LETTERS
https://doi.org/10.1038/s41587-019-0296-7
nature biotechnology

Dose-dependent activation of gene expression is achieved using CRISPR and small molecules that recruit endogenous chromatin machinery

Anna M. Chiarella[1], Kyle V. Butler[2], Berkley E. Gryder[3], Dongbo Lu[1], Tiffany A. Wang[1], Xufen Yu[2], Silvia Pomella[3,4], Javed Khan[3], Jian Jin[2*] and Nathaniel A. Hathaway[1*]

Gene expression can be activated or suppressed using CRISPR–Cas9 systems. However, tools that enable dose-dependent activation of gene expression without the use of exogenous transcription regulatory proteins are lacking. Here we describe chemical epigenetic modifiers (CEMs) designed to activate the expression of target genes by recruiting components of the endogenous chromatin-activating machinery, eliminating the need for exogenous transcriptional activators. The system has two parts: catalytically inactive Cas9 (dCas9) in complex with FK506-binding protein (FKBP) and a CEM consisting of FK506 linked to a molecule that interacts with cellular epigenetic machinery. We show that CEMs upregulate gene expression at target endogenous loci up to 20-fold or more depending on the gene. We also demonstrate dose-dependent control of transcriptional activation, function across multiple diverse genes, reversibility of CEM activity and specificity of our best-in-class CEM across the genome.

The eukaryotic genome is organized and packaged into chromatin with varying degrees of compaction, which contributes to the regulation of gene expression. A network of protein–protein and protein–DNA interactions regulates the proper levels of gene expression. Disruptions to this regulatory network drive many human diseases, including cancer[1,2]. An important contributing factor that sculpts the chromatin landscape is post-translational histone tail modification. Lysine acetylation is one such modification that has both biophysical and indirect protein-recruitment effects. Protein families of writers (histone acetyltransferases (HATs)), erasers (histone deacetylases (HDACs)) and readers (for example, bromodomains and chromodomains) intricately control gene expression[3,4].

Several groups have demonstrated the power of recruiting exogenous chromatin-modifying machinery as a way to control expression levels in a gene-specific manner[5–11]. With major advances in Cas9 and dCas9 technology, the ability to precisely induce changes in expression has rapidly evolved. Pioneering work by Liszczak and colleagues demonstrated the ability to recruit endogenous machinery to a reporter locus using a dCas9 system combined with conjugated inhibitors of chromatin regulatory proteins[12]. Other work by Ansari and colleagues used programmable DNA-binding ligands coupled with bromodomain inhibitors to modulate transcription[13]. Inspired by these studies, we sought to develop a system capable of modulating gene expression of endogenous mammalian genes in a specific, dose-dependent manner using chemical entities.

We previously demonstrated the ability of CEMs to modify chromatin and subsequently repress gene expression at engineered reporter loci[14]. In this study, we report CEM-activating (CEMa) molecules that recruit endogenous gene-activating machinery. Our CEMa family includes CEM87, CEM88 and CEM114, which each bind to different chromatin-modifying enzymes from previously published bromodomain inhibitors of HATs or acetylated lysine reader proteins (Fig. 1a). CEM87 was created with iBet762, shown to bind BRD2, BRD3 and BRD4 (ref. [15]; Fig. 1b and Supplementary Note). CEM88 was created with a 1,3-dimethyl benzimidazolone, previously shown to bind the BRPF1 bromodomain (Fig. 1b and Supplementary Note)[16]. Lastly, CEM114 was created with 'compound 33', previously shown to bind CBP (Fig. 1b and Supplementary Note)[17]. Here we show that our CEMa family is compatible with dCas9–FKBP-based systems, allowing us to direct CEMa activity to any gene.

To test for changes in gene expression, we transfected HEK293T cells with a green fluorescent protein (GFP) reporter gene downstream of the TRE3G promoter and performed flow cytometry on cells coexpressing the guide RNA (gRNA) and dCas9 machinery[9]. The same targeting gRNA was used throughout our experiments with the TRE3G-GFP reporter and comprised a sequence that binds six interspaced sites within the TRE3G promoter. As a benchmark control for gene activation, we used a dCas9 protein (from *Streptococcus pyogenes*) fused to p300, TET or VP64–p65–Rta (VPR), all previously shown to increase expression (Supplementary Fig. 1a)[11,18,19]. We then tested activating the reporter gene with a plasmid expressing dCas9–FKBP×1 or dCas9–FKBP×2 and one of our three predicted CEMa molecules. After 48 h of exposure to 200 nM of the indicated CEM, normalized GFP reporter expression was significantly increased in all cases as compared to untreated cells (Fig. 1c). We next tested another commonly used dCas9 system adapted from *Staphylococcus aureus*. Cells expressing dCas9–FKBP×1 or dCas9–FKBP×2 with *S. aureus* dCas9 were treated with each CEM. As with dCas9 from *S. pyogenes*, cells showed increased GFP expression, demonstrating that the CEM technology is adaptable to multiple species of dCas9 (Fig. 1d). We performed further experiments with dCas9 from *S. pyogenes*. To confirm that the CEMa system was activating GFP in a controlled, FKBP-dependent

[1]Division of Chemical Biology and Medicinal Chemistry, Center for Integrative Chemical Biology and Drug Discovery, UNC Eshelman School of Pharmacy, Chapel Hill, NC, USA. [2]Mount Sinai Center for Therapeutics Discovery, Departments of Pharmacological Sciences and Oncological Sciences, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [3]Oncogenomics Section, Genetics Branch, National Cancer Institute, National Institutes of Health (NIH), Bethesda, MD, USA. [4]Department of Pediatric Hematology and Oncology, Bambino Gesu' Children's Hospital, Rome, Italy. *e-mail: jian.jin@mssm.edu; hathaway@unc.edu

