplementary Fig. 7 Evaluating the landscape of targeting by Cas9-gRNA complexes.	10
Supplementary Fig. 8 Validations, specificity of reporter assay.	12
Supplementary Fig. 9 Validations, single and double-base gRNA mismatches.	13
Supplementary Fig. 10 Validations, 5' gRNA truncations.	14
Supplementary Fig. 11 Validations, <i>S. pyogenes</i> PAM.	15
Supplementary Fig. 12 Validations, TALE mutations.	16
Supplementary Fig. 13 Evaluating the landscape of targeting by TALEs of different sizes.	17
Supplementary Fig. 14 TALE monomer specificity versus TALE protein specificity.	18
Supplementary Fig. 15 Cas9D10A nickase mediated NHEJ.	20
Supplementary Fig. 16 Off-set nicking, native locus.	21
Supplementary Fig. 17 Off-set nicking, NHEJ profiles.	22
Supplementary Table 1 gRNA targets for endogenous gene regulation.	23
Supplementary Table 2 Summary of sequence processing conducted as part of this study.	24
Supplementary Table 3. Summary of statistical analysis of Cas9-gRNA and TALE specificity data.	26
Supplementary Note 1 Determination of Cas9 amino acids involved in Mg ²⁺ coordination.	28
Supplementary Note 2 Exponential relationship between Cas9 nuclease and Cas9-TF activity.	29
Supplementary Note 3 Sequences of proteins and RNAs used in this study.	33

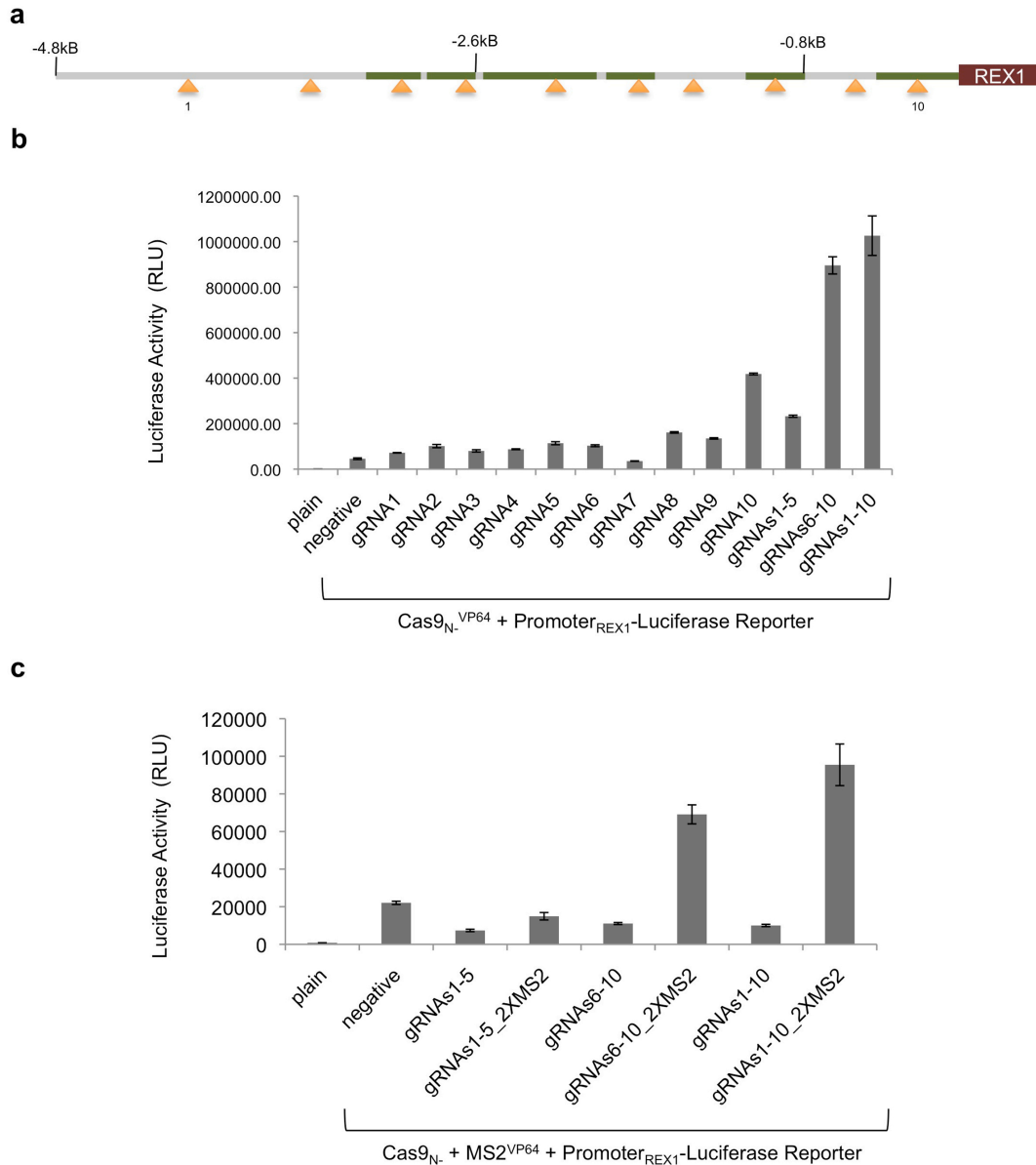


Supplementary Fig. 1 Generation of nuclease deficient Cas9. (a) Metal coordinating residues in known protein structures with homology to Cas9. Residues are labeled based on position in Cas9 sequence. *Left:* RuvC structure, PDB ID: 4EP4 (blue) position D7, which corresponds to D10 in the Cas9 sequence, is highlighted in a Mg-ion coordinating position. *Middle:* Structures of HNH endonuclease domains from PDB IDs: 3M7K (orange) and 4H9D (cyan) including a

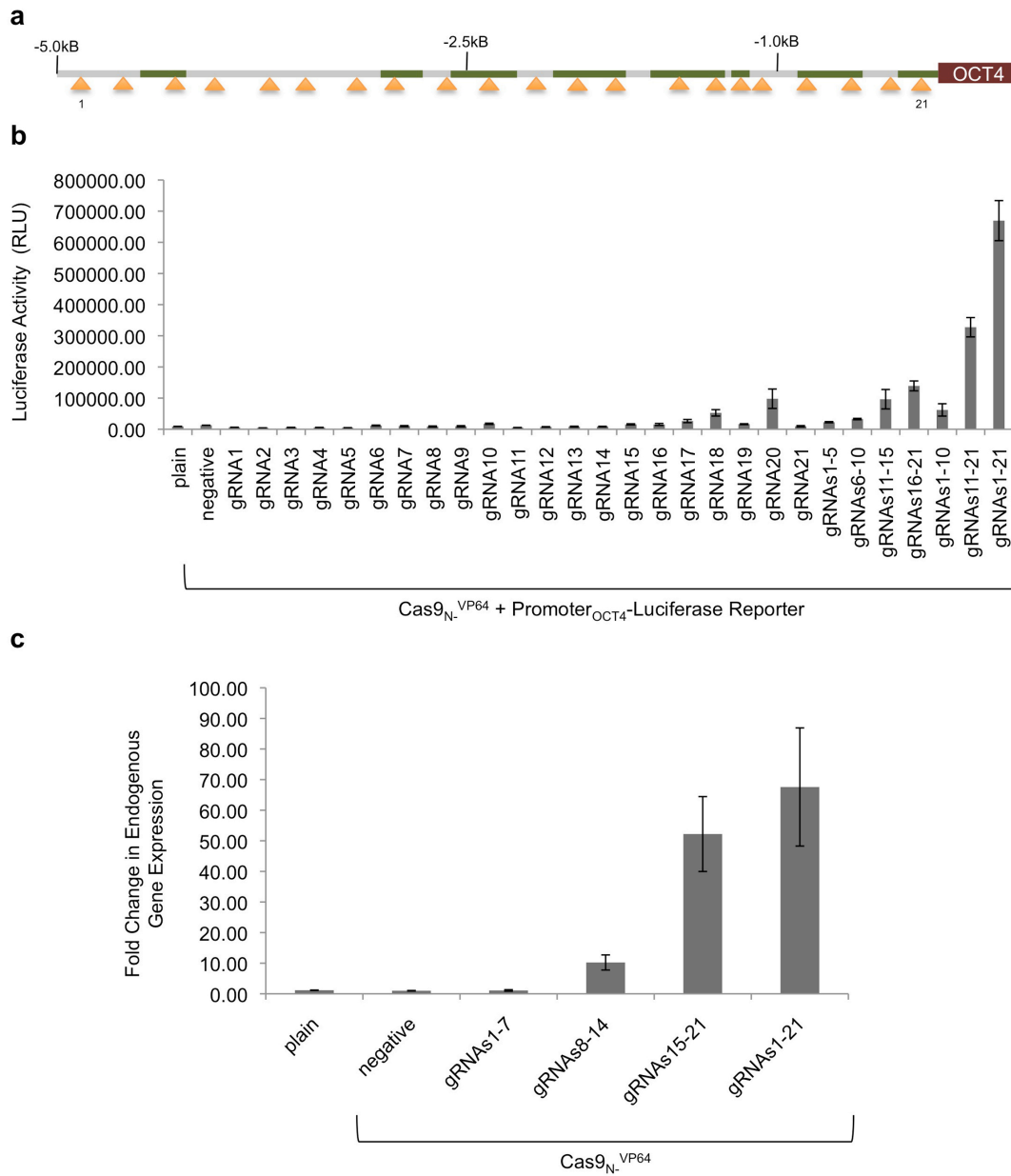
coordinated Mg-ion (gray sphere) and DNA from 3M7K (purple). Residues D92 and N113 in 3M7K and 4H9D positions D53 and N77, which have sequence homology to Cas9 amino acids D839 and N863, are shown as sticks. *Right:* List of mutants analyzed. **(b)** The Cas9 mutants showed undetectable nuclease activity upon deep sequencing at targeted loci. The plots show the mutation frequency versus genomic position, with the red lines demarcating the gRNA target (Cas9_N, here refers to the Cas9_{m4} mutant). **(c)** All three Cas9 mutants: m2, m3 and m4, retain RNA-guided DNA binding activity as assayed here by the ability of their corresponding activator proteins to sequence specifically activate the reporter constructs described in **Fig. 1c**.



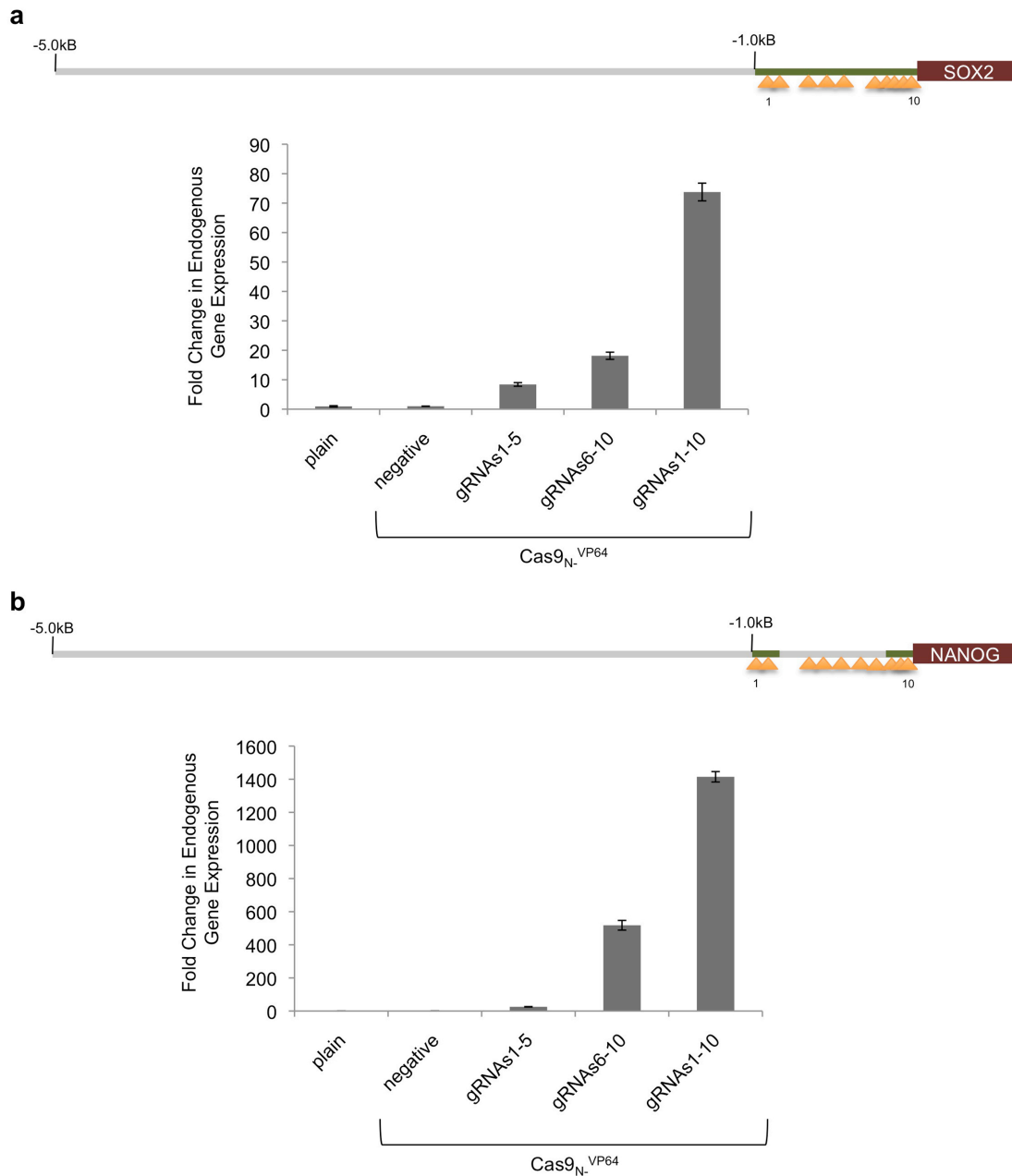
Supplementary Fig. 2 gRNA flexibility to random base insertions. (a) Schematic of HR assay to determine Cas9-gRNA activity. **(b)** gRNAs bearing random sequence insertions at either the 5' end of the crRNA portion or the 3' end of the tracrRNA portion of a chimeric gRNA retain functionality, while insertions into the tracrRNA scaffold portion of the chimeric gRNA result in loss of function. The points of insertion in the gRNA sequence are indicated by red nucleotides. We conjecture that the increased activity upon random base insertions at the 5' end could be due to increased half-life of the longer gRNA. Data are means \pm SEM (N=3).



