

**Table 2 Galaxy development projects**

Contributing institution(s)	Hosting URL	Applications emphasized
Netherlands Proteomics Centre (Utrecht, The Netherlands); Netherlands Bioinformatics Centre (Nijmegen, The Netherlands); University of Groningen (Groningen, The Netherlands); Academic Medical Center (Amsterdam, The Netherlands)	<a href="http://galaxy.nbic.nl/">http://galaxy.nbic.nl/</a>	MS-based proteomic and metabolomic software integration; interactomics, proteogenomics
La Trobe University (Melbourne, Australia)	Galaxy Tool Shed under 'Proteomics' <a href="http://toolshed.g2.bx.psu.edu">http://toolshed.g2.bx.psu.edu</a>	Tools for general analysis and visualization of MS-proteomic data
University of Minnesota (Minneapolis)	<a href="https://usegalaxy.org/">https://usegalaxy.org/</a>	Tools for general analysis and visualization of MS-based proteomic data; integration for metaproteomic and proteogenomic applications
Plant Research International, Wageningen University and Research Center (Wageningen, The Netherlands)	<a href="http://galaxy.wur.nl/">http://galaxy.wur.nl/</a>	Tools for MS-based proteomics and metabolomics; software integration for metabolo-proteomic applications

studies. Given the one constant across all the omics fields—that technologies and data analysis needs continually change—this flexibility toward new software and data types should prove beneficial.

Galaxy's transparency and shareability also facilitates reproducible and publicly available analyses of the 'Big Data' produced in omics studies. Coupled with emerging efforts to make workflow frameworks interoperable<sup>13,14</sup>, the sharing functions inherent to frameworks such as Galaxy could transform the way in which large-scale molecular data are exchanged, wherein raw data along with the complete workflow used for its analysis would be deposited and made publicly available. With this vision in mind, we hope that this article will stimulate a much-needed discussion on the best ways to meet the challenges of multi-omic data analysis and move us closer to realizing its potential for biological discovery.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper ([doi:10.1038/nbt.3134](https://doi.org/10.1038/nbt.3134)).

#### ACKNOWLEDGMENTS

The authors gratefully acknowledge funding for P.L.H. from the Netherlands Bioinformatics Centre SP5.12.2.1 Bioassist and 2.2.3 Biorange and the Netherlands Proteomics Centre NPC II E4.2 programs. P.D.M. acknowledges support from the Netherlands Bioinformatics Centre and Netherlands Proteomics Centre (NPC-GM WP3.2). J.B. and J.L. are supported by grants from the Swedish Research Council, Bioinformatics Infrastructure for Life Sciences (BILS) Sweden, Swedish Cancer Foundation and EU FP7 GlycoHit Project. P.L. acknowledges support from the Consortium for Improving Plant Yield (CIPY) and the 7th Framework Program FUEL4ME (FP7-ENERGY-2012-1-2stage grant number 308983). P.D.J., J.M.C. and T.J.G. acknowledge support from US National Science Foundation grant 1147079.

#### COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Jorrit Boekel<sup>1,2</sup>, John M Chilton<sup>3,11</sup>,  
Ira R Cooke<sup>4,5</sup>, Peter L Horvatovich<sup>6</sup>,

Pratik D Jagtap<sup>3,7</sup>, Lukas Käll<sup>8</sup>, Janne Lehtiö<sup>1</sup>,  
Pieter Lukasse<sup>9</sup>, Perry D Moerland<sup>10</sup> & Timothy  
J Griffin<sup>7</sup>

<sup>1</sup>Department of Oncology-Pathology, Science for Life Laboratory, Karolinska Institute, Stockholm, Sweden. <sup>2</sup>Bioinformatics Infrastructure for Life Sciences (BILS), Stockholm, Sweden. <sup>3</sup>Minnesota Supercomputing Institute, University of Minnesota, Minneapolis, USA. <sup>4</sup>Department of Biochemistry, La Trobe Institute for Molecular Science, La Trobe University, Melbourne, Australia. <sup>5</sup>Life Sciences Computation Centre, Victorian Life Sciences Computation Initiative, Melbourne, Australia. <sup>6</sup>Analytical Biochemistry, Department of Pharmacy, University of Groningen, Groningen, The Netherlands. <sup>7</sup>Department of Biochemistry, Molecular Biology and Biophysics, University of Minnesota, Minneapolis, USA. <sup>8</sup>School of Biotechnology, Science for Life Laboratory, Royal Institute of Technology - KTH, Stockholm, Sweden. <sup>9</sup>Plant Research International, Wageningen University and Research Center, Wageningen, The Netherlands. <sup>10</sup>Bioinformatics Laboratory, Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center,

