RNA-guided gene activation by CRISPR-Cas9–based transcription factors

Pablo Perez-Pinera[1], D Dewran Kocak[1], Christopher M Vockley[2,3], Andrew F Adler[1], Ami M Kabadi[1], Lauren R Polstein[1], Pratiksha I Thakore[1], Katherine A Glass[1,4], David G Ousterout[1], Kam W Leong[1,5], Farshid Guilak[1,4], Gregory E Crawford[2,6], Timothy E Reddy[2,7] & Charles A Gersbach[1,2,4]

Technologies for engineering synthetic transcription factors have enabled many advances in medical and scientific research. In contrast to existing methods based on engineering of DNA-binding proteins, we created a Cas9-based transactivator that is targeted to DNA sequences by guide RNA molecules. Coexpression of this transactivator and combinations of guide RNAs in human cells induced specific expression of endogenous target genes, demonstrating a simple and versatile approach for RNA-guided gene activation.

Synthetic transcription factors have been engineered to control gene expression for many medical and scientific applications in mammalian systems, including stimulating tissue regeneration[1], compensating for genetic defects[2], activating silenced tumor suppressors[3], controlling stem-cell differentiation[4,5], performing genetic screens[6,7] and creating synthetic gene circuits[8,9]. These transcription factors can target promoters or enhancers of endogenous genes[10] or be purposefully designed to regulate transgenes[11]. To date, the most common strategies for engineering new transcription factors targeted to user-defined sequences have been based on the programmable DNA-binding domains of zinc-finger proteins[10,12] and transcription activator–like effectors[13,14]. Although these methods have been widely successful for many applications, the protocols necessary to engineer proteins with new DNA-binding specificity can be laborious and require specialized expertise.

Recently, an engineered form of the clustered, regularly interspaced, short palindromic repeat (CRISPR) system[15] of *Streptococcus pyogenes* has been shown to function in human cells for genome engineering[16–19]. In this type II CRISPR system, the Cas9 protein is directed to genomic target sites by short RNAs, where it functions as an endonuclease. In the naturally occurring system, Cas9 is directed to its DNA target site by two noncoding CRISPR RNAs (crRNAs), including a trans-activating crRNA (tracrRNA) and a precursor crRNA (pre-crRNA)[15]. In the synthetically reconstituted system, these two short RNAs can be fused into a single chimeric guide RNA (gRNA)[15]. A Cas9 mutant with undetectable endonuclease activity (dCas9)[15] has been recently targeted to genes in bacteria, yeast and human cells by gRNAs to silence gene expression[20,21]. Here we demonstrate a strategy to activate the expression of endogenous human genes by targeting a fusion protein of dCas9 and a transactivation domain to human promoters via combinations of gRNAs.

To create a CRISPR-Cas9–based transcriptional activation system, we mutated the endonuclease catalytic residues of Cas9 (D10A and H840A) to create dCas9 (refs. 15,20,21) and genetically fused it with a C-terminal VP64 acidic transactivation domain (**Fig. 1a**,**b**)[10]. We observed robust expression of dCas9-VP64 from the transfected plasmid in human embryonic kidney 293T (HEK293T) cells by western blot (**Supplementary Fig. 1**). We and others have recently shown that combinations of synthetic transcription factors targeted to endogenous human promoters result in synergistic and robust activation of gene expression[22,23]. Therefore we identified four gRNA target sites of 20 bp followed by the protospacer-adjacent motif sequence[15–20] NGG, where N is any nucleotide, in the promoter of the *IL1RN* gene (**Supplementary Fig. 2** and **Supplementary Table 1**). To compare crRNA-based and gRNA-based targeting strategies, we introduced the four target site sequences into crRNA and gRNA expression plasmids[17] and cotransfected them with the dCas9-VP64 expression plasmid into HEK293T cells. Although we observed substantial induction of *IL1RN* expression by quantitative reverse transcription–PCR (qRT-PCR) in samples treated with the combination of crRNAs, we observed much higher expression with the combination of gRNAs (**Fig. 1c**). There were no changes to *IL1RN* gene expression in cells treated with gRNAs and an expression plasmid for dCas9 without VP64, demonstrating the critical role of the activation domain in modulating gene expression (**Fig. 1c**). Using the Surveyor assay to detect DNA-repair events, we observed no nuclease activity at the target sites in samples treated with dCas9-VP64 (**Supplementary Fig. 3**). By transfecting each of the four gRNAs individually or in combination, we observed