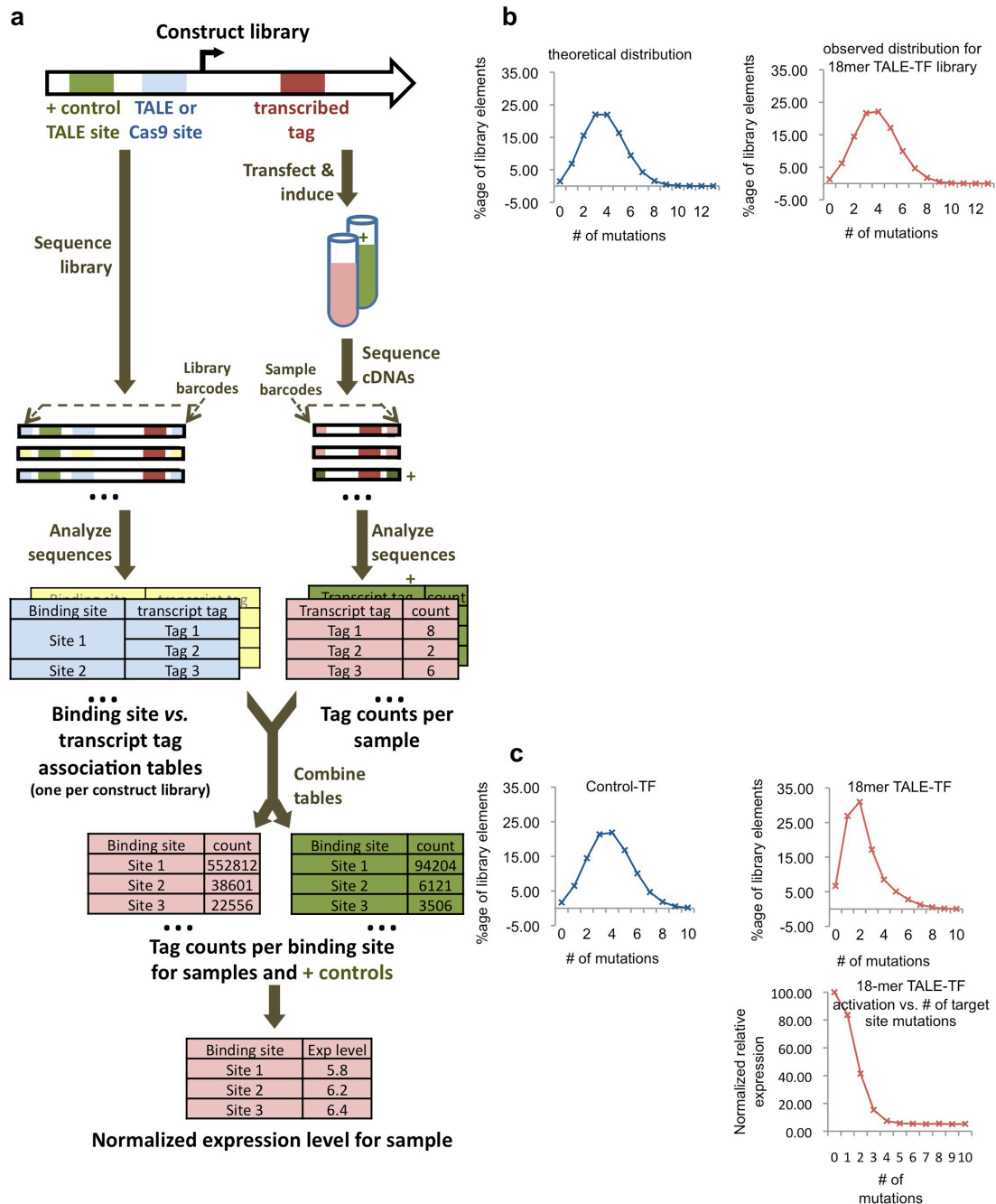
Supplementary Fig. 3 RNA-guided REX1 reporter regulation. (a) As described in Fig. 1f, for the REX1 gene we designed 10 gRNAs targeting a ~5kb stretch of DNA upstream of the transcription start site. Introduction of individual gRNAs modestly stimulated transcription, while multiple gRNAs acted synergistically to stimulate robust multi-fold transcriptional activation in both modalities: (b) Cas9_N-VP64 tethering and (c) gRNA-MS2-VP64 tethering. Note that in the absence of the 2X-MS2 aptamers on the gRNA we do not observe transcriptional activation. Data are means +/- SEM (N=3).



Supplementary Fig. 4 RNA-guided OCT4 regulation using Cas9_N-VP64. (a) For the OCT4 gene we designed 21 gRNAs targeting a ~5kb stretch of DNA upstream of the transcription start site. The DNase hypersensitive sites are highlighted in green. We assayed transcriptional activation using either (b) a promoter-luciferase reporter construct or (c) directly via qPCR of the endogenous genes. While introduction of individual gRNAs modestly stimulated transcription, multiple gRNAs acted synergistically to stimulate robust multi-fold transcriptional activation. Data are means +/- SEM (N=3).

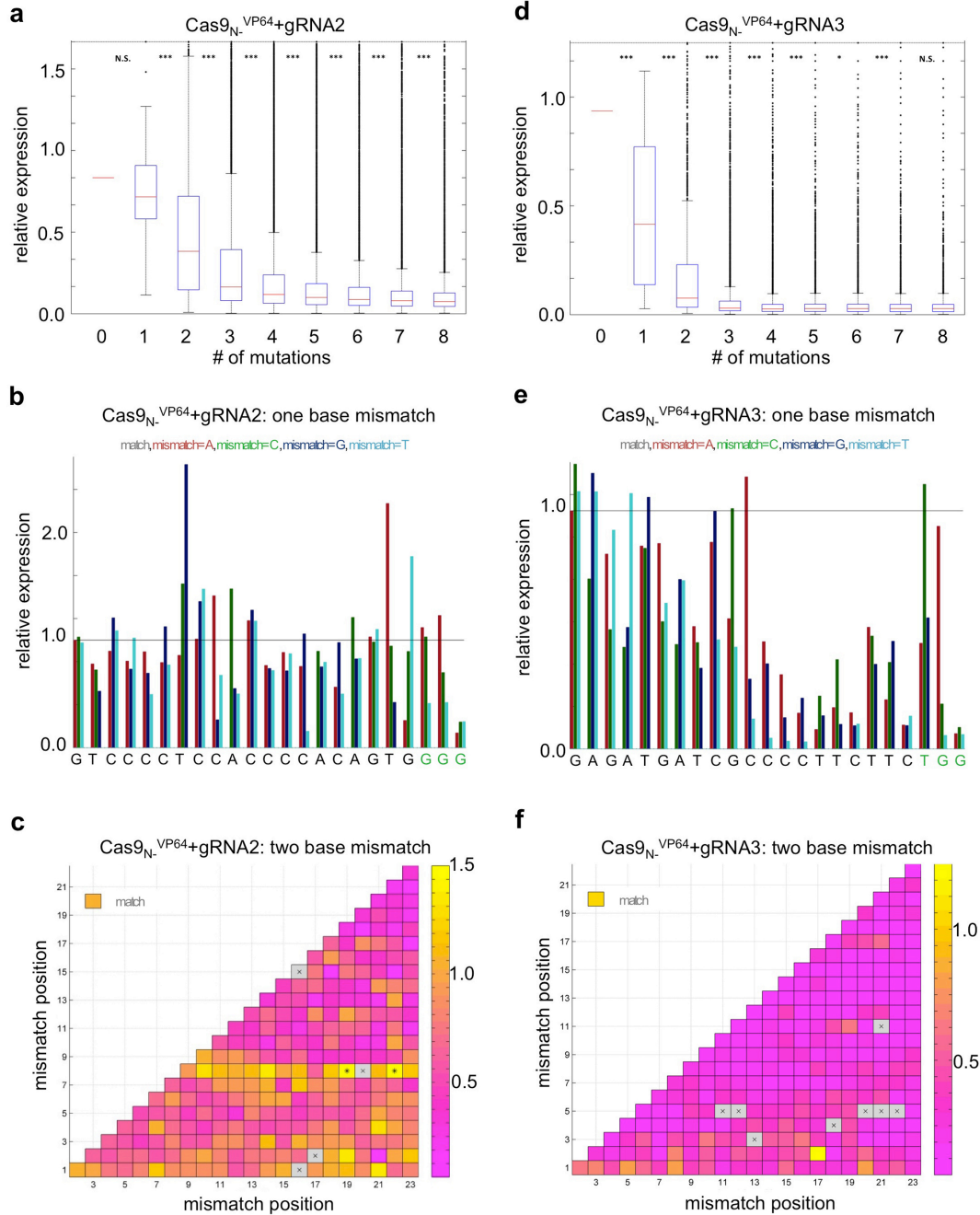


Supplementary Fig. 5 RNA-guided SOX2 and NANOG regulation using Cas9_N-VP64. For the **(a)** SOX2 and **(b)** NANOG genes we designed 10 gRNAs targeting a ~1kb stretch of DNA upstream of the transcription start site. The DNase hypersensitive sites are highlighted in green. We assayed transcriptional activation via qPCR of the endogenous genes. In both instances, while introduction of individual gRNAs modestly stimulated transcription, multiple gRNAs acted synergistically to stimulate robust multi-fold transcriptional activation. Data are means +/- SEM (N=3).



Supplementary Fig. 6 High level specificity analysis processing flow for calculation of normalized expression levels illustrated with examples from experimental data. (a) High level processing flow: Construct libraries are generated with a biased distribution of binding site sequences and random sequence 24bp tags that will be incorporated into reporter gene transcripts (top). The transcribed tags are highly degenerate so that they should map many-to-one to Cas9 or TALE binding sequences. The construct libraries are sequenced (3rd level, left) to establish which tags co-occur with binding sites, resulting in an association table of binding sites vs. transcribed tags (4th level, left). Multiple construct libraries built for different binding sites

may be sequenced at once using library barcodes (indicated here by the light blue and light yellow colors; levels 1-4, left). A construct library is then transfected into a cell population and a set of different Cas9/gRNA or TALE transcription factors are induced in samples of the populations (2nd level, right). One sample is always induced with a fixed TALE activator targeted to a fixed binding site sequence within the construct (top level, green box); this sample serves as a positive control (green sample, also indicated by a + sign). cDNAs generated from the reporter mRNA molecules in the induced samples are then sequenced and analyzed to obtain tag counts for each tag in a sample (3rd and 4th level, right). As with the construct library sequencing, multiple samples, including the positive control, are sequenced and analyzed together by appending sample barcodes. Here the light red color indicates one non-control sample that has been sequenced and analyzed with the positive control (green). Because only the transcribed tags and not the construct binding sites appear in each read, the binding site vs. tag association table obtained from construct library sequencing is then used to tally up total counts of tags expressed from each binding site in each sample (5th level). The tallies for each non-positive control sample are then converted to normalized expression levels for each binding site by dividing them by the tallies obtained in the positive control sample. Examples of plots of normalized expression levels by numbers of mismatches are provided in Figures 2b and e, and in S7a and b and S8b and c. Not covered in this overall process flow are several levels of filtering for erroneous tags, for tags not associable with a construct library, and for tags apparently shared with multiple binding sites (see section on *Computational and sequence analysis for calculation of Cas9-TF and TALE-TF reporter expression levels* in our Supplemental Methods). **(b)** Example distributions of percentages of binding sites by numbers of mismatches generated within a biased construct library. *Left:* Theoretical distribution. *Right:* Distribution observed from an actual TALE construct library. **(c)** Example distributions of percentages of tag counts aggregated to binding sites by numbers of mismatches. *Left:* Distribution observed from the positive control sample. *Right:* Distribution observed from a sample in which a non-control TALE was induced. As the positive control TALE binds to a fixed site in the construct, the distribution of aggregated tag counts closely reflects the distribution of binding sites in b, while the distribution is skewed to the left for the non-control TALE sample because sites with fewer mismatches induce higher expression levels. *Below:* Computing the relative enrichment between these by dividing the tag counts obtained for the target-TF by those obtained for the control-TF reveals the average expression level versus the number of mutations in the target site.



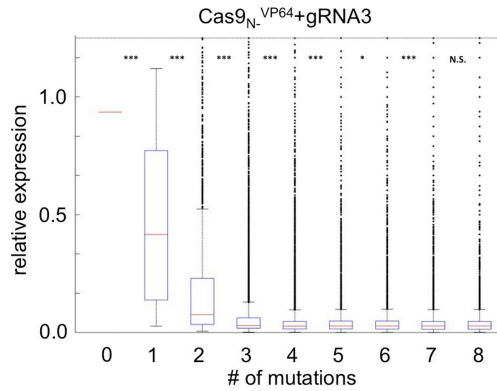
Supplementary Fig. 7 Evaluating the landscape of targeting by Cas9-gRNA complexes. Using the approach described in Fig. 2 we analyzed the targeting landscape of two additional Cas9-gRNA complexes (a-c) and (d-f). Notably we find that these two gRNAs have vastly different specificity profiles with gRNA2 tolerating up to 2-3 mismatches and gRNA3 only up to 1. These aspects are reflected in both the one base mismatch (b,e) and two base mismatch plots (c,f). In (c,f), base mismatch pairs for which insufficient data were available to calculate a normalized expression level are indicated as gray boxes containing an 'x', while, to improve data display, mismatch pairs whose normalized expression levels are outliers that exceed the top of the color scale are indicated as yellow boxes containing an asterisk '*'. Statistical significance symbols

are: *** for $P < .0005/n$, ** for $P < .005/n$, * for $P < .05/n$, and N.S. (Non-Significant) for $P \geq .05/n$, where n is the number of comparisons (refer **Supplementary Table 3**).