University of Amsterdam, Amsterdam, The Netherlands. <sup>11</sup>Present address: Department of Biochemistry and Molecular Biology, Pennsylvania State University, State College, Pennsylvania, USA. e-mail: [tgriffin@umn.edu](mailto:tgriffin@umn.edu)

- Lappalainen, T. *et al. Nature* **501**, 506–511 (2013).
- Junot, C., Fenaille, F., Colsch, B. & Becher, F. *Mass Spectrom. Rev.* **33**, 471–500 (2013).
- Low, T.Y. *et al. Cell Reports* **5**, 1469–1478 (2013).
- Castellana, N.E. *et al. Mol. Cell. Proteomics* **13**, 157–167 (2014).
- Armengaud, J. *Environ. Microbiol.* **15**, 12–23 (2013).
- Martinez-Outschoorn, U.E. *et al. Cell Cycle* **10**, 1271–1286 (2011).
- Palsson, B. & Zengler, K. *Nat. Chem. Biol.* **6**, 787–789 (2010).
- Goecks, J., Nekrutenko, A. & Taylor, J. *Genome Biol.* **11**, R86 (2010).
- Hull, D. *et al. Nucleic Acids Res.* **34**, W729–W732 (2006).
- Jagla, B., Wiswedel, B. & Coppee, J.Y. *Bioinformatics* **27**, 2907–2909 (2011).
- Hunter, A.A., Macgregor, A.B., Szabo, T.O., Wellington, C.A. & Bellgard, M.I. *Source Code Biol. Med.* **7**, 1 (2012).
- Kacsuk, P. *et al. J. Grid Comput.* **10**, 601–630 (2012).
- Abouelhoda, M., Issa, S.A. & Ghanem, M. *BMC Bioinformatics* **13**, 77 (2012).
- Goble, C.A. *et al. Nucleic Acids Res.* **38**, W677–W682 (2010).

## A split-Cas9 architecture for inducible genome editing and transcription modulation

#### To the Editor:

The RNA-guided CRISPR-associated (Cas) endonuclease Cas9 has been harnessed as a tool for genome editing in mammalian cells<sup>1,2</sup>. In addition, strategies employing catalytically inactive Cas9 can direct effector proteins to genomic targets<sup>3–5</sup> to modulate transcription. Here, we demonstrate that Cas9 can be split into two fragments and rendered chemically inducible by rapamycin-binding dimerization domains for controlled reassembly to mediate genome editing and transcription modulation.

To develop a split-Cas9 system, we identified 11 potential split sites based on a crystal structure of Cas9 in complex with a single guide RNA (sgRNA) and complementary target DNA<sup>6</sup> (Fig. 1a and Supplementary Fig. 1a). The resulting C-terminal Cas9 fragment Cas9(C) and N-terminal Cas9 fragment Cas9(N) were fused to FK506 binding protein 12 (FKBP) and FKBP rapamycin binding (FRB) domains<sup>7</sup> of the mammalian target of rapamycin (mTOR), respectively, to make 11 split-Cas9 sets (split-1 through split-11)

(Supplementary Fig. 1b and Fig. 1b). We tested all split-Cas9 sets by targeting the *EMX1* locus in human embryonic kidney 293FT (HEK293FT) cells. Using the SURVEYOR nuclease assay, we detected insertion/deletion (indels) mutations mediated by all split-Cas9 sets in cells treated with rapamycin (Supplementary Methods, for a list of all primers and sgRNAs see Supplementary Tables 2-5). In addition, moderate levels of indels could also be detected in the absence of rapamycin (Supplementary Fig. 1c,d). The observed background activity was not due to residual nuclease activity of individual split pieces (data not shown). Using a small number of split-Cas9 sets lacking dimerization domains, we found that Cas9 split fragments can auto-assemble in cells (Supplementary Fig. 1e,g), which explains our observed background activity.

