

Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease

James E Dahlman^{1,2,7}, Omar O Abudayyeh^{1,3-5,7},
Julia Joung^{1,3-5}, Jonathan S Gootenberg^{1,3-6},
Feng Zhang^{1,3-5} & Silvana Konernmann^{1,3-5}

We have developed a CRISPR-based method that uses catalytically active Cas9 and distinct single guide (sgRNA) constructs to knock out and activate different genes in the same cell. These sgRNAs, with 14- to 15-bp target sequences and MS2 binding loops, can activate gene expression using an active *Streptococcus pyogenes* Cas9 nuclease, without inducing double-stranded breaks. We use these ‘dead RNAs’ to perform orthogonal gene knockout and transcriptional activation in human cells.

The RNA-guided Cas9 nuclease from the microbial, clustered, regularly interspaced, short palindromic repeats (CRISPR) and the CRISPR-associated protein 9 (Cas9) system holds promise as a system that can be used for both gene knockout and activation^{1,2}. Once bound to its target DNA, the active sgRNA:Cas9 complex can induce a double-stranded break through its RuvC and HNH domains. By mutating these nuclease domains, Cas9 can be made catalytically inactive^{3,4} and repurposed for genetic perturbation beyond DNA editing. For example, this ‘dead’ Cas9 (dCas9) has been combined with protein domains that suppress or activate gene expression⁵⁻⁹ or change the epigenetic state of a target locus¹⁰. The simplest CRISPR-based orthogonal gene control system would function with only catalytically active wild-type Cas9, in contrast to previous work that used distinct Cas9 proteins from different organisms². Here we show that by reducing the length of the RNA targeting sequence to 14–15 nucleotides (nt), and by adding MS2 binding loops into the sgRNA backbone¹¹, sgRNAs can guide catalytically active Cas9 to activate transcription without inducing double-stranded breaks (Fig. 1a). After systematically studying how the design of these ‘dead sgRNAs’ (dRNAs) influences transcriptional activation and indel formation, and characterizing off-target transcriptional perturbations, we show that this system can simultaneously knock out and upregulate target genes in melanoma cells. These data demonstrate that sgRNAs can be engineered to exert

transcriptional control using active Cas9, and that a single active Cas9 protein can be used for orthogonal gene control in mammalian cells.

Changes in sgRNA structure and mismatches between the sgRNA targeting sequence and DNA can prevent Cas9-mediated DNA cleavage¹²⁻¹⁴. However, it is unknown whether these modified sgRNAs still allow binding of Cas9 to the DNA target. To test this, we designed sgRNAs with two structural characteristics. First, we added two aptamers that selectively bind dimerized MS2 bacteriophage coat proteins to the tetraloop and stem loop two of the sgRNA, as previously described (sgRNA-MS2)¹¹. Second, we shortened the length of the sgRNA guide sequence from 20 nt to 11 nt. We reasoned that these changes could result in a dRNA that would still enable binding of Cas9 while preventing nuclease activity (Fig. 1a).

We transfected 80 sgRNA-MS2s targeting four DNA sequences within 200 bp of the transcriptional start site of human hemoglobin 1 (*HBG1*), together with active Cas9 and the MS2:P65:HSF1 (MPH) activation complex. The MPH complex has previously been reported to mediate efficient target upregulation by binding to MS2 loops in the sgRNA¹¹. We observed that guides from 16–20 nt resulted in indel formation, whereas shorter guides (11–15 nt) did not have detectable levels of indel formation in most cases (Fig. 1b). Notably, guides truncated to 11–15 nt of complementarity to the target DNA were able to increase *HBG1* mRNA expression by as much as 10,000-fold (Fig. 1c and Supplementary Fig. 1). We then investigated guides with mismatches on the 5' end of the sgRNA analogous to our truncation experiments (Supplementary Fig. 2). In accordance with our results from truncated guides, we observed that guides with only 15 bp complementarity to the target DNA were still able to mediate efficient activation in all four cases.