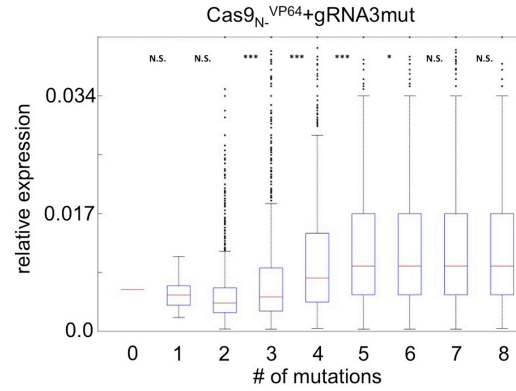
a

Target : GAGATGATCGCCCCTTCTTC TGG
gRNA3 : GAGATGATCGCCCCTTCTTC
gRNA3mut : GTGATGACCGCCGTTCTTC

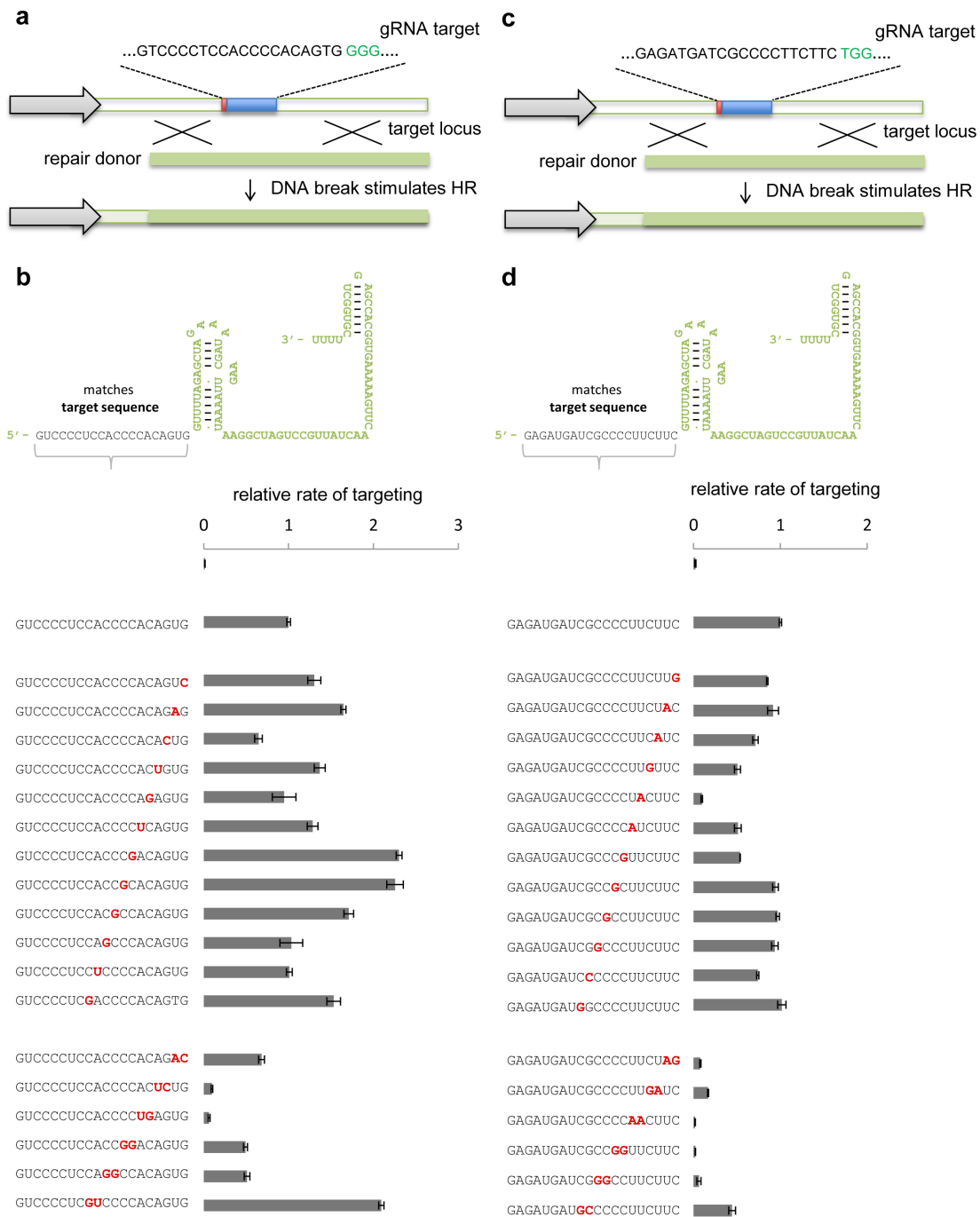
b



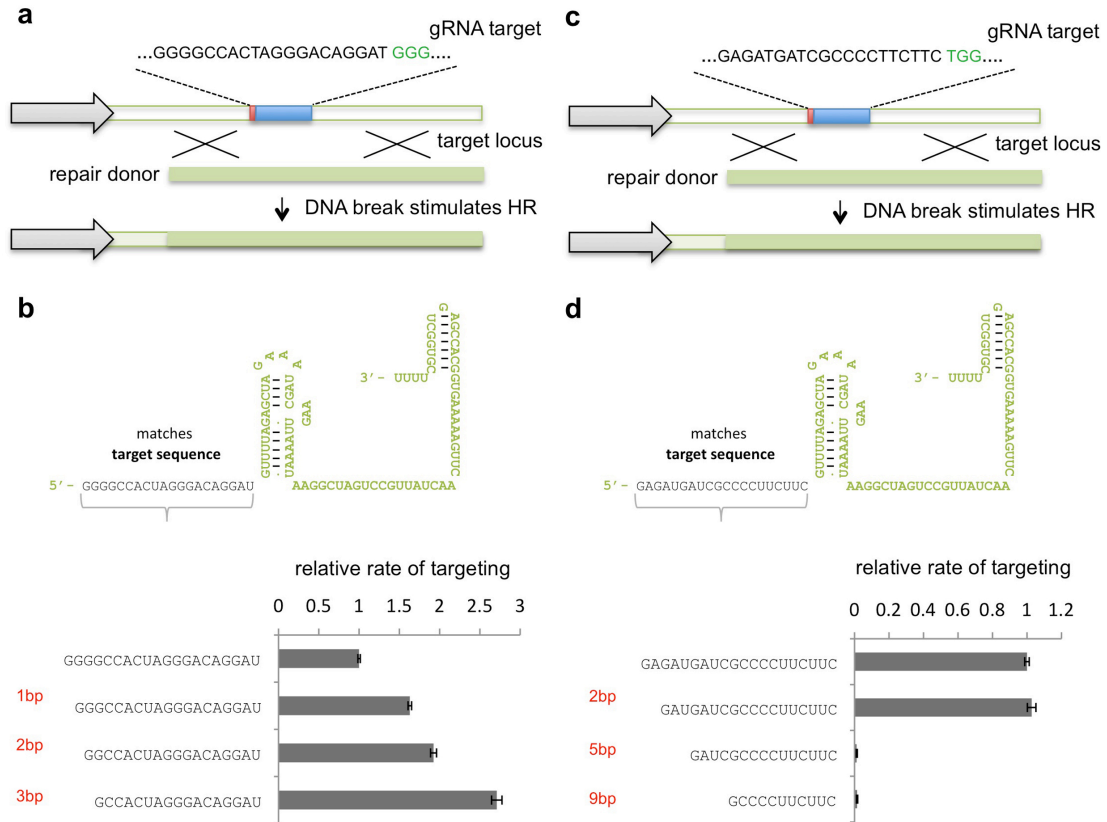
c



Supplementary Fig. 8 Validations, specificity of reporter assay. (a) We evaluated the specificity profile of two gRNAs (wild-type and mutant; sequence differences are highlighted in red) using a reporter library designed against the wild-type gRNA target sequence. **(b)** We confirmed this assay is indeed specific for the gRNA being evaluated (data re-plotted from **Supplementary Fig. 7d**), as the corresponding mutant gRNA is unable to stimulate the reporter library. Statistical significance symbols are: *** for $P < .0005/n$, ** for $P < .005/n$, * for $P < .05/n$, and N.S. (Non-Significant) for $P \geq .05/n$, where n is the number of comparisons (refer **Supplementary Table 3**).

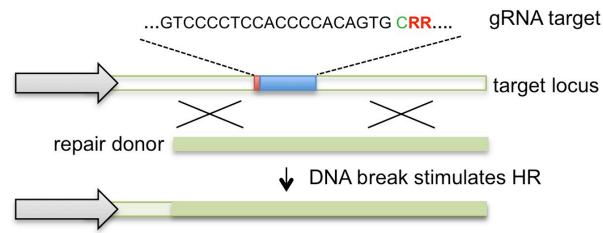


Supplementary Fig. 9 Validations, single and double-base gRNA mismatches. Using a nuclease assay we tested 2 independent gRNAs: gRNA2 (**a,b**) and gRNA3 (**c,d**) bearing single or double-base mismatches (highlighted in red) in the spacer sequence versus the target. These experiments confirmed that single-base mismatches within 12bp of the 3' end of the spacer in the assayed gRNAs indeed still result in detectable targeting, however 2bp mismatches in this regions result in rapid loss of activity. These results further highlight the differences in specificity profiles between different gRNAs consistent with the results in **Supplementary Fig. 7**. Data are means \pm SEM (N=3).

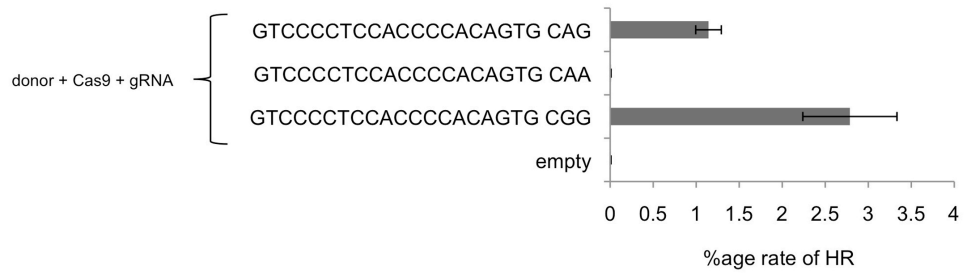


Supplementary Fig. 10 Validations, 5' gRNA truncations. Using a nuclease assay we tested 2 independent gRNAs: gRNA1 (a,b) and gRNA3 (c,d) bearing truncations at the 5' end of their spacer. We observed that 1-3bp 5' truncations are indeed well tolerated, but larger deletions lead to loss of activity. Data are means \pm SEM (N=3).

a

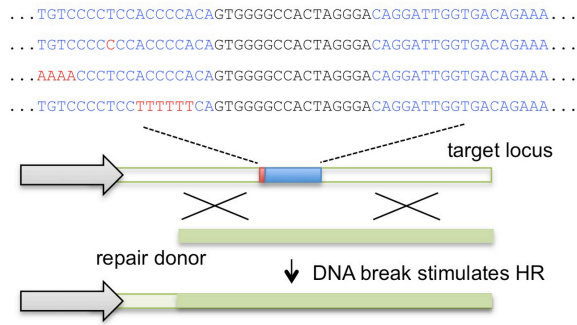


b

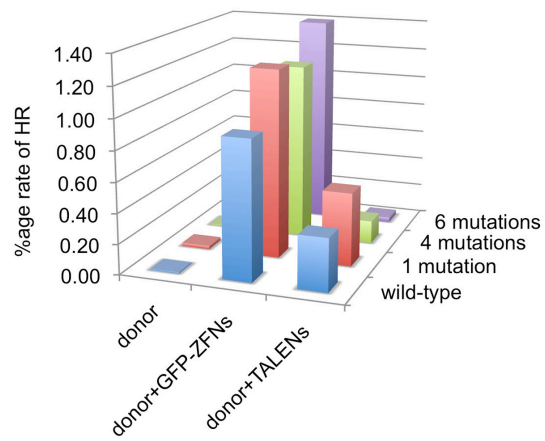


Supplementary Fig. 11 Validations, *S. pyogenes* PAM. We confirmed using a nuclease mediated HR assay (**a,b**) that the PAM for the *S. pyogenes* Cas9 is NGG and also NAG. Data are means +/- SEM (N=3).