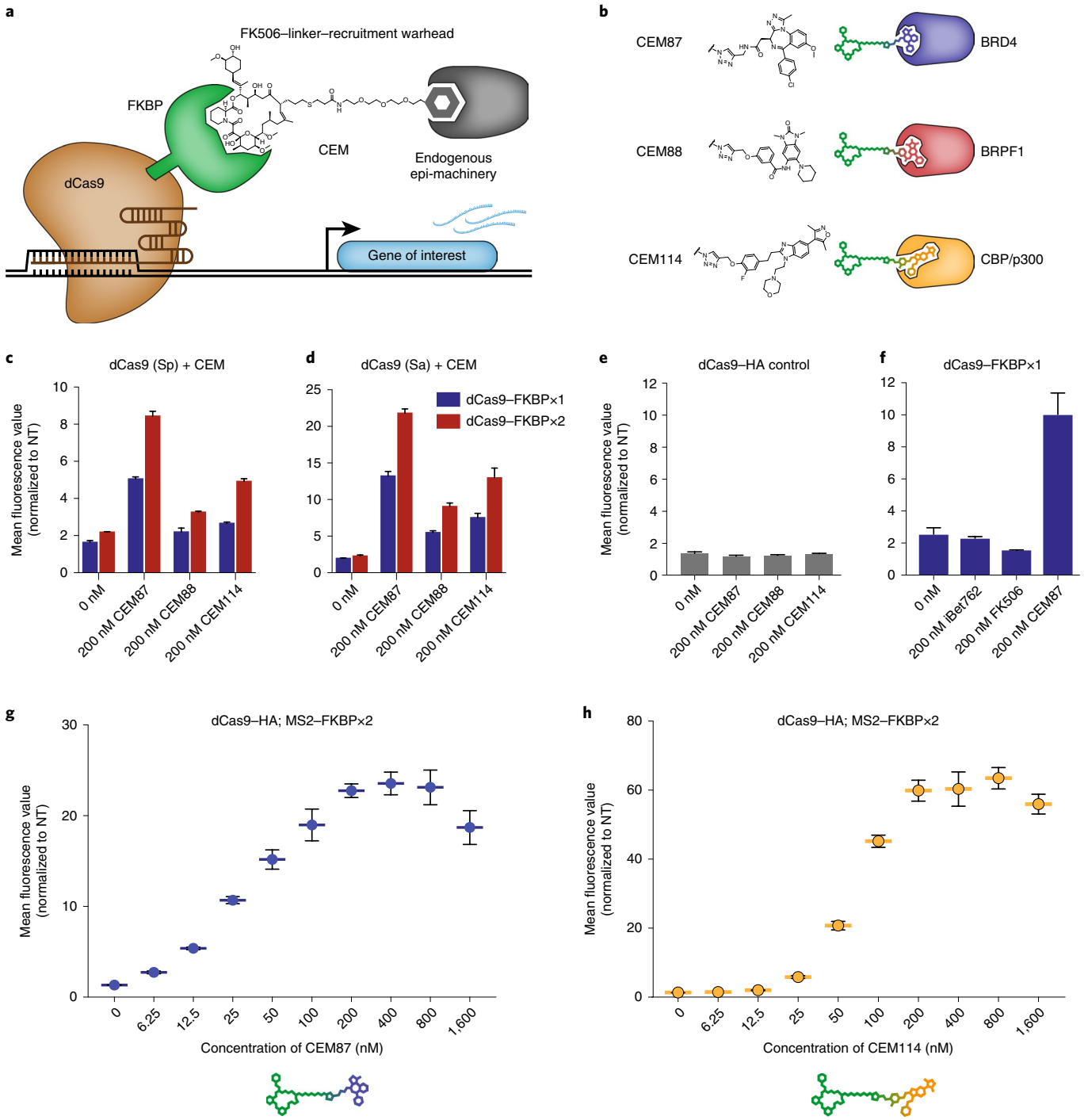

Fig. 1 | Using CEMs to increase gene expression. a, A dCas9–FKBP fusion protein is used to target the CEMs to our gene of interest for activation. **b**, CEM87, CEM88 and CEM114 are predicted to bind and recruit BRD4, BRPF1 and CBP/p300, respectively. **c**, dCas9 (from *S. pyogenes*) fused to one (blue bars) or two (red bars) copies of FKBP was recruited to a GFP reporter, and flow cytometry was performed. CEM87, CEM88 and CEM114 significantly increased expression (as quantified by mean fluorescence value) as compared to untreated cells. **d**, dCas9 (from *S. aureus*) fused to one (blue bars) or two (red bars) copies of FKBP was recruited to a GFP reporter, and flow cytometry was performed. CEM87, CEM88 and CEM114 significantly increased expression as compared to untreated cells. **e**, When dCas9 was fused to an HA tag instead of FKBP, CEM87, CEM88 and CEM114 did not significantly change expression as compared to untreated cells. **f**, Treatment with the individual recruitment components of CEM87 (iBet762 and FK506) did not significantly change gene expression; activation was observed only with the bifunctional CEM87. **g**, A dose curve of CEM87 was performed with CEM concentration ranging from 6.25 to 1,600 nM. Mean fluorescence value was normalized to that in cells with NT gRNA. In comparison to untreated cells, expression was increased after treatment with 6.25 nM CEM87 for 48 h. **h**, A dose curve of CEM87 was performed with CEM concentration ranging from 6.25 to 1,600 nM. Mean fluorescence value was normalized to that in cells with NT gRNA. In comparison to untreated cells, expression was increased after treatment with 6.25 nM CEM114 for 48 h. In **c**–**h**, significance was determined with three different samples using a two-tailed Student's *t*-test; fold change and significance are reported in Supplementary Table 1. Error bars, s.d.

manner, we analyzed cells expressing dCas9 alone treated with 200 nM CEMa for 48 h. As predicted, CEM treatment did not significantly change GFP expression (Fig. 1e). We also sought to validate that the activation was a result of the CEMa molecule as a whole, rather than any one component of the molecule. To test this, we expressed dCas9–FKBP×2 in cells and treated the cells with 200 nM iBet762 (the inhibitor from which CEM87 was synthesized), FK506 (the FKBP-binding moiety) or CEM87. CEM87 treatment was the only condition that increased GFP expression (Fig. 1f).

Optimization of the dCas9–CEMa system was performed by incorporating several dCas9-related systems that we adapted from the work of others, specifically the MS2 system and the dCas9–SunTag system[7,20]. The MS2 gRNAs have a modified stem loop, capable of recruiting both a dCas9 fusion and a bacteriophage MS2 coat protein (MCP) fusion[7,21]. By using an MCP–FKBP fusion, we were able to increase the number of CEMs recruited to the chromatin. The MS2(N55K) protein used, referred to hereafter as MS2, was a previously published MCP mutant that exhibits strong binding of the coat protein to the stem loops[22]. To optimize our CEMa gene activation with the MS2 system, we transfected cells to express dCas9 alone or in combination with targeting gRNA and either MS2–FKBP×1 or MS2–FKBP×2. Of these approaches, the one including MS2–FKBP×2 was the most effective (Supplementary Fig. 1b).

We next adapted the dCas9–SunTag system, in which an array of 10 or 24 yeast-specific gene control protein 4 (GCN4) peptides is fused to the C terminus of dCas9 (refs. [20,23]). We cotransfected cells to express a single-chain variable fragment (scFv), made to be specific for GCN4, fused to one copy of FKBP. Our results showed that the dCas9–SunTag×24 fusion was more effective (Supplementary Fig. 1c). Then, we tested dCas9–SunTag×24 with scFv–FKBP×1 or scFv–FKBP×2. Cells were again treated with 200 nM of each CEM. With most of the CEMs, scFv–FKBP×2 provided a slight advantage over scFv–FKBP×1 (Supplementary Fig. 1d). Using a combinatorial approach to the SunTag and MS2 systems, we created and tested fusions of MCP to the SunTag peptide, in an attempt to recruit additional scFv–FKBP fusions closer to the targeted chromatin (Supplementary Fig. 1e).

To directly compare these optimized dCas9 systems, we transfected cells with equal amounts of the gRNA and plasmids encoding the dCas9 fusion, scFv fusion, MS2 fusion or 'filler' DNA. Cells were treated with 200 nM CEM87, CEM88 or CEM114 for 48 h, and mean fluorescence values were normalized to those of untreated cells with the same transfection mixture (Supplementary Fig. 2a). The same transfection was conducted on cells using non-targeting (NT) gRNA. As expected, the GFP signal was consistently lower in control cells with NT gRNA than in cells with targeting gRNA (Supplementary Fig. 2b). Together, these data suggest that recruiting FKBP fusions more strategically, rather than simply increasing the number of FKBP molecules, is important for improving the dCas9–CEMa system. With this in mind, we chose to proceed with dCas9 alone, bringing in the CEM binding site only through MS2–FKBP×2.