We proceeded to investigate the gene activation efficiency of 14-nt and 15-nt dRNAs at three loci. In all cases, when HEK293FT cells were co-transfected with the dRNAs, active Cas9 and the MPH complex, the dRNAs increased target mRNA expression of all three human genes (*HBG1*, interleukin 1B (*IL1B*) and zinc finger protease 42 (*ZFP42*)) without inducing significant indel formation (Fig. 1d). Notably, dRNA activation was similar to the recently reported system using dCas9 in combination with a 20-nt sgRNA-MS2 (ref. 11). At all three loci 20-nt sgRNAs cut target DNA and did not activate gene expression when combined with active Cas9. This was true for sgRNAs with and without the MS2 binding loops (Fig. 1d).

Biological studies using activators will require specific target upregulation. Therefore, it is important to understand the specificity of Cas9-mediated gene activation. In addition, recent work has demonstrated that a single sgRNA can bind many sites in the genome, but the relationship between binding and transcriptional control is not clear¹⁵. Specificity may also change when the sgRNA is shortened to

¹Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. ²Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology, Atlanta, Georgia, USA. ³McGovern Institute for Brain Research, MIT, Cambridge, Massachusetts, USA. ⁴Department of Brain and Cognitive Sciences, MIT, Cambridge, Massachusetts, USA. ⁵Department of Biological Engineering, MIT, Cambridge, Massachusetts, USA. ⁶Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, USA. ⁷These authors contributed equally to this work. Correspondence to J.E.D. (james.dahlman@bme.gatech.edu) or S.K. (silvanakonernmann@gmail.com).

