An enhanced CRISPR repressor for targeted mammalian gene regulation

Nan Cher Yeo [1,2,14], Alejandro Chavez [3,14]*, Alissa Lance-Byrne[1], Yingleong Chan[1,2], David Menn[4], Denitsa Milanova[1,2], Chih-Chung Kuo[5,6], Xiaoge Guo[1,2], Sumana Sharma[7], Angela Tung[1], Ryan J. Cecchi[1], Marcelle Tuttle[1], Swechchha Pradhan[4], Elaine T. Lim[1,2], Noah Davidsohn[1,2], Mo R. Ebrahimkhani[4,8], James J. Collins[1,9,10,11,12], Nathan E. Lewis [5,6,13], Samira Kiani [4]* and George M. Church [1,2]*

The RNA-guided endonuclease Cas9 can be converted into a programmable transcriptional repressor, but inefficiencies in target-gene silencing have limited its utility. Here we describe an improved Cas9 repressor based on the C-terminal fusion of a rationally designed bipartite repressor domain, KRAB–MeCP2, to nuclease-dead Cas9. We demonstrate the system's superiority in silencing coding and noncoding genes, simultaneously repressing a series of target genes, improving the results of single and dual guide RNA library screens, and enabling new architectures of synthetic genetic circuits.

The clustered regularly interspaced short palindromic repeats (CRISPR)–Cas9 system, which mediates adaptive immunity in bacteria and archaea, has emerged as a powerful tool for genome engineering[1–7]. Cas9 is an RNA-guided endonuclease that can be directed to specific DNA sequences through complementarity between a Cas9-associated guide RNA (gRNA) and the target locus, provided that a protospacer-adjacent motif is proximal to the target. Because it is possible to change the target locus by altering only the delivered gRNA with Cas9, its use has been quickly adopted for selective gene ablation and for unbiased genome-wide screens[8–11]. However, Cas9 cutting can lead to cellular toxicity due to the formation of DNA double-strand breaks, and Cas9-generated modifications are irreversible, both of which limit its applications[12].

Within Cas9, the amino acids critical for DNA catalysis can be mutated to generate a nuclease-dead Cas9 (dCas9), which remains competent for DNA binding but lacks endonuclease activity[13]. When directed toward the transcriptional start site (TSS) of a gene, dCas9 can physically block RNA polymerase passage, thereby leading to gene silencing[13]. Further improvement in transcriptional inhibition can be achieved with the addition of repression domains such as the Krüppel-associated box (KRAB) to dCas9, with the resultant dCas9–KRAB fusion protein being the current gold standard for dCas9-based repression studies[14–17]. Although it has been widely adopted, the dCas9–KRAB system suffers from inefficient knockdown and poor performance compared with that of Cas9 nuclease-based methods[14,18,19].

Previous work has shown that the fusion of several transcriptional regulators to dCas9 in tandem can lead to a synergistic increase in activity[20–23], and initial efforts focused on building more potent transcriptional activators. Here we assembled and screened combinations of potent repressor domains to engineer a highly effective dCas9–KRAB–MeCP2 transcriptional repressor.

Results

Identification of dCas9–KRAB–MeCP2. To begin to design a more potent Cas9 repressor, we separately fused more than 20 different effector domains with known roles in transcriptional regulation and gene silencing to the C terminus of dCas9. We then transfected the resulting dCas9 fusion proteins into HEK293T cells, along with a gRNA targeting the promoter of a gene encoding an enhanced yellow fluorescent protein reporter (*EYFP*). The majority of dCas9 fusions were able to repress *EYFP* expression, with a few leading to greater repression (up to eightfold) compared with that observed with dCas9 alone (Supplementary Fig. 1). Next we generated a library of dCas9 bipartite repressors consisting of the commonly used KRAB repressor and the six top-performing domains from our initial screen (MeCP2, SIN3A, HDT1, MBD2B, NIPP1, and HP1A). Our library contained all pairwise repeating and nonrepeating combinations of the seven selected domains. As expected, many bipartite fusion proteins showed strong improvement, ranging from 5-fold to 60-fold greater repression of *EYFP* compared with that achieved by dCas9 alone (Supplementary Fig. 2).

Having done our initial studies with a synthetic reporter, we next determined whether our most potent repressors could also downregulate endogenous target genes. We selected nine bipartite repressors for further characterization. We cotransfected each of

[1]Wyss Institute for Biologically Inspired Engineering, Harvard University, Cambridge, MA, USA. [2]Department of Genetics, Harvard Medical School, Boston, MA, USA. [3]Department of Pathology and Cell Biology, Columbia University College of Physicians and Surgeons, New York, NY, USA. [4]School of Biological and Health Systems Engineering, Ira A. Fulton Schools of Engineering, Arizona State University, Tempe, AZ, USA. [5]Department of Bioengineering, University of California, San Diego, San Diego, CA, USA. [6]Novo Nordisk Foundation Center for Biosustainability, University of California, San Diego, San Diego, CA, USA. [7]Cell Surface Signalling Laboratory, Wellcome Trust Sanger Institute, Cambridge, UK. [8]Division of Gastroenterology and Hematology, Mayo Clinic College of Medicine and Science, Phoenix, AZ, USA. [9]Institute for Medical Engineering & Science, Massachusetts Institute of Technology, Cambridge, MA, USA. [10]Synthetic Biology Center, Massachusetts Institute of Technology, Cambridge, MA, USA. [11]Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA, USA. [12]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [13]Department of Pediatrics, University of California, San Diego, San Diego, CA, USA. [14]These authors contributed equally: Nan Cher Yeo, Alejandro Chavez. *e-mail: ac4304@cumc.columbia.edu; samira.Kiani@asu.edu; gchurch@genetics.med.harvard.edu
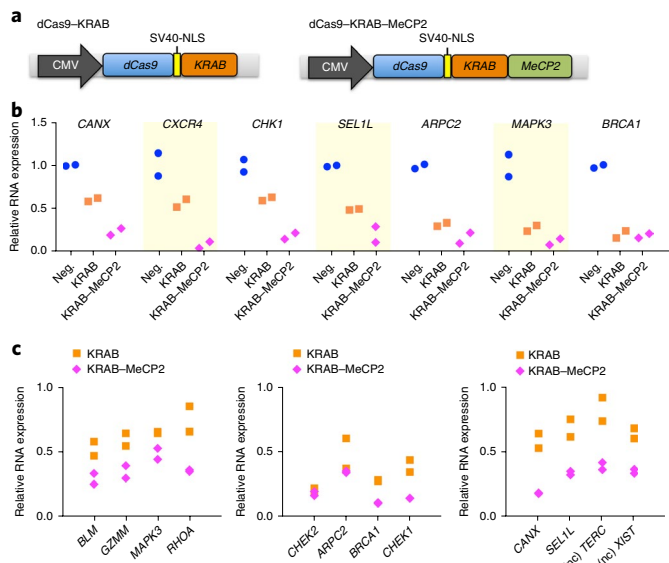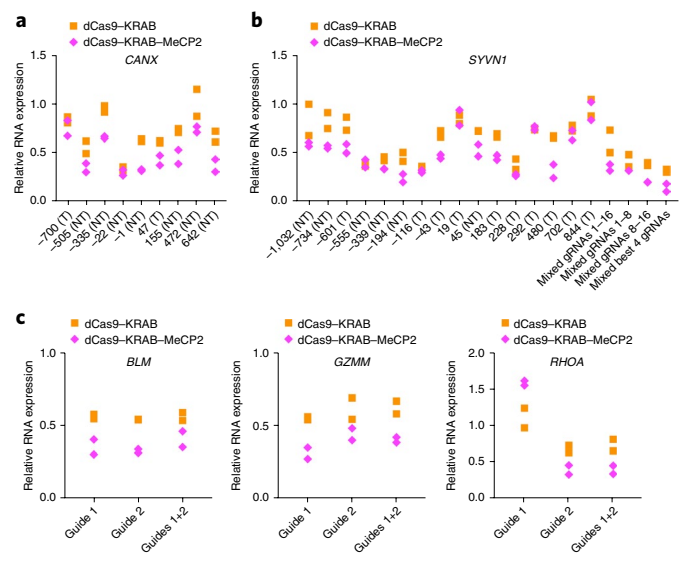
Fig. 1 | Repression of endogenous genes by dCas9–KRAB–MeCP2.
a, Schematic of dCas9–KRAB and dCas9–KRAB–MeCP2 repressors. NLS, nuclear localization signal. b, RNA expression of targeted single genes. $n = 2$ biologically independent samples (cell cultures). Neg., negative control. c, RNA expression during three separate multiplex repression studies. In each study, four different genes were targeted simultaneously. Two noncoding (nc) genes, XIST and TERC, were studied. $n = 2$ biologically independent samples (cell cultures).

