© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

Inducible and multiplex gene regulation using CRISPR–Cpf1-based transcription factors

Y Esther Tak[1,2], Benjamin P Kleinstiver[1,2], James K Nuñez[3], Jonathan Y Hsu[1,4], Joy E Horng[1], Jingyi Gong[1], Jonathan S Weissman[3,5] & J Keith Joung[1,2]

Targeted and inducible regulation of mammalian gene expression is a broadly important capability. We engineered drug-inducible catalytically inactive Cpf1 nuclease fused to transcriptional activation domains to tune the expression of endogenous genes in human cells. Leveraging the multiplex capability of the Cpf1 platform, we demonstrate both synergistic and combinatorial gene expression in human cells. Our work should enable the development of multiplex gene perturbation library screens for understanding complex cellular phenotypes.

Sequence-specific RNA-guided CRISPR–Cas nucleases are simple to program[1,2]; the widely used *Streptococcus pyogenes* Cas9 (SpCas9) can be targeted to specific DNA sequences by an associated complementary guide RNA (gRNA), provided that a protospacer-adjacent motif (PAM) of the form NGG is also present. Catalytically inactive SpCas9 (dSpCas9) has been fused to transcriptional activation or repression domains to alter the expression of individual genes or to perform genome-wide library screens in mammalian cells[3]. Both small-molecule- and light-inducible dSpCas9-based fusions have been developed[4–6], which has enabled researchers to regulate the activity of this gene regulatory platform. Recently described CRISPR–Cpf1 nucleases offer additional capabilities beyond those of SpCas9—including shorter length CRISPR RNAs (crRNAs) for guiding Cpf1 to targets, the ability to target T-rich PAMs[7,8], and RNase processing of multiple crRNAs from a single transcript by RNase activity of Cpf1 (refs. 9,10). However, to our knowledge, 'dead' Cpf1 (dCpf1)-based gene regulators have thus far only been shown to repress gene expression in bacteria[11,12] and plants (*Arabidopsis*)[13].

Based on preliminary experiments suggesting its higher activity in a GFP reporter gene experiment (**Supplementary Note 1** and **Supplementary Fig. 1**), we fused dCpf1 from *Lachnospiraceae* bacterium (dLbCpf1) to the strong synthetic VPR activator (herpes simplex virus-derived VP16 activator, the human NF-KB p65 activation domain, and the Epstein–Barr-virus-derived R transactivator (Rta))[14]. We targeted these dLbCpf1–VPR fusions to the promoters of three different endogenous genes that are either epigenetically silenced according to ENCODE data (*HBB*) or expressed at low levels (*AR* and *NPY1R*) in HEK293 cells by designing three crRNAs for each promoter located at various distances within 1 kb upstream of the transcription start site (TSS). At least one crRNA for each of the three target genes achieved robust transcriptional activation, as assayed by quantitative reverse transcription PCR (RT-qPCR); whereas a dLbCpf1–p65 fusion alone showed little or no transcriptional activation of the target gene promoter (**Fig. 1a**), which confirmed previous observations with dSpCas9 (ref. 14). We also tested a larger series of 32 crRNAs positioned within 1 kb upstream or 500 bp downstream of the TSSs of two additional endogenous genes, *CD5* and *CD22*, which encode cell surface proteins. Most of the 32 crRNAs tested could significantly ($P < 0.05$) activate the target gene promoter when positioned between ≈600 bp upstream and ≈400 bp downstream of the TSSs (**Supplementary Fig. 2**), and this was consistent with results obtained using dSpCas9 activators[15]. The levels of activation observed with dLbCpf1-based activators are comparable to what might be observed in naturally occurring biological systems and are similar to what has been previously reported for analogous dCas9-based activators[16].

To construct drug-regulated versions of dCpf1–VPR, we used DmrA and DmrC domains, which only interact in the presence of a rapamycin analog known as the A/C heterodimerizer[4,17]. We created split dLbCpf1 activators consisting of two fusions: dLbCpf1 fused to a DmrA domain and a DmrC domain fused to VPR or p65 (**Fig. 1b**); this made activation dependent on the A/C drug. Although single crRNAs failed to reveal inducible activation of the *HBB*, *AR*, or *NPY1R* genes (**Fig. 1c**), dLbCpf1 fusions harboring two, three, or four tandem copies of the DmrA domain in the presence of DmrC–VPR and the A/C heterodimerizer drug increased gene activation at two of the three endogenous gene promoters (*HBB* and *NPY1R*; **Fig. 1c**). The degree of activation observed correlated with the number of DmrA domains, with maximum levels reaching approximately half of those observed with direct dLbCpf1–VPR fusions (**Fig. 1c**). We found that using a DmrC–p65 fusion with dLbCpf1–DmrA fusions led to drug-dependent transcriptional upregulation from all three target gene promoters (**Fig. 1c**)—an unexpected result given the lack of activation observed with dLbCpf1–p65 direct fusions and these same crRNAs (**Fig. 1a**).