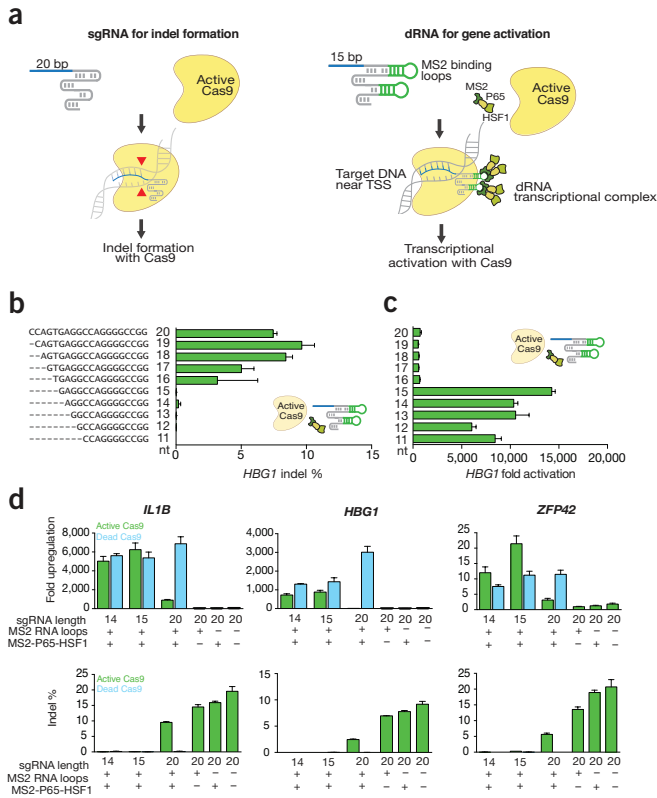


Figure 1 dRNAs can mediate robust gene activation using an active Cas9. (a) dRNA-mediated gene activation. 15-bp dRNA with MS2 loops on the scaffold can bind the MPH transcriptional activation complex and activate gene expression without inducing Cas9-mediated DNA indel formation. (b,c) The length of the RNA targeting sequence was varied from 11 nt to 20 nt. *HBG1* indel frequencies (b) were quantified with *HBG1* mRNA levels (c) (normalized to *GAPDH*, and compared to cells transfected with GFP plasmid). No indel formation was observed when sgRNAs had less than 16 bp of homology to target DNA. In all cases, guides were designed with MS2 binding loops in the tetraloops and stemloop two¹¹ and were transfected with active Cas9 and the MPH transcriptional activation complex. (d) Three dRNAs targeting the promoter regions of *IL1B*, *HBG1* and *ZFP42* were tested for activation and indel formation. dRNAs with 14 bp or 15 bp of homology to target DNA did not induce detectable indel formation. dRNAs transfected with Cas9 and MPH activated transcription to a similar extent as 20-nt sgRNA-MS2 transfected with dCas9 and MPH. (In all cases, mean ± s.e.m. is plotted. *N* = 2–3 replicates/group.) Guide RNAs targeting *IL1B*, *HBG1* and *ZFP42* were derived from the 20-bp spacers AAAGGGGAAAAGAGTATTGG, GGCAAGGCTGGCCAACCCAT and ACCCTGGCGGAGCTGATGGG, respectively.

15 bp. To assess the difference in specificity between 20-nt sgRNA-MS2 and 15-nt dRNAs, we compared whole transcriptome mRNA levels in HEK293FT cells. Cells were co-transfected with dCas9, the MPH complex and a 20-nt activator sgRNA-MS2, or active Cas9, the MPH complex and 15-nt dRNA targeting the same sequence in the human *HBG1/2* promoter. We previously determined that *HBG1/2* upregulation induces limited downstream effects that could confound our analysis in HEK293FT cells. RNA-seq results showed that both the sgRNA-dCas9 and dRNA systems significantly ($P < 0.05$) activated *HBG1/2* only, demonstrating that dRNAs can specifically upregulate target genes (Fig. 2a). We next performed off-target analysis on a second 15-nt dRNA and 20-nt sgRNA targeting the same *HBG1/2* promoter and found they had 32 and 55 perturbed transcripts,

respectively, and 20 of these were shared for both gRNAs (Fig. 2b). We expanded this analysis by studying whole transcriptome analysis on ten additional sgRNAs targeting the proximal promoter of *HBG1/2* (Supplementary Fig. 3). Overall, 4 out of 12 20-bp guides exhibited high specificity (<3 significant genome-wide off-targets), confirming activators can be specific. Notably, a previously published algorithm that predicts off-target indels did not correlate with the number of nontargeted transcripts that were altered by each guide ($R = 0.12$, $P = 0.7$)¹³.

Finally, we tested whether dRNAs in combination with sgRNAs could mediate orthogonal gene control (activation and knockout) using only active Cas9. We previously used CRISPR-Cas9 loss-of-function¹⁶ and gain-of-function¹¹ screens to identify genetic modifiers that promote resistance of A375 melanoma cells to the BRAF inhibitor PLX-4720. We performed orthogonal gene regulation with hits selected from these screens. We first transduced and selected A375 cells with two lentiviral constructs encoding active Cas9 and the MPH complex, respectively (Fig. 2c). We then transduced these cells with lentiviral constructs encoding a dRNA targeting *LPAR5* for activation and/or sgRNAs targeting *MED12* or *TADA2B* for gene knockout. *LPAR5* mRNA expression increased over 600-fold when cells were treated with dRNA targeting *LPAR5*, even when combined with sgRNAs targeting other genes. In all conditions, average *LPAR5* indel rates equal to 0.85% or less were detected (Fig. 2d). Average indel rates of 0.6% and 0.05% at targeted loci were also measured after cells were treated with two additional dRNAs targeting *EGFR* and *ITGA9*, respectively (Supplementary Fig. 4). By contrast, the loci targeted by *MED12* and *TADA2B* showed robust indel formation of 33–36% and 67.4–91.5%, respectively, even in orthogonal conditions (Fig. 2e).

dCas9 proteins have been engineered for diverse functions including gene suppression, gene activation and epigenetic modification^{5–10}. Here we have demonstrated that guide RNAs can be engineered to bind target DNA and successfully recruit a transcriptional activation complex, without inducing measurable indel formation using an active Cas9 nuclease.