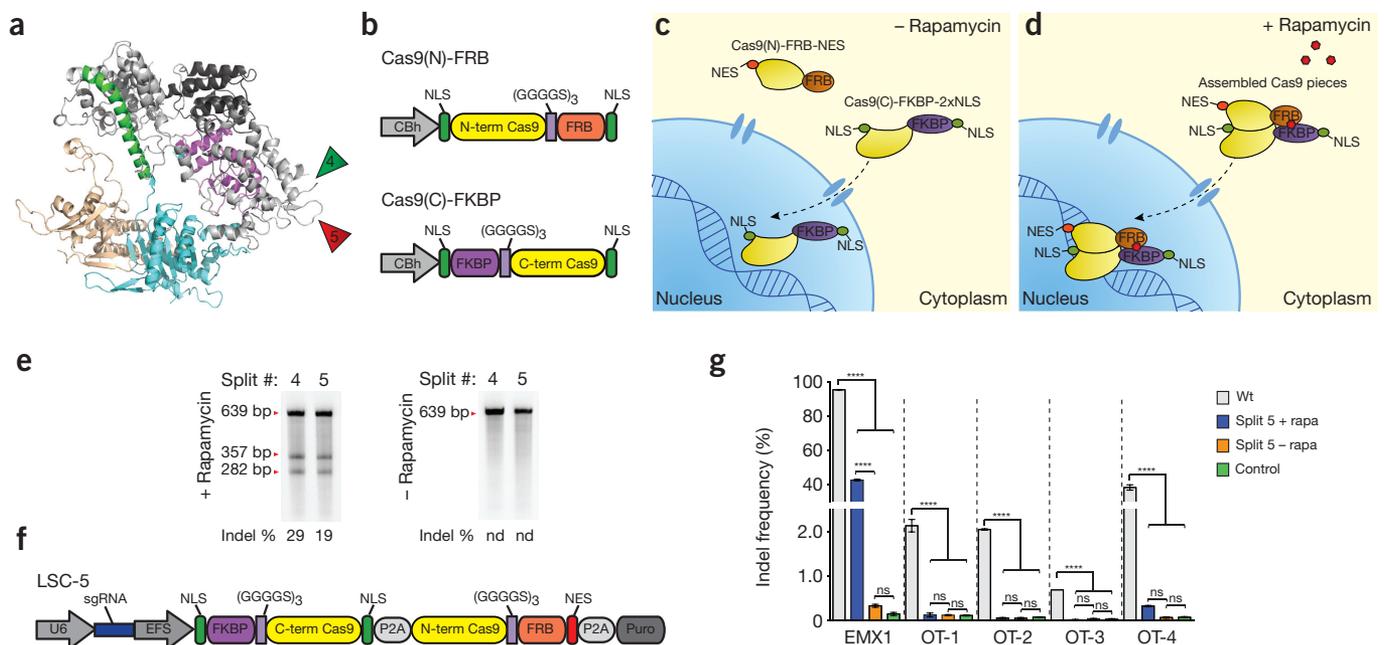
After establishing that background activity in the split-Cas9 system is due to spontaneous auto-assembly of Cas9, we hypothesized that keeping each Cas9 fragment spatially separated might reduce background activity<sup>8</sup>. To sequester the Cas9(N)-FRB fragment in the cytoplasm,

where it is less likely to dimerize with the nuclear-localized Cas9(C)-FKBP fragment, we replaced the two nuclear localization sequences (NLSs) on Cas9(N)-FRB with a single nuclear export sequence (NES) (Cas9(N)-FRB-NES). In the presence of rapamycin, Cas9(N)-FRB-NES dimerizes with Cas9(C)-FKBP-2 × NLS to reconstitute a complete Cas9 protein, which shifts the balance of nuclear trafficking toward nuclear import and allows DNA targeting (Fig. 1c,d). We tested our strategy with split-4 and split-5 (Fig. 1a) and found that a single NES is sufficient to reduce background activity below the detection limit of the SURVEYOR assay (Fig. 1e). Our data show that spatial sequestration of Cas9(N)-FRB and Cas9(C)-FKBP split fragments inside the cell, combined with rapamycin-activated dimerization, allows inducible activation of the Cas9 nuclease.