Compared with Cas9, one major advantage of Cpf1 is its ability to encode two or more crRNAs in a multiplex single transcript (MST), and subsequent processing of individual crRNAs by the

[1]Molecular Pathology Unit, Center for Cancer Research, and Center for Computational and Integrative Biology, Massachusetts General Hospital, Charlestown, Massachusetts, USA. [2]Department of Pathology, Harvard Medical School, Boston, Massachusetts, USA. [3]Department of Cellular & Molecular Pharmacology University of California, San Francisco, San Francisco, California, USA. [4]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. [5]Howard Hughes Medical Institute, San Francisco, California, USA. Correspondence should be addressed to J.K.J. (jjoung@mgh.harvard.edu).

RECEIVED 14 JUNE; ACCEPTED 24 SEPTEMBER; PUBLISHED ONLINE 30 OCTOBER 2017; DOI:10.1038/NMETH.4483

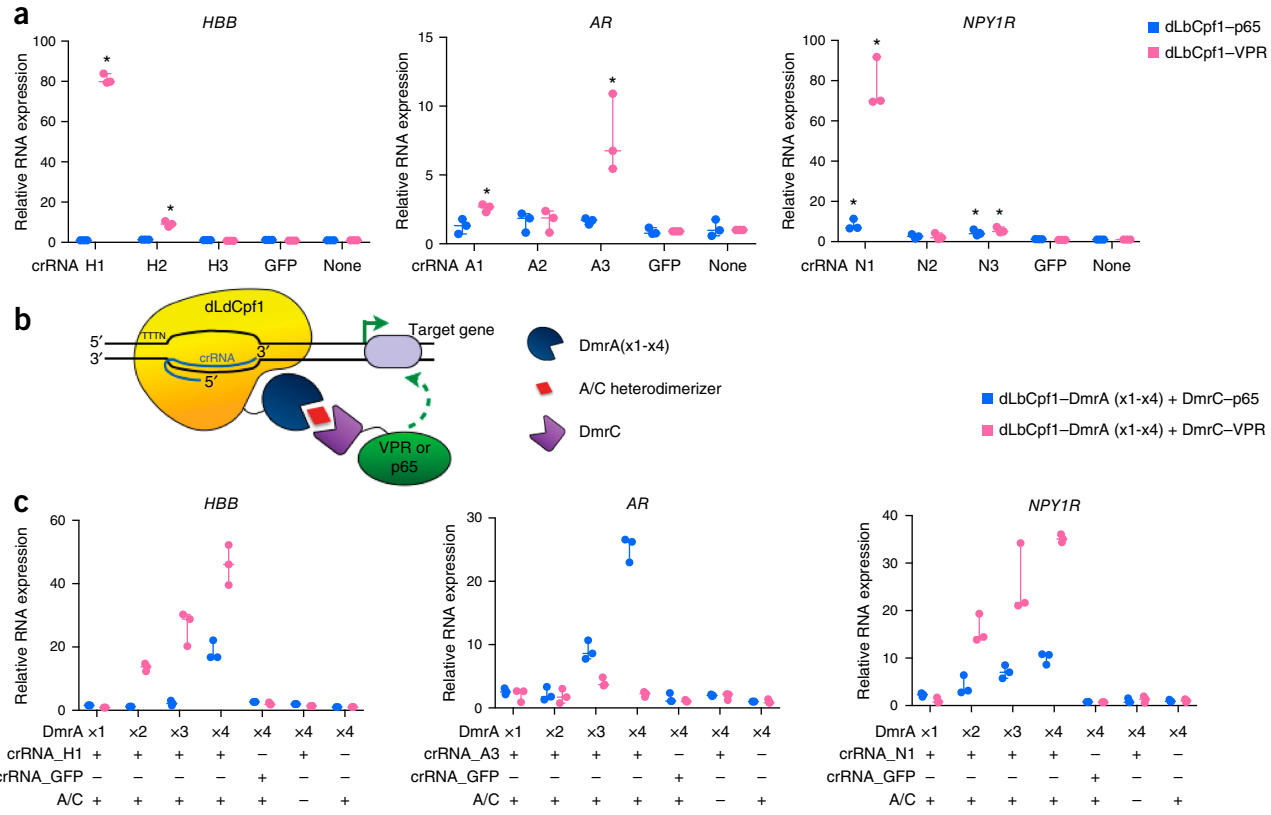


Figure 1 | Targeted human endogenous gene regulation using individual crRNAs with dLbCpf1-based activators. (**a**) Activities of dLbCpf1–p65 or dLbCpf1–VPR using single crRNAs at three endogenous human genes (*HBB*, *AR*, and *NPY1R*) in HEK293 cells. Relative mRNA expression is calculated by comparison to the control sample in which no crRNA is expressed. *, statistically significant difference compared to the control sample by Student *t*-test (two-tailed test assuming equal variance, $P < 0.05$). (**b**) Schematic illustrating drug-dependent bipartite dLbCpf1-based activator fusion proteins. (**c**) Activities of drug-dependent bipartite dLbCpf1-based activators using single crRNAs at three endogenous human genes (*HBB*, *AR*, and *NPY1R*) in HEK293 cells. The crRNA used for each promoter was the one that showed the highest activity in **a**. Data shown in **a** and **c** represent three biological independent replicates; ×1, one copy of DmrA; ×2, two copies of DmrA; ×3, three copies of DmrA; ×4, four copies of DmrA; +, presence; –, absence.

RNase-processing activity of Cpf1 (refs. 9 and 10) (**Fig. 2**) potentially simplifies the implementation of synergistic and multiplex gene activation. We tested the use of MST crRNAs for synergistic activation by encoding sets of the three crRNAs targeting the promoters of *HBB*, *AR*, or *NPY1R* genes on single transcripts (**Fig. 2a**). With dLbCpf1–VPR, we observed synergistic activation for *HBB* and *AR* but not for *NPY1R* (**Supplementary Fig. 3**). We also failed to observe synergistic activation of any of the three genes with dLbCpf1–p65 with either MST crRNAs or pooled single crRNAs (**Supplementary Fig. 3**). We also observed drug-dependent synergy with