With the optimized dCas9 system, we conducted a time course with the two most effective CEMa molecules: CEM87 and CEM114. We transfected cells with plasmids expressing dCas9, MS2–FKBP×2 and GFP (from the TRE3G promoter) together with either targeting or NT gRNA. Cells were treated with 200 nM CEM87 and expression was then assessed by flow cytometry. The data were normalized to those for cells with NT gRNA. In comparison to untreated cells, treatment of cells with both CEM87 and CEM114 yielded high GFP expression after 24 and 48 h (Supplementary Fig. 2c,d). We next performed a dose curve with the same CEMa molecules, measuring expression at 48 h. CEM87 and CEM114 treatment significantly increased GFP expression in direct response to CEM addition up to a concentration of ~200 nM (Fig. 1g,h). Dose curves from CEM87 and CEM114 treatment showed a hook effect near treatments of 400–800 nM, where the CEMa treatment became less effective. This could be a result of active inhibition of the desired chromatin regulator machinery meant for recruitment. Alternatively, there could be toxicity at high concentrations (morphology indicating potential cytotoxicity was observed at concentrations above 400 nM; data not shown). Thus, the dCas9–CEMa platform can be used to exert dose-dependent control over gene activity by varying the compound dose between 6.25 and 200 nM. This could be useful for target validation studies, which need new tools to improve clinical success rates[24,25].

To confirm that CEMa-mediated activation of GFP is a result of CEMa binding directly to the MS2–FKBP×2 protein fusions through interaction with an FK506-binding site, we tested whether excess FK506 could outcompete CEMa binding, thereby diminishing activation. Cells were treated with 2,000 nM FK506, 200 nM CEM114 or a combination of 2,000 nM FK506 and 200 nM CEM114 for 48 h. After flow cytometry, data were normalized to those for cells given the same treatment but transfected with NT gRNA. As expected, only the treatment with 200 nM CEM114 significantly activated the GFP signal, suggesting that excess FK506 is able to outcompete CEM114 from the FKBP binding site (Supplementary Fig. 2e). We next tested whether a combination of CEMa molecules could yield a stronger level of gene activation. We treated cells for 48 h, performed flow cytometry and normalized the results to those for cells with NT gRNA. GFP expression was significantly increased in cells treated with 66.6 or 200 nM each of CEM87, CEM88 and CEM114, but neither condition increased expression to the extent seen with CEM114 alone (Supplementary Fig. 2f). These data suggest that CEM114 yields the most effective activation in this particular assay.

To determine how stable CEM87 and CEM114 activation is without constant CEMa treatment, we exchanged the CEM-containing medium with untreated medium or with medium containing excess FK506 (2,000 nM). As a baseline, cells were treated with 200 nM CEM87 or CEM114 for 48 h. Flow cytometry was performed on a subset of the cells, and the data were normalized to those for untreated cells. The remaining cells were then re-seeded with continued CEMa treatment or received medium with excess FK506 or untreated regular growth medium. Initially, cells treated with CEM87 or CEM114 for 48 h had significantly increased GFP expression. After 2 d of continued treatment, no treatment or exposure to excess FK506, the cells under each condition continued to exhibit significantly activated GFP expression. After 4 d, the cells under all conditions continued to have significant activation, except for cells treated with CEM87 for 2 d and then exposed to excess FK506 for 4 d. After 6 d of excess FK506 in cells treated with CEM87 or CEM114 and after 6 d of no treatment in cells treated with CEM87, the increase in expression was no longer significant (Supplementary Fig. 2g). These data suggest that the dCas9–CEMa system is reversible.

To target endogenous genes, we adapted our optimized dCas9–CEMa system to be delivered via lentiviral infection of the dCas9 machinery (dCas9 and MS2–FKBP) into HEK293T cells. After the cells had been stably selected for integration of the dCas9 and FKBP constructs, we transfected them with gRNAs for the desired gene target. As a positive control, we transfected cells expressing dCas9 fused to activating domains with gRNAs targeted to the myogenic differentiation 1 (*MYOD1*) gene, extracted the RNA and performed quantitative RT–PCR (qRT–PCR) (Supplementary Fig. 3a). *MYOD1* was an ideal initial target for this question because it was previously shown to be capable of modulation by transiently expressed dCas9–p300 (ref. [11]). To determine the optimal time at which to observe activation, we transfected cells expressing dCas9 and MS2–FKBP×2 with gRNA and conducted RNA extraction at 24, 48, 72, 96 and 120 h after treatment. At 48 h after treatment, cells with *MYOD1*-targeting gRNA had significantly increased expression with 200 nM CEM87 (22.7-fold increase; $P < 0.01$). CEM87 yielded the greatest

