Functional annotation of native enhancers with a Cas