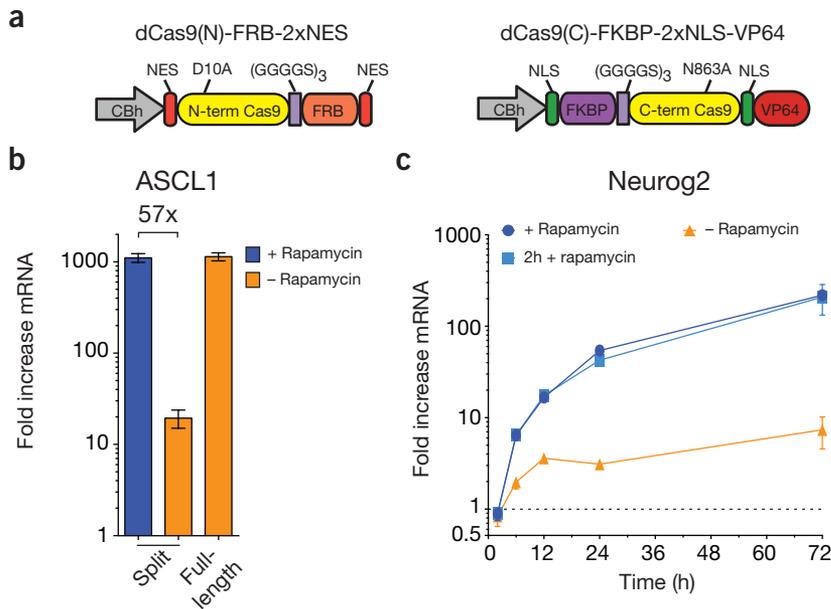
High dosage of Cas9 can exacerbate indel frequencies at off-target sequences<sup>9</sup>. We speculated that induction of low levels of Cas9 activity could be used to reduce off-target indels compared to constitutively active Cas9, which may exhibit high on-target activity but also elevated levels of off-target activity.

Therefore, we generated a lentivirus construct for split-5 (LSC-5 for lentivirus split-Cas9 split-5; Fig. 1f) and transduced HEK293FT cells with a multiplicity of infection (MOI) of  $\leq 0.3$  followed by puromycin selection for 5 days.

DNA from wild-type (wt)-Cas9-transduced HEK293FT cells was analyzed by deep sequencing 4 weeks after transduction, whereas DNA from LSC-5-transduced cells was analyzed after 6 weeks, to allow for 12 consecutive days of treatment with 200 nM rapamycin (Fig. 1g). In cells transduced with a lentivirus carrying both wt-Cas9 and an *EMX1*-targeting sgRNA, we detected ~95% indel frequency at the on-target site as well as mutations at four validated off-target sites (OT-1, 2, 3 and 4) OT-1 = 2%, OT-2 = 2%, OT-3 0.7% and OT-4 = 38%. In comparison, on-target indel frequency in cells transduced with LSC-5 was ~43% after 12 days of rapamycin treatment. In untreated cells, no significant difference in *EMX1* on-target indels could be detected between LSC-5 and control samples. Notably, no significant modification of off-target sites could be detected in cells transduced with LSC-5, regardless of rapamycin treatment (one-way



**Figure 1** Generation and optimization of inducible split-Cas9 fragments. (a) Ribbon representation of Cas9. Triangles indicate split sites for split-4 (green) and split-5 (red). (b) Diagram of inducible split-Cas9 fusions. N- and C-terminal pieces of human codon-optimized *Streptococcus pyogenes* Cas9 are fused to FRB and FKBP dimerization domains, respectively. (c,d) Strategy for optimizing the split-Cas9 system. In the absence of rapamycin (c), the Cas9(N)-FRB-NES piece is sequestered in the cytoplasm owing to the addition of a NES. The Cas9(C)-FKBP piece contains two NLSs and is actively imported into the nucleus. In the presence of rapamycin (d), Cas9(N)-FRB-NES binds to Cas9(C)-FKBP. NLSs of the resulting reassembled Cas9 mediate nuclear importation and subsequent binding to the targeted locus. (e) Representative SURVEYOR assay for split-4- and split-5-mediated indels at the human *EMX1* locus, with (left) and without (right) rapamycin. Arrowheads indicate expected SURVEYOR fragments. Nd, not detected. (f) Schematic of lentiviral split-Cas9 plasmid containing U6 promoter-driven sgRNA, EFS promoter-driven split-Cas9 pieces and puromycin resistance gene (puro). 2A self-cleaving peptides (P2A) separate both split-Cas9 pieces and puro. (g) Indel frequencies measured by deep sequencing at the *EMX1* locus and four annotated OT sites. Indels were measured 4 weeks (wt-Cas9;  $n = 2$  biological replicates) or 6 weeks (split-Cas9;  $n = 3$  biological replicates) after transduction (\*\*\*\* $P < 0.0001$ ). Rapamycin treatments lasted 12 days. Mean  $\pm$  s.e.m. in all panels. Ns, not significant.