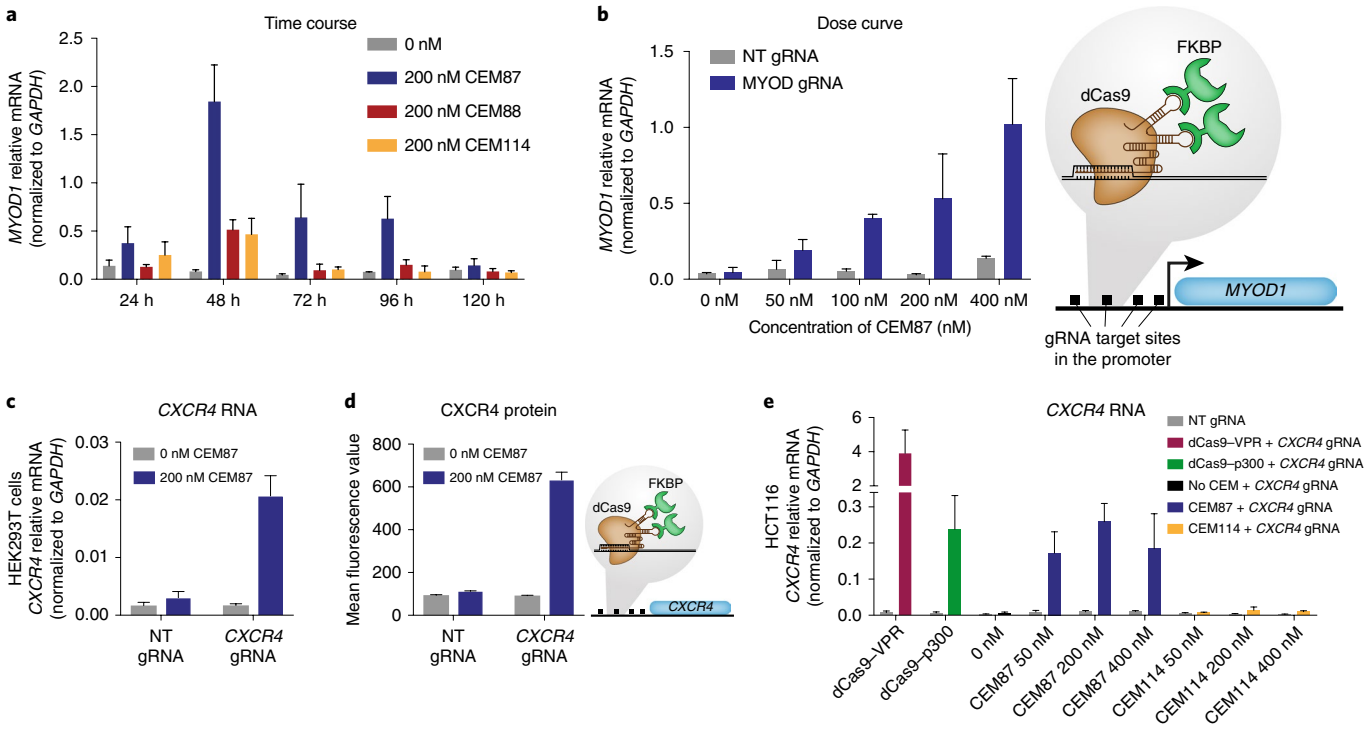

Fig. 2 | Evaluating the dCas9–CEMa system over time and a dose range, and benchmarking the system against current dCas9 activating technologies. **a**, A time course was performed after transfection of *MYOD1*-targeting gRNAs into cells that were virally induced to express dCas9–HA and MS2–FKBP×2. Cells were treated with 200 nM CEM87, CEM88 or CEM114, mRNA was isolated and qRT–PCR was performed every 24 h (following the initial medium change after transfection) for 5 d. Relative mRNA levels were compared to those in untreated cells. CEM87 activated endogenous *MYOD1* most effectively at 48 h. **b**, A dose curve of CEM87-mediated activation was performed with concentrations ranging from 50 to 400 nM. mRNA was isolated after 48 h of treatment. **c**, HEK293T cells were transfected with *CXCR4*-targeting gRNA or NT gRNA and treated with 200 nM CEM87 for 48 h. RNA was isolated, and qRT–PCR was performed. Significant levels of *CXCR4* activation were observed only in cells with the targeting gRNA. **d**, HEK293T cells were transfected with *CXCR4*-targeting gRNA or NT gRNA and treated with 200 nM CEM87 for 48 h. Flow cytometry was performed with a fluorescently conjugated anti-CXCR4 antibody. **e**, HCT116 cells were infected to express dCas9–VPR, dCas9–p300 or dCas9–HA together with MS2–FKBP×2. Cells were transfected with *CXCR4*-targeting gRNA or NT gRNA. Infected dCas9–CEMa cells were treated with 0, 50, 200 or 400 nM CEM87 or CEM114 for 48 h. RNA was isolated, and qRT–PCR was performed. Significant levels of *CXCR4* activation were observed in cells expressing dCas9–VPR or dCas9–p300 with all doses of CEM87. Significant levels of *CXCR4* activation but with a lower fold change were also observed in cells treated with 200 or 400 nM CEM114. No significant *CXCR4* activation was observed in cells transfected with *CXCR4*-targeting gRNA and not treated with CEM. In **a**–**e**, significance was determined with three different samples using a two-tailed Student's *t*-test; fold change and significance are reported in Supplementary Table 1. Error bars, s.d.

level of *MYOD1* overexpression after 48 and 72 h. After 96 h, CEM87 activation was no longer significant, likely owing to high variability and/or plasmid loss after transfection (Fig. 2a and Supplementary Fig. 3b). To determine the concentration at which CEM87 was most effective in this assay, we performed a dose curve with CEM87 concentrations ranging from 50 to 400 nM. Cells with targeting gRNA showed dose-dependent activation of *MYOD1* after 48 h of treatment for doses ranging from 50 to 400 nM. However, cells with NT gRNA showed lower nonspecific activation of *MYOD1* after 48 h of treatment at a dose of 400 nM as compared to untreated cells (Fig. 2b). Thus, CEM87 was most effective over a concentration range of 50–200 nM. With increased activity coupled to more nonspecific activation at 400 nM, usage over 400 nM led to more variable results (data not shown).

To evaluate the versatility of the dCas9–CEMa system, we tested the dCas9–CEMa system with gRNAs targeting the interleukin-1 receptor antagonist (*IL1RN*) locus, a weakly expressed gene. In comparison to untreated cells, cells treated with 200 nM CEM87 had *IL1RN* expression that was increased by 92.5-fold ($P < 0.05$), whereas control cells with NT gRNA did not show a significant change in *IL1RN* expression (Supplementary Fig. 4a). We also tested activating the octamer-binding transcription factor 4 (*OCT4*; also known as *POU5F1*) gene. After 48 h of treatment with 200 nM CEM87, *OCT4* mRNA levels were increased 4.9-fold ($P < 0.01$), whereas the levels did not significantly change in control cells with the NT gRNA (Supplementary Fig. 4b). Next, we targeted the *MYC1* locus, an area of the genome that is more highly expressed than the other targets. CEM87 did not significantly increase expression in the *MYC1*-targeting cells (Supplementary Fig. 4c). As a fourth test of the general utility of the system, we designed a series of gRNAs targeting the super-enhancer (SE) network controlling *MYOD1*. We created one set of four gRNAs targeting the SE boundaries and a second set of six gRNAs targeting the epicenters of this SE network. These gRNAs were induced in a rhabdomyosarcoma cancer cell line (RH4) that expresses *MYOD1* and possesses these SE networks (Supplementary Fig. 4d)[26]. After introducing dCas9 and MS2–FKBP into RH4 cells, we virally infected cells with the gRNA sets overnight, replaced the medium and added CEM87 to recruit additional BRD4. We observed, after 24 h of CEM87 treatment, that *MYOD1* expression was significantly increased, more so when CEMa was guided to the SE epicenters (Supplementary Fig. 4d). Because BRD4 is already present at these epicenters, as confirmed in bulk chromatin immunoprecipitation and sequencing (ChIP–seq), we envision two possible alternatives: either these population-average data mask underlying heterogeneity of BRD4 binding to this SE and CEM87 activates *MYOD1* in cells that have incomplete BRD4