[1]Department of Biomedical Engineering, Duke University, Durham, North Carolina, USA. [2]Institute for Genome Sciences and Policy, Duke University, Durham, North Carolina, USA. [3]Department of Cell Biology, Duke University Medical Center, Durham, North Carolina, USA. [4]Department of Orthopaedic Surgery, Duke University Medical Center, Durham, North Carolina, USA. [5]King Abdulaziz University, Jeddah, Saudi Arabia. [6]Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, North Carolina, USA. [7]Department of Biostatistics and Bioinformatics, Duke University Medical Center, Durham, North Carolina, USA. Correspondence should be addressed to C.A.G. (charles.gersbach@duke.edu).

RECEIVED 17 MAY; ACCEPTED 18 JULY; PUBLISHED ONLINE 25 JULY 2013; DOI:10.1038/NMETH.2600

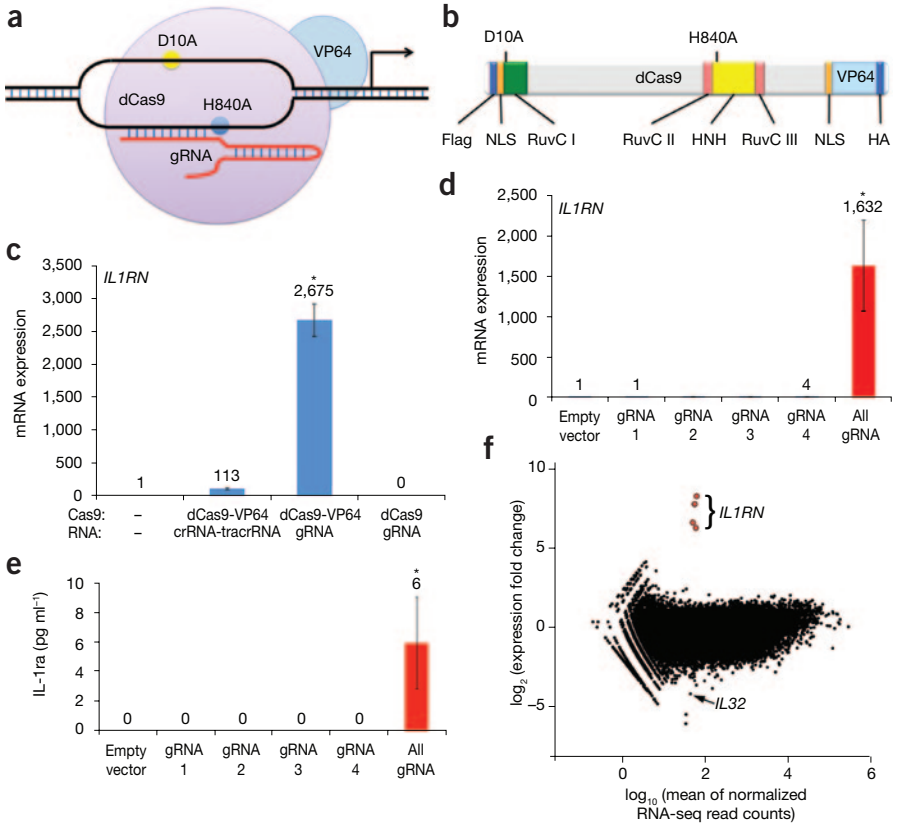

Figure 1 | RNA-guided activation of the human *IL1RN* gene by dCas9-VP64. (**a**) Schematic of an RNA-guided transcriptional activator, created by fusing dCas9 (D10A,H840A) to the VP64 transactivation domain. dCas9-VP64 recognizes genomic target sites through the hybridization of a gRNA to a 20-base-pair target sequence. (**b**) Schematic of dCas9-VP64 construct, which includes Flag and hemagglutinin (HA) epitope tags, two nuclear localization signals (NLSs) and inactivated RuvC- and HNH-like endonuclease domains. (**c**,**d**) *IL1RN* expression as assessed by qRT-PCR in HEK293T cells transfected with the indicated dCas9 or dCas9-VP64 expression plasmids, gRNA or crRNA-tracrRNA expression plasmids and empty vector controls (–). In **d**, the four gRNA expression plasmids were cotransfected with dCas9-VP64 expression plasmid individually or in combination. Data are expressed as fold increase relative to the control sample. (**e**) Immunoassay of secretion of IL-1ra into the medium in samples from HEK293T cells transfected with the indicated gRNAs in combination with dCas9-VP64 vector. Data in **c**–**e** are shown as the mean ± s.e.m. ($n = 3$ independent experiments) and mean values are included above each bar. *P* values were determined by Tukey's test (*, versus empty vector, $P = 0.0001$ (**c**), $P = 0.006$ (**d**) and $P = 0.05$ (**e**)). (**f**) Fold changes in genome-wide gene expression as a function of expression levels assessed by RNA-seq on samples treated with empty expression vector ($n = 2$ independent experiments) or cotransfected with the expression plasmids for dCas9-VP64 and the four gRNAs targeting *IL1RN* ($n = 2$). The only significant changes in gene expression between these treatments were an increase in the four *IL1RN* isoforms (false discovery rate ≤ 3×10^{-4}; $P = 3.6 \times 10^{-8}$) and a decrease in *IL32* (false discovery rate = 0.03; $P = 4.3 \times 10^{-6}$).