Fig. 2 | dCas9–KRAB–MeCP2 paired with gRNA at various positions. a, An array of gRNAs was designed to target the region from 1 kb upstream to 1 kb downstream of the TSS for CANX. Shown is the RNA expression of CANX when different gRNAs were used. T, template strand; NT, nontemplate strand. $n = 2$ biologically independent samples (cell cultures). **b,** RNA expression of SYVN1 when individual or combinations of different gRNAs were used. $n = 2$ biologically independent samples (cell cultures). **c,** RNA expression of the indicated target genes with one or two gRNAs used. $n = 2$ biologically independent samples (cell cultures).

the dCas9 variants into HEK293T cells along with a set of gRNAs targeting four different endogenous genes. Although we observed varying degrees of gene repression depending on the target gene, the dCas9 repressor consisting of KRAB and the transcription-repression domain of MeCP2, referred to here as dCas9–KRAB–MeCP2 (Fig. 1a and Supplementary Table 1), was the most potent across all targets (Supplementary Fig. 3). We also generated a series of tripartite fusion proteins to test whether we could achieve further improvements in repression by using three different effector domains (Supplementary Fig. 4). Compared with the gene silencing observed with the dCas9–KRAB–MeCP2 protein, none of the designed tripartite repressors demonstrated any improvement in activity (Supplementary Fig. 5). The lack of improved repression with the tripartite repressors could be due to the domains recruiting identical secondary effectors. It is also possible that the extent to which the domains fold and function properly decreases as greater numbers of effectors are fused together.

To understand the contributions of KRAB and MeCP2 to the overall effect, we carried out a side-by-side comparison of different dCas9 fusion proteins containing KRAB or MeCP2 (Supplementary Fig. 6). The dCas9–KRAB–MeCP2 fusion outperformed both KRAB and MeCP2 as single or double fusions to dCas9, which suggests that it is the combined effect of both domains that leads to increased gene repression.

Improved repression of endogenous genes by dCas9–KRAB–MeCP2. We next systematically compared the activity of dCas9–KRAB–MeCP2 to that of the current gold-standard repressor, dCas9–KRAB, by targeting a wide range of endogenous loci in HEK293T cells. For the majority of single genes tested, dCas9–KRAB–MeCP2 showed improved repression compared with that of dCas9–KRAB (Fig. 1b). To test whether dCas9–KRAB–MeCP2 could simultaneously downregulate the expression of multiple genes more effectively, we cotransfected three sets of four gRNAs, each targeting a different locus, into HEK293T cells (Fig. 1c).

dCas9–KRAB–MeCP2 showed improved multiplexed repression for all genes tested except two, for which it showed activity similar to that of dCas9–KRAB.

We next designed an array of gRNAs targeting both template and nontemplate strands ranging from 1 kb upstream to 1 kb downstream of the TSS for two different genes (CANX and SYVN1). We found that 15 out of 25 gRNAs tested showed greater repression with dCas9–KRAB–MeCP2 than with dCas9–KRAB. These results were independent of the DNA strand targeted and whether the gRNA was directed outside of the previously characterized optimal targeting window for repression[15] (Fig. 2a,b). Initial studies with CRISPR repressors suggested that the use of multiple gRNAs targeting the same locus leads to marked improvement in gene knockdown[13]. In contrast to those results, neither dCas9–KRAB nor dCas9–KRAB–MeCP2 showed improved repression when we used multiple guides against the same target; rather, they exhibited activity that appeared to be dictated by the most potent guide in the set tested, in agreement with recent observations[24] (Fig. 2b,c).

The effect of dCas9–KRAB–MeCP2 is highly specific. Effector domains that recruit chromatin modifiers can cause widespread epigenetic changes over large regions of DNA[25–27]. We evaluated the targeting specificity of dCas9–KRAB–MeCP2 by probing the expression of neighboring genes when either CXCR4 or SYVN1 was targeted (Supplementary Fig. 7a,b). We did not observe any significant off-target effects on the neighboring genes examined.

We next targeted CXCR4 and carried out whole-transcriptome RNA sequencing (RNA-seq) to evaluate the specificity of dCas9–KRAB–MeCP2 on a genome-wide scale. We compared the results with those obtained from cells transfected with either dCas9 or dCas9–KRAB. dCas9–KRAB–MeCP2 showed the strongest repression signal for the target gene, CXCR4. The global transcriptome profiles of all dCas9 repressors showed high correlation with that of the negative control, cells transfected with gRNA alone (Fig. 3
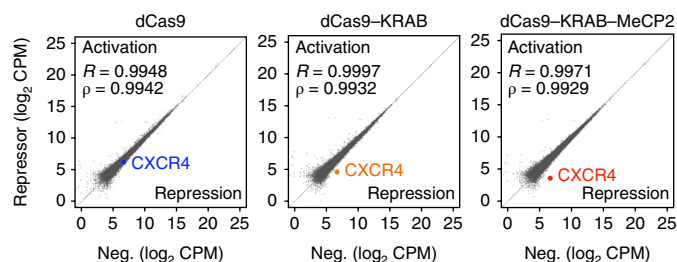
Fig. 3 | dCas9–KRAB–MeCP2-mediated repression is highly specific in human cells. Results of RNA-seq analyses of HEK293T cells transfected with a gRNA targeting *CXCR4* along with dCas9, dCas9–KRAB, or dCas9–KRAB–MeCP2 repressors. Data are normalized, and $\log_2$-transformed values in counts per million (CPM) are plotted for each repressor (*y*-axis) versus values for a negative control (Neg.) transfected with gRNA alone (*x*-axis). Pearson's and Spearman's correlation coefficients are provided for each repressor. $n = 2$ biologically independent samples (cell cultures).