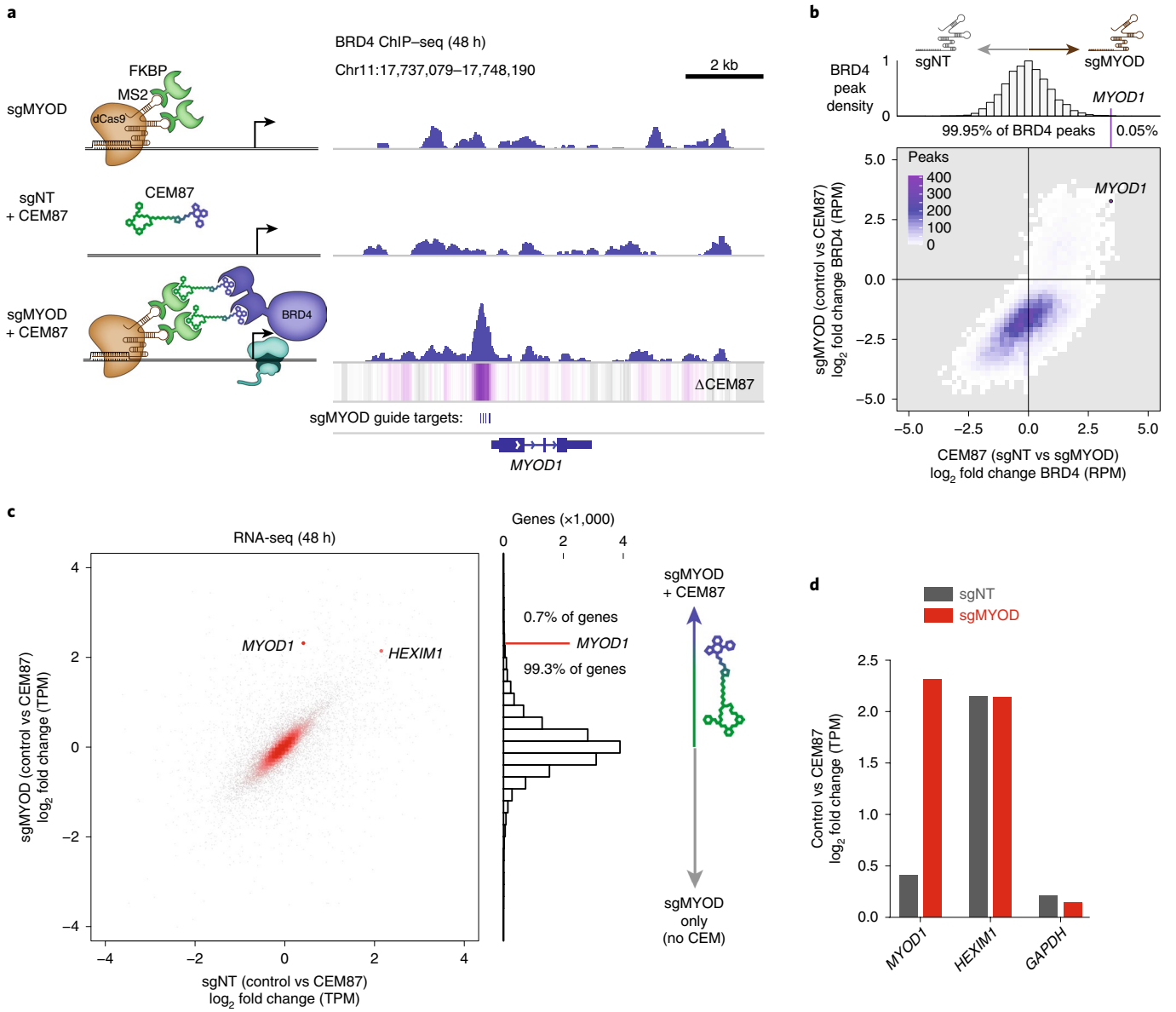

Fig. 3 | Whole-genome analysis of the dCas9–CEMa system. a, Genome-wide binding of BRD4 assayed by ChIP–seq in HEK293T cells expressing dCas9–HA and MS2–FKBP that were treated with four sgRNAs targeting the *MYOD1* promoter (sgMYOD) and the BRD4 recruitment molecule CEM87, alone and in combination, for 48 h. Control cells were transfected with NT sgRNA (sgNT). The genomic heat map was generated by subtracting the BRD4 ChIP–seq signal (reads per million mapped reads, RPM) in cells treated with sgMYOD and CEM87 by the signal in cells treated with only sgMYOD. **b**, Quantification of changes in ChIP–seq signal ($\log_2$ fold change in RPM) for each called BRD4 peak, compared for changes induced by CEM87 (*y* axis) and sgMYOD (*x* axis). A histogram of BRD4 peak density (normalized to 1) for sgMYOD-specific changes in the presence of CEM87 (compared to sgNT) is shown above. The BRD4 peak at the *MYOD1* promoter is highlighted. **c**, RNA-seq data quantified as transcripts per million (TPM) in HEK293T cells treated as in a, showing CEM87-induced transcriptional changes with sgMYOD (*y* axis) compared to sgNT (*x* axis). A histogram of CEM87-induced gene expression changes in cells with sgMYOD is shown on the right. *MYOD1* and *HEXIM1* are highlighted. **d**, Bar charts of the $\log_2$ fold change in TPM for *MYOD1* and *HEXIM1* from the experiment in **c**.

binding or, alternatively, even when an SE has endogenous BRD4, binding may not have reached saturation. Taken together, these data suggest that dCas9–CEMa technology can robustly activate weakly expressed genes; it can also activate highly expressed genes, but to a lesser extent.

To compare RNA expression with single-cell protein analysis, we targeted C-X-C chemokine receptor type 4 (*CXCR4*) and analyzed the level of expression in two ways. After 48 h of exposure to 200 nM CEM87, we conducted RNA extraction and qRT–PCR, which showed a 12.2-fold increase in expression as compared to untreated cells ($P < 0.05$). Cells expressing the NT gRNA did not show a significant change in *CXCR4* expression (Fig. 2c). In addition, we used a fluorescent anti-CXCR4 antibody and performed flow cytometry to test for changes in protein expression. In comparison to untreated cells, the results showed a 5.6-fold ($P < 0.005$) increase in CXCR4, whereas cells with NT gRNA did not show any changes in CXCR4 protein levels (Fig. 2d). To complement our work in HEK293T cell line models, we also performed a *CXCR4*-targeting assay in HCT116 human colorectal carcinoma tissue. The *CXCR4*-targeting gRNAs that we used in the HEK293T cell line showed modest activation (data not shown). To improve the activation, we tested a different set of *CXCR4*-targeting gRNAs in a

multiplexed gRNA backbone. We virally infected HCT116 cells to express dCas9 and MS2–FKBP×2. To use this cell line to benchmark our dCas9–CEMa system against two widely used dCas9-based activators, VPR and p300, we also virally infected HCT116 cells to express dCas9–VPR and dCas9–p300. Cells were transfected with the new *CXCR4*-targeting gRNA or with NT gRNA. After 48 h of treatment with CEM87 or CEM114, we found that the dCas9–CEMa system also showed dose-dependent activation in this cell line. Activation peaked at 200 nM CEM87, which was in the range of what we saw with dCas9–p300 but less than the activation achieved with the VPR system (Fig. 2e). No control cells with NT gRNA showed a significant change in *CXCR4* expression. The ability to control expression levels was unique to dCas9–CEMa. For any given new cell application, the CEM87 concentration should be titrated between 0 and 200 nM.

To investigate the extent of potential off-target or indirect effects of CEM activation by recruitment of BRD4 to the *MYOD1* promoter, we assayed the whole transcriptome (by RNA-seq) and genome-wide BRD4 binding profiles (by ChIP–seq). BRD4 was recruited to the promoter of *MYOD1* only in the presence of both CEM87 and sgMYOD1 (Fig. 3a). BRD4 strongly positively regulates *MYOD1* in myogenic contexts[26], where thousands of downstream targets of MYOD are activated from BRD4-bound *cis*-regulatory elements, yet, at 48 h, *MYOD1* was one of the very few sites with a gain of BRD4 binding (Fig. 3b). Indeed, nearly all other BRD4 sites saw a reduction in binding, which is a known effect of small molecules engaging BET bromodomains[27]. Transcriptionally, *MYOD1* was selectively upregulated (Fig. 3c). Downstream myogenic targets of MYOD remained silent, potentially because of incomplete dosage and short duration. BRD4 inhibition also strongly upregulates *HEXIM1* regardless of biological context[28], which we observed in the RNA-seq data from cells treated with CEM87 (Fig. 3c,d). To validate our ChIP–seq results, we performed ChIP–qPCR with primers targeting the *MYOD1* locus. Using a primer set 1,000 base pairs (bp) upstream of the *MYOD1* transcriptional start site (TSS), BRD4 enrichment was found to be increased 2.4-fold ($P < 0.05$) in cells with *MYOD1*-targeting gRNA as compared to those with NT gRNA after 48 h of treatment with 50 nM CEM87. When we used primers targeting a region 350 bp upstream of the TSS, BRD4 enrichment was increased 10.7-fold ($P < 0.05$) in cells with targeting versus NT gRNA under the same treatment. Lastly, using a primer set targeting a region 50 bp downstream of the TSS, BRD4 enrichment was increased 5.5-fold (Supplementary Fig. 5a; $P < 0.05$).

In summary, we designed, synthesized and optimized a class of CEMa molecules capable of activating endogenous genes in a dose-dependent, gene-specific manner. By adapting the CEMa technology to dCas9 targeting constructs, we can use this system to theoretically target any gene in the genome by strategic gRNA design. We have demonstrated the ability to control the chromatin landscape and induce changes in the expression of endogenous mammalian disease-related genes in a direct, biologically relevant manner. This dCas9–CEMa technology paves the way for targeting disease-relevant genes to ask specific, pointed questions about the mechanisms of action in disease. Because our gene activation platform is chemically activated in a dose-dependent manner, it will be useful in target validation work for visualization of trends between phenotype and gene dosage over a wide range of target gene concentrations.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-019-0296-7.