dLbCpf1–DmrA(×4) (dLbCpf1 fused to four copies of DmrA) and DmrC–VPR at the *HBB* and *AR* promoters but again not at the *NPY1R* promoter (even though synergy was observed with pooled single crRNAs introduced together) (**Fig. 2b**). A similar pattern of results for these three gene promoters was observed with dLbCpf1–DmrA(×4) and DmrC–p65 (**Fig. 2b**), and higher activation was observed on the *HBB* and *AR* genes relative to comparable experiments performed with DmrC–VPR (**Fig. 2b**). For the *HBB* gene, this higher activation was not seen with the single crRNA that worked the best among the three single crRNAs tested (**Fig. 1c**). Although the reason for this difference is not known, it is possible that it relates to the number and configuration of transcriptional activation domains recruited to the promoter.

We also tested the use of MST crRNAs for multiplex activation of different gene promoters in the same cells (**Fig. 2c**). We found that MST crRNAs could be used together with (i) dLbCpf1–VPR, (ii) dLbCpf1–DmrA(×4) and DmrC–VPR fusions, or (iii) dLbCpf1–DmrA(×4) and DmrC–p65 fusions to simultaneously activate the endogenous *HBB*, *AR*, and *NPY1R* gene promoters in HEK293 cells (**Fig. 2d**). The activities of MST crRNAs were generally comparable relative to single individually expressed crRNAs, and more than half of the gene/activator configurations we examined showed no statistically significant difference ($P < 0.05$) (**Fig. 2d**). For the cases where a difference was observed, the MST crRNAs generally showed only ≈2-fold lower activation compared to single individually expressed crRNAs, and the absolute level of fold activation was still quite high, ranging from ≈9-fold to ≈40-fold. Attempts to further increase the activities of MST crRNAs did not yield substantial improvements (**Supplementary Note 1** and **Supplementary Fig. 4**). To extend our findings to another human cell line, we also tested the direct VPR activator fusions and drug-regulated VPR and p65 activators in human U2OS cells, and we targeted the same genes (*HBB*, *AR*, and *NPY1R*) with MST crRNAs and single crRNAs (**Supplementary Fig. 5**). We observed that *HBB* and *NPY1R* were highly upregulated, and that *HBB* activation with DmrC–p65 was even greater than it was in HEK293s. We did not observe activation of the *AR* gene, most likely because *AR* is already highly expressed in U2OS cells (with a basal RT-qPCR threshold cycle (*Ct*) value of ≈28). Relative differences in the efficacies of crRNAs expressed singly or in MSTs were similar to what we observed in HEK293s.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