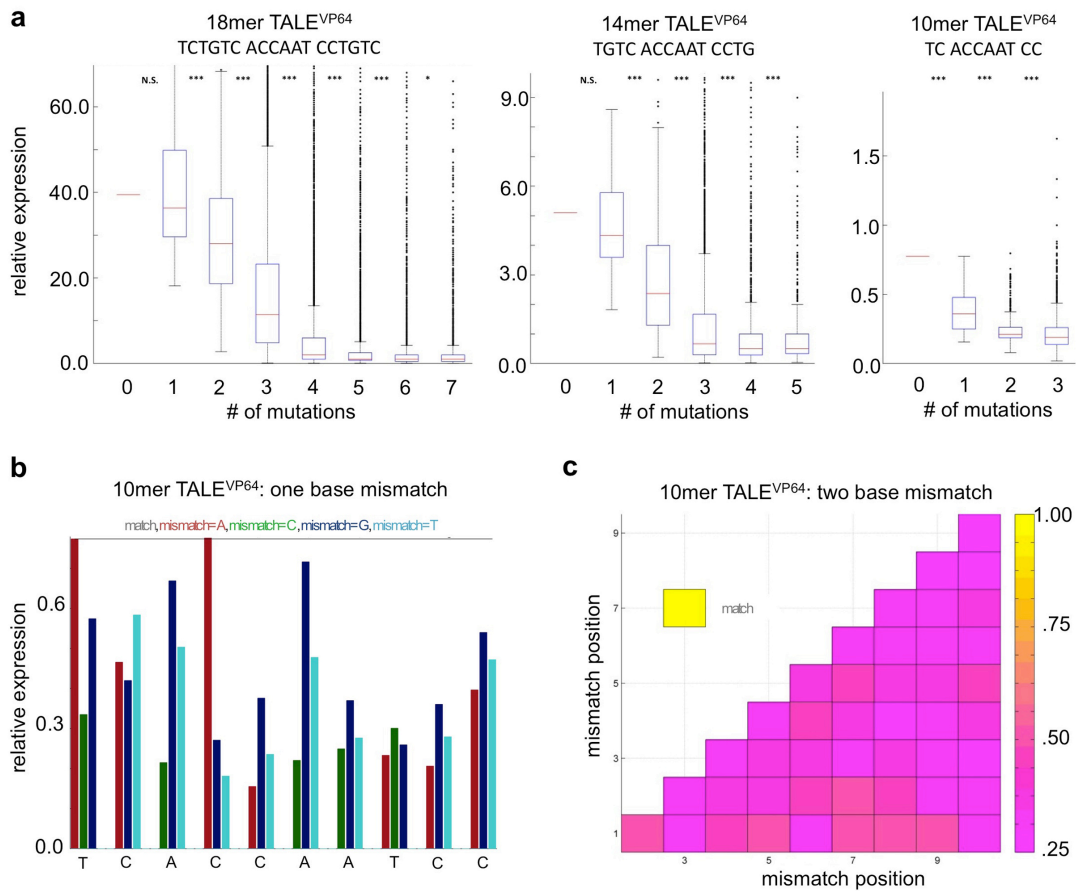
a



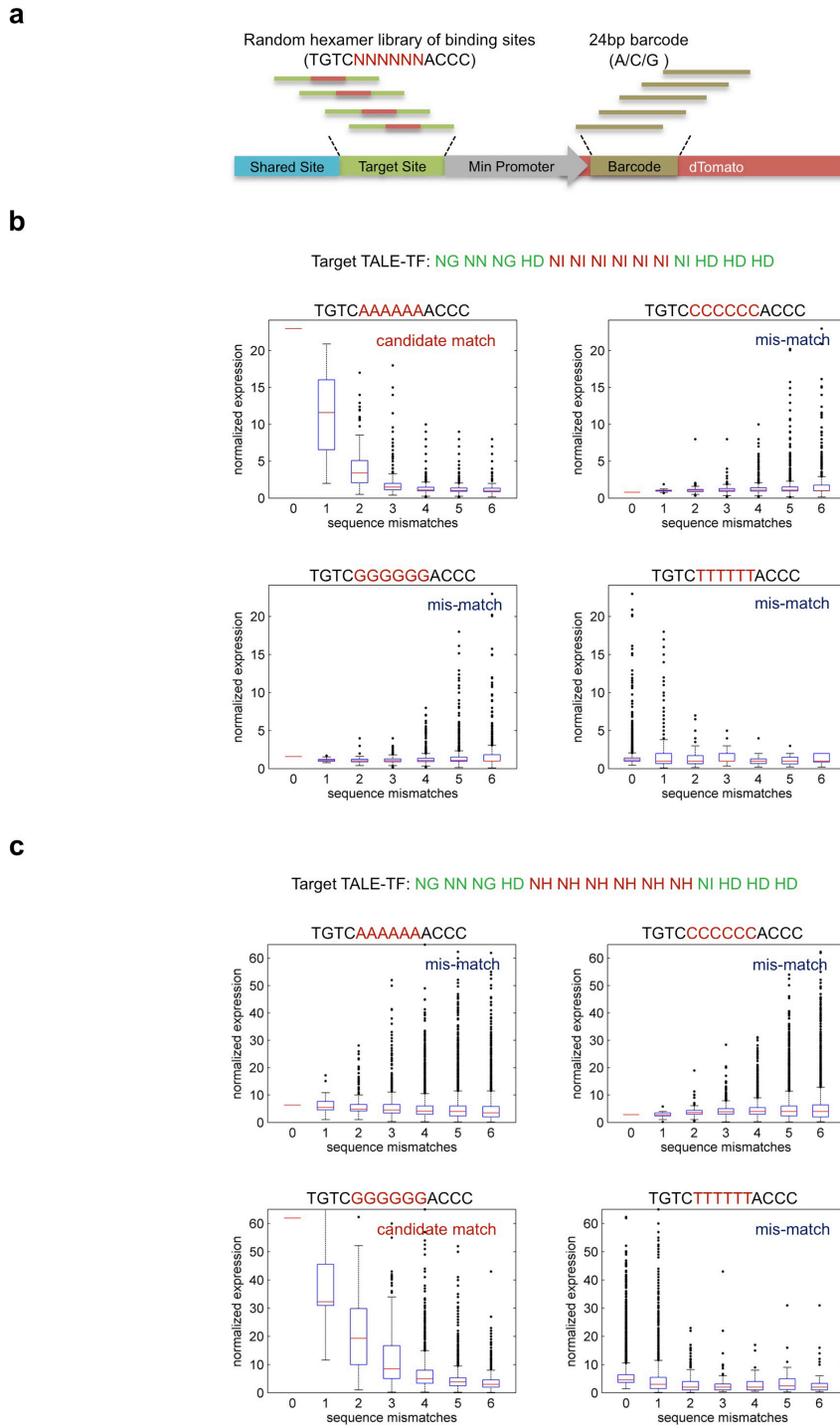
b



Supplementary Fig. 12 Validations, TALE mutations. Using a nuclease mediated HR assay (**a,b**) we confirmed that 18-mer TALEs indeed tolerate multiple mutations in their target sequences.

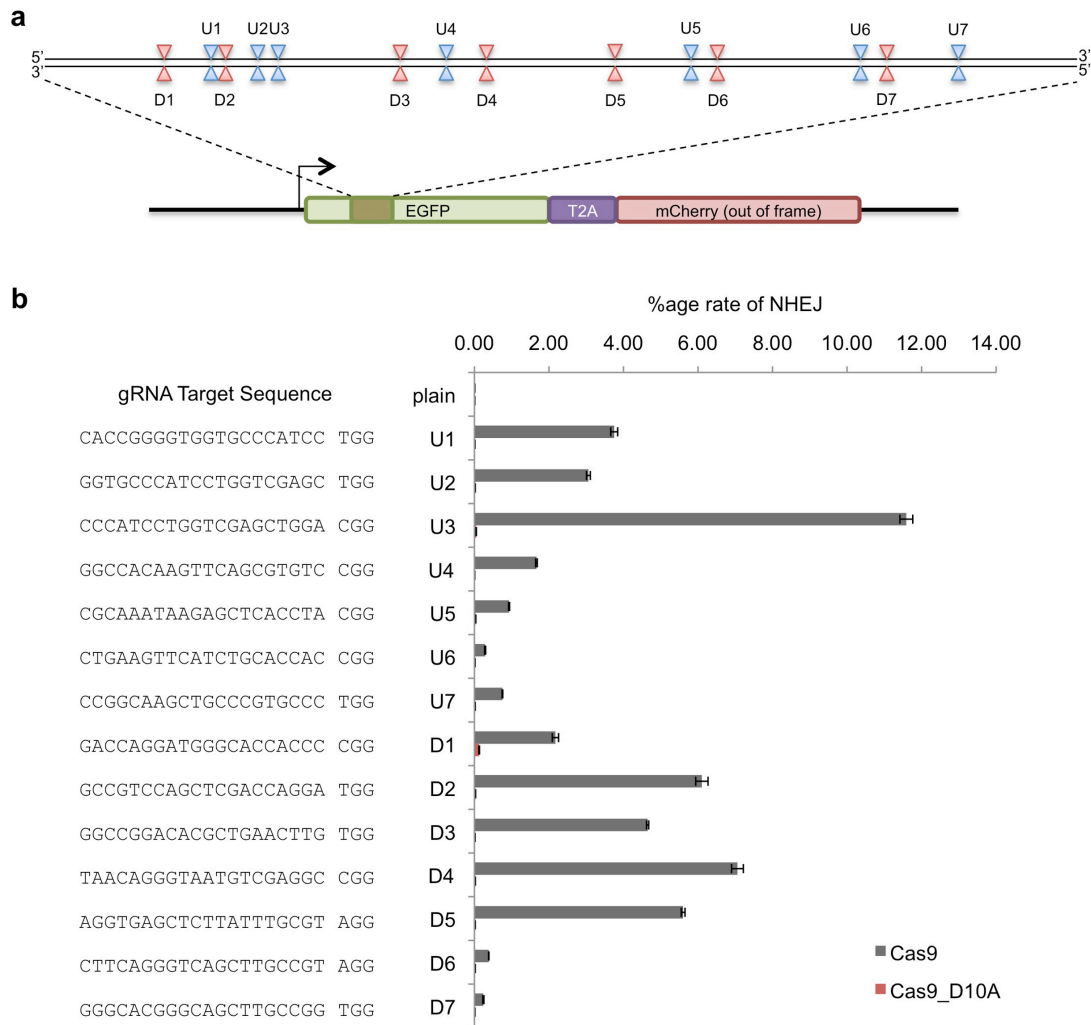


Supplementary Fig. 13 Evaluating the landscape of targeting by TALEs of different sizes. (a) Using the approach described in **Fig. 2** we analyzed the targeting landscape of TALEs of 3 different sizes (18-mer, 14-mer and 10-mer). We find that shorter TALEs (14-mer and 10-mer) are progressively more specific in their targeting but also reduced in activity by nearly an order of magnitude (plotted here are the relative expression levels as measured by observed frequency of barcodes versus the control-TF for normalization). **(b,c)** In fact, 10-mer TALEs show near single-base mismatch resolution, with almost complete loss of activity against targets bearing 2 mismatches (in the heat plot the target sequence positions are labeled from 1-10 starting from the 5' end). Statistical significance symbols are: *** for $P < .0005/n$, ** for $P < .005/n$, * for $P < .05/n$, and N.S. (Non-Significant) for $P \geq .05/n$, where n is the number of comparisons (refer **Supplementary Table 3**).

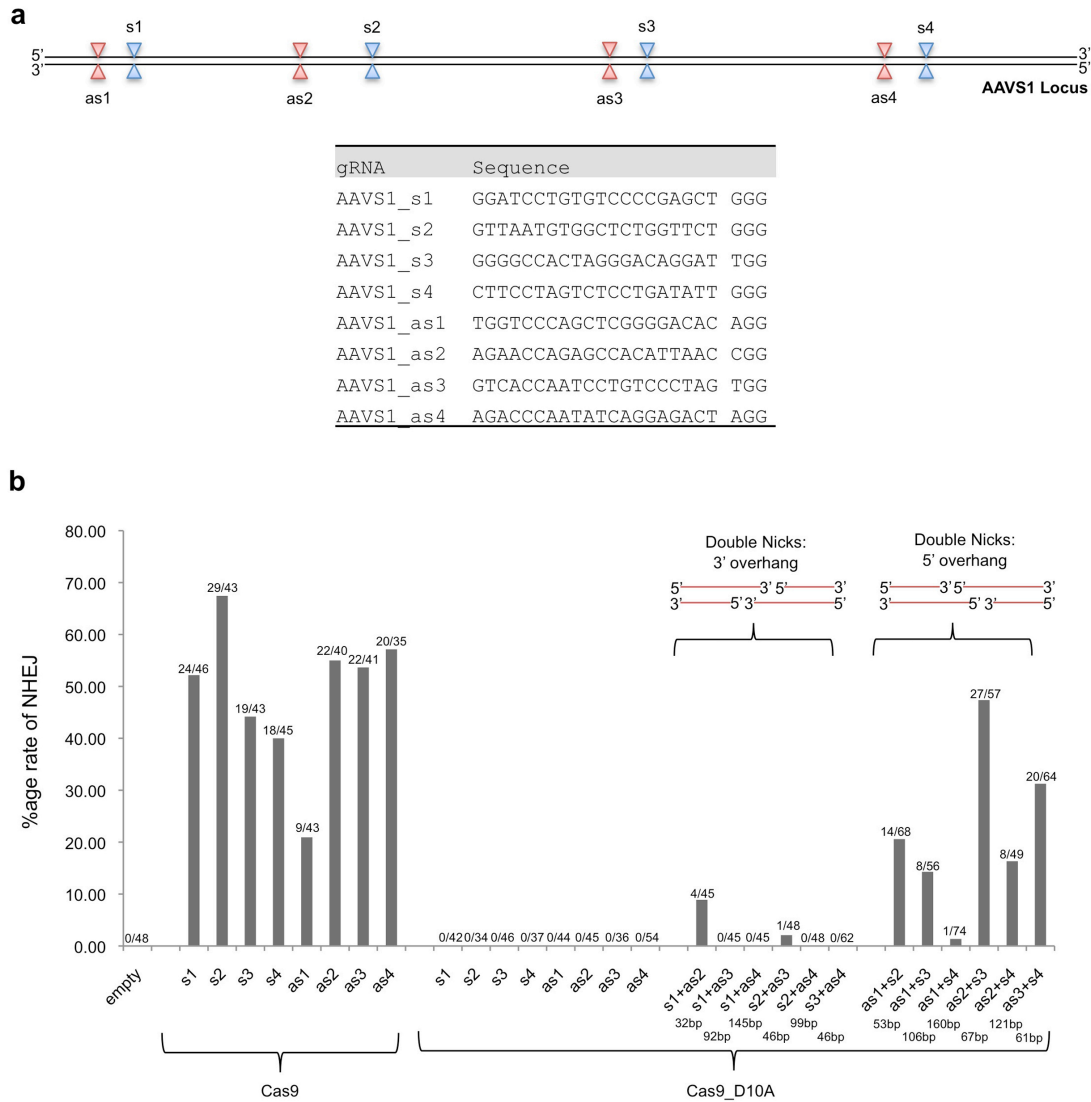


Supplementary Fig. 14 TALE monomer specificity versus TALE protein specificity. (a) Using a modification of approach described in **Fig. 2** we analyzed the targeting landscape of 2 14-mer TALE-TFs bearing a contiguous set of 6 NI or 6 NH repeats. In this approach a reduced library of reporters bearing a degenerate 6-mer sequence in the middle was created and used to assay the TALE-TF specificity. **(b, c)** In both instances one notes that the expected target sequence is indeed enriched for (i.e. one bearing 6 As for NI repeats, and 6 Gs for NH repeats). Notably

though each of these TALEs still tolerate 1-2 mismatches in the central 6-mer target sequence. These results indicate that while choice of monomers does contribute to base specificity, TALE specificity is also a function of the binding energy of the protein as a whole.