**Figure 2** Inducible transcriptional activation using split dCas9-VP64 fusions. (a) Schematic of dCas9(N)-FRB-2 × NES and dCas9(C)-FKBP-2 × NLS-VP64 fusions used for transcriptional activation. Each piece harbors an annotated point mutation (D10A or N863A), which reconstitutes a catalytically inactive dCas9 upon rapamycin-induced assembly. A VP64 transcriptional activator domain is fused to the C-terminal end of the dCas9(C)-FKBP-2 × NLS-VP64 piece. (b) *ASCL1* expression measured by qPCR in HEK293FT cells transfected with split-4 (Split) and four sgRNAs per gene. Expression was measured in cells with and without rapamycin ( $n = 4$  biological replicates), compared with full-length dCas9-VP64 (full-length;  $n = 3$  biological replicates). Untransfected cells were used as baseline. (c) *Neurog2* expression in N2A cells measured by qPCR 2, 6, 12, 24 and 72 h after rapamycin treatment ( $n = 3$  biological replicates for each time point). Cells were treated continuously with rapamycin (dark blue circles), only treated for 2 h (light blue squares) or untreated (orange triangles). Untransfected cell were used as baseline. Mean ± s.e.m. in all panels.

ANOVA,  $P > 0.9999$ ). Transient transfection experiments using the same guide sgRNA and wt-Cas9 performed previously by us and others<sup>10</sup> showed that at a 32–50% on-target mutation rate, frequency of indels at off-target site OT-4 was 4–20% (Supplementary Table 1). Taken together, these data indicate that stable, low-copy expression of split-Cas9 fragments can be used to induce substantial indels at a targeted locus without high mutation at off-target sites.

Next, we sought to explore whether the split-Cas9 architecture can be applied to catalytically inactive Cas9 (dCas9) to mediate inducible transcription activation. We cloned split-4 fragments harboring a D10→A point mutation in the FRB fusion (dCas9(N)-FRB-2 × NES) and an N863→A point mutation in the FKBP fusion and added a VP64 transactivation domain to Cas9(C)-FKBP-2 × NLS (dCas9(C)-FKBP-2 × NES-VP64) (Fig. 2a). These fragments reconstitute a catalytically inactive Cas9-VP64 fusion (dCas9-VP64).

We tested split dCas9-VP64 by activating *ASCL1*, *MYOD1* or *IL1RN* transcription in HEK293FT cells, using four previously validated sgRNAs<sup>11</sup> per gene. Cells were

treated with rapamycin 24 h after transfection and maintained in 200 nM rapamycin until harvested for RNA at 48 h after transfection. A significant increase in mRNA levels, compared with untransfected HEK293FT, could be detected using quantitative real-time PCR (qPCR) for all three genes (one-way ANOVA, *ASCL1*  $P < 0.0001$ , *MYOD1*  $P < 0.0001$ , *IL1RN*  $P < 0.0001$ ) (Fig. 2b and Supplementary Fig. 2a). Background activity was low compared with rapamycin-induced cells (+rapamycin/–rapamycin ratio: *ASCL1* = 57, *MYOD1* = 27, *IL1RN* = 552) and not significant compared with untransfected cells (one-way ANOVA,  $P > 0.99$ ).