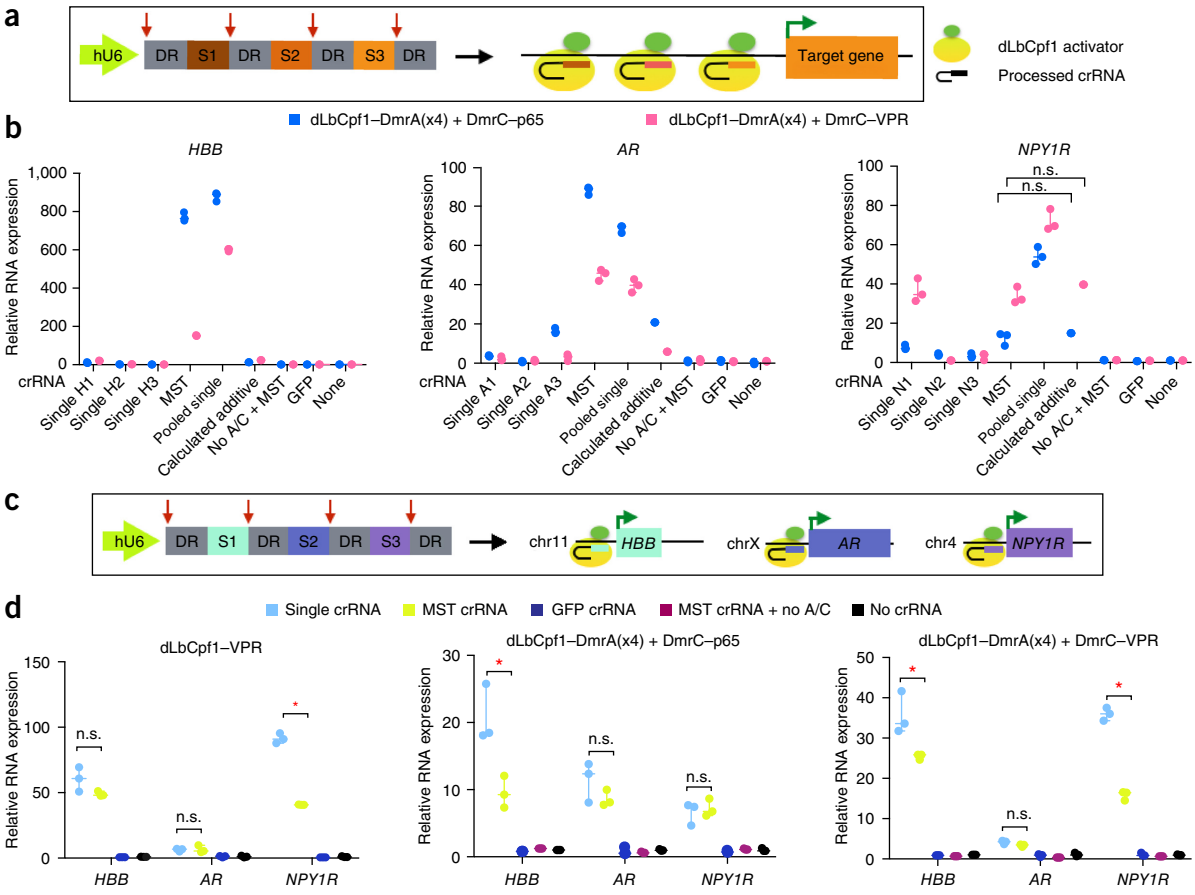


Figure 2 | Multiplex and synergistic regulation of endogenous human genes by dLbCpf1-based activators. (**a**) Schematic illustrating multiplex expression of three crRNAs, each targeted to the same endogenous gene promoter in the same cell. hU6, human U6 polymerase III promoter; DR, direct repeat sequence. S1, Spacer1; S2, Spacer2; S3, Spacer3. (**b**) Activities of dLbCpf1–DmrA(x4) and DmrC–VPR fusions (pink dots) or of dLbCpf1–DmrA(x4) and DmrC–p65 fusions (blue dots) with three single crRNAs, pooled sets of single crRNAs, and an MST encoding all three crRNAs on the *HBB*, *AR*, or *NPY1R* endogenous gene promoters. Synergistic effects of dLbCpf1–VPR/p65 fusions with MST crRNAs or individual pooled crRNAs are all statistically significant (Student *t*-test, two-tailed test assuming equal variance, $P < 0.05$) except for cases where not significant (n.s.) is indicated. (**c**) Schematic illustrating multiplex expression of three crRNAs each targeted to a different endogenous gene promoter in a single cell. (**d**) Simultaneous activation of three endogenous human genes using crRNAs expressed singly or from a MST together with dLbCpf1–VPR direct fusions (left panel), dLbCpf1–DmrA(x4) and DmrC–p65 fusions (middle panel), or dLbCpf1–DmrA(x4) and DmrC–VPR fusions (right panel). For the direct-fusion experiment, a total of 250 ng of plasmids encoding the indicated crRNAs was used. For each drug-regulated fusion experiment, a total of 400 ng of plasmids encoding the indicated crRNAs was used. Transcripts were measured in HEK293 cells using RT-qPCR with relative mRNA expression calculated by comparison to the control sample in which no crRNA is expressed. Representative data shown in **b** and **d** are of three biological independent replicates.; n.s., not significant by Student *t*-test ($P > 0.05$); *, statistically significant by Student *t*-test (two-tailed test assuming equal variance, $P < 0.05$).

In addition, we assessed the kinetics of activator effects on the addition and withdrawal of A/C heterodimerizer. We found that maximum activation of the *HBB* and *AR* genes was observed ≈25–35 h after drug addition (**Supplementary Fig. 6a**), and return of activated gene expression to baseline occurred ≈35–45 h after drug withdrawal (**Supplementary Fig. 6b**). We envision that drug inducibility could be easily extended to other orthologs, and we have successfully used these same strategies to regulate and tune dSpCas9-based activators (**Supplementary Fig. 7**), as have others who published similar drug-regulated dCas9-based activators while this work was in progress[5].