and Supplementary Fig. 7c), although an overlapping set of differentially expressed genes was also observed (Supplementary Figs. 8 and 9, Supplementary Tables 2 and 3, and Supplementary Data 1). Of the few differentially expressed genes that showed downregulation, none exhibited a near-sequence match to the *CXCR4*-targeting gRNA, which suggests that these changes did not result from inappropriate targeting of repressors to the loci with altered expression.

dCas9–KRAB–MeCP2 efficiently suppresses genes when used at library scales. One of the most powerful uses of CRISPR–Cas9 technology is to enable facile genome-wide screens. To determine whether our tool was amenable to such screening, we generated heterogeneous populations of human haploid (HAP1) cells stably expressing dCas9, dCas9–KRAB, or dCas9–KRAB–MeCP2. RNA expression levels of the dCas9–KRAB and dCas9–KRAB–MeCP2 repressors were similar in these haploid lines but were significantly lower than that of dCas9 alone (Supplementary Fig. 10a). When endogenous genes were targeted, cells containing dCas9–KRAB–MeCP2 showed stronger repression compared with that in cells expressing other dCas9 constructs (Supplementary Fig. 10b).

Genes that are essential for cellular function serve as a useful set of targets for comparing the relative performance of different screening platforms[19]. With this in mind, we infected each of our dCas9-expressing lines, as well as wild-type HAP1 cells, with a lentiviral single-guide RNA (sgRNA) library targeting an assortment of essential and nonessential genes. We then passaged the cells over a period of 14 d and quantified the extent to which the various sgRNAs were depleted over time. In the screen, cells expressing dCas9–KRAB–MeCP2 showed the strongest depletion for guides targeting essential genes versus nonessential genes ($P = 3.52 \times 10^{-80}$ with dCas9–KRAB–MeCP2 versus 5.41×10^{-19} with dCas9–KRAB at day 14) (Fig. 4a, Supplementary Table 4, and Supplementary Data 2). In addition, we observed strong depletion signals (up to 256-fold depletion) with dCas9–KRAB–MeCP2 as early as day 7, compared with the mostly weak signals exhibited by dCas9–KRAB (up to twofold depletion). We did not observe any depletion in sgRNAs targeting essential genes in wild-type cells, which indicates that our results were not due to technical artifacts (Supplementary Fig. 10c).

To test the generality of our system, we repeated the above screen in SH-SY5Y, a near-diploid human neuroblastoma cell line (Fig. 4b and Supplementary Data 3), and in HEK293T cells (Fig. 4c and Supplementary Data 4). Although the overall depletion signal was not as strong as that observed in HAP1 cells, cell lines containing dCas9–KRAB–MeCP2 showed a greater degree of depletion for sgRNAs targeting essential genes at all times of
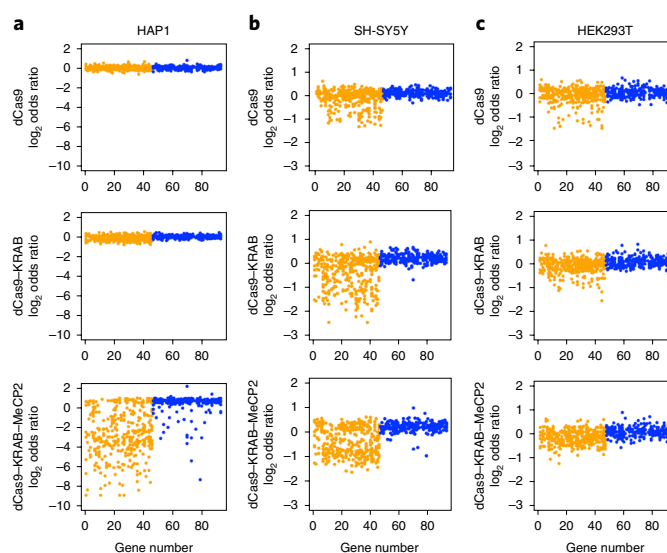
Fig. 4 | dCas9–KRAB–MeCP2 outperforms previous tools in screens of gene essentiality. a–c, $\log_2$ odds ratios of all sgRNA constructs compared with values for wild-type HAP1 cells (**a**), SH-SY5Y cells (**b**), and HEK293T cells (**c**) at day 14. sgRNAs targeting essential genes are represented in orange, and sgRNAs targeting nonessential genes are represented in blue.

measurement compared with that obtained with previous technologies (Supplementary Tables 5 and 6).

We plotted sgRNA depletion as a function of position from the TSS for the several hundred essential gene-targeting sgRNAs used (Supplementary Fig. 11a and Supplementary Table 7). As expected, sgRNAs positioned within the previously identified optimal targeting window (−50 bp to +200 bp from the TSS) showed a higher likelihood of being depleted compared with sgRNAs positioned outside of the window (Supplementary Fig. 11b). Regardless of targeting position, dCas9–KRAB–MeCP2 outperformed dCas9–KRAB (Supplementary Fig. 11c).

In the dCas9–KRAB–MeCP2 screen, a few of the sgRNAs designed to target nonessential genes also showed marked depletion. We found that a subset of these sgRNAs also showed depletion when combined with either dCas9–KRAB or dCas9 alone (Supplementary Fig. 12 and Supplementary Data 5), indicating that the observed off-target binding is not only a property of our improved repressor. Furthermore, a few sgRNAs that showed unexpected depletion in the HAP1 screen also showed depletion in the SH-SY5Y screen for either dCas9–KRAB–MeCP2 or dCas9–KRAB (Supplementary Data 5). These data suggest that there are consistent off-target sites that these unique sgRNAs are binding to that affect growth. We hypothesize that because dCas9–KRAB–MeCP2 is a more potent repressor, signals from these off-target binding events are more readily observed.