that targeting multiple sites in the promoter with combinations of gRNAs was necessary for robust increases in gene expression (**Fig. 1d**), as we have seen with other classes of engineered transcription factors[22,23]. Similarly, we observed production of the IL-1 receptor antagonist (IL-1ra) protein, encoded by the *IL1RN* gene, in three of the six samples treated with the combination of gRNAs across three different experiments, whereas we never detected it in samples treated with single gRNAs or control plasmid (**Fig. 1e**). To examine the specificity of gene activation by dCas9-VP64, we assessed global gene expression of HEK293T cells treated with the combination of four gRNAs by RNA high-throughput sequencing (RNA-seq; **Fig. 1f**). Notably, the only genes with significantly increased expression relative to the control (false discovery rate ≤ 3×10^{-4}, $P = 3.6 \times 10^{-8}$) were the four isoforms expressed from the *IL1RN* locus (**Supplementary Fig. 2**), indicating a high level of specificity of gene activation.

To demonstrate the general applicability of this system, we designed four gRNAs targeted to each of the promoters of eight other genes, including *ASCL1*, *NANOG*, *HBG1* and *HBG2*, *MYOD1*, *VEGFA*, *TERT*, *IL1B* and *IL1R2* (**Supplementary Fig. 2** and **Supplementary Table 1**). Cotransfection of HEK293T cells with expression plasmids for dCas9-VP64 and the four gRNAs enhanced expression of each of these genes, as determined by qRT-PCR (**Fig. 2**). In some cases expression of a single gRNA was sufficient to induce gene expression, but in all cases cotransfection of the four gRNAs led to synergistic effects (**Fig. 2a–d**). Chromatin accessibility, as determined by sequencing of DNase I hypersensitive sites (DNase-seq), was not a predictor of successful gene activation (**Supplementary Fig. 2**), consistent with our previous observations[22]. We performed RNA-seq on cells transfected with vectors encoding dCas9-VP64 and the four gRNAs targeting *HBG1*, three of which also perfectly target *HBG2*. This revealed specific and reproducible increases in expression of both *HBG1* and *HBG2*, which could not be distinguished by RNA-seq (**Supplementary Fig. 4**). However, the increase was not significant after multiple-hypothesis correction (unadjusted $P = 0.054$), likely owing to low total expression. We also confirmed increases in protein expression of Ascl1 and γ-globin after treatment with dCas9-VP64 and the four gRNAs (**Supplementary Fig. 5**).