These dRNAs can be designed to specifically and potently upregulate gene expression (Figs. 1 and 2). We anticipate that dRNAs may also be designed to perform other genetic functions, for example, epigenetic modification or mRNA suppression. dRNAs can be used to simplify orthogonal gene control experiments, as nuclease-mediated gene knockout and transcriptional activation can be achieved in the same cell population with one Cas9 protein (Fig. 2). By activating *LPAR5*, a gene previously described to promote drug resistance when upregulated¹¹, and knocking out known tumor suppressors¹⁶, we showed that orthogonal gene modifications can be performed in a biologically meaningful context (Fig. 2d–e). Streamlining orthogonal gene control is particularly important for *in vivo* experiments, as concurrently delivering two distinct Cas9 complexes to the same cell is challenging. Given the fact that mice have been engineered to constitutively and inducibly express Cas9 (refs. 17,18), we anticipate that dRNAs can be used for *in vivo* orthogonal experiments, as the MPH activator complex combined with a dRNA and sgRNA fits within the 4.7-kb packaging limit of AAV vectors.

Finally, our data align with previously reported work¹⁹ that show sgRNAs with 17 or more nucleotides reliably and efficiently cut target DNA. It is notable that by reducing the guide length down to 15 nt, we abrogate indel formation, but do not abrogate ‘functional binding’. Further evidence supporting this effect was recently published; active Cas9 fused to the transcriptional activator VPR and targeted to DNA by 14-bp guide RNAs resulted in transcriptional activation²⁰. This may point to a more fundamental mechanism, whereby the interactions