Supplementary Fig. 15 Cas9D10A nickase mediated NHEJ. (a) As described in Fig. 3 we employed the traffic light reporter to assay NHEJ events upon introduction of targeted nicks or double-stranded breaks. Briefly upon introduction of DNA cleavage events, if the break goes through mutagenic NHEJ the GFP is translated out of frame and the downstream mCherry sequences are rendered in frame resulting in red fluorescence. As described earlier, we designed 14 gRNAs covering a 200bp stretch of DNA: 7 targeting the sense strand (U1-7) and 7 the antisense strand (D1-7). (b) We observe that unlike the wild-type Cas9 which results in DSBs and robust NHEJ across all targets, most nicks (using the Cas9D10A mutant) seldom result in NHEJ events. It is interesting to note that albeit all the 14 sites are located within a contiguous 200bp stretch of DNA we observe over 10-fold differences in targeting efficiencies using these. We conjecture that likely the Cas9-gRNA system may have nucleotide preferences that as yet remain to be deciphered. Data are means \pm SEM (N=3).



Supplementary Fig. 16 Off-set nicking, native locus. (a) We targeted the native AAVS1 locus with 8 gRNAs covering a 200bp stretch of DNA: 4 targeting the sense strand (s1-4) and 4 the antisense strand (as1-4). Using the Cas9D10A mutant, which nicks the complementary strand, we used different two-way combinations of the gRNAs to induce a range of programmed 5' or 3' overhangs. **(b)** Using a Sanger sequencing based assay, we observed that while single gRNAs did not induce detectable NHEJ events, inducing off-set nicks to generate DSBs is highly effective at inducing gene disruption. Notably off-set nicks leading to 5' overhangs result in more NHEJ events as opposed to 3' overhangs. The number of Sanger sequencing clones is highlighted above the bars, and the predicted overhang lengths are indicated below the corresponding x-axis legends.

gRNA Name	gRNA Target
REX1 1	ctggcggatcactcgcgggtt agg
REX1 2	cctcggcctccaaaagtgct agg
REX1 3	acgctgattcctgcagatca ggg
REX1 4	ccaggaatacgtatccacca ggg
REX1 5	gccacacccaagcgatcaaa tgg
REX1 6	aaataatacattctaaggta agg
REX1 7	gctactggggaggctgaggc agg
REX1 8	tagcaatacagtcacattaa tgg
REX1 9	ctcatgtgatcccccgctc cgg
REX1 10	ccgggcagagagtgaacgcg cgg
OCT4 1	ttccttcctctcccgtgct tgg
OCT4 2	tctctgcaaagcccctggag agg
OCT4 3	aatgcagttgccgagtgcag tgg
OCT4 4	cctcagcctcctaaagtgct ggg
OCT4 5	gagtccaaatcctctttact agg
OCT4 6	gagtgtctggatttgggata agg
OCT4 7	cagcacctcatctcccagtg agg
OCT4 8	tctaaaaccaggaatcat ggg
OCT4 9	cacaaggcagccagggatcc agg
OCT4 10	gatggcaagctgagaaacac tgg
OCT4 11	tgaaatgcacgcatacaatt agg
OCT4 12	ccagtccagacctggccttc tgg
OCT4 13	cccagaaaaacagaccctga agg
OCT4 14	aagggttgagcacttgttta ggg
OCT4 15	atgtctgagttttggttgag agg
OCT4 16	ggtcccctgaagggaagta ggg
OCT4 17	tggcagctactcttgaaga tgg
OCT4 18	ggcacagtgccagaggtctg tgg
OCT4 19	taaaaataaaaaactaaca ggg
OCT4 20	tctgtggggacctgcactg agg
OCT4 21	ggccagaggtcaaggctagt ggg
SOX2 1	cacgaccgaaacccttctta cgg
SOX2 2	gttgaatgaagacagtctag tgg
SOX2 3	taagaacagagcaagttacg tgg
SOX2 4	tgtaaggtaagagaggagag cgg
SOX2 5	tgacacaccaactcctgcac tgg
SOX2 6	tttaccacttccttcgaaa agg
SOX2 7	gtggctggcaggctggctct ggg
SOX2 8	ctccccggcctccccgcg cgg
SOX2 9	caaaacccggcagcaggct ggg
SOX2 10	aggagccgcccgcgctgat tgg
NANOG 1	cacacacaccacacgagat ggg
NANOG 2	gaagaagctaaagagccaga ggg
NANOG 3	atgagaatttcaataacctc agg
NANOG 4	tcccgctctgttggccaggc tgg
NANOG 5	cagacacccaccacatgcg tgg
NANOG 6	tccaatttactgggattac agg
NANOG 7	tgatttaaaagttggaacg tgg
NANOG 8	tctagtccccacctagtct ggg
NANOG 9	gattaactgagaattcacia ggg
NANOG 10	cgccaggggggtgggtcta agg

Supplementary Table 1. gRNA targets for endogenous gene regulation. Targets in the REX1, OCT4, SOX2 and NANOG promoters used in Cas9-gRNA mediated activation experiments is listed.

Sequencing	Expt	Construct library	TF induced	Target sequence	Read pairs from sequencer	Merged read pairs	Validated read pairs	Recognized tags	Assigned	Figure
(Notes)	(1)		(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Construct libraries	81a	gRNA1	N/A	GTCCCTCCACCCACAGTGGGG	5416685	N/A	3412936	N/A	1130572	N/A
		gRNA2		GGGGCCACTAGGGACAGGATTGG					2282364	
	81d	gRNA3	N/A	GAGATGATCGCCCTTCTICTGG	13381751	N/A	9428891	N/A	9428891	N/A
	69a	TALE-A	N/A	TCTGTCCCAATCCTGTC	23078287	N/A	20380894	N/A	8683562	N/A
TALE-B		TCTGTCCCCTCCACCCCA		5370242						
TALE-MS		TGTCNNNNNNACCC		6327090						
Reporter gene cDNA libraries	81b	gRNA1	Cas9-VP64+gRNA1	GTCCCTCCACCCACAGTGGGG	15229928	10967927	10402225	7702244	2229814	S7
			PosControl (11)	PosControl sequence (11)					5387069	
			NegControl (12)	no induction					85361	
	81c	gRNA2	Cas9-VP64+gRNA2	GGGGCCACTAGGGACAGGATTGG	15811347	12239539	11625733	10012421	4699948	2b-d
			PosControl (11)	PosControl sequence (11)					5286782	
			NegControl (12)	no induction					25691	
	81e	gRNA3	Cas9-VP64+gRNA3	GAGATGATCGCCCTTCTICTGG	14842480	11588456	10924785	9084909	2240937	S7, S8
			Cas9-VP64+gRNA3mut	GAGATGATCGCCCTTCTICTGG					245828	
			PosControl (11)	PosControl sequence (11)					6322200	
			NegControl (12)	no induction					275944	
	69c	TALE-A	TALE-A-TF18	TCTGTCCCAATCCTGTC	46553956	13920473	13135728	12639049	10021862	S13
			TALE-A-TF14	--TGTCACCAATCCTG--					1422311	
			TALE-A-TF10	----TCACCAATCC----					252770	
			PosControl (11)						788867	
			NegControl (12)						153239	
	69d	TALE-A	TALE-A-TF18	TCTGTCCCAATCCTGTC	47576536	16210680	15272143	14721253	10708190	2e-g
			TALE-A-TF14	--TGTCACCAATCCTG--					1972081	
			PosControl (11)	PosControl sequence (11)					1189598	
			NegControl (12)	no induction					174346	
			other (10)						677038	
	69g	TALE-B	TALE-B-TF18	TCTGTCCCCTCCACCCCA	38675868	16874674	15716970	15218496	3719497	NA
			TALE-B-TF14	--TGTCCCCCTCCACCC--					4998807	
			TALE-B-TF10	----TCCCCTCCAC----					970556	
			PosControl (11)	PosControl sequence (11)					3274386	
NegControl (12)			no induction	357599						
69a	TALE-B	other (10)						1897651		
		TALE-B-TF18	TCTGTCCCCTCCACCCCA	47283720	15799461	14657441	14075322	8351454	NA	
		PosControl (11)	PosControl sequence					5611822		
NegControl (12)	no induction	112046								
69e	TALE-MS	TALE-MS-TF-NI	TGTCNNNNNNACCC	7724540	5242545	4582369	3980469	463905	S14	
		TALE-MS-TF-HD	TGTCNNNNNNACCC					45270		
		TALE-MS-TF-NH	TGTCNNNNNNACCC					2280801		
		PosControl (11)	PosControl sequence					361235		
		NegControl (12)	no induction					116795		
		other (10)						712463		

Supplementary Table 2. Summary of sequence processing conducted as part of this study, as described in the section entitled Computational and sequence analysis for calculation of Cas9-TF and TALE-TF reporter expression levels in Supplemental Methods. The processed data for all of these libraries and samples is provided as a supplemental data file at: http://arep.med.harvard.edu/Cas9_activators/.

Notes

- (1) Experiment label ("expt") given to the batch of barcoded samples processed together in a single sequencing run. Also used to identify data file within supplemental data provided with this study.
- (2) Refers to transcriptional activator induced in a sample of a cell population transfected with a construct library. Not applicable (N/A) to construct library sequencing.
- (3) Refers to main binding site sequence of biased construct libraries, but to programmed transcriptional activator binding site in reporter gene cDNA libraries.
- (4) Obtained from Illumina MiSeq for Cas9-gRNA libraries, and from Illumina HiSeq for TALE libraries.

- (5) Read pairs merged by SeqPrep. Only applicable to reporter gene cDNA libraries, not applicable (N/A) to construct libraries.
- (6) Refers to read pairs or merged reads passing validation checks described in Supplemental Methods section cited above.
- (7) Refers to number of transcribed tags found in construct library binding site vs. transcript tag table (see Supplemental Methods section cited above). Does not include "novel" tags. Not applicable (N/A) to construct library sequencing.
- (8) For construct library sequencing, number of valid read pairs assigned to construct library. For reporter gene cDNA sequencing, number of recognized tags assigned to TF induced.
- (9) Data used for indicated Figure or Supplemental Figure, otherwise not applicable (N/A).
- (10) Other construct library or cDNA sample included in the sequencing run but not analyzed or provided as part of this study.
- (11) PosControl refers to the fixed binding site included in every construct, and to the fixed TALE-TF used to activate a positive control sample (see Supplemental Methods section cited above). The PosControl binding site was based on Isce-I and so the PosControl may be referred to as ISce-I in some supplemental data files.
- (12) NegControl refers to a sample in which no transcriptional activator was induced. A NegControl sample was processed with every batch of induced samples.

a	Figure	Expression level comparison:		t-test	P-value	Symbol
		mutations vs.	mutations			
2b	0	1	1-samp	7.8E-05	**	
	1	2	2-samp	1.4E-06	***	
	2	3	2-samp	4.0E-61	***	
	3	4	2-samp	0	***	
	4	5	2-samp	0	***	
	5	6	2-samp	1.0E-217	***	
	6	7	2-samp	1.7E-43	***	
	7	8	2-samp	3.7E-02	N.S.	
2e	0	1	1-samp	8.9E-01	N.S.	
	1	2	2-samp	1.9E-06	***	
	2	3	2-samp	5.0E-147	***	
	3	4	2-samp	0	***	
	4	5	2-samp	0	***	
	5	6	2-samp	4.2E-62	***	
	6	7	2-samp	1.6E-03	*	
	7	8	2-samp	4.7E-01	N.S.	
S7a	0	1	1-samp	5.2E-02	N.S.	
	1	2	2-samp	2.8E-05	***	
	2	3	2-samp	3.5E-21	***	
	3	4	2-samp	1.4E-58	***	
	4	5	2-samp	8.3E-101	***	
	5	6	2-samp	6.8E-94	***	
	6	7	2-samp	1.8E-61	***	
	7	8	2-samp	8.1E-24	***	
S7d and S8d	0	1	1-samp	2.3E-18	***	
	1	2	2-samp	2.4E-08	***	
	2	3	2-samp	6.2E-54	***	
	3	4	2-samp	4.0E-141	***	
	4	5	2-samp	1.9E-20	***	
	5	6	2-samp	1.2E-03	*	
	6	7	2-samp	3.8E-05	***	
	7	8	2-samp	9.4E-01	N.S.	
S8c	0	1	1-samp	7.2E-03	N.S.	
	1	2	2-samp	5.0E-01	N.S.	
	2	3	2-samp	3.9E-84	***	
	3	4	2-samp	8.5E-153	***	
	4	5	2-samp	8.6E-76	***	
	5	6	2-samp	1.6E-03	*	
	6	7	2-samp	7.1E-01	N.S.	
	7	8	2-samp	7.8E-02	N.S.	
S13a (left)	0	1	1-samp	7.3E-01	N.S.	
	1	2	2-samp	2.4E-06	***	
	2	3	2-samp	7.2E-140	***	
	3	4	2-samp	0	***	
	4	5	2-samp	0	***	
	5	6	2-samp	1.0E-72	***	
	6	7	2-samp	4.0E-03	*	
S13a (middle)	0	1	1-samp	9.4E-02	N.S.	
	1	2	2-samp	5.2E-09	***	
	2	3	2-samp	7.9E-86	***	
	3	4	2-samp	2.9E-53	***	
S13a (right)	4	5	2-samp	3.5E-10	***	
	0	1	1-samp	1.3E-13	***	
	1	2	2-samp	1.1E-04	***	
	2	3	2-samp	3.7E-08	***	

b	seed start position	Number position pairs		-log10 P-value
		both in seed	not both in seed	
	2	171	19	3.11
	3	153	37	1.46
	4	136	54	2.01
	5	120	70	3.34
	6	105	85	5.65
	7	91	99	7.34
	8	78	112	6.61
	9	66	124	7.10
	10	55	135	9.72
	11	45	145	9.83
	12	36	154	10.44
	13	28	162	10.72
	14	21	169	8.97
	15	15	175	5.61
	16	10	180	3.34
	17	6	184	2.26
	18	3	187	1.16

Supplementary Table 3. Summary of statistical analysis of Cas9-gRNA and TALE specificity data. (a) P-values for comparisons of normalized expression levels of TALE or Cas9-VP64 activators binding to target sequences with particular numbers of target site mutations. Normalized expression levels have been indicated by boxplots in the figures indicated in the Figure column, where the boxes represent the distributions of these levels by numbers of

mismatches from the target site. P-values were computed using t-tests for each consecutive pair of numbers of mismatches in each boxplot, where the t-tests were either one sample or two sample t-tests (see Methods). Statistical significance was assessed using Bonferroni-corrected P-value thresholds, where the correction was based on the number of comparisons within each boxplot. Statistical significance symbols are: *** for $P < .0005/n$, ** for $P < .005/n$, * for $P < .05/n$, and N.S. (Non-Significant) for $P \geq .05/n$, where n is the number of comparisons. – **(b)** Statistical characterization of seed region in **Fig. 2d**: $\log_{10}(P\text{-values})$ indicating the degree of separation between expression values for Cas9_N^{VP64+}gRNA binding to target sequences with two mutations for those position pairs mutated within candidate seed regions at the 3' end of the 20bp target site vs. all other position pairs. The greatest separation, indicated by the largest $-\log_{10}(P\text{-values})$ (highlighted above), is found in the last 8-9bp of the target site. These positions may be interpreted as indicating the start of the "seed" region of this target site. See the section "Statistical characterization of seed region" in Methods for information on how the P-values were computed.