As evidence that the dCas9-VP64–gRNA system can activate gene expression in other cell types, we transfected expression plasmids for dCas9-VP64 and the four gRNAs targeting *Ascl1* into mouse embryonic fibroblasts (MEFs; **Supplementary Fig. 6**). Because the gRNA target sites are conserved in the human *ASCL1* and mouse *Ascl1* promoters, we also observed activation of *Ascl1* expression 4 d after transfection in MEFs treated with plasmids encoding dCas9-VP64 and the four gRNAs (**Supplementary Fig. 6**).

RNA-guided transcriptional activation is fundamentally different from previously described methods based on engineering sequence-specific DNA-binding proteins and may provide new opportunities for targeted gene regulation. Because the generation of new gRNA expression plasmids simply involves synthesizing two short custom oligonucleotides and one cloning step,

it is possible to generate many new gene activators quickly and economically. The gRNAs can also be transfected directly into cells after *in vitro* transcription[18]. Here we show multiple gRNAs targeted to single promoters, but simultaneous targeting of multiple promoters could also be envisioned[16]. It is likely that future efforts for combining multiple gRNA expression cassettes into single delivery vectors will facilitate this type of approach. Recognition of genomic target sites with RNAs, rather than proteins, may also circumvent limitations of targeting epigenetically modified sites, such as methylated DNA. Notably, the activation of *IL1RN* by dCas9-VP64 in this study was weaker than what we have reported for transcription factors based on transcription activator–like effectors targeting the same promoter[22], and we have observed a similar trend for other targets, but additional studies are necessary to understand the differences in these nascent technologies.

Using RNA-seq, we showed that targeted gene activation was exquisitely specific with no detectable off-target gene activation (**Fig. 1f** and **Supplementary Fig. 4**). We chose *IL1RN* and *HBG1-HBG2* for this specificity analysis because we expected the gene products, IL-1ra and γ-globin, would not generate secondary effects on gene expression in HEK293T cells. We hypothesize that exploiting the synergistic activity of multiple weak transcriptional activators as we have done here, in contrast to using a single strong activator, may increase the specificity of gene regulation because it is unlikely that multiple adjacent off-target sites would exist at another locus. Although these results show that gene activation by dCas9-VP64 and combinations of gRNAs can be very specific, this is not necessarily a surrogate for DNA-binding, and additional studies are necessary to define the role of the gRNA target sequence in determining genome-wide Cas9 activity in mammalian cells. The RNA-seq results for both *IL1RN* and *HBG1-HBG2* showed moderate downregulation of the *IL32* gene (false discovery rate < 0.03) in the samples treated with dCas9-VP64 and gRNA expression plasmids compared to control samples treated with only an empty expression plasmid (**Fig. 1f** and **Supplementary Fig. 4**). Because both the *IL1RN*-targeted and *HBG1-HBG2*–targeted samples were similarly affected, it is unlikely that this is the result of off-target dCas9-VP64 activity related to the identity of the target sequences. We confirmed these results by qRT-PCR for *IL32*, which showed that this downregulation is a general response to dCas9-VP64, even in the absence of gRNAs (**Supplementary Fig. 7**).

The potential applications of RNA-guided transcription factors are diverse. Here we provided examples of activation of genes related to cell and gene therapy, genetic reprogramming and regenerative medicine, including the regulators of cell lineage specification *ASCL1*, *NANOG* and *MYOD1* (**Fig. 2**). We hypothesize that with continued optimization, the activation of endogenous genes encoding key regulators of cell fate, rather than forced overexpression of these factors, could lead to improved methods for genetic reprogramming and transdifferentiation. Finally, dCas9 fusions to other domains, including repressive and epigenetic-modifying domains[24], could provide a greater diversity of RNA-guided transcriptional regulators for mammalian cell engineering.