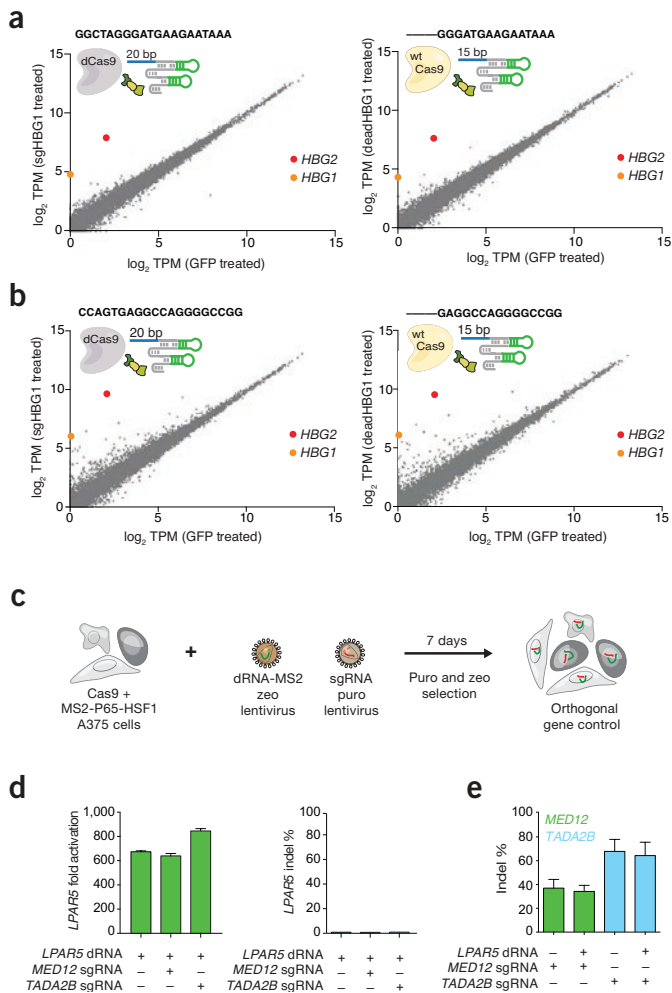


Figure 2 dRNAs can specifically upregulate gene expression, have a specificity profile similar to 20bp sgRNA activators, and can be used for orthogonal gene control in human cells. Sequences targeted to the *HBG1/2* promoter were tested for off-target transcriptional activation using RNA-seq. 20-nt sgRNAs with MS2 binding loops were co-transfected with dCas9 and the MPH activation complex. These were compared to dRNAs co-delivered with active Cas9 and the MPH activation complex. Both systems showed similar off-target profiles. (a) Zero significantly upregulated genes apart from *HBG1/2* were observed for both the 20-nt dCas9- and dRNA:Cas9-treated cells. (b) A second guide showed 55 significantly upregulated genes apart from *HBG1/2* for the 20-nt dCas9-treated cells, whereas 31 significantly upregulated genes were measured for dRNA-treated cells. (In all cases, $N = 3$ replicates/group.) (c) Orthogonal gene control in melanoma A375 cells expressing an active Cas9 and the MS2-P65-HSF1 fusion protein. Cells were transduced with lentivirus containing a dRNA targeting one gene and an sgRNA targeting a second gene. Selected cells were subsequently treated with BRAF-inhibitor PLX4720 and their survival was quantified. (d) Activation and indel % were measured for individually and orthogonally controlled genes. Left: *LPAR5* transcriptional upregulation mediated by dRNA was robust in the presence and absence of sgRNAs targeting *MED12* or *TADA2B*. Right: *LPAR5* indel formation was $\leq 0.85\%$ at the dRNA target site. (e) Robust indel formation was detected at DNA sites targeted by *MED12* and *TADA2B* sgRNAs alone and when delivered together with a dRNA targeting *LPAR5*.

Simons, Paul G. Allen and New York Stem Cell Foundations, and David R. Cheng and Bob Metcalfe. The authors plan to make the reagents widely available to the academic community through Addgene and to provide software tools via the Zhang laboratory website (<http://www.genome-engineering.org/>).

AUTHOR CONTRIBUTIONS

J.E.D., O.O.A., F.Z. and S.K. conceived this study and designed the experiments. J.E.D., O.O.A., S.K., J.J. and J.S.G. performed experiments. J.E.D., O.O.A., F.Z. and S.K. wrote the manuscript with input from all authors.

COMPETING FINANCIAL INTERESTS

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

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

between the protein, sgRNA and DNA dictate whether the Cas9 protein cuts the target DNA.

METHODS

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

Accession codes. SRA: PRJNA296872.

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

ACKNOWLEDGMENTS

The authors thank K. Zheng and I. Slaymaker for their support and input. J.E.D. is supported by a Life Science Research Foundation post-doctoral fellowship of the Cystic Fibrosis Foundation. O.O.A. is supported by a Friends of the McGovern Institute Fellowship. J.S.G. is supported by a Department of Energy (DOE) Computational Science Graduate Fellowship. F.Z. is supported by the National Institute of Mental Health (NIMH) (1DP1-MH100706), the Poitras, Vallee,