To quantify Cpf1's ability to complement the targeting range of SpCas9, we performed an informatics analysis of sequences located between 600 bp upstream and 400 bp downstream of 32,696 TSSs in the human genome for sites with TTTV, TYCV, and TATV PAMs[7,8] (see Online Methods). We identified 'desert' regions of ≥20 bp within these sequences in which no SpCas9 target sites could be found and summed these for each TSS. We then determined, for each TSS, what percentage of these summed desert regions could be targeted only by wild-type Cpf1 (**Supplementary Fig. 8a**) and the two Cpf1 PAM recognition variants with wild-type Cpf1 (**Supplementary Fig. 8b**). Wild-type Cpf1 permits the targeting of 66% of the sum of all desert regions, while the two PAM variants together with wild-type Cpf1 permit the targeting of 88% of these regions. As might be expected, the targeting capability increases as the size of the desert increases (**Supplementary Fig. 8a**,**b**).

Overall, our findings should enable and encourage the use of dCpf1-based gene regulatory proteins for performing both focused and genome-wide gene-perturbation screens in mammalian cells. More broadly, our results should also motivate researchers to engineer dCpf1-based direct and drug-regulated fusions

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

to other heterologous regulatory domains (e.g., p300, DNMT, TET1) as has previously been done with dSpCas9 (refs. 18–20). Successful development of such fusions could enable recruitment of two different effector domains in a controlled manner using dCas9 and dCpf1 scaffolds targeted to the same promoter.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

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

ACKNOWLEDGMENTS

We thank M. Welch and A. Sousa for technical assistance with constructing certain plasmids. B.P.K. acknowledges support from Banting (Natural Sciences and Engineering Research Council of Canada) and Charles A. King Trust Postdoctoral Fellowships. This work was supported by the National Institutes of Health R35 GM118158 (J.K.J.), NIH R01 GM107427 (J.K.J.), R01 DA036858 (J.S.W.) and U01 CA168370 (J.S.W.), the Howard Hughes Medical Institute (J.S.W.), and the Desmond and Ann Heathwood Massachusetts General Hospital Research Scholar Award (J.K.J.).

AUTHOR CONTRIBUTIONS

Y.E.T., B.P.K., J.K.N., J.Y.H., J.S.W., and J.K.J. conceived of and designed experiments. Y.E.T., B.P.K., J.K.N., J.Y.H., J.E.H., and J.G. performed experiments. Y.E.T., B.P.K., J.K.N., J.Y.H., J.S.W., and J.K.J. wrote the manuscript.