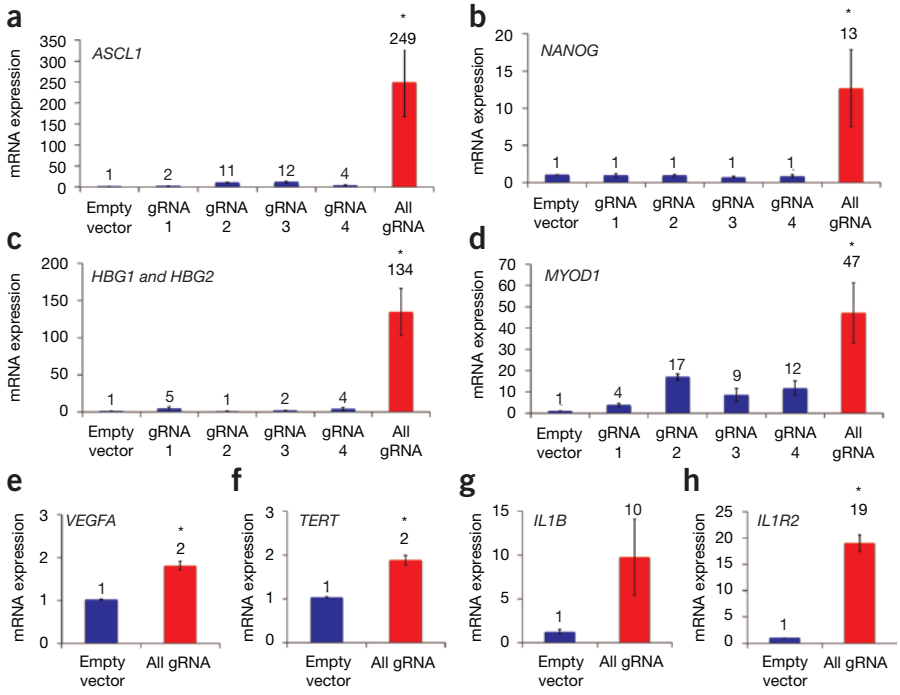

Figure 2 | RNA-guided activation of human genes. (**a**–**h**) Target gene expression, measured by qRT-PCR and normalized to *GAPDH* mRNA levels, in HEK293T cells transfected with the expression plasmids for dCas9-VP64 and four gRNAs individually or in combination, targeting the indicated genes. Data are shown as the mean ± s.e.m. ($n = 3$ independent experiments). P values were determined by Tukey's test (* versus empty vector, $P = 0.002$ (**a**), $P = 0.02$ (**b**), $P = 0.0001$ (**c**), $P = 0.002$ (**d**), $P = 0.001$ (**e**), $P = 0.002$ (**f**) and $P = 0.0003$ (**h**)).

METHODS

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

Accession codes. Gene Expression Omnibus: GSE47114.

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

ACKNOWLEDGMENTS

This work was supported by a US National Institutes of Health (NIH) Director's New Innovator Award (DP2OD008586), National Science Foundation (NSF) Faculty Early Career Development (CAREER) Award (CBET-1151035), NIH R03AR061042, and an American Heart Association Scientist Development Grant (10SDG3060033) to C.A.G., and grants from the NIH to G.E.C. (U54HG004563), K.W.L. (EB015300 and HL109442), and F.G. (R01AR48852). K.A.G. and P.I.T. were supported by NSF Graduate Research Fellowships. L.R.P. was supported by an NIH Biotechnology Training Grant to the Duke Center for Biomolecular and Tissue Engineering (T32GM008555). D.G.O. was supported by a predoctoral fellowship from the American Heart Association. C. Grigsby (Duke University) provided the pABOL polymer used in MEF transfections.