- Zalatan, J.G. *et al. Cell* **160**, 339–350 (2015).
- Esvelt, K.M. *et al. Nat. Methods* **10**, 1116–1121 (2013).
- Jinek, M. *et al. Science* **337**, 816–821 (2012).
- Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. *Proc. Natl. Acad. Sci. USA* **109**, E2579–E2586 (2012).
- Perez-Pinera, P. *et al. Nat. Methods* **10**, 973–976 (2013).
- Mali, P. *et al. Nat. Biotechnol.* **31**, 833–838 (2013).
- Maeder, M.L. *et al. Nat. Methods* **10**, 977–979 (2013).
- Konermann, S. *et al. Nature* **500**, 472–476 (2013).
- Gilbert, L.A. *et al. Cell* **154**, 442–451 (2013).
- Hilton, I.B. *et al. Nat. Biotechnol.* **33**, 510–517 (2015).
- Konermann, S. *et al. Nature* **517**, 583–588 (2015).
- Nishimasu, H. *et al. Cell* **156**, 935–949 (2014).
- Hsu, P.D. *et al. Nat. Biotechnol.* **31**, 827–832 (2013).
- Briner, A.E. *et al. Mol. Cell* **56**, 333–339 (2014).
- Wu, X. *et al. Nat. Biotechnol.* **32**, 670–676 (2014).
- Shalem, O. *et al. Science* **343**, 84–87 (2014).
- Platt, R.J. *et al. Cell* **159**, 440–455 (2014).
- Dow, L.E. *et al. Nat. Biotechnol.* **33**, 390–394 (2015).
- Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. *Nat. Biotechnol.* **32**, 279–284 (2014).
- Kiani, S. *et al. Nat. Methods*. doi:10.1038/nmeth.3580 (7 September 2015).

ONLINE METHODS

Generation of sgRNA:Cas9 and dRNA:Cas9 constructs. All DNA constructs used were delivered to cells as PCR products or plasmids. sgRNAs and dRNAs used in **Figure 1** and **Supplementary Figs. 1** and **2** were designed to be expressed from a U6 promoter and delivered as a dsDNA PCR product. sgRNAs/dRNA used in **Figure 2** and **Supplementary Figures 3** and **4** were also expressed from U6 promoters but in a plasmid backbone (**Supplementary Notes**, section Guide sequences). The active Cas9, dCas9 and MPH constructs were cloned into plasmids and used an EF1- α promoter. All constructs are available from Addgene. (accession numbers: 61424, 61423, 48137 and 42230 for transient transfection experiments and 61427, 61426, 52963 and 52962 for lentiviral delivery).

Transient transfection. HEK293FT cells (Life Technologies) were maintained in high-glucose DMEM with GlutaMax and sodium pyruvate (Life Technologies) supplemented with 10% heat-inactivated characterized HyClone FBS (Thermo Scientific) and 1% penicillin/streptomycin (Life Technologies). For gene activation experiments, 20,000 HEK293FT cells/well were plated in 100 μ l media in poly-D-lysine-coated, 96-well plates (BD BioSciences). 24 h after plating, cells were transfected with a 1:1:1 mass ratio of:

- sgRNA/dRNA PCR product/plasmid with gene-specific targeting sequence or an EGFP control plasmid;
- MS2-effector plasmid or pUC19;
- dCas9 plasmid, activeCas9 plasmid or pUC19.

The total plasmid transfected per well was 300 ng and was transfected with Lipofectamine 2000 (Life Technologies) according to the manufacturer's instructions. Culture medium was changed 4 h after transfection.

Lentivirus production. HEK293T cells were cultured as described above. 1 day before transfection, cells were seeded at ~40% confluency in a T25 flask. Cells were transfected when they reached 80–90% confluency. For each flask, 3.4 μ g of plasmid containing the vector of interest, 1.7 μ g of pMD2.G, and 2.6 μ g of psPAX (Addgene) were transfected using 30 μ l of Lipofectamine 2000 and 33 μ l Plus Reagent (Life Technologies). Media was changed 4 hours after transfection. After 24 h, 2 mM (final conc.) caffeine was added to the media to boost viral titer. Virus supernatant was isolated 48 h later and filtered with a 0.45 μ m polyvinylidene fluoride filter (Millipore), and stored at –80 °C.