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. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Komor, A.C., Badran, A.H. & Liu, D.R. *Cell* **168**, 20–36 (2017).
2. Wright, A.V., Nuñez, J.K. & Doudna, J.A. *Cell* **164**, 29–44 (2016).
3. Dominguez, A.A., Lim, W.A. & Qi, L.S. *Nat. Rev. Mol. Cell Biol.* **17**, 5–15 (2016).
4. Zetsche, B., Volz, S.E. & Zhang, F. *Nat. Biotechnol.* **33**, 139–142 (2015).
5. Bao, Z., Jain, S., Jaroenpuntaruk, V. & Zhao, H. *ACS Synth. Biol.* **6**, 686–693 (2017).
6. Polstein, L.R. & Gersbach, C.A. *Nat. Chem. Biol.* **11**, 198–200 (2015).
7. Zetsche, B. *et al. Cell* **163**, 759–771 (2015).
8. Gao, L. *et al. Nat. Biotechnol.* **35**, 789–792 (2017).
9. Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A. & Charpentier, E. *Nature* **532**, 517–521 (2016).
10. Zetsche, B. *et al. Nat. Biotechnol.* **35**, 31–34 (2017).
11. Kim, S.K. *et al. ACS Synth. Biol.* **6**, 1273–1282 (2017).
12. Zhang, X. *et al. Cell Discov.* **3**, 17018 (2017).
13. Tang, X. *et al. Nat. Plants* **3**, 17018 (2017).
14. Chavez, A. *et al. Nat. Methods* **12**, 326–328 (2015).
15. Gilbert, L.A. *et al. Cell* **159**, 647–661 (2014).
16. Chavez, A. *et al. Nat. Methods* **13**, 563–567 (2016).
17. Rivera, V.M., Berk, L. & Clackson, T. *Cold Spring Harb. Protoc.* **2012**, 821–824 (2012).
18. Hilton, I.B. *et al. Nat. Biotechnol.* **33**, 510–517 (2015).
19. Liu, X.S. *et al. Cell* **167**, 233–247.e17 (2016).
20. Amabile, A. *et al. Cell* **167**, 219–232.e14 (2016).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

ONLINE METHODS

Plasmids and oligonucleotides. A diagram of constructs and a list of plasmids and related sequences used in this study are found in **Supplementary Note 2**; LbCpf1 target sites, multiplex crRNA sequences, and SpCas9 target sites are found **Supplementary Table 1**. Plasmids dLbCpf1-p65 (JG1202) and dLbCpf1-VPR (JG1211) were constructed by cloning p65 and VPR domains into dLbCpf1 (D832A) (MMW1578) using BstZ17I and NotI sites through isothermal assembly. VPR was amplified from dSpCas9-VPR, which was a gift from G. Church (Harvard Medical School, Boston, Massachusetts; Addgene plasmid 63798)[14]. Plasmids encoding dLbCpf1-DmrA(×1), dLbCpf1-DmrA(×2), dLbCpf1-DmrA(×3), and dLbCpf1-DmrA(×4) (JG674, JG676, JG693, and YET1000, respectively) were generated by subcloning dLbCpf1(D832A) into AgeI- and XhoI-digested constructs that have different numbers of DmrA domains (BPK1019, BPK1033, BPK1140, and BPK1179 for dSpCas9-DmrA(×1) dSpCas9-DmrA(×2), dSpCas9-DmrA(×3), and dSpCas9-DmrA(×4), respectively) using isothermal assembly. Plasmids encoding dSpCas9 (D10A/H840A) effector fusions to VP64, p65, or DmrA were cloned via isothermal assembly (for a complete list of plasmids, please see **Supplementary Note 2**). The DmrC entry vector was digested with NruI, and p65 or VPR were inserted via isothermal assembly to generate DmrC-p65 (BPK1169) and DmrC-VPR (MMW948). Single LbCpf1 crRNA expression plasmids were constructed by ligating annealed oligo duplexes into BsmBI-digested BPK3082 (Addgene 78742)[21]. MST LbCpf1 crRNA plasmids were assembled by annealing, phosphorylating, and ligating three pairs of oligonucleotides into BsmBI- and HindIII-digested BPK3082. Sequences for all oligo pairs are listed in **Supplementary Table 1**. The NF-KB p65 domain encoded in JG1202, BPK1169, BPK1160, and BPK1163 encompasses amino acids 285 to 551 of the native protein; while domain encoded in constructs containing VPR encompass amino acids 428 to 546.