AUTHOR CONTRIBUTIONS

P.P., C.M.V., A.F.A., D.G.O., G.E.C., T.E.R. and C.A.G. designed experiments. P.P., D.D.K., C.M.V., A.F.A., A.M.K., L.R.P., P.I.T., K.A.G. and D.G.O. performed the

experiments. P.P., D.D.K., C.M.V., A.F.A., A.M.K., L.R.P., P.I.T., K.A.G., D.G.O., K.W.L., F.G., G.E.C., T.E.R. and C.A.G. analyzed the data. P.P. and C.A.G. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Rebar, E.J. *et al. Nat. Med.* **8**, 1427–1432 (2002).
2. Graslund, T., Li, X., Magnenat, L., Popkov, M. & Barbas, C.F. III. *J. Biol. Chem.* **280**, 3707–3714 (2005).
3. Beltran, A. *et al. Oncogene* **26**, 2791–2798 (2007).
4. Bartsevich, V.V., Miller, J.C., Case, C.C. & Pabo, C.O. *Stem Cells* **21**, 632–637 (2003).
5. Bultmann, S. *et al. Nucleic Acids Res.* **40**, 5368–5377 (2012).
6. Blancafort, P., Magnenat, L. & Barbas, C.F. III. *Nat. Biotechnol.* **21**, 269–274 (2003).
7. Park, K.S. *et al. Nat. Biotechnol.* **21**, 1208–1214 (2003).
8. Lohmueller, J.J., Armel, T.Z. & Silver, P.A. *Nucleic Acids Res.* **40**, 5180–5187 (2012).
9. Li, Y., Moore, R., Guinn, M. & Bleris, L. *Scientific Rep.* **2**, 897 (2012).
10. Beerli, R.R., Dreier, B. & Barbas, C.F. III. *Proc. Natl. Acad. Sci. USA* **97**, 1495–1500 (2000).
11. Garg, A., Lohmueller, J.J., Silver, P.A. & Armel, T.Z. *Nucleic Acids Res.* **40**, 7584–7595 (2012).
12. Beerli, R.R. & Barbas, C.F. III. *Nat. Biotechnol.* **20**, 135–141 (2002).
13. Zhang, F. *et al. Nat. Biotechnol.* **29**, 149–153 (2011).
14. Miller, J.C. *et al. Nat. Biotechnol.* **29**, 143–148 (2011).
15. Jinek, M. *et al. Science* **337**, 816–821 (2012).
16. Cong, L. *et al. Science* **339**, 819–823 (2013).
17. Mali, P. *et al. Science* **339**, 823–826 (2013).
18. Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. *Nat. Biotechnol.* **31**, 230–232 (2013).
19. Jinek, M. *et al. eLife* **2**, e00471 (2013).
20. Qi, L.S. *et al. Cell* **152**, 1173–1183 (2013).
21. Gilbert, L.A. *et al. Cell* doi:10.1016/j.cell.2013.06.044 (9 July 2013).
22. Perez-Pinera, P. *et al. Nat. Methods* **10**, 239–242 (2013).
23. Maeder, M.L. *et al. Nat. Methods* **10**, 243–245 (2013).
24. de Groote, M.L., Verschure, P.J. & Rots, M.G. *Nucleic Acids Res.* **40**, 10596–10613 (2012).

ONLINE METHODS

Cell culture and transfection. HEK293T cells were obtained from the American Tissue Collection Center (ATCC) through the Duke University Cancer Center Facilities and were maintained in DMEM supplemented with 10% FBS and 1% penicillin-streptomycin at 37 °C with 5% CO_2. HEK293T cells were transfected with Lipofectamine 2000 (Invitrogen) according to manufacturer's instructions. Transfection efficiencies were routinely higher than 80%, as determined by fluorescence microscopy after delivery of a control eGFP expression plasmid. dCas9-VP64 expression plasmid was transfected at a mass ratio of 3:1 to either the individual gRNA expression plasmids or the identical amount of gRNA expression plasmid consisting of a mixture of equal amounts of the four gRNAs.