Received: 25 July 2018; Accepted: 24 September 2019;
Published online: 11 November 2019

References

1. Dawson, M. A. The cancer epigenome: concepts, challenges, and therapeutic opportunities. *Science* **355**, 1147–1152 (2017).
2. MacDonald, I. A. & Hathaway, N. A. Epigenetic roots of immunologic disease and new methods for examining chromatin regulatory pathways. *Immunol. Cell Biol.* **93**, 261–270 (2015).
3. Zeng, L. & Zhou, M. M. Bromodomain: an acetyl-lysine binding domain. *FEBS Lett.* **513**, 124–128 (2002).
4. de Ruijter, A. J. M., van Gennip, A. H., Caron, H. N., Kemp, S. & van Kuilenburg, A. B. P. Histone deacetylases (HDACs): characterization of the classical HDAC family. *Biochem. J.* **370**, 737–749 (2003).
5. Gao, Y. et al. Complex transcriptional modulation with orthogonal and inducible dCas9 regulators. *Nat. Methods* **13**, 1043–1049 (2016).
6. Chen, T. et al. Chemically controlled epigenome editing through an inducible dCas9 system. *J. Am. Chem. Soc.* **139**, 11337–11340 (2017).
7. Braun, S. M. G. et al. Rapid and reversible epigenome editing by endogenous chromatin regulators. *Nat. Commun.* **8**, 560 (2017).
8. Gao, D. & Liang, F. S. Chemical inducible dCas9-guided editing of H3K27 acetylation in mammalian cells. *Methods Mol. Biol.* **1767**, 429–445 (2018).
9. Ma, D., Peng, S. & Xie, Z. Integration and exchange of split dCas9 domains for transcriptional controls in mammalian cells. *Nat. Commun.* **7**, 13056 (2016).
10. Shrimp, J. H. et al. Chemical control of a CRISPR–Cas9 acetyltransferase. *ACS Chem. Biol.* **13**, 455–460 (2018).
11. Hilton, I. B. et al. Epigenome editing by a CRISPR–Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat. Biotechnol.* **33**, 510–517 (2015).
12. Liszczak, G. P. et al. Genomic targeting of epigenetic probes using a chemically tailored Cas9 system. *Proc. Natl Acad. Sci. USA* **114**, 681–686 (2017).
13. Erwin, G. S. et al. Synthetic transcription elongation factors license transcription across repressive chromatin. *Science* **6370**, 1617–1622 (2017).
14. Butler, K. V., Chiarella, A. M., Jin, J. & Hathaway, N. A. Targeted gene repression using novel bifunctional molecules to harness endogenous histone deacetylation activity. *ACS Synth. Biol.* **7**, 38–45 (2018).
15. Chung, C. et al. Discovery and characterization of small molecule inhibitors of the BET family bromodomains. *J. Med. Chem.* **54**, 3827–3838 (2011).
16. Demont, E. H. et al. 1,3-Dimethyl benzimidazolones are potent, selective inhibitors of the BRPF1 bromodomain. *ACS Med. Chem. Lett.* **5**, 1190–1195 (2014).
17. Hay, D. A. et al. Discovery and optimization of small-molecule ligands for the CBP/p300 bromodomain. *J. Am. Chem. Soc.* **26**, 9308–9319 (2014).
18. Chavez, A. et al. Highly efficient Cas9-mediated transcriptional programming. *Nat. Methods* **12**, 326–328 (2015).
19. Morita, S. et al. Targeted DNA demethylation in vivo using dCas9–peptide repeat and scFv–TET1 catalytic domain fusions. *Nat. Biotechnol.* **34**, 1060–1065 (2016).
20. 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).
21. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature* **517**, 583–588 (2015).
22. Lim, F. et al. Altering the RNA binding specificity of a translational repressor. *J. Biol. Chem.* **12**, 9006–9010 (1994).
23. Chavez, A. et al. Comparison of Cas9 activators in multiple species. *Nat. Methods* **13**, 563–567 (2016).
24. Begley, C. G. & Ellis, L. M. Raise standards for preclinical cancer research. *Nature* **483**, 531–533 (2012).
25. Hay, M., Thomas, D. W., Craighead, J. L., Economides, C. & Rosenthal, J. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **32**, 40–51 (2014).
26. Gryder, B. E. et al. PAX3–FOXO1 establishes myogenic super enhancers and confers BET bromodomain vulnerability. *Cancer Discov.* **7**, 884–899 (2017).
27. Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
28. Lin, X. et al. *HEXIM1* as a robust pharmacodynamic marker for monitoring target engagement of BET family bromodomain inhibitors in tumors and surrogate tissues. *Mol. Cancer Ther.* **16**, 388–396 (2017).

Methods

Chemical synthesis. See the Supplementary Note for details. CEMa compounds were diluted in DMSO (Sigma, D2650) and kept dry at −20 °C.

Statistical analysis. For all flow cytometry and RNA extractions followed by qRT–PCR, a Student's *t*-test (two tailed) was used to determine significance. In each experiment, three different cell culture replicates were used. Error bars represent the s.d. Further information can be found in the Life Sciences Reporting Summary.

Plasmid design. To create gRNAs compatible with *S. aureus* dCas9, we used Addgene plasmid 64710. To create gRNAs compatible with *S. pyogenes* dCas9–MS2, we used Addgene plasmid 61427. For the multiplexed gRNAs, constructs were modified from Addgene kit 1000000055 to express gRNAs in a lentiviral backbone, containing MS2-compatible stem loops and lacking a Cas9 protein. All targeting gRNA sequences are listed in Supplementary Table 2.

All dCas9, MS2 and scFv fusions were reconstructed to ensure proper comparison between effectors: Ef1α-dCas9(Sp)-HA; Ef1α-dCas9(Sp)-FKBP×1; Ef1α-dCas9(Sp)-FKBP×2; Ef1α-dCas9(Sp)-SunTag×10; Ef1α-dCas9(Sp)-SunTag×24; Ef1α-dCas9(Sp)-p300; Ef1α-dCas9(Sp)-TET; Ef1α-dCas9(Sp)-VPR; Ef1α-dCas9(Sa)-FKBP×1; Ef1α-dCas9(Sa)-FKBP×2; Ef1α-scFv-FKBP×1; Ef1α-scFv-FKBP×2; Ef1α-MS2(N55K)-FKBP×1; Ef1α-MS2(N55K)-FKBP×2; Ef1α-MS2(N55K)-SunTag×10; and Ef1α-MS2(N55K)-SunTag×24. Constructs were adapted from Hathaway et al.[29], Braun et al.[7], Tanenbaum et al.[20], Thakore et al.[30], Liu et al.[31], Gao et al.[5] and Hilton et al.[11].

Cell culture. Low-passage (passage 30–40) HEK293T cells were cultured with high-glucose DMEM (Corning, 10-013-CV) supplemented with 10% FBS (Atlantic Biologicals, S11550), 10 mM HEPES (Corning, 25-060-Cl), NEAA (Gibco, 11140-050), penicillin-streptomycin and 55 µM of 2-mercaptoethanol. Cells were passaged every 2–5 d and maintained at 10–90% confluency in an incubator at 37 °C and 5% CO_2.

Cell transfection. The day after splitting cells, polyethlenimine (PEI; Polysciences, 23966-1) transfection was performed with a 1:3 ratio of DNA (in µg) to PEI (in µl) in OptiMEM (Gibco, 31985070) diluted 1:10 in total medium. Specific ratios of DNA are shown in Supplementary Table 3. For six-well plates, cells were cultured and transfected with 2 ml of medium. For 12-well plates, cells were cultured and transfected with 1 ml of medium. For cells grown in a 15-cm plate, 20 ml of medium was used. For each transfection experiment, the medium was changed 16 h later[32][js1] . [js1]I think the query to the CE is saying that they've added the reference citation here as a placeholder so that the references would be cited in order. The later citation by STAR is the correct location for this reference though, so please delete here and add a TS query asking for the remaining references to be renumbered.