Supplementary Note 1. Determination of Cas9 amino acids involved in Mg²⁺ coordination

We searched for sequence homologous to Cas9 with known structure to identify candidate mutations in Cas9 that could ablate the natural activity of its RuvC and HNH domains. Using HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>), we queried the full sequence of Cas9 against the full Protein Data Bank (January 2013). This search returned two different HNH endonucleases that had significant sequence homology to the HNH domain of Cas9; Pacl and a putative endonuclease (PDB IDs: 3M7K and 4H9D respectively). We examined these proteins to find residues involved magnesium ion coordination, and then looked for the corresponding residues in the sequence alignment to Cas9. We discovered two Mg-coordinating side-chains in each structure that aligned to the same amino acid type in Cas9. They are 3M7K D92 and N113, and 4H9D D53 and N77. These residues corresponded to Cas9 D839 and N863. It was also reported that mutations of Pacl residues D92 and N113 to alanine rendered the nuclease catalytically deficient. We chose to make the Cas9 mutations D839A and N863A based on this analysis. Additionally, HHpred also predicts homology between Cas9 and the N-terminus of a *Thermus thermophilus* RuvC (PDB ID: 4EP4). This sequence alignment covers the previously reported mutation D10A which eliminates function of the RuvC domain in Cas9. To confirm that this was a mutation we wanted to incorporate we investigated the metal binding residues as before. In 4EP4, D7 helps to coordinate a magnesium ion. This position has sequence homology to corresponds to Cas9 D10, confirming that this mutation helps remove metal binding, and thus catalytic activity from the Cas9 RuvC domain. Unfortunately, there was no significant homology to other sections of Cas9 to allow us to rationally choose additional mutations.

Supplementary Note 2. Exponential relationship between Cas9 nuclease and Cas9-TF activity

Here we derive a mathematical relationship between the mutation rate induced by a Cas9 nuclease and the expression level of a gene driven by a Cas9-TF that is loaded with the same guide RNA. The derivation is based on two key assumptions:

- (1) All other things being equal, when co-expressed and loaded with the same guide RNA(s), a Cas9 nuclease-gRNA complex, and a Cas9-TF-gRNA complex, will occupy the same target sites and exhibit the same residence times at these sites. In the following, the expression “Cas9-gRNA” will be used to refer equally to either of these two kinds of complexes.
- (2) The activities of Cas9-gRNA complexes, whether formed from Cas9 nucleases or Cas9-TFs, are described by expressions in which residence times are multiplicatively separable from the other factors governing their kinetic behavior.

For simplicity, let us assume there is a single target site. Then by (1) we can assume that there is a constant θ such that, over a period of time T in seconds, the Cas9-gRNAs are resident at their target sites for the amount of time θT per second per genome.

Let us start by considering Cas9-TFs. Let us assume that the Cas9-TF-gRNA’s target site is upstream of a gene that codes for a transcript that we will designate as R , and that the Cas9-TF-gRNA activates expression of R . As a first order approximation, we can assume that the kinetics of R generation are controlled by a 0th order rate constant α and a 1st order decay rate constant δ , so that:

$$\frac{d[R]}{dt} = \alpha - \delta[R]$$

Now, by (2), we can assume that $\alpha = \gamma\theta$, so that

$$\frac{d[R]}{dt} = \gamma\theta - \delta[R]$$

At steady state, we then have $\frac{d[R]}{dt} = \gamma\theta - \delta[R] = 0$, so that

$$[R]_{ss} = \frac{\gamma\theta}{\delta}$$

where $[R]_{ss}$ denotes the steady state level of R that can be measured by RNAseq or other means. This allows us to express θ as proportional to $[R]_{ss}$; i.e.,

$$\theta = \frac{\delta}{\gamma}[R]_{ss}$$

Turning now to Cas9 nucleases, a Cas9 nuclease can be assumed to generate a dsDNA cut at its target site with a constant probability density κ , such that given an infinitesimally small amount of residence time $\Delta\tau$, a cut will be effected with probability $\kappa\Delta\tau$, which (again by (1) and (2)) can be expressed as $\kappa\theta\Delta t$ for a small unit of time Δt in seconds. Let us assume that $\Delta t = \frac{T}{N}$ for some large integer N , where T represents the amount of time that the Cas9-TF and gRNA are coordinately expressed in a given experiment. Then:

$$P(\text{target uncut in time } T) = \lim_{N \rightarrow \infty} (1 - \kappa\theta\Delta t)^N = \lim_{N \rightarrow \infty} \left(1 - \frac{\kappa\theta T}{N}\right)^N = e^{-\kappa\theta T}$$

Therefore, if f_{uncut} is the fraction of genomes that are uncut in time interval T ,

$$f_{\text{uncut}} = e^{-\kappa\theta T} = e^{-\kappa\frac{\delta}{\gamma}[R]_{ss}T} = e^{-\left(\frac{\delta}{\gamma}T\right)[R]_{ss}} = e^{-K[R]_{ss}}$$

or

$$\ln f_{\text{uncut}} = -K[R]_{ss}$$

for constant $K = \kappa\frac{\delta}{\gamma}T$. In other words, if a Cas9-TF and a Cas9-nuclease are set up such that the TF activates gene R by binding a particular sequence when loaded by a gRNA, and the nuclease targets and cuts this same sequence using the same gRNA, and the target sequences and/or gRNAs are changed equivalently for both cases while all other parameters are left unchanged, the steady-state expression level $[R]_{ss}$ achieved by the Cas9-TF, and the fraction of sites cut by the Cas-9 nuclease, will vary according to an exponential law $f_{\text{uncut}} = e^{-K[R]_{ss}}$.

The argument above has only considered the degree of cutting of DNA induced by a Cas9 nuclease, but has not yet considered the introduction of mutations into the target site through dsDNA cut repair by error-prone Non-Homologous End Joining (NHEJ). Let us assume that after any dsDNA cut the target site is repaired by NHEJ, and that the probability that a cut is repaired by NHEJ *without error* is p . The probability of NHEJ repair *with a mutation* will then be $q = 1 - p$.

Now, the same logic that was used to show that $P(\text{target uncut in time } T) = e^{-\kappa\theta T}$ for Cas9 nucleases can be extended to show that the number of Cas9-nuclease cuts in time T follows a Poisson distribution with parameter $\kappa\theta T$.

$$\begin{aligned} P(n \text{ cuts in time } T) &= \lim_{N \rightarrow \infty} \binom{N}{n} (\kappa\theta\Delta t)^n (1 - \kappa\theta\Delta t)^{N-n} \\ &= \lim_{N \rightarrow \infty} \frac{N(N-1)\cdots(N-n+1)}{n!} \left(\frac{\kappa\theta\Delta t}{1 - \kappa\theta\Delta t}\right)^n (1 - \kappa\theta\Delta t)^N \\ &\approx \lim_{N \rightarrow \infty} \frac{(N\kappa\theta\Delta t)^n}{n!} (1 - \kappa\theta\Delta t)^N = \frac{(\kappa\theta T)^n}{n!} e^{-\kappa\theta T} \end{aligned}$$

The probability that a target will be cut repeatedly and repaired error-free by NHEJ each time is then:

$$P(\text{no mutation}) = \sum_{n=0}^{\infty} p^n \frac{(\kappa\theta T)^n}{n!} e^{-\kappa\theta T} = e^{p\kappa\theta T} e^{-\kappa\theta T} = e^{-\kappa\theta T(1-p)} = e^{-q\kappa\theta T}$$

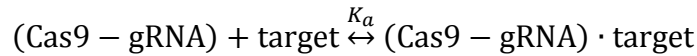
Recalling that $\theta = \frac{\delta}{\gamma} [R]_{ss}$, we then have

$$P(\text{no mutation}) = e^{-q\kappa\frac{\delta}{\gamma}[R]_{ss}T} = e^{-qK[R]_{ss}}$$

Thus, essentially the same exponential law holds between $[R]_{ss}$ and dsDNA cutting as holds between $[R]_{ss}$ and NHEJ-induced mutation, except for an adjustment to the exponential rate constant from in the base of the power law from K to qK .

This exponential relationship cannot be directly extended to one between TALENs vs. TALE-TFs, or ZFNs vs. ZF-TFs, because in these cases the nucleases involve dimerizations that do not play a role in the TF forms of the proteins, which act as monomers. This invalidates assumption (1) because the nuclease and TF forms of the proteins will no longer have equivalent residence times at their targets.

Finally, we note that arguments above do not assume that there is any simple relation between Cas9-gRNA residence time at, and its biochemical affinity to, its target. However, a simple model of residence time suggests that there could be such a relation. Let us assume the equilibrium



where K_a is the affinity constant

$$K_a = \frac{[(\text{Cas9} - \text{gRNA}) \cdot \text{target}]}{[\text{Cas9} - \text{gRNA}][\text{target}]} = \frac{k_{on}}{k_{off}}$$

and where k_{on} and k_{off} are the reaction rate constants for the Cas9-gRNA target-binding and -dissociation half reactions in this equilibrium. If we further assume that a single instance of the bound molecule pair $(\text{Cas9} - \text{gRNA}) \cdot \text{target}$ complex dissociates with a probability density in time λ , such that the probability of dissociating in a time dt is λdt , then by the same argument as we used above, the probability density of this complex dissociating at time t is $\lambda e^{-\lambda t}$. The mean time span during which the complex remains associated, which is equivalent to the residence time of the Cas9-gRNA at its target, is then $\langle t \rangle = 1/\lambda$. Meanwhile, in any given interval dt $\lambda dt [(\text{Cas9} - \text{gRNA}) \cdot \text{target}]$ instances of the complex $(\text{Cas9} - \text{gRNA}) \cdot \text{target}$ will dissociate per unit volume, so that $\lambda = k_{off}$. A closely parallel argument shows that $\eta = k_{on}$,

where η is the probability density over time of a molecule of target associating with a molecule of Cas9-gRNA, and that the mean time during which a target molecule remains *unbound* by Cas9-gRNA is $1/\eta[\text{Cas9} - \text{gRNA}]$. This allows us to derive $\eta/\lambda = k_{on}/k_{off} = K_a$, and also

allows us to establish an estimate of θ as

$$\begin{aligned} \theta &= \frac{\text{mean time target is bound}}{\text{mean time target is bound} + \text{mean time target is free}} = \frac{\frac{1}{\lambda}}{\frac{1}{\lambda} + \frac{1}{\eta[\text{Cas9} - \text{gRNA}]}} \\ &= \frac{1}{1 + \frac{\lambda}{\eta[\text{Cas9} - \text{gRNA}]}} = \frac{1}{1 + \frac{1}{K_a[\text{Cas9} - \text{gRNA}]}} = \frac{K_a[\text{Cas9} - \text{gRNA}]}{1 + K_a[\text{Cas9} - \text{gRNA}]} \end{aligned}$$

Note that when $K_a[\text{Cas9} - \text{gRNA}] = 1$, $\theta = 1/2$. The condition $K_a[\text{Cas9} - \text{gRNA}] = 1$ also implies that $[(\text{Cas9} - \text{gRNA}) \cdot \text{target}] = [\text{target}]$, i.e., that exactly half of the target is bound by the Cas9-gRNA. Indeed, eliminating K_a from the expression above, we see that

$$\theta = \frac{[(\text{Cas9} - \text{gRNA}) \cdot \text{target}]}{[\text{target}] + [(\text{Cas9} - \text{gRNA}) \cdot \text{target}]}$$

so that θ is simply the fraction of the target that is bound by the Cas9-gRNA.

Supplementary Note 3. Sequences of proteins and RNAs used in this study.

A. Sequences of the Cas9_N-VP64 activator constructs based on the m4 mutant are displayed below. Three versions were constructed with the Cas9_{m4}^{VP64} and Cas9_{m4}^{VP64}N fusion protein formats showing highest activity. Corresponding vectors for the m3 and m2 mutants (**Supplementary Fig. 1a**) were also constructed (**NLS** and **VP64** domains are highlighted).