To test whether transcriptional activation is reversible upon withdrawal of rapamycin, we activated *Neurog2* expression in N2A cells and *ASCL1* in HEK293FT cells (Fig. 2c and Supplementary Fig. 2b). Cells were treated with rapamycin 24 h after transfection. Rapamycin was either withdrawn after 2 h or replaced every 24 h for continuous induction. Cells were harvested at 2, 6, 12, 24 and 72 h after rapamycin treatment, and mRNA levels were analyzed by qPCR. *Neurog2* and *ASCL1* levels increased during the entire study, with

no significant difference between continuous rapamycin treatment and a 2-h treatment (correlation coefficient: *Neurog2* = 1, *ASCL1* = 1). Given the persistent activation after rapamycin withdrawal, this system will be useful for experiments where synchronized activation is beneficial, such as cellular differentiation or development or modulation of genes that adversely affect the health or growth of the cell.

Taken together, we have demonstrated that Cas9 can be split into two distinct fragments, which form a functional full-length Cas9 nuclease when brought back together by chemical induction. The split-Cas9 architecture will be useful for a variety of applications. For example, split-Cas9 systems may enable genetic strategies for restricting Cas9 activity to intersections of cell populations by putting each fragment under a different tissue-specific promoter. Additionally, different chemically inducible dimerization domains, such as abscisic acid or gibberellin-sensing domains, may also be employed to generate an array of inducible Cas9 molecules, fused to different modulatory domains, to construct synthetic transcriptional networks.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper (doi:10.1038/nbt.3149).

#### ACKNOWLEDGMENTS

We would like to thank N.E. Sanjana, S. Konermann, M.D. Brigham, A. Trevino, F.A. Ran and W. Yan for technical assistance, S. Jones, M. Heidenreich and K. Zheng for editing and members of the Zhang laboratory for discussion, support and advice. F.Z. is supported by the National Institute of Mental Health (NIMH) through a National Institutes of Health (NIH) Director's Pioneer Award (DPI-MH100706), the National Institute of Neurological Disorders and Stroke (NINDS) through an NIH Transformative R01 grant (R01-NS 07312401), National Science Foundation (NSF) Waterman Award, the Keck, Damon Runyon, Searle Scholars, Klingenstein, Vallee, Merkin and Simons Foundations, and Bob Metcalfe, New York Stem Cell Foundation. F.Z. is a New York Stem Cell Foundation Robertson Investigator. CRISPR reagents are available to the academic community through Addgene, and associated protocols, support forum and computational tools are available via the Zhang laboratory website (<http://www.genome-engineering.org>).

#### AUTHOR CONTRIBUTIONS

B.Z. and F.Z. conceived the project and designed the experiments. B.Z. and S.E.V. performed experiments and analyzed data. B.Z. and F.Z. wrote the paper with help from all authors.

#### COMPETING FINANCIAL INTERESTS

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

Bernd Zetsche<sup>1–5</sup>, Sara E Volz<sup>1–4</sup> & Feng Zhang<sup>1–4</sup>

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. <sup>2</sup>McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA. <sup>3</sup>Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology Cambridge, Massachusetts, USA. <sup>4</sup>Department of Biological Engineering, Massachusetts Institute of Technology Cambridge, Massachusetts, USA. <sup>5</sup>Department of Developmental Pathology, Institute of Pathology, Bonn Medical School, Bonn, Germany.  
e-mail: zhang@broadinstitute.org

Published online 2 February 2015  
doi:10.1038/nbt.3149

1. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
2. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
3. Konermann, S. *et al.* Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* doi:10.1038/nature14136 (10 December 2014).
4. Gilbert, L.A. *et al.* Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
5. Qi, L.S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
6. Nishimasu, H. *et al.* Crystal structure of cas9 in complex with guide RNA and target DNA. *Cell* **156**, 935–949 (2014).
7. Banaszynski, L.A., Liu, C.W. & Wandless, T.J. Characterization of the FKBP-rapamycin-FRB ternary complex. *J. Am. Chem. Soc.* **127**, 4715–4721 (2005).
8. Konermann, S. *et al.* Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).
9. Hsu, P.D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat. Biotechnol.* **31**, 827–832 (2013).
10. Fu, Y., Sander, J.D., Reyon, D., Cascio, V.M. & Joung, J.K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. *Nat. Biotechnol.* **32**, 279–284 (2014).
11. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat. Methods* **10**, 973–976 (2013).