Human cell culture and transfection. HEK293 cells (Invitrogen) and HEK293T cells (ATCC) were grown at 37 °C, in 5% CO_2 in Dulbecco's Modified Eagle Medium (DMEM) with 10% heat-inactivated fetal bovine serum (FBS) and 1% penicillin and streptomycin. U2OS cells (obtained from T. Cathomen, University of Freiburg) were grown at 37 °C in 5% CO_2 in advanced DMEM supplemented with 10% heat-inactivated FBS, 2 mM GlutaMax, and 1% penicillin and streptomycin. Media supernatant was analyzed biweekly for the presence of mycoplasma. For the GFP-activation assays shown in **Supplementary Figure 1**, 500 ng of plasmid DNA expressing dCas9-VPR-BFP (JKNp44), dAsCpf1-VPR-BFP (JKNp49), or dLbCpf1-VPR-BFP (JKNp47) and 500 ng of crRNA plasmid were cotransfected into HEK293T cells that stably encode GFP under a Tet-On 3G doxycycline-inducible promoter (Takara Clontech). Both the dLbCpf1 and dAsCpf1 protein-coding sequences were codon optimized for expression in human cells. Median GFP fluorescence was measured 3 d post-transfection using a LSR II Flow Cytometer (BD), and values were normalized to HEK293T-GFP cells with no plasmids transfected. For LbCpf1 direct-fusion experiments on endogenous genes, 750 ng of dLbCpf1-p65 or dLbCpf1-VPR plasmids with 250 ng of LbCpf1 crRNA plasmids were cotransfected using 3 µl of TransIT-LT1 Transfection Reagent (Mirus, cat. no. MIR2300) into HEK293 cells in a 12-well plate. For experiments with dLbCpf1 fusions to DmrA domains, 400 ng of dLbCpf1-DmrA fusions plasmids, 200 ng of DmrC-p65 or DmrC-VPR plasmids, and 400 ng of LbCpf1 crRNA plasmids were cotransfected using 3 µl of LT1 into HEK293 cells in a 12-well plate. A complete media containing 500 µM A/C heterodimerizer (Takara Clontech) was used at the time of transfection. For LbCpf1 experiments in U2OS cells, the same amount of plasmids used in HEK293 cells was cotransfected by nucleofection using the DN-100 program on a Lonza 4-D Nucleofector with the s.e.m. Cell Line Kit (Lonza). For the experiments shown in **Supplementary Figure 6**, 1,000 ng of dLbCpf1-DmrA fusions plasmids, 500 ng of DmrC-p65 plasmid, and 1,000 ng of LbCpf1 MST crRNA plasmids targeting *HBB* promoter or *AR* promoter were cotransfected using 7.5 µl of LT1 into HEK293 cells in a six-well plate. At 34 h post-transfection, cells were split into 24-well plates to harvest at different time points.

Quantitative reverse-transcription polymerase chain reaction. Total RNA was extracted from the transfected cells 72 h post-transfection using the NucleoSpin RNA Plus (Clontech, cat. no. 740984.250), and 250 ng of purified RNA was used for cDNA synthesis using High-Capacity RNA-cDNA kit (ThermoFisher, cat. no. 4387406). cDNA was diluted 1:20, and 3 µl of cDNA was used for quantitative PCR (qPCR). qPCR reaction samples were prepared using cDNA, SYBR (ThermoFisher, cat. no. 4385612), and primers detecting each target transcript. Primer sequences are listed in **Supplementary Table 2**. qPCR was performed using Roche LightCycler480 with the following cycling protocols (**Supplementary Table 2**). When (*Ct*) values were over 35, we considered them to be 35, because (*Ct*) values fluctuate for very weakly expressed transcripts. Samples that were transfected with LbCpf1 crRNA backbone plasmid (BPK3082) were used as negative controls; and the levels of fold activation over negative controls were normalized to the expression of *HPRT1*.

Flow cytometry. Cells were washed with cell-staining buffer (BioLegends, cat. no. 420201) after 72 h post-transfection and incubated with CD22 (BioLegends, cat. no. 363511) or CD5 antibody (BioLegends, cat. no. 364008) for 15 min; this was followed by two washes with cell-staining buffer. PE/Cy-7 for CD5-positive cells and BV421 for CD22-positive cells were measured by a LSR Fortessa X-20 flow cytometer (BD).