To assess the overall performance of dCas9–KRAB–MeCP2 in screening environments, we used the conventional MAGeCK analysis pipeline[28]. MAGeCK takes into account the behavior of all sgRNAs against a given gene when determining whether that gene is subject to selection during the screen. In HAP1 cells, dCas9–KRAB–MeCP2 correctly identified 21 essential genes, compared with only 3 identified by dCas9–KRAB at day 14. Similarly, in HEK293T cells, dCas9–KRAB–MeCP2 identified 11 essential genes, compared with 5 identified by dCas9–KRAB. In SH-SY5Y cells, dCas9–KRAB–MeCP2 showed similar performance by identifying 11 essential genes, compared with 10 identified by dCas9–KRAB. No nonessential genes were deemed significant in any of the experimental groups (Supplementary Fig. 13 and Supplementary Data 6). These
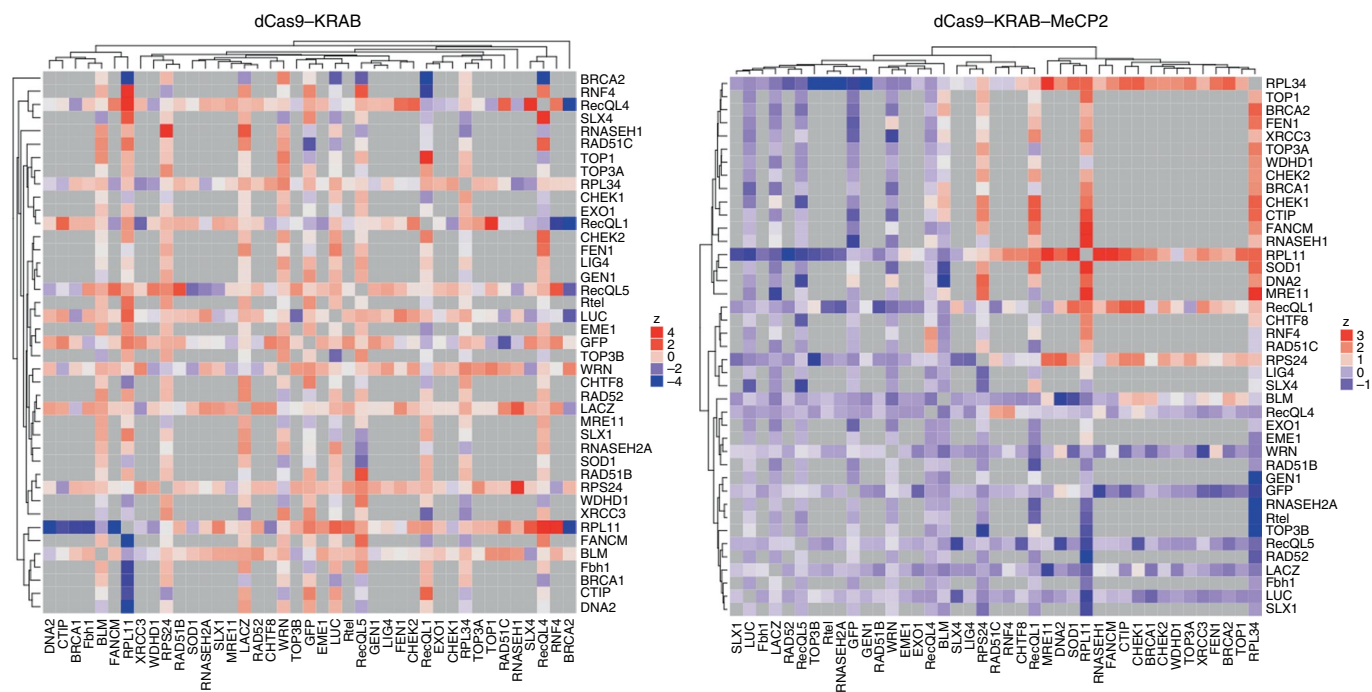
Fig. 5 | dCas9–KRAB–MeCP2 improves genetic interaction mapping. Hierarchical clustered heat maps of genetic interactions for dCas9–KRAB and dCas9–KRAB–MeCP2. Only the screen using dCas9–KRAB–MeCP2 showed a discernible clustering structure.

results support the idea that dCas9–KRAB–MeCP2 is a more potent tool than dCas9–KRAB for screening gene essentiality.

dCas9–KRAB–MeCP2 improves genetic interaction mapping. To further assess the capabilities of dCas9–KRAB–MeCP2, we conducted a combinatorial repression screen. Our screening library consisted of dual guides targeting genes involved in DNA repair, as well as a set of positive and negative controls. In our library each construct contained two gRNAs, and the majority of gRNA pairs targeted two different genes (Supplementary Table 8). Similar to what we observed in the single-gene targeting screens, samples for the dual-guide screen that contained dCas9–KRAB–MeCP2 showed improved selection for or against specific gRNA pairs over time compared with that in samples containing dCas9–KRAB (Supplementary Fig. 14a).

We next estimated the fitness effects for each individual gRNA and quantified genetic interactions (indicated by pi scores) between gene pairs[29] (details on the interpretation of pi scores are included in Supplementary Note 1). Specifically, we tested whether distant gene pairs tend to engage more in negative genetic interactions, and whether gene pairs that form protein complexes tend to have positive pi scores[30,31]. For the negative control and dCas9–KRAB screens, we did not observe any clear correlation between gene distance in the protein complex network and pi scores. In contrast, we did observe the expected effect in dCas9–KRAB–MeCP2 screens (Supplementary Fig. 14b,c).

Clustering of genetic interaction profiles provides a quantitative measure of functional similarity[32]. Among the samples, only dCas9–KRAB–MeCP2-containing cells showed a discernible clustering structure (Fig. 5, Supplementary Fig. 14d,e, and Supplementary Data 7). We subsequently looked at the gene pairs with the strongest interactions in the dCas9–KRAB–MeCP2 dataset. One of the most significant negative interactions was between *BLM* and *SOD1*, in line with previous data showing this to be a synthetically lethal interaction[33]. We also detected a negative genetic interaction between *BLM* and *DNA2*, consistent with results from yeast showing that the yeast *BLM* homolog *SGS1* can rescue *DNA2* deficiency, and that

the absence of both genes causes enhanced DNA-damage sensitivity[34,35]. Analyses of the positive genetic interactions revealed a strong interaction between *CHEK1* and *RECQL*. This result is consistent with previous knowledge that a loss of RECQL1 leads to the activation of CHEK1 signaling, which causes cell-cycle arrest. Thus, in cells that lack RECQL1, the growth arrest caused by CHEK1 activation should be alleviated after its removal, thus enabling the double mutants to grow better[36]. When we examined the same interactions in cells expressing dCas9–KRAB, we observed either no interaction (*BLM–DNA2*) or an interaction that was the opposite of what was expected (*BLM–SOD1* and *CHEK1–RECQL*).

Superiority of dCas9–KRAB–MeCP2 in synthetic gene circuits. We next performed five separate experiments highlighting the benefit of dCas9–KRAB–MeCP2 in the context of synthetic gene circuits. First, we constructed a simple repressor circuit in which *EYFP* was repressed by a U6-driven gRNA in combination with different dCas9 repressors. We observed approximately 400-fold repression with dCas9–KRAB–MeCP2, whereas earlier dCas9 variants led to <60-fold repression (Supplementary Fig. 15a).

Next, we added these U6-driven gRNAs to a two-layer repressor circuit. Here, a gRNA–dCas9 pair repressed the expression of a TALE repressor, which in turn repressed expression of *EYFP*. *EYFP* should be repressed in the absence of gRNA, but should be derepressed after the addition of gRNA to the circuit, depending on the strength of the dCas9 repressor. As expected, dCas9–KRAB–MeCP2 led to higher levels of derepression of *EYFP*, to the extent that it was indistinguishable from *EYFP* expressed in the absence of TALE repressors (Supplementary Fig. 15b).