Flow cytometry. Flow cytometry was performed with Attune Nxt as previously described[33]. The data presented represent three separate transfections, and significance was determined by Student's *t*-test. To compare CXCR4 protein levels, we used a protocol with a fluorophore-conjugated primary antibody (anti-CXCR4-APC; BioLegend, 306510). After washing with 1× PBS, we incubated the cells with the antibody (for a final concentration of 1.6 µg ml^{-1}) in the dark at room temperature for 1 h. An example of the gating strategy can be found in Supplementary Fig. 4f.

Lentiviral infection. Lentivirus production for HEK293T cell infection was done using Lenti-X 293T cells (Clontech). Low-passage cells (passage 8–20) were plated onto 15-cm plates such that they were 70% confluent 24 h later. Each plate was transfected with 18 µg of the plasmid of interest (dCas9-X), 13.5 µg of plasmid expressing Gag–Pol (Addgene, 12260) and 4.5 µg of plasmid expressing VSV-G envelope protein (Addgene, 12259). PEI transfection was performed, and 60 h after transfection the virus was spun down at 20,000 r.p.m. for 2.5 h at 4 °C and then added to HEK293T cells in combination with 10 µg ml^{-1} polybrene (Santa Cruz Biotechnology, sc-134220). Selection of lentiviral constructs was performed with either hygromycin (200 µg ml^{-1}) or blasticidin (12 µg ml^{-1}).

RNA extraction. For Figs. 2 and 3, and Supplementary Figs. 3 and 4, cells from 12-well or 6-well plates were isolated. Cells were washed with 1× FpbasPBS, disassociated with 0.05% trypsin, quenched with medium, centrifuged and washed with 1× PBS. RNA extraction was performed with an RNeasy Plus Mini kit (Qiagen, 74134), and the relative enrichment of mRNA was quantified with the RNA-to-C_T 1-Step kit (Thermo Fisher Scientific, 4389986). The primers used are shown in Supplementary Fig. 4e.

RNA sequencing. RNA extracted as above was quantified and profiled for quality using an Agilent Bioanalyzer. Then, poly(A)-enriched and Illumina-barcoded libraries were made and sequenced on an Illumina NextSeq 500 to a depth of 30 million base pairs (150-bp paired-end reads). Reads were mapped to the UCSC reference for hg19 using STAR[32], and read abundance for each gene was counted (TPM method) using RSEM. Data visualization was performed using custom R scripts (available at https://github.com/GryderArt/VisualizeRNAseq/tree/master/RNAseq_Pipeline).

ChIP–qPCR. Chromatin isolation was performed on HEK293T cells as described in Chiarella et al.[33]. Immunoprecipitation was performed with an anti-BRD4 antibody (ab128874). qPCR was performed on the isolated DNA with SYBRGreen as previously described using the primers described in Supplementary Fig. 5b.

ChIP–seq. Chromatin bound to BRD4 was immunoprecipitated from fixed and sheared nuclei, with or without CEM87 treatment, in HEK293T cells bearing sgNT or sgMYOD1 (promoter). Enriched DNA from ChIP was barcoded and libraries were amplified, followed by size selection (fragment ranges between 250 and 1,000 bp) using AMPure XP beads (Beckman Coulter). Libraries were sequenced to a depth of 30 million base pairs (75-bp single-end reads) on an Illumina NextSeq 500. Reads were mapped to hg19 using BWA, indexed with SAMtools and converted to TDF format for visualization of browser tracks in IGV using igvtools. Peaks were identified using MACS2 with statistical thresholds of 1×10^{-5} for CEM87-treated samples and 1×10^{-7} for untreated samples, owing to the reduced signal-to-noise ratio under CEM87 conditions. Read coverage under all BRD4 peaks was calculated using BEDTools (multicov mode; see details at https://bedtools.readthedocs.io/en/latest/content/tools/multicov.html) and visualized in R (scripts available from https://github.com/CBIIT/ChIP_seq).

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

Data availability

Genome-wide data generated herein are publicly available through the Gene Expression Omnibus with accession number GSE129407. All data presented in this manuscript are available from the corresponding authors upon reasonable request.

Code availability

Readers can view our code through the public link (https://github.com/GryderArt/ChIPseqPipe). There are no access restrictions.

References

29. Hathaway, N. A. et al. Dynamics and memory of heterochromatin in living cells. *Cell* **149**, 1447–1460 (2012).
30. Thakore, P. I. et al. RNA-guided transcriptional silencing in vivo with *S. aureus* CRISPR–Cas9 repressors. *Nat. Commun.* **9**, 1674 (2018).
31. Liu, S. et al. Editing DNA methylation in the mammalian genome. *Cell* **167**, 233–247 (2016).
32. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
33. Chiarella, A. M. et al. Cavitation enhancement increases the efficiency and consistency of chromatin fragmentation from fixed cells for downstream quantitative applications. *Biochemistry* **57**, 2756–2761 (2018).

Acknowledgements

We thank the members of the Hathaway, Jin and Khan laboratories for many helpful discussions. We acknowledge E. Chory (Stanford University), S. Braun (Stanford University), G. Crabtree (Stanford University), R. Vale (University of California, San Francisco), F. Zhang (Massachusetts Institute of Technology), Z. Xie (Tsinghua University) and C. Gersbach (Duke University) for sharing plasmids used or adapted in this study. This work was supported in part by grant R01GM118653 from the US National Institutes of Health (to N.A.H.) and by grants R01GM122749 and R01HD088626 from the US National Institutes of Health (to J.J.). This work was further supported by a Tier 3 grant and a Student Grant from the UNC Eshelman Institute for Innovation (to N.A.H. and A.M.C., respectively). T32GM007092 from the US National Institutes of Health (to A.M.C.) also supported early portions of this work. Additional support was provided by American Cancer Society postdoctoral fellowship PF-14-021-01-CDD (to K.V.B.). Flow cytometry data were obtained at the UNC Flow Cytometry Core Facility, funded by the P30CA016086 Cancer Center Core Support Grant to the UNC Lineberger Comprehensive Cancer Center.

Author contributions

N.A.H., J.J., K.V.B. and A.M.C. conceived the project. N.A.H., J.J., K.V.B., A.M.C., D.L., B.E.G. and J.K. contributed to the experimental design. N.A.H., J.J., K.V.B., A.M.C., D.L., B.E.G., J.K., T.A.W., X.Y. and S.P. contributed experimentally. K.V.B. synthesized the compounds. B.E.G. and J.K. carried out ChIP–seq and RNA-seq experiments and analysis. N.A.H., J.J., K.V.B., A.M.C., D.L. and B.E.G. wrote the manuscript. All authors edited the manuscript.

Competing interests

N.A.H., J.J., K.V.B. and A.M.C. are inventors on U.S. provisional patent application no. 62/654,958, "Bifunctional chemical epigenetic modifiers and methods of use."

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41587-019-0296-7.

Correspondence and requests for materials should be addressed to J.J. or N.A.H.