Lentiviral transduction. A375 cells (Sigma) were cultured in RPMI 1640 (Life Technologies) supplemented with 9% FBS (Seradigm) and 1% penicillin/streptomycin (Life Technologies). An A375 cell line was transduced with both an active Cas9 lentivirus and an MPH lentivirus and selected for 7 days (with blasticidin and hygromycin, respectively). For the addition of sgRNAs/dRNAs, cells were transduced by spinfection in 12-well plates. 3×10^6 cells in 2 ml of media supplemented with 8 μ g/ml polybrene (Sigma) were added to each well, supplemented with lentiviral supernatant at a multiplicity of infection (MOI) of 0.5 and centrifuged for 2 h at 1,000g. 24 h later, cells were detached with TrypLE (Life Technologies) and counted. Cells were replated at low density in a T75 flask before a selection agent was added (zeocin and/or puromycin, all Life Technologies). Media was replaced 2 days later, and cells were passaged every other day starting

4 days after replating. The duration of selection was 4 days for puromycin and 7 days for zeocin, lentiviral titers were determined by spinfecting cells with six different volumes of lentivirus ranging from 0 to 400 μ l and counting the number of surviving cells after a complete selection. Cell lines were not tested for mycoplasma contamination.

RT-qPCR. 48 h after transfection, cells were lysed and reverse transcription were performed using a Cells-to-Ct kit (Life Technologies). RNA expression was quantified by quantitative PCR (qPCR) using TaqMan qPCR probes (Life Technologies) and Fast Advanced Master Mix (Life Technologies). RNA expression was normalized to GapDH and compared to wells transfected with 300 ng of GFP plasmid. qPCR was carried out in 5 μ l multiplexed reactions and 384-well format using the LightCycler 480 Instrument II.

Indel analysis. 48 h after transient transfection with Lipofectamine 2000 and 7 days after lentiviral transduction, cells were lysed and genomic DNA was extracted from transfected cells using QuickExtract DNA Extraction Solution (Epicentre). Target regions for indel analysis were PCR amplified (which adds Illumina P5 adapters) and pooled into a library for next-generation sequencing. Libraries were quantified using a Qubit 2.0 Fluorometer and sequenced using an Illumina MiSeq instrument using a 300 cycle v2 kit. Indels were analyzed computationally, as previously described²¹. MiSeq reads were first filtered by requiring an average Phred quality (Q score) of at least 23. Reads were analyzed for potential indels by using the Python difflib package (based on the Ratcliff and Obershelp algorithm) on a 30-bp flanking region around the target site (80 bp total). Reads were counted as indels if an insertion or deletion operation was detected. Using the negative control samples, maximum-likelihood estimation was used to calculate the true indel rate at each target site, as previously described¹³. DNA indel data were deposited into SRA with the accession code PRJNA296872.

RNA-seq. Samples were prepped for RNA sequencing using a TruSeq Stranded mRNA Sample Prep Kit (Illumina) and deep-sequenced on Illumina sequencing machines. RSEM v1.27 (ref. 22) was run on the sequencing reads to estimate expression levels using the following parameters:–no-bam-output–estimate-rspd–bowtie–chunkmbs 512–paired. RSEM's gene level expression estimates (transcript per million (TPM)) were transformed to log-space by taking $\log_2(\text{TPM}+1)$. All genes detected were used to construct scatter plots comparing each condition to the control GFP condition, using the average across biological replicates ($\log_2(\text{mean}(\text{TPM})+1)$ value per gene).

To find differentially expressed genes, we performed Student's *t*-test against cells treated with 300 ng of EGFP plasmid. The *t*-test was run on all genes that had expression levels above $\log_2(\text{TPM}+1) > 2.5$ in at least two samples. This threshold was chosen as the minimal threshold for which the number of detected genes across all libraries was constant. Only genes that were significant (*P*-value passes 0.05 FDR correction) and had at least 1.5-fold change were reported. RNA-seq data were deposited into SRA with the accession code PRJNA296872.

21. Ran, F.A. *et al. Cell* **154**, 1380–1389 (2013).

22. Li, B. & Dewey, C.N. *BMC Bioinformatics* **12**, 323 (2011).