Because inducible circuits are desirable in many synthetic gene networks, we reconstructed two previously described circuits in which gRNA expression was driven by a doxycycline-inducible RNA polymerase II (Pol II) promoter[37]. Although similar constructs have previously been shown to be functional, their activity has been inferior to that of constructs in which gRNA expression is under the control of Pol III promoters[37]. In the context of a simple repressor circuit, dCas9–KRAB–MeCP2 substantially increased the
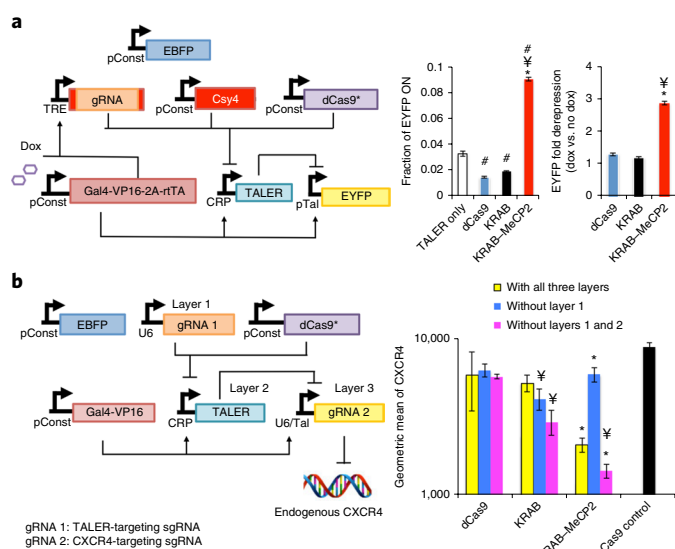
Fig. 6 | Superiority of dCas9–KRAB–MeCP2 in regulating complex synthetic circuits. a, Inducible Pol-II-expressed gRNA edited by Csy4 and in complex with dCas9–KRAB–MeCP2. $n = 4$ biologically independent samples (cell cultures). Dox, doxycycline. **b**, With the full circuit, dCas9–KRAB–MeCP2 decreased CXCR4 protein levels. In the absence of layer 1, dCas9–KRAB–MeCP2 mediated derepression of CXCR4. In the absence of layers 1 and 2, the repressor surpassed dCas9 and dCas9–KRAB in repressing CXCR4. $n = 3$ biologically independent samples (cell cultures). **a,b**, Data are presented as mean ± s.e.m. One-sided Student's t-test was applied for all statistical comparisons. #$P < 0.05$ versus TALER-only control; ¥$P < 0.05$ versus dCas9; *$P < 0.05$ versus dCas9–KRAB. In the circuit diagrams, dCas9* is meant to represent dCas9, dCas9–KRAB, or dCas9–KRAB–MeCP2.

efficiency of Pol-II-driven gRNA repression (Supplementary Fig. 15c). We next used it in a Pol-II-driven two-layer cascade. Unlike in our experiments with previous dCas9 tools, we observed, for the first time, a clear transfer of information with dCas9–KRAB–MeCP2, which showed the expected changes in *EYFP* expression between induced and uninduced states (Fig. 6a).

We next sought to determine whether dCas9–KRAB–MeCP2 could be used to create a functional three-layer cascade that interfaces with an endogenous gene. We constructed a circuit in which a U6-driven gRNA–dCas9 complex repressed a TALE repressor (layer 1). The TALE repressor suppressed another gRNA (layer 2), which targeted the endogenous *CXCR4* locus and mediated repression of this gene when combined with dCas9 repressors (layer 3). We tested the different repressors in circuits containing all three layers, layers 2 and 3, or layer 3 alone, and we measured the surface expression of CXCR4. Our results show that only dCas9–KRAB–MeCP2 facilitated the transfer of information in all settings, with CXCR4 levels reflecting the expected expression patterns (Fig. 6b).

Discussion

As in our work with transcriptional activators[20], the improved performance of our dCas9–KRAB–MeCP2 repressor is probably due to the distinct mechanisms by which each of the fused domains functions. The KRAB domain represses transcription via interaction with KAP1, which functions as a scaffold to recruit corepressors including heterochromatin protein 1 (HP1), histone deacetylases, and SETDB1[38–40]. The transcription-repression domain of MeCP2 binds to a different set of transcriptional regulators including the DNA methyltransferase DNMT1 and the SIN3A–histone deacetylase corepressor complex[41–44].

RNA-seq data suggest that dCas9–KRAB–MeCP2 does not induce additional gross differences in the cellular transcriptome aside from those already produced by current methods. It is worth pointing out that in our screen targeting essential genes, the dCas9–KRAB–MECP2-expressing cells exhibited much more robust depletion, but a few of the guides designed to target nonessential genes also showed marked depletion. Further investigations are needed to clarify the source of these effects. Although not used in these studies, various methods to improve Cas9 specificity have been reported in the literature, such as the use of a 'high-fidelity' Cas9 protein or truncated sgRNAs, each of which have been shown to help mitigate off-target activity[45,46].

For the majority of loci tested, dCas9–KRAB–MeCP2 achieved greater degrees of gene repression than dCas9–KRAB. Yet there were a few loci for which we observed only modest repression with either tool. Potential causes for variations in gene silencing include poorly functional guides, insufficient time between targeting and measurement of gene expression, local chromatin effects, competition for binding between Cas9 and endogenous transcriptional regulators, and interference from already present epigenetic marks that prevented further modification by our tool[47–49]. For the most part, researchers can overcome these inefficiencies in repression by simply targeting the same gene with an array of different sgRNAs. The utility of this strategy is demonstrated in our essential gene screen. For most essential genes tested, at least one of the targeting guides exhibited the expected levels of depletion, with the most robust effects observed in samples expressing dCas9–KRAB–MeCP2.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41592-018-0048-5.

References

1. Cho, S. W., Kim, S., Kim, J. M. & Kim, J.-S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat. Biotechnol.* **31**, 230–232 (2013).
2. Cong, L. et al. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
3. Mali, P. et al. RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
4. Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821 (2012).
5. DiCarlo, J. E. et al. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. *Nucleic Acids Res.* **41**, 4336–4343 (2013).
6. Hwang, W. Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. *Nat. Biotechnol.* **31**, 227–229 (2013).
7. Wang, H. et al. One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* **153**, 910–918 (2013).
8. Zhou, Y. et al. High-throughput screening of a CRISPR/Cas9 library for functional genomics in human cells. *Nature* **509**, 487–491 (2014).
9. Shalem, O. et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
10. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343**, 80–84 (2014).
11. Koike-Yusa, H., Li, Y., Tan, E.-P., del Castillo Velasco-Herrera, M. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).
12. Mandegar, M. A. et al. CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. *Cell Stem Cell* **18**, 541–553 (2016).
13. Qi, L. S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173–1183 (2013).
14. Gilbert, L. A. et al. CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442–451 (2013).
15. Gilbert, L. A. et al. Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* **159**, 647–661 (2014).
16. Thakore, P. I., Black, J. B., Hilton, I. B. & Gersbach, C. A. Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat. Methods* **13**, 127–137 (2016).

17. Konermann, S. et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).

18. La Russa, M. F. & Qi, L. S. The new state of the art: Cas9 for gene activation and repression. *Mol. Cell Biol.* **35**, 3800–3809 (2015).

19. Evers, B. et al. CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat. Biotechnol.* **34**, 631–633 (2016).

20. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).

21. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).

22. Zalatan, J. G. et al. Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).

23. Tanenbaum, M. E., Gilbert, L. A., Qi, L. S., Weissman, J. S. & Vale, R. D. A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* **159**, 635–646 (2014).

24. Shao, S. et al. Multiplexed sgRNA expression allows versatile single non-repetitive DNA labeling and endogenous gene regulation. *bioRxiv* Preprint at https://www.biorxiv.org/content/early/2017/03/29/121905 (2017).

25. Stolzenburg, S. et al. Stable oncogenic silencing in vivo by programmable and targeted de novo DNA methylation in breast cancer. *Oncogene* **34**, 5427–5435 (2015).