>Cas9_{m4}^{VP64}

```
gccaccATGGACAAGAAGTACTCCATTGGGCTCGTATCGGCACAAACAGCGTCGGCTGGGCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAA
AATTCAAAGTTCGGCAATACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCGACTCCGGGAGACGGCCGAAGCCACGCG
GCTCAAAAAGAACAGCACGGCCGATATACCCGCAGAAAGAAATCGGATCTGTACACGAGGATCTTTAGTAATGAGATGGCTAAGGTGGATGAC
TCTTTCTCCATAGGCTGGAGGAGTCTTTTGGTGGAGGAGGATAAAAAGCAGCAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGT
ACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTGTGACAGTACTGATAAGGCTGACTTGCGGTGTATCTATCTCGCGCTGGCGCA
TATGATCAAATTCGGGGACACTTCCTCATCGAGGGGACCTGAACCCAGACAACAGCGATGTCGACAAAATCTTTATCCAATGGTTCAGACTTAC
AATCAGCTTTTCGAAGAGAACCAGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAAATCCCGCGCGCTCGAAAACC
TCATCGCACAGCTCCCTGGGAGAGAAGAAGCGCCTGTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCAACTTTAAATCTAACTTCGA
CCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGACAATCTGCTGGCCAGATCGGCGACAGTACGCAGACCTT
TTTTGGCGGCAAGAAGCTGTGAGAGCCATCTGTGAGTGATTTCTGCGAGTGAACACGGAGATCACCAGAGCTCCGCTGAGCGCTAGTATGA
TCAAGCGCTATGATGAGCGCCAGACTTGACTTTGCTGAAGCGCTTGTACACGCAACTGCCTGAGAAGTACAAGAAAATTTTCTCGACTCA
GTCTAAAATGGCTACGCCGATACATTGACGGCGGACAAGCCAGGAAATTTTACAAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACC
GAGGAGCTGTGGTAAAGCTTAAACAGAGAAGATCTGTGCGCAACAGCGCACTTTCGACAATGGAAGCATCCCCACAGATTCACCTGGGCGAAC
TGACCGTATCCTCAGCGGCAAGAGGATTTCTACCCCTTTTGAAGATAACAGGGAAAAGATTGAGAAAATCCTCACATTCGGATACCCCTACTA
TGTAGCCCCCTCGCCGGGAAATTCAGATTCGCGTGGATGACTCGCAATCAGAAGAGACCATCCTCCCTGGAATTCGAGGAAGCTGCTGGAT
AAGGGGGCTCTGCCAGTCTTCATCGAAAGGATGACTAATTTGATAAAAAATCTGCCTAACGAAAAGGTGCTTCTAAACACTCTCTGCTGTACG
AGTACTTCACAGTTTATAACGAGCTCACCAAGTCAAATACGTACAGAAGGGATGAGAAAGCCAGCATTCCTGTCTGGAGAGCAGAAGAAGCTAT
CTGGGACCTCTTCAAGACGAAACCGGAAAGTTACCTGAAACAGCTCAAAGAAAGATTTTCAAAGAAATGAAATTTTTCGACTCTGTTGAAATG
AGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGCAAGGACTCCTGGCAATGAGGAGAACG
AGCATTTCTGAGGACATTTGCTCTCACCCCTTACCTTTGTAAGATAGGGAGATGATGAAAGACGCTTGAAGAACTTGAAGAACTTCACTCTTCGACGA
CAAAGTCAAGAACAGCTCAAGAGCGCCGATATACAGGATGGGGCGGCTGTCAAGAAAATGATCAATGGGATCCGAGACAAGCAGAGTGGAAAG
ACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAACTTCATGCAAGTGTATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGA
AAGCAAGTTTCTGGCCAGGGGACAGTCTTACGAGCACATCGTAACTTTCGAGGTAGCCAGCTATCAAAAAGGAATACTGCAGACCGTTAA
GGTCTGGATGAACTCGTCAAAGTAAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAGATGGCCCGAGAGAACCACAACTACCCAGAAGGACAG
AAGAACAGTAGGAAAGGATGAAGAGGATGAAGAGGATATAAAGAAGTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTTC
AGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACCTGGACATCAATCGGCTCTCCGACTACGACGTGGC
TGCTATCGTGGCCAGTCTTTCTCAAAGATGATTTCTATTGATAAATGATGTTGACAAGATCCGATAAAgcTAGAGGGAAGAGTGATAACGTCCCC
TCAGAAGAAGTTGTCAGAAAATGAAAATTTATGGCGGCAGTGTGTAAGCAGCAACTGATCACACAACCGGAAGTTGATAATCTGACTAAGGCTG
AACGAGGTGGCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAGGCAGCTTGTGAGACACGCCAGATCACAAGCACGTGGCCCAAAATTCGCA
TTCACCGATGAACACCAAGTACGATGAAAATGACAACTGATTCGAGAGGTGAAAGTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTCAAGAAAG
GACTTTTCAGTTTATAAGGTGAGAGATCAACAATACCACATCGCATGATGCTTACCTGAATGCAGTGGTAGGCACTGCCTTATCAAAAAT
ATCCAAAGCTTGAATTTGTTTTACGGAGACTATAAAGTGTACGATTTTGGCAAAGTGTAGGAAAATGATCGCAAAGCTGACGAGGAAAATAGCCAAAGC
CGTAAGTACTTCTTTTACGCAATATATGAAATTTTTTCAAGACCGAGATTACTGGCCAAATGGAGAGATTCGGAAGCGACCACTTATCGAAACA
AACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTTCGCGACAGTCCGGAAAGTCTGTCCATGCGCAGGTGAACATCGTTAAAAGA
CCGAAGTACAGACCGGAGGCTTCTCAAAGAAAGTATCTCCGAAAAGGAACAGCGACAAGTGTATCGCACGCAAAAAGATTGGGACCCCAAGAA
ATACGGCGGATTCGATCTTCTACAGTCTACAGTGTGTTGGCAAAGTGTGGGAAAAGGAAAGTCTAAAAGAAAGTCTAAAAGAAAGCTCAAGGAA
CTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCATCGACTTCTCGAGGCGAAAGGATAATAAGAGGTCAAAAAGACCTCA
TCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAGACGGCCGGAACGAAATGCTCGCTAGTGCAGGCGAGCTGCAGAAAAGGTAACGAGCTGGC
ACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCATATGAAAAGTCAAAAGGCTCCTCCGAAAGATAATGAGCAGAAGCAGCTGTTCTGTG
GAACAACACAAACACTACCTTGATGAGATCATCGAGCAAAATAAGCGAATTTCTCCAAAAGAGTATCCTCGCCGACGCTAACCTCGATAAAGGTGCTTT
CTGCTTACAATAAGCACAGGATAAGCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCTGCAGCCTT
CAAGTACTTCGACACCACATAGACAGAAAGCGGTACACCTTACAAAAGGAGTCTGGACGCCACACTGATTCATCAGTCAATTACGGGGCTCTAT
GAAACAAGAAATCGACCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGGAGGTTGGAGGCCAGCGTTCCGGACGGGCTGACG
CATTTGGACGATTTGATCTGGATATGCTGGGAAGTGAAGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCCGATGCCCTTGGATGACTTTGACCT
CGACATGCTCGGCAGTGAAGCCCTTGTATGATTTTCGACCTGGACATGCTGATTAACCTAGATGA
```

>Cas9_{m4}^{VP64}N Sequences

```
gccaccATGCCCAAGAAGAAGGAAAGGTTGGGAAGGGGGATGGACAAGAAGTACTCCATTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGGGCCG
TCATTACGGACGAGTACAAAGTCCGAGCAAAAATTCAAAGTTCGGCAATACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCT
GTTCTGACTCCGGGGAGACGGCCGAAGCCACGGCGCTCAAAGAAGCAGCAGCGCGCAGATATACCCGCAGAAAGAAATCGGATCTGTACCTGCAGGAG
ATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGCTGGAGGAGTCTTTTGGTGGAGGAGGATAAAAAGCAGCAGCGCCACC
CAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGC
```

TGACTTGGCGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTTGGGGGACTTCTCATCGAGGGGACCTGAACCCAGACAACAGCGATGTC
GACAACTCTTTTAACTGGTTTCAGACTTACAATCAGCTTTTCGAAAGGAAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCTCTGAGCG
CTAGGCTGTCAAATCCCGCGCTCGAAACCTCATCGCACAGTCCCTCGGGGAGAAAGAACCGGCTGTTTGGTAAATCTTATCGCCCTGTCACT
CGGGCTGACCCCAACTTTAAATCTAATCTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCGACAATCTG
CTGGCCAGATCGGGCAGCAGTACGCAGACCTTTTTTTGGCGGCAAAGAACCTGTGACAGCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGG
AGATCACCAAAGCTCCGTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCCCAAGACTTGACTTTGCTGAAGGCCCTTGTGACAGACAAC
GCCTGAGAAGTACAAGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGATACATTTGACGGCGGAGCAAGCCAGGAAATTTTCAAATTT
ATTAAGCCATCTTGGAAAAATGGACGGCACCAGGAGCTGTGGTAAAGCTTAAACAGAGAAGATCTGTTGCGCAACAGCGCACCTTCGACAATG
GAAGCATCCCCACCAGATTACCTGGGCGAAGTGCAGCTATCCTCAGCGGCAAGAGGATTCTACCCCTTTTTGAAAGATAACAGGGAAAGAT
TGAGAAAATCCTCACATTTCCGATACCTACTATGTAGGCCCTCGCCCGGGGAAATTCAGATTGCGGTGGATGACTCGCAATCAGAAGAGACC
ATCACTCCCTGGAATTCGAGGAAGTCTGGATAAGGGGCGCTGTCGCCAGTCTTCACTGAAAGGATGACTAACTTTGATAAAAAATCTGCCTAACG
AAAAGGTCTTCCAAACACTCTCTGCTGTACAGTACTTACAGTTTAAACAGAGCTCACCAAGGTCAAATACGTACACAGAAGGGATGAGAAAAGCC
AGCATTCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCTCTTCAAGCAGAACCGGAAAGTTACCGTGAACAGCTCAAAGAAAGACTATTT
AAAAGATTGAATGTTTTCGACTCTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGATACAGATCTCTCGAAAATCATT
AAGACAAGGACTTCTGGACAATGAGGAGAACGAGGACATTTGAGGACATTTGCTCACCCCTACGTTGTTGAAGATAGGGAGATGATTGAAGA
ACCTTTGAAAACCTTACCTCTTTCGACACAAGTCTTTCGACGACAAAGCTGTAAGAACAGCTCAAGAGCGCCGATATACAGGATGAGGCGGCTGTCAA
ATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCTGGATTTTCTTAAAGTCCGATGGATTGGCAACCGGAACTTCAATGCAGTTGATCCATG
ATGACTCTCTACCTTTAAGGAGGACATCCAGAAAAGCAAGTTTCTGGCCAGGGGACAGTCTTACGAGACATCGCTAATCTTGCAGGTAGCC
AGCTATCAAAAAGGAACTACTCGACACCGTTAAGGTCGTGGATGAATCTGCAAAGTAATGGAAAGCATAAGCCCGAGAATATCGTTATCGAGATG
GCCGAGAGAACCACCAACTCCAGGAAGGACAGAAGAACAGTGAAGGAACTGCTGAGTTGGATAAAAGCCGGCTTCAAAAAGCCAGCTTGTGAGACAG
TTAAGGAACCCAGTTGAAAACACCCAGCTTCAAGATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACT
GGACATCAATCGGCTCTCCGACTACGACGTGGCTGCTATCGTCCCGAGTCTTTTCTCAAAGATGATCTATTTGATAATAAAGTGTGACAAGATCC
GATAAAgcTAGAGGGAAGAGTATAACCTCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTTATGGCGGACAGTCTGACCCCAAACTGATCA
CACAAAGGAAAGTTTCAATCTGACTAAGGCTGAACGAGGTGGCCGTGCTGAGTTGGATAAAAGCCGGCTTCAAAAAGCCAGCTTGTGAGACAG
CCAGATCACCAAGCAGTGGCCCAAATTTCTCGATTACGCATGAACACCAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACT
CTGAAGTCTAAGTGGTCTCAGATTTCAAGAAAGACTTTCAGTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCTTACTGTA
ATGCAGTGGTAGGCATGCACTTATCAAAAATATCCCAAGCTGAACTGAAATTTGTTTACGGAGACTATAAAGGTGACGATGTTAGGAAAATGAT
CCGAAAGTCTGACGAGGAACTCAGTAAGGCTAAGGCTGACGAGTCTTCTTAAAGTCTTCTTAAAGTCTTCTTAAAGTCTTCTTAAAGTCTTCT
GGAGAGATTGGGAAGCGACCACTTATCGAAACAACCGGAGAAACAGGAGAAATCGTGTGGACAAGGGTAGGGATTTCGCGACAGTCCGGAAGGTCC
TGTCATGCCGAGGTGAACATCGTTAAAAGACCGAAGTACAGACCGGAGGCTTCTCAAGGAAAGTATCTCCCGAAAAGGAAACAGCGACAGCT
GATCGCACGAAAAAGATTGGGACCCCAAGAAATACGGCGGATTCGATTCTCTACAGTCTGTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAA
GGGAAGTCTAAAAAAGCTCAAAGCGTCAAGGAACTGTGGGATCACAAATCATGGAGGATCAAGCTTCGAAAAAAGCCCACTGCGACTTCTCGAGG
CGAAAGGATATAAAGAGGTCAAAGACCTCATCATTAAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAGAACGGCCGAAACGAATGCTCGCTAG
TGCGGGCAGCTCGAGAAAGGTAAAGACTGGCACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGTTCT
CCCGAAGATAATGAGCAGAAGCAGTGTTCGTGGAACAACACAACACTACCTTGATGAGATCATCGAGCAATAAGCGAATTTCTCCAAAAGAGTGA
TCTCAGCGCAGCTAACCTCGATAAGGTGCTTCTGCTTACAATAAGCACAGGGATAAGCCCAATCAGGGAGCAGGCAGAAAACATTTACCTTGT
TACTCTGACCAACTTGGGCGCGCTGCAGCCTTCAAGTACTTCGACACCACATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCTGGACGCC
ACACTGATTCATCAGTCAATTACGGGGCTCTATGAAAACAAGAAATCGACTCTCTCAGCTCGTGGAGACAGCAGGGCTGACCCCAAGAAAGAGGA
AGGTGGAGGCCAGCGTTCCGGACGGGCTGACGCATTTGGACGATTTGATCTGGATATGCTGGGAAGTGAACGCCCTCGATGATTTGACCTTGACAT
GCTTGGTTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGAACGCCCTTGATGATTTGACCTGGACATGCTGATTAACCTAGATGA