Primary MEFs (PMEF-HL, Millipore) were seeded (75,000 per well) in 24-well TCPS plates (BD) and maintained at 37 °C and 5% CO_2 in complete MEF medium consisting of high-glucose DMEM supplemented with 10% Premium Select FBS (Atlanta Biologicals), 25 μg ml^{-1} gentamicin (Invitrogen), 1× GlutaMAX, nonessential amino acids, sodium pyruvate and β-mercaptoethanol (Invitrogen). MEF transfections were performed with a single 1 μg cm^{-2} dose of total plasmid DNA, delivered as cationic nanocomplexes after electrostatic condensation with poly(CBA-ABOL) in serum-free and antibiotic-free OptiMEM, as described previously[25]. OptiMEM was replaced with complete MEF medium 4 h after transfection. MEFs were processed for qRT-PCR 4 d after transfection. A GFP reporter vector (pmax-GFP, 3,486 bp; Amaxa) was used to optimize transfection conditions. dCas9-VP64 expression plasmid was transfected at a mass ratio of 3:1 or 1:1 to an equal mixture of four gRNA expression plasmids.

Plasmids. The plasmids encoding wild-type Cas9 and Cas9(H840A) were obtained from Addgene (plasmid 39312 and plasmid 39316)[15]. Sequence encoding Cas9(H840A) was cloned into the vector pcDNA3.1 in frame with a Flag epitope tag and a NLS at the N terminus with a primer pair that introduced the mutation resulting in the D10A substitution. Sequences encoding the VP64 domain[10,26], an NLS and an HA epitope tag were cloned in frame with the open reading frame encoding Cas9 at the 3′ end (**Fig. 1a** and **Supplementary Fig. 8**). The tracrRNA and crRNA expression cassettes were ordered as gBlocks (IDT) based on published sequences[16] and cloned into a pZDonor plasmid (Sigma) with KpnI and SacII sites. A previously described chimeric guide RNA expression cassette[17] was also ordered as gBlocks with modifications to include a BbsI restriction site[16] to facilitate rapid cloning of new guide RNA spacer sequences (**Supplementary Fig. 9**). The oligonucleotides containing the target sequences were obtained from IDT, hybridized, phosphorylated and cloned in the appropriate plasmids using BbsI sites. The target sequences and positions relative to the transcriptional start site are provided in **Supplementary Table 1**. dCas9, dCas9-VP64 and gRNA expression plasmids have been deposited in Addgene (plasmids 47106, 47107 and 47108).

Western blot. Cells were lysed in 50 mM Tris-Cl (pH 7.4), 150 mM NaCl, 0.5% Triton X-100 and 0.1% SDS. Lysates were mixed with loading buffer and boiled for 5 min; equal volumes of protein were run in NuPAGE Novex 4–12% or 10% Bis-Tris Gel polyacrylamide gels and transferred to nitrocellulose membranes. Nonspecific antibody binding was blocked with TBS-T (50 mM Tris, 150 mM NaCl and 0.1% Tween-20) with 5% nonfat milk for 30 min. The membranes were incubated with primary antibodies (HRP-conjugated anti-Flag (Cell Signaling, 2044) in 5% BSA in TBS-T, diluted 1:1,000, overnight; anti-GAPDH (Cell Signaling, clone 14C10) in 5% milk in TBS-T, diluted 1:5,000, for 30 min; anti-ASCL1 (Santa Cruz, clone sc-48449) in 5% BSA, diluted 1:500; or anti–γ-globin (Santa Cruz, clone 51-7) in 5% milk diluted 1:500), and the membranes were washed with TBS-T for 30 min. Membranes labeled with primary antibodies were incubated with anti-rabbit HRP-conjugated antibody (Sigma-Aldrich) diluted 1:5,000 for 30 min, anti-goat (1:3,000) or anti-mouse (1:5,000) and washed with TBS-T for 30 min. Membranes were visualized using the Immun-Star WesternC Chemiluminescence Kit (Bio-Rad) and images were captured using a ChemiDoc XRS+ System and processed using ImageLab software (Bio-Rad).