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

nature research

Corresponding author(s): Nate Hathaway

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	The software for flow cytometry acquisition is Attune NxT 2.6. For sequencing analysis: Illumina barcoded libraries were made and sequenced on an Illumina NextSeq 500, to a depth of 30 million base pairs (150 bp paired end).
Data analysis	The software for flow cytometry analysis is FlowJo 10.4.2. For sequencing analysis reads were mapped to UCSC reference for hg19 using STAR34, then read abundance for each gene counted (TPM method) using RSEM. Data visualization was performed using custom R scripts (available here: https://github.com/GryderArt/VisualizeRNAseq/tree/master/RNAseq_Pipeline).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

No large data sets were generated or analyzed during this current study, other than ChIP-seq and RNA-seq, availability is described in that section below.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculation was performed. Each condition was analyzed with 3 biological replicates. This is typically the standard for the experiments performs so that a reasonable range of variability between samples can be accounted for.
Data exclusions	All biological replicates are presented as indicated, with the number defined for each experiment.
Replication	All attempts of replication showed a consistent pattern and were deemed successful. Multiple independent experiments were conducted as reported to ensure reproducibility of the experimental findings were validated.
Randomization	Samples of cells were sorted into pre-labeled tubes. Following sorting into tubes, the order of sample preparation was randomized and not chosen in a specific order.
Blinding	Allocation was not blinded to ensure that full sets of biological replicates were analyzed together, rather than multiple biological replicates of a few sample conditions.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	APC anti-human CD184 (CXCR4) was used for flow cytometry (from Biolegend catalog: 306510). BRD4 CHIP-grade antibody was used for ChIP qPCR and ChIP sequencing (Abcam catalogue #ab128874).
Validation	For the CXCR4 antibody, we optimized the incubation conditions in the 293t cells through serial dilutions and determining the concentration at which signal plateaus. For the BRD4 antibody, we used the amount of antibody recommended by the ChIP-IT High Sensitivity Kit from Active Motif (catalogue #53040) and used negative control qPCR primer sets.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	293 [HEK-293] (ATCC® CRL-1573™) Obtained from the UNC Tissue Culture Facility ; Lenti-X™ 293T cell line obtained from Clontech ; HCT116 (ATCC® CCL-247™) Obtained from the UNC Tissue Culture Facility. RH4 human cancer cell line obtained from Dr. Peter Houghton.
Authentication	The cell lines used were not authenticated in our hands.
Mycoplasma contamination	Cell lines in Hathaway lab are tested for mycoplasma contamination by PCR every 6 months or at least annually. All lines to date tested negative.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	https://github.com/GryderArt/ChIPseqPipe
Files in database submission	Sample_HEK_sgMYOD_cntrl_BRD4_037_C_HMJ2HBGX7 Sample_HEK_sgMYOD_CEM87_BRD4_037_C_HMJ2HBGX7 Sample_HEK_sgNT_CEM87_BRD4_037_C_HMJ2HBGX7
Genome browser session (e.g. UCSC)	https://genome.ucsc.edu

Methodology

Replicates	One replicate was used for each condition.
Sequencing depth	Instrument mode: Illumina NextSeq 475 Read length: 75 Single end
Antibodies	BRD4 CHIP-grade antibody was used for ChIP qPCR and ChIP sequencing (Abcam catalogue #ab128874).
Peak calling parameters	GEO.name bed.path fastq.path bed fastq HEK sgMYOD cntrl BRD4 /khanlab/projects/ChIP_seq/DATA/Sample_HEK_sgMYOD_cntrl_BRD4_037_C_HMJ2HBGX7/MACS_Out_p_1e-07/Sample_HEK_sgMYOD_cntrl_BRD4_037_C_HMJ2HBGX7_peaks.narrowPeak.nobl.GREAT.bed /khanlab/projects/ChIP_seq/DATA/Sample_HEK_sgMYOD_cntrl_BRD4_037_C_HMJ2HBGX7/ Sample_HEK_sgMYOD_cntrl_BRD4_037_C_HMJ2HBGX7_R1.fastq.gz Sample_HEK_sgMYOD_cntrl_BRD4_037_C_HMJ2HBGX7_peaks.narrowPeak.nobl.GREAT.bed Sample_HEK_sgMYOD_cntrl_BRD4_037_C_HMJ2HBGX7_R1.fastq.gz HEK sgMYOD CEM87 BRD4 /khanlab/projects/ChIP_seq/DATA/Sample_HEK_sgMYOD_CEM87_BRD4_037_C_HMJ2HBGX7/MACS_Out_p_1e-07/Sample_HEK_sgMYOD_CEM87_BRD4_037_C_HMJ2HBGX7_peaks.narrowPeak.nobl.GREAT.bed /khanlab/projects/ChIP_seq/DATA/Sample_HEK_sgMYOD_CEM87_BRD4_037_C_HMJ2HBGX7/ Sample_HEK_sgMYOD_CEM87_BRD4_037_C_HMJ2HBGX7_R1.fastq.gz Sample_HEK_sgMYOD_CEM87_BRD4_037_C_HMJ2HBGX7_peaks.narrowPeak.nobl.GREAT.bed Sample_HEK_sgMYOD_CEM87_BRD4_037_C_HMJ2HBGX7_R1.fastq.gz HEK sgNT CEM87 BRD4 /khanlab/projects/ChIP_seq/DATA/Sample_HEK_sgNT_CEM87_BRD4_037_C_HMJ2HBGX7/MACS_Out_p_1e-07/Sample_HEK_sgNT_CEM87_BRD4_037_C_HMJ2HBGX7_peaks.narrowPeak.nobl.GREAT.bed /khanlab/projects/ChIP_seq/DATA/Sample_HEK_sgNT_CEM87_BRD4_037_C_HMJ2HBGX7/ Sample_HEK_sgNT_CEM87_BRD4_037_C_HMJ2HBGX7_R1.fastq.gz Sample_HEK_sgNT_CEM87_BRD4_037_C_HMJ2HBGX7_peaks.narrowPeak.nobl.GREAT.bed Sample_HEK_sgNT_CEM87_BRD4_037_C_HMJ2HBGX7_R1.fastq.gz
Data quality	SampleFile PEAKS_PVALUE PEAKS_FOLD_CHANGE5 Sample_HEK_sgMYOD_cntrl_BRD4_037_C_HMJ2HBGX7 21901 21612 Sample_HEK_sgMYOD_CEM87_BRD4_037_C_HMJ2HBGX7 1879 1851 Sample_HEK_sgNT_CEM87_BRD4_037_C_HMJ2HBGX7 523 479
Software	ChIP enriched DNA reads were mapped to reference genome using BWA version 0.7.10 The ChIP-seq and DNase-seq peaks were called by MACS2 version 2.1.0. Enrichment of known and de novo motifs were found using HOMER version 4.8.2

https://github.com/GryderArt/ChIPseqPipe

Flow Cytometry

Plots

Confirm that:

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

☒ 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).

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

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

Methodology

Sample preparation	Cells were washed with 1X PBS, dissociated, quenched with media. Cells were spun down, washed with 1X PBS, and spun down, and lastly resuspended the cell pellet in 400 uL of FACS buffer. For the experiments with APC-CXCR4, cells were incubated in 8 uL per mL of blocking buffer for 1 hour in the dark at room temperature.
Instrument	The Thermo Fisher Attune NxT is equipped with a 96/384 well plate Autosampler and 4 lasers (405, 488, 561, and 637nm) and can detect up to 14 fluorescent parameters. Its acquisition software is Attune NxT 2.6.
Software	The analysis software that was used is FlowJo.
Cell population abundance	At least 100,000 cells were obtained for each sample and technical replicate.
Gating strategy	Gating was performed using FSC/SSC area, followed by SSC-height/SSC-area. The same gate was applied to all samples of cells within the same experiment.

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