>Cas9_{m4}^{VP64C}

gccaccATGGACAAGAAGTACTCCATTGGGCTCGTATCGGCACAAACAGCGTCGGCTGGGCCGTCTATTACGGACGAGTACAAGGTGCCGAGCAAAA
AATTCAAAGTTCGGGCAATCCGATCGCCACAGCATAAAGAAAGAACCTCTATTCGGCGCCCTCTGTTTCGACTCCGGGAGAGCGGCCAAGCCACCGC
GCTCAAAGAACACGACCGGCGAGATATACCCGCAAGAAATCGGATCTGCTAGCAGGAGATCTTTAGTAATGATGCTGCGAAGGTTGATGAC
TCTTTCTTCCATAGGCTGGAGGAGTCTTTTTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCAATCTTTGGCAATATCGTGGACGAGGTGGCGT
ACCATGAAAAGTACCCAAACATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGGCGTTGATCTATCTCGCGCTGGCGCA
TATGATCAAATTTCCGGGACACTTCTCATCGAGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAATGGTTTCAGACTTAC
AATCAGCTTTTCGAAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCTGAGCGCTAGGCTGTCCAAATCCCGCGGCTCGAAAAC
TCATCGCACAGTCCCTGGGAGAGAAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCAACTTTAAATCTAACTTCGA
CCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTAGATGATGATCTCGACAATCTGCTGGCCAGATCGGGCAGCAGTAGCAGACCTT
TTTTTGGCGGCAAGAACCTGTGAGACCCATCTGCTGAGTATCTGCGAGTGAACACGGAGATCACAAAGCTCCGCTGAGCGCTAGTATGA
TCAAGCGCTATGATGAGCAACCAAGACTTACTTTGCTGAAAGGCTGTGACAGCAAGCAACTGCTGAGCAAGCAACTGCTGAGGAAATTTTCTCGATCA
GTCTAAAAATGGCTACGCCGATACATTTGACGGCGGAGCAAGCCAGGAGAAATTTTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCAC
GAGGAGTCTGGTAAAGCTTAAACAGAGAAGATCTGTTGCGCAACAGCGCACTTTCGACAATGGAAGCATCCCCACCAGATTCACCTGGGCGAAC
TGACAGCTATCTCAGCGGCAAGAGGATTTTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGAAAATCTCTACATTTTCGGATACCTACTA
GTAGGCCCCCTCGCCGGGAAATTCAGATTTCGCTGGATGACTCGCAATCAGAAGACACTCACTCCCTGGAATTCGAGGAAGTCTGTGAT
AAGGGGGCTCTGCCGCTCTTCAATCGAAAAGGATGACTAATTTGATAAAAAATCTGCCAAACGAAAGGTTGCTTCTTAAACACTCTCTGCTGACG
AGTACTTACAGTTTATAACGAGCTACCAAGTCAAATACGTACAGAAGGATGAGAAAGCCAGCATCTCTGCTGGAGAGCAGAAGAAAGCTAT
CTGGACCTCTCTTCAAGACGAAACCGGAAAGTTACCGTGAACAGCTCAAAGAAAGACTATTTCAAAGAAATGAAATGTTTCGACTCTGTTGAAAT
AGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGATACAGATCTCTGAAAATCATTAAGACAAGGACTTCTGGACAATGAGGAGAAGC
AGGACATTTGAGGACATTTGCTTACCCCTTACGTTGTTGAAAGATAGGGAGATGATTTGAAGAACCGTTGAAAACCTTACGCTTCTTCCGACGA
CAAAGTCAAGAACAGCTCAAGAGCGCGGATATACAGGATGGGGCGGCTGTCAAGAAAATGATCAATGGGATCCGAGACAAGCAGAGTGGAAAG
ACAATCTGGATTTTCTTAAAGTCCGATGGATTTGCCAACCGGAACTTCAATGCAAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCGA
AAGCAAAGTTTCTGGCCAGGGGACAGTCTTACGAGCACATCGTAAATCTGCAAGTACGCCAGCTATCAAAAAGGAAATCTCCGAGACCGTTAA
GGTCTGGATGAACTCGTCAAAGTAAAGGAGGCAAGCCGAGGATATCGTTGAAAGATAGGGAGATGATTTGAAGAACCGTTGAAAACCTTACGCTTCTTCCGACGA
AAGAACAGTAGGAAAGGATGAAGAGGATTAAGAGGGTATAAAAAGACTGGGGTCCCAATCTTAAAGAACACCCAGTTGAAAACACCCAGCTTC

AGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGC
 TCGTATCGTGCCCGAGTCTTTCTCAAAGATGATTCTATTGATAATAAAGTGTGACAAGATCCGATAAAgcTAGAGGGAAGAGTGATAACGTCCCC
 TCAGAAGAAGTTGTCAAGAAAATGAAAATTTATGGCGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAGGCTG
 AACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAGGCAGCTTGTGAGACACGCCAGATCACAAGCACGTGGCCAAAATTTCTCGA
 TTACGCGATGAACACCAAGTACGATGAAAATGACAACTGATTGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTAGAAAAG
 GACTTTGAGTTTATAAGGTGAGAGAGATCAACAATACCACCATGCCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAAT
 ATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCAC
 CGCTAAGTACTTTTACAGCAATATTATGAATTTTTCAAGACCGAGATTACACTGGCCAAATGGAGAGATTGGAAAGCGACCCTTATCGAAAACA
 AACGGAGAACAGGAGAAATCGTGTGGGACAAGGTTAGGATTTCCGACAGTCCGGAAGTCTGTCCATGCCGAGGTGAACATCGTTAAAAGA
 CCGAAGTACAGACCGGAGGCTTCTCCAGGAAAGTATCTCCGAAAAGGAAACAGCGACAAGCTGATCGCACGCAAAAAGATTTGGGACCCCAAGAA
 ATACGGCGGATTCGATTTCTTACAGTCTGTTACAGTGTACTGGTGTGGCCAAAGTGGAGAAAGGGAAGTCTAAAAAATCAAAGCGTCAAAGGAA
 CTGCTGGCATCACAAATCATGGAGCGATCAAGCTTCGAAAAAACCCCATCGACTTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAGACCTCA
 TCATTAAGCTTCCCAAGTACTCTCTTTGAGCTTGAAAACGGCCGAAACGAATGCTCGCTAGTGGCGGCGAGCTGCAGAAAAGGTAACGAGCTGGC
 ACTGCCCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCATATGAAAAGCTCAAAGGCTCTCCGAAAGATAATGAGCAGAAAGCAGCTGTTCTGTG
 GAACAACAAAACACTACCTTGATGAGATCATCGAGCAATAAGCGAATTTCCAAAAGAGTGATCTCGCCGACGCTAACCTCGATAAGGTGCTTT
 CTGCTTACAATAAGCAGGGTAAGCCATCAGGGACAGGAGCAAAAACATTATCCACTTGTGTTACTCTGACCAACTGGGCGCCCTGCAGCCCTT
 CAAGTACTTCGACACCACATAGACAGAAAGCGGTACACCTTACAAAAGGAGTCTGGACGCCACACTGATTCATCAGTCAATTACGGGCTCTAT
 GAAACAAGAATCGACCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAGAGGAAAGGTGGAGGCCAGCGGTTCCGGACGGGCTGACG
 CATTGGACGATTTGATCTGGATGCTGGGAAGTGACGCCCTCGATGATTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCT
 CGACATGCTCGGCAGTGCAGCCCTTGATGATTTGACCTGGACATGCTGATTAACCTAGAGCGGCCCGAGATCCAAAAAGAAGAGAAAGGTAGAT
 CCAAAAAAGAAGAGAAAGGTAGATCCAAAAAGAAGAGAAAGGTAGATACGGCCGATAG

B. Sequences of the MS2-activator constructs and corresponding gRNA backbone vector with 2X MS2 aptamer domains is provided below (NLS, VP64, gRNA spacer, and MS2-binding RNA stem loop domains are highlighted). Two versions of the former were constructed with the MS2_{VP64}N fusion protein format showing highest activity.

>MS2_{VP64}N

gccaccATGGGACCTAAGAAAAAGAGGAAGGTGGCGGCGCTTCTAGAATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTG
 GCGACGTGACTGTGCCCCAAGCAACTTCGCTAACGGGATCGTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGT
 TCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTGCAGGTGCCTAAAGGCGCCTGGCGTTCGTACTTAAATATGGAACCTAACCATCCA
 ATTTTCGCCACGAATTCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCCTCTAAAAGATGGAACCCGATTCCTCAGCAATCGCAGCAAACT
 CCGGCATCTACGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCT
 TGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGCAGCCCTTGATGATTTGACCTGGACATGCTGATTAACCTCT
 AGATGA

>MS2_{VP64}C

gccaccATGGGACCTAAGAAAAAGAGGAAGGTGGCGGCGCTTCTAGAATGGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTG
 GCGACGTGACTGTGCCCCAAGCAACTTCGCTAACGGGATCGTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCGT
 TCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAAAGTGCAGGTGCCTAAAGGCGCCTGGCGTTCGTACTTAAATATGGAACCTAACCATCCA
 ATTTTCGCCACGAATTCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCCTCTAAAAGATGGAACCCGATTCCTCAGCAATCGCAGCAAACT
 CCGGCATCTACGAGGCCAGCGGTTCCGGACGGGCTGACGCATTGGACGATTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCT
 TGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGCAGCCCTTGATGATTTGACCTGGACATGCTGATTAACCTCT
 AGAGCGGCCCGAGATCCAAAAAGAAGAGAAAGGTAGATCCAAAAAGAAGAGAAAGGTAGATCCAAAAAGAAGAGAAAGGTAGATACGGCCGAT
 AG

>gRNA_{2xMS2}

TGTACAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAGGTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCTTTCAT
 ATTTGCATATACGATACAAGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAACACAAAGATATTAGTACAAAATACGTGACGTAGAAAGTA
 ATAAATTTCTGGTAGTTGTCAGTTTTAAAAATATGTTTTAAATGGACTATCATATGCTTACCCTAACTTGAAGTATTTTCGATTTCTGGCTTTA
 TATATCTTGTGGAAGGACGAAACACCCNNNNNNNNNNNNNNNNNNNTTTTAGAGCTAGAAATAGCAAGTAAAATAAGGCTAGTCCGTTATCAA
 CTTGAAAAAGTGGCACCGAGTCCGTTGCTCTGCAGGTCGACTTAGAAAACATGAGGATCACCCATGCTGTCAGTATTTCCGGGTTTCATTAGATCCTA
 AGGTACCTAATTCCTAGAAACATGAGGATCACCCATGCTGTCAGGTCGACTCTAGAAATTTTTCTAGAC

C. dTomato fluorescence based transcriptional activation reporter sequences are listed below (**ISceI** control-TF target, **gRNA** targets, **minCMV** promoter and **FLAG tag + dTomato** sequences are highlighted).

>TF Reporter 1

```
TAGGGATAACAGGGTAATAGTGTCCCTCCACCCACAGTGGGGCGAGGTAGGCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAA
CCGTCAGATCGCCTGGAGAATTGccaccatgGACTACAAGGATGACGACGATAAAACTTCCGGTGGCGGACTGGGTTCACCGTGAGCAAGGGCGA
GGAGGTCAATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCC
TACGAGGGCACCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGGCCCCCTGCCCTTCGCTGGGACATCCTGTCCCCCAGTTCATGTACGGTCCA
AGGCGTACGTGAAGCACCCCGCGACATCCCCGATTACAAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGG
CGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCAGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCGACGGCCCC
GTAATGCAGAAGAAGACCATGGGCTGGGAGGCCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGTGAAGGGCGAGATCCACCAGGCCCTGAAGC
TGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCACACTGCCGGCTACTACTACGTGGACACCAAGCT
GGACATCACCTCCCAACAGGAGACTACACCATCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGTTCTGTACGGCATGGACGAGCTG
TACAAGTAA
```

>TF Reporter 2

```
TAGGGATAACAGGGTAATAGTGGGGCCACTAGGGACAGGATTGGCGAGGTAGGCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAA
CCGTCAGATCGCCTGGAGAATTGccaccatgGACTACAAGGATGACGACGATAAAACTTCCGGTGGCGGACTGGGTTCACCGTGAGCAAGGGCGA
GGAGGTCAATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCC
TACGAGGGCACCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGGCCCCCTGCCCTTCGCTGGGACATCCTGTCCCCCAGTTCATGTACGGTCCA
AGGCGTACGTGAAGCACCCCGCGACATCCCCGATTACAAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGG
CGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCAGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCGACGGCCCC
GTAATGCAGAAGAAGACCATGGGCTGGGAGGCCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGTGAAGGGCGAGATCCACCAGGCCCTGAAGC
TGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCACACTGCCGGCTACTACTACGTGGACACCAAGCT
GGACATCACCTCCCAACAGGAGACTACACCATCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGTTCTGTACGGCATGGACGAGCTG
TACAAGTAA
```

D. General format of the reporter libraries used for TALE and Cas9-gRNA specificity assays is provided below (**ISceI** control-TF target, **gRNA/TALE** target site (23bp for gRNAs and 18bp for TALEs), **minCMV** promoter, **RNA barcode**, and **dTomato** sequences are highlighted).

> Specificity Reporter Libraries

```
TAGGGATAACAGGGTAATAGTNNNNNNNNNNNNNNNNNNNNNNCGAGGTAGGCGGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAA
CCGTCAGATCGCCTGGAGAATTGccaccatgGACTACAAGGATGACGACGATAAANNNNNNNNNNNNNNNNNNNNNNACTTCCGGTGGCGGACT
GGGTTCACCGTGAGCAAGGGCGAGGAGGTCAATCAAAGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACGGCCACGAGTTCGAGATC
GAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGGCCCCCTGCCCTTCGCTGGGACATCCTGT
CCCCCAGTTCATGTACGGTCCAAGGCGTACGTGAAGCACCCCGCGACATCCCCGATTACAAGAAGCTGTCTTCCCCGAGGGCTTCAAGTGGGA
GCGCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACGCTGATCTACAAGGTGAAGATGCGCGGC
ACCAACTTCCCCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCCTCCACCGAGCGCCTGTACCCCCGCGACGGCGTGCTGAAGG
GGAGATCCACCAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAGACCATCTACATGGCCAAGAAGCCCGTGCACACTGCCCGG
CTACTACTACGTGGACACCAAGCTGGACATCACCTCCCAACAGGAGACTACACCATCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTG
TTCTGTACGGCATGGACGAGCTGTACAAGTAAGAATTTC
```