Enzyme-linked immunosorbent assay. Serum-free culture media (OPTI-MEM) was collected and frozen at –80 °C. Human IL-1ra secretion into culture medium was quantified via enzyme-linked immunosorbent assay (ELISA), according to the manufacturer's protocols (R&D Systems, DY280). The standard curve was prepared by diluting recombinant human IL-1ra in OPTI-MEM, and the IL-1ra in culture medium was measured undiluted. The samples were concentrated ~8-fold via centrifugation through 3 kDa MWCO filters for 20 min (Amicon Ultra, UFC500396). Reported values were corrected by the concentration factor for each sample.

Optical density was measured at 450 nm, with a wavelength correction at 540 nm. Each standard and sample was assayed in duplicate. The duplicate readings were averaged and normalized by subtracting the average zero standard optical density. A standard curve was generated by log-transforming the data and performing a linear regression of the IL-1ra concentration versus the optical density. Reported values are the mean and s.e.m. from three independent experiments ($n = 3$) that were performed on different days with technical duplicates that were averaged for each experiment.

Quantitative reverse transcription–PCR. Total RNA was isolated using the RNeasy Plus RNA isolation kit (Qiagen). cDNA synthesis was performed using the SuperScript VILO cDNA Synthesis Kit (Invitrogen). Real-time PCR using PerfeCTa SYBR Green FastMix was performed with the CFX96 Real-Time PCR Detection System (Bio-Rad) with oligonucleotide primers reported in **Supplementary Table 2** that were designed using Primer3Plus software and purchased from IDT. Primer specificity was confirmed by agarose gel electrophoresis and melting curve analysis. Reaction efficiencies over the appropriate dynamic range were calculated to ensure linearity of the standard curve (**Supplementary Fig. 10**). The results are expressed as fold-increase mRNA expression of the gene of interest normalized to *GAPDH* expression by the ΔΔCt method. Reported values are the mean and s.e.m. from three independent experiments performed on different days ($n = 3$) with technical duplicates that were averaged for each experiment.

RNA-seq. RNA-seq libraries were constructed as previously described[27]. Briefly, first-strand cDNA was synthesized from oligo(dT) Dynabead (Invitrogen) captured mRNA using SuperScript VILO cDNA Synthesis Kit (Invitrogen). Second-strand cDNA was

synthesized using DNA polymerase I (New England Biolabs). cDNA was purified using Agencourt AMPure XP beads (Beckman Coulter) and Nextera transposase (Illumina; 5 min at 55 °C) was used to simultaneously fragment and insert sequencing primers into the double-stranded cDNA. Transposition reactions were halted using QG buffer (Qiagen) and fragmented cDNA was purified on AMPure XP beads. Indexed sequencing libraries were generated by six cycles of PCR.

Libraries were sequenced using 50-bp single-end reads on two lanes of an Illumina HiSeq 2000 instrument, generating between 29 million and 74 million reads per library. Reads were aligned to human RefSeq transcripts using Bowtie[28]. The significance of differential expression, including correction for multiple hypothesis testing, was calculated using DESeq[29].

Statistics. Statistical analysis was performed by Tukey's test with alpha equal to 0.05 in JMP 10 Pro.

25. Adler, A.F. *et al. Mol.Ther. Nucleic Acids* **1**, e32 (2012).
26. Beerli, R.R., Segal, D.J., Dreier, B. & Barbas, C.F. III. *Proc. Natl. Acad. Sci. USA* **95**, 14628–14633 (1998).
27. Gertz, J. *et al. Genome Res.* **22**, 134–141 (2012).
28. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. *Genome Biol.* **10**, R25 (2009).
29. Anders, S. & Huber, W. *Genome Biol.* **11**, R106 (2010).

doi:10.1038/nmeth.2666