VEGFA ELISA. HEK293 cells were seeded in 24-well plates roughly 20 h prior to transfection using Lipofectamine 3000 (Thermo Fisher Scientific). Unless otherwise indicated, 250 ng of dCas9-DmrA plasmid was cotransfected with 250 ng of sgRNA plasmid (single VEGFA site 3 sgRNA or pooled VEGFA sites 1, 2, and 3 sgRNAs[22]) and 125 ng of DmrC-effector plasmid. A complete media exchange was performed approximately 20 h post-transfection, with the exchanged media containing 500 µM A/C heterodimerizer (Takara Clontech), unless otherwise noted. Approximately 42 h post-transfection, supernatant media were removed and clarified before analysis using a Human VEGF Quantikine ELISA kit (R&D Systems). Optical density of stopped ELISA reactions was determined using a Model 860 Microplate reader (Bio-Rad).

Western blot. HEK293T cells were transfected in six-well plates with 3 µg of plasmids encoding VPR-fused dCas9, dAsCpf1,

and dLbCpf1 and 3 μg of guide-RNA-encoding plasmid. Cells were harvested 48 h post-transfection with PBS and lysed in RIPA buffer (Thermo Fisher Scientific) supplemented with protease inhibitors (Sigma-Aldrich, cat. no. 11867431001). Soluble lysates were quantified by BCA protein assay (Thermo Scientific), and 40 μg of each sample was loaded into a 4–12% Bis–Tris protein gel (Thermo Scientific). HA-tagged VPR fusions were detected with rat anti-HA antibody (Sigma-Aldrich, cat. no. 11867431001). GAPDH was detected with rabbit anti-GAPDH antibody (Cell Signaling Technology, cat. no. 2118S). Blots were imaged by LI-COR Odyssey CLx.

Computational analysis of Cpf1-targetable sites in the human genome. Cpf1-targeting range was characterized by analyzing sequences −600 bp and +400 bp relative to the TSS of 32,696 genes in the human genome (GRCh37/hg19). To assess the complementarity between the targeting ranges of Cpf1 and SpCas9, 'desert' regions of ≥20 bp that are not targetable by SpCas9 (20 bp spacer + NGG PAM) were enumerated and summed for each gene. These desert regions were then analyzed to quantify the proportion that could be targetable by wild-type Cpf1 alone or with the two Cpf1 PAM recognition variants (TYCV and TATV)[8] together with wild-type Cpf1. The desert region sums for each gene were binned into the following groups to assess the distributions: 20–100 bp, 101–200 bp, 201–300 bp, 301–400 bp, 401–500 bp, and 501–1,000 bp. Codes for the analysis performed are available from the corresponding author upon reasonable request.

Data availability statement. The data that support the findings of this study are available from the corresponding author upon reasonable request.

A **Life Sciences Reporting Summary** is available.

21. Kleinstiver, B.P. *et al. Nat. Biotechnol.* **34**, 869–874 (2016).
22. Maeder, M.L. *et al. Nat. Methods* **10**, 977–979 (2013).

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.

doi:10.1038/nmeth.4483

nature research

Corresponding author(s): J. Keith Joung

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

Sample size of n =3 was chosen for experiments where statistical analyses were performed

2. Data exclusions

Describe any data exclusions.

N/A

3. Replication

Describe whether the experimental findings were reliably reproduced.

All attempts at replication were successful.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

N/A

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

N/A

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Nature Methods: doi:10.1038/nmeth.4483

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	python for supplementary figure7 and FlowJo for supplementary figure 2

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No restrictions

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	CD5-PE/Cy7 and CD22-BV421 were used in flow cytometry.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	HEK293 cell line was purchased from Invitrogen, HEK293T from ATCC, U2OS were obtained from from Toni Cathomen, Freiburg
b. Describe the method of cell line authentication used.	N/A
c. Report whether the cell lines were tested for mycoplasma contamination.	Yes, bi-weekly (indicated in the Methods section)
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	No

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	N/A

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	N/A

Nature Methods: doi:10.1038/nmeth.4483

nature research

Corresponding author(s): J. Keith Joung

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	Sample preparation listed in Methods section.
6. Identify the instrument used for data collection.	Fortessa X-20
7. Describe the software used to collect and analyze the flow cytometry data.	BD FACSDIVA for collection and Flow Jo (version1.1.0) for analysis
8. Describe the abundance of the relevant cell populations within post-sort fractions.	1 million cells were analyzed by flow cytometry.
9. Describe the gating strategy used.	Control (no crRNA transfected) samples stained with antibodies were used to determine the gating.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒

Nature Methods: doi:10.1038/nmeth.4483