26. Li, F. et al. Chimeric DNA methyltransferases target DNA methylation to specific DNA sequences and repress expression of target genes. *Nucleic Acids Res.* **35**, 100–112 (2007).

27. Stepper, P. et al. Efficient targeted DNA methylation with chimeric dCas9-Dnmt3a-Dnmt3L methyltransferase. *Nucleic Acids Res.* **45**, 1703–1713 (2017).

28. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).

29. Shen, J. P. et al. Combinatorial CRISPR-Cas9 screens for *de novo* mapping of genetic interactions. *Nat. Methods* **14**, 573–576 (2017).

30. Dixon, S. J., Costanzo, M., Baryshnikova, A., Andrews, B. & Boone, C. Systematic mapping of genetic interaction networks. *Annu. Rev. Genet.* **43**, 601–625 (2009).

31. Menche, J. et al. Uncovering disease-disease relationships through the incomplete interactome. *Science* **347**, 1257601 (2015).

32. Costanzo, M. et al. A global genetic interaction network maps a wiring diagram of cellular function. *Science* **353**, aaf1420 (2016).

33. Sajesh, B. V. & McManus, K. J. Targeting SOD1 induces synthetic lethal killing in BLM- and CHEK2-deficient colorectal cancer cells. *Oncotarget* **6**, 27907–27922 (2015).

34. Imamura, O. & Campbell, J. L. The human Bloom syndrome gene suppresses the DNA replication and repair defects of yeast *dna2* mutants. *Proc. Natl Acad. Sci. USA* **100**, 8193–8198 (2003).

35. Budd, M. E. & Campbell, J. L. The pattern of sensitivity of yeast *dna2* mutants to DNA damaging agents suggests a role in DSB and postreplication repair pathways. *Mutat. Res.* **459**, 173–186 (2000).

36. Popuri, V., Croteau, D. L., Brosh, R. M. Jr. & Bohr, V. A. RECQ1 is required for cellular resistance to replication stress and catalyzes strand exchange on stalled replication fork structures. *Cell Cycle* **11**, 4252–4265 (2012).

37. Kiani, S. et al. CRISPR transcriptional repression devices and layered circuits in mammalian cells. *Nat. Methods* **11**, 723–726 (2014).

38. Friedman, J. R. et al. KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev.* **10**, 2067–2078 (1996).

39. Kim, S. S. et al. A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins. *Proc. Natl. Acad. Sci. USA* **93**, 15299–15304 (1996).

40. Moosmann, P., Georgiev, O., Le Douarin, B., Bourquin, J. P. & Schaffner, W. Transcriptional repression by RING finger protein TIF1 beta that interacts with the KRAB repressor domain of KOX1. *Nucleic Acids Res.* **24**, 4859–4867 (1996).

41. Jones, P. L. et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19**, 187–191 (1998).

42. Nan, X. et al. Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393**, 386–389 (1998).

43. Wade, P. A. et al. Histone deacetylase directs the dominant silencing of transcription in chromatin: association with MeCP2 and the Mi-2 chromodomain SWI/SNF ATPase. *Cold Spring Harb. Symp. Quant. Biol.* **63**, 435–445 (1998).

44. Kimura, H. & Shiota, K. Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *J. Biol. Chem.* **278**, 4806–4812 (2003).

45. Kleinstiver, B. P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490–495 (2016).

46. 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).

47. Polstein, L. R. et al. Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. *Genome Res.* **25**, 1158–1169 (2015).

48. Wu, X. et al. Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. *Nat. Biotechnol.* **32**, 670–676 (2014).

49. Kuscu, C., Arslan, S., Singh, R., Thorpe, J. & Adli, M. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat. Biotechnol.* **32**, 677–683 (2014).

Author contributions

Y.C., D. Menn, D. Milanova, and C-.C.K. contributed equally to the manuscript. A.C. and N.C.Y. conceived the study. N.C.Y. and A.C. designed experiments. A.C. and A.T. designed, built, and tested the initial set of dCas9 repressor fusions. N.C.Y. performed the majority of endogenous gene targeting experiments with assistance from A.L.-B. and additional technical contributions from R.J.C., M.T., and A.C. Library screens and next-generation sequencing were performed by N.C.Y. Analysis of the essential gene library data was performed by Y.C. with contributions from N.C.Y. and A.C. S.S. performed MAGeCK analysis. RNA-seq analysis was led by D. Milanova with assistance in library preparation from N.C.Y. and data interpretation by N.C.Y. and A.C. C-.C.K. analyzed the dual-guide epistasis experiment with oversight from N.E.L. and interpreted data with assistance from N.C.Y. and A.C. X.G. aided in next-generation sequencing and performed preliminary analysis of the sequencing data. N.D. provided technical experience and insight. E.T.L. helped analyze a portion of the library screening data. D. Menn and S.P. performed the synthetic circuit experiments under the guidance of M.R.E. and S.K. S.K. designed all synthetic circuits used in the study. Research was performed in the laboratory of G.M.C. with oversight from both J.J.C. and G.M.C. N.C.Y. and A.C. wrote the manuscript, with contributions from all other authors.

Competing interests

G.M.C. is a founder of and advisor for Editas Medicine. G.M.C. has equity in Editas and Caribou Biosciences (for the full disclosure list, please see http://arep.med.harvard.edu/gmc/tech.html).

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-018-0048-5.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to A.C. or S.K. or G.M.C.

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

Methods

Repressor and gRNA plasmid construction. Repressor fusions were initially cloned into a modified Gateway-compatible dCas9 plasmid backbone[50]. The bipartite and tripartite dCas9 fusions were cloned in a modified Golden Gate–compatible version of the dCas9-m4 vector (Addgene plasmid #47316). DNA fragments containing the specific domains of interest were then PCR-amplified and cloned into each of our vectors via either Gateway or Golden Gate assembly methods. For bipartite and tripartite repressors, a glycine-serine-rich linker was placed between the different domains. The sequences, as well as species origin, of all protein domains used to construct the different repressors are listed in Supplementary Data 8. The sequences of dCas9–KRAB (Addgene plasmid #110820) and dCas9–KRAB–MeCP2 (Addgene plasmid #110821) are provided in Supplementary Table 1. All other vectors are available from the corresponding author upon request.

All gRNAs for endogenous gene repression were selected to bind within −50 to +200 bp around the gene TSS, unless a different position was specified. Target genes were selected on the basis of their use in previous studies or because they were of particular interest to our research group, such as DNA-repair and cell-motility genes[14,15]. To generate sgRNA expression plasmids, we cloned oligonucleotides containing gRNA sequence into a pSB700 vector (Addgene plasmid #64046) or variants with different selection markers downstream of a U6 promoter via Golden Gate assembly methods. Sequences for gRNAs are listed in Supplementary Table 9.

Cell culture and transfections. HEK293T cells (gift from P. Mali, University of California, San Diego) were maintained in DMEM (Life Technologies) with 10% heat-inactivated FBS (Life Technologies) and penicillin–streptomycin (Life Technologies) as previously described[50]. Approximately 50,000 cells were seeded per well in 24-well plates, and the next day they were transfected using Lipofectamine 2000 (Life Technologies) as previously described[50]. 200 ng of dCas9 repressors, 50 ng of sgRNA, and 60 ng of EYFP reporter along with 50 ng of Gal4–VP16 (reporter assay only) were delivered to each well. We cotransfected 50 ng of puromycin-resistant plasmids (endogenous gene study) or 25 ng of EBFP-expressing plasmids (reporter assay) to select for transfected cells. 10 ng of each sgRNA per gene was used during multiplex repression. For the endogenous gene study, cells were treated with 3 μg/ml puromycin at 24 h post-transfection to enrich for transfected cells. 48 or 72 h after transfection, cells were collected for assay by flow cytometry or lysed for RNA purification, for reporter and endogenous experiments, respectively. Cells were tested every 3 months for mycoplasma contamination and consistently tested negative.

Flow cytometry for reporter assays. For reporter assays we targeted dCas9 fusion proteins to a Gal4–VP16-regulated EYFP reporter gene. The reporter plasmid contained an sgRNA-binding sequence (tacctcatcaggaacatgt) followed by a protospacer-adjacent motif (PAM; tgg). HEK293T cells were transfected with the reporter, Gal4–VP16 activator, sgRNA, and the indicated dCas9 fusion proteins, along with an EBFP-expressing plasmid to aide in analysis of only cells that were transfected. Cells were assayed by flow cytometry 48 h after transfection. We analyzed cells expressing >10³ arbitrary units of EBFP2 and quantified the median EYFP intensity within the gated population by using FlowJo.

Quantitative real-time polymerase chain reaction (PCR) to analyze endogenous gene expression. Total RNA was extracted with the RNeasy Plus mini kit (Qiagen). 500 ng of RNA was used to make cDNA using the qScript cDNA synthesis kit (Quanta Bio). KAPA SYBR Fast universal 2× quantitative PCR master mix (KAPA Biosystems) with 0.5 μl of cDNA and 0.4 μl each of forward and reverse primers at 10 μM were used for qPCR, with the following cycling conditions: 95 °C for 3 min, and 40 cycles of 95 °C for 10 s, 55 °C for 20 s, and 72 °C for 30 s. RNA expression was normalized to that of the housekeeping gene *ACTB*, and relative gene expression was calculated via the $2^{-\Delta\Delta C_t}$ method[51]. Sequences for qPCR primers are listed in Supplementary Table 10.

Statistical analysis. For studies targeting reporter and endogenous genes, at least two biologically independent samples (independent transfections) per group were used. Statistical comparisons were carried out in experiments using sample sizes (n) of 3 or 4 biologically independent samples, using one-tailed Student's t-test with a P value < 0.05 as the threshold for significance. The exact n values used to calculate statistics are described in the associated figure legends. Statistical analyses for RNA-seq, gRNA library screening, and circuit experiments are described in Supplementary Notes 2–5.

Whole-transcriptome RNA sequencing for analysis of repressor specificity. For each sample, total RNA was extracted with an RNeasy mini kit (Qiagen) and treated with on-column RNase-free DNase I (Qiagen) according to the manufacturer's instructions. 1 μg of RNA from each sample was used for library preparation. RNA-seq libraries were constructed with the TruSeq Stranded Total RNA library prep kit with Ribo-Zero Gold (Illumina) designed for cytoplasmic and mitochondrial rRNA depletion. All coding RNA and certain forms of noncoding RNA were isolated by bead-based rRNA depletion followed by cDNA synthesis and PCR amplification as per the manufacturer's protocol. Final libraries were analyzed on TapeStation,

quantified by qPCR, pooled, and run on one lane of an Illumina HiSeq 2500 using 2 × 100-bp paired-end reads. The Illumina paired-end adaptor sequences were removed from the raw reads with Cutadapt v1.8.1. The TruSeq adaptor sequence 5′- AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC-3′ was used for read 1, and its reverse complement, 3′-AGATCGGAAGAGCGTCGTGTAGGGAAAGAG TGTAGATCTCGGTGGTCGCCGTATCATT-5′, was used for read 2.

Next, RNA libraries were processed via a pipeline that includes STAR-HtSeq-EdgeR for alignment, count generation, and gene expression. Briefly, STAR aligner (v. 2.4.0j) was used to map the reads to hg19, and HtSeq was used to generate gene expression counts. For gene expression and differential expression analyses, edgeR, limma, and custom R scripts were used to filter out genes with very low expression (with a cutoff of one count in at least two samples), calculate normalization factors, and compute effective library sizes using TMM (trimmed mean of M values) normalization. The gene count was then reported in counts per million, and correlations were calculated on $\log_2$-transformed data. Finally, to determine the most biologically significant differentially expressed genes, we assessed relative gene expression by fold-change thresholding ($\log_2$ FC > 1.5) and ranking by P value. Supplementary Note 2 presents details on differential expression analysis. A small set of genes in addition to the target gene *CXCR4* showed decreases ($\log_2$ FC < 1.5) in their transcript expression (Supplementary Table 3). We analyzed these genes further to assess whether the observed differential expression was caused by nonspecific binding of our gRNA. We examined genomic sequences of regions 2 kb upstream or downstream from the TSS of those genes by searching for the presence of a full-length gRNA binding site (up to six mismatches for near matches) and for the seed region of the gRNA alone (8 bp proximal to PAM).

Cell culture and generation of repressor-expressing stable cell lines. HAP1 cells (Horizon Discovery) were maintained in Iscove's modified Dulbecco's medium with 10% FBS (Life Technologies) and penicillin–streptomycin (Life Technologies) according to the manufacturer's instructions. SH-SY5Y (ATCC) was maintained in a 1:1 mixture of Eagle's minimum essential medium and F12 medium (ATCC) with 10% FBS and penicillin–streptomycin according to the manufacturer's instructions. To generate stable dCas9-repressor-expressing cell lines, we transfected approximately 30,000–50,000 cells in 24-well plates with 400 ng of dCas9-repressor-containing PiggyBac expression plasmids (Addgene plasmids #110822, #110823, #110824) and 100 ng of transposase vector using Lipofectamine 3000 (Life Technologies) as previously described[50]. Media was changed after 24 h. Cells were allowed to recover for 2 d and then were treated with 5 μg/ml blasticidin. Cells were passaged regularly in drug media for more than 2 weeks to select for heterogeneous populations of dCas9 repressor integrant-containing cells.

Production of single-gene-targeting gRNA lentivirus and cell transduction. HEK293T cells were seeded at 200,000 cells per well in six-well plates 1 d before transfection. Cells were transfected with 450 ng of pSB700 sgRNA expression plasmid (with puromycin-resistant marker), 600 ng of psPAX2 (a gift from Didier Trono; Addgene plasmid # 12260), and 150 ng of pCMV-VSV-G (a gift from Bob Weinberg; Addgene plasmid # 8454) using Lipofectamine 2000 (Life Technologies). We collected viral supernatants 72 h after transfection by centrifuging the medium at 400*g* for 5 min to remove cell debris. HAP1 and SH-SY5Y repressor stable cell lines were seeded at ~15,000 and 35,000 cells, respectively, per well in 24-well plates. The following day each sample was infected with 100 μl of sgRNA-containing lentiviruses. We treated cells with 0.5 μg/ml (HAP1) or 2.5 μg/ml (SH-SY5Y) puromycin to select for transductants at 48 h after transduction. Cells stably expressing sgRNA were passaged for 2 weeks and collected for RNA extraction and qPCR analysis. Sequences for qPCR primers are listed in Supplementary Table 10.

Production of lentiviral single-guide and dual-guide RNA libraries. The plasmid containing an sgRNA library targeting essential genes was a gift from Dr. Rene Bernards[19]. To generate the dual-guide library, we designed a series of oligonucleotides such that the forward oligo created the first gRNA in the array and the reverse oligo was used to introduce the second gRNA into the array (a list of oligos used for library construction is provided in Supplementary Table 8). A template containing a modified sgRNA tail sequence fused in cis to the 7SK Pol III promoter was then used as a PCR template (the sequence of the sgRNA2-7SK template is given in Supplementary Table 11). To generate the dual-guide library, we carried out a PCR reaction in which all forward and reverse primers were mixed together. The resulting ~475-bp PCR product was run on a gel, extracted, and inserted into the pSB700 gRNA expression backbone by Golden Gate cloning. To produce lentiviruses expressing each gRNA library, we plated approximately 1 million HEK293T cells on a 10-cm dish. The following day cells were transfected with each of the gRNA library plasmids mixed with psPAX2 and pCMV-VSV-G at a 4:3:2 ratio using a total of 7–8 μg of DNA via the following protocol. Total plasmid DNA was diluted in 1 ml of serum-free media. Polyethylenimine (PEI; Polysciences) was added to the diluted DNA based on a 3:1 ratio of PEI (μg):total DNA (μg). Mixtures were incubated for 15 min at room temperature and then added to the cells. Viral supernatants were collected at 72 h and concentrated with a PEG virus precipitation kit (BioVision) according to the manufacturer's instructions.

CRISPR repressor screens. To compare the ability of different repressors in screening, we seeded a series of stable cell lines, each containing a unique repressor, along with a control cell line without a repressor integrated into the genome in six-well plates and allowed the cells to grow to 30–50% confluency in preparation for transduction the following day. Lentiviruses expressing each gRNA library were produced and used to infect experimental cells so the multiplicity of infection was <0.5. Cells were treated with 0.5 µg/ml (HAP1, HEK293T) or 2.5 µg/ml (SH-SY5Y) puromycin at 48 h (HAP1) or 72 h (SY-SH5Y, HEK293T) after virus transduction. After drug selection, 50% of the cells were collected immediately for DNA extraction with Epicentre Quick Extract Solution, and 50% of the cells were seeded into a set of 15-cm dishes for subsequent passaging. Cells were regularly passaged via standard protocols and collected again at 7, 14, and 22 d (SH-SY5Y screen only) after drug selection for DNA extraction. The number of cells manipulated was kept sufficiently large to maintain ~500–1,000-fold coverage of the library at each stage of passaging. For PCR, 25 µg (lethality screen) or 60 µg (gene epistasis screen) of genomic DNA divided over 25 or 60 reactions, respectively, was amplified with the KAPA2G robust PCR kit (KAPA Biosystems) along with primer set PCR 1 (Supplementary Table 12). The products of all first-round PCR reactions from the same sample were then pooled. We used 1 µl of the pooled product for sample indexing in preparation for next-generation sequencing using either Illumina TruSeq or Nextera indexing primers. PCR cycling conditions are listed in Supplementary Table 13. Supplementary Notes 3 and 4 describe bioinformatics analyses of screen data.

Circuit experiments. HEK293FT cells were transfected as previously described[37] with PEI reagents. For inducible circuits, 2 µg/µl doxycycline was added to samples and changed daily after transfection until flow cytometry. All samples were processed for flow cytometry at 72 h post-transfection, and data were analyzed by FlowJo. Supplementary Note 5 provides detailed methods and materials used for circuit experiments.

Software. FlowJo (version 7) was used to analyze data generated from flow cytometry experiments. MAGeCK (0.5.7) was used to analyze single gRNA library screens to determine gene essentiality.

Reporting Summary. Further information on experimental design is available in the Nature Research Reporting Summary linked to this article.

References

50. Chavez, A. et al. Comparison of Cas9 activators in multiple species. *Nat. Methods* **13**, 563–567 (2016).
51. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C(T)}$ method. *Methods* **25**, 402–408 (2001).

nature research

Corresponding author(s): Drs. Alejandro Chavez, Samira Kiani, and George M. Church

☐ 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.

No methods were used to predetermine sample size. Sample sizes were chosen due to being able to show reproducibility and statistical significance.

2. Data exclusions

Describe any data exclusions.

No data were excluded from the analyses.

3. Replication

Describe whether the experimental findings were reliably reproduced.

Yes, all attempts at replication were successful. Methods and materials used in our experiments were described in the manuscripts to allow reliable replication of our studies.

4. Randomization

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

No randomization was used for samples as samples with particular genetic constituents were needed for the experiments.

5. Blinding

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

Blinding was not relevant to the studies as samples with particular genetic constituents were needed for the experiments. Labeling of samples was used to prevent mixed up of experimental samples.

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.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

> FlowJo (version 7) was used to analyze data generated from flow-cytometry experiments. MAGeCK (0.5.7) was used to analyze single gRNA library screens to determine gene essentiality. All custom scripts are available upon request.

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.

> Unique plasmids such as dCas9-KRAB and dCas9-KRAB-MeCP2 expression plasmids are deposited with Addgene (plasmid # 110820-110824) and are described in the Methods section. All other vectors are available upon request.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

> No antibodies were used.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

> HEK293T cells were gift from P. Mali, University of California, San Diego. HAP1 cells were purchased from commercial company Horizon Discovery. SH-SY5Y cells were purchased from commercial company ATCC.

b. Describe the method of cell line authentication used.

> The method of cell line authentication was defined by commercial company from which the cell line was purchased. Horizon Discovery validated all HAP1 cell lines by PCR amplification and Sanger Sequencing to confirm the mutation at the genomic level. ATCC uses methods including an assay to detect species specific variants of the cytochrome C oxidase I gene (COI analysis) to rule out inter-species contamination and short tandem repeat (STR) profiling to distinguish between individual human cell lines and rule out intra-species contamination.

c. Report whether the cell lines were tested for mycoplasma contamination.

> Yes, cells were tested every 3 months for mycoplasma contamination and consistently tested negative.

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 commonly misidentified cell lines were used.

▶ 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.

> No animals were used in our studies.

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.

> No human research participants were used in our studies

nature research

Corresponding author(s): Drs. Alejandro Chavez, Samira Kiani, and George M. Church

☐ 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.	Cell cultures were treated with trypsin and diluted in complete media or PBS for flow cytometry experiments.
6. Identify the instrument used for data collection.	BD LSRFortessa™ was used for data collection.
7. Describe the software used to collect and analyze the flow cytometry data.	All cytometry data were analyzed by FlowJo (version 7).
8. Describe the abundance of the relevant cell populations within post-sort fractions.	No cell sorting was performed.
9. Describe the gating strategy used.	>80% of viable and intact cells were gated from FSC/SSC for analysis. Within the population, >50% were transfected cells that were selected for downstream analysis by gating cells expressing > 10^3 arbitrary units of EBFP2 (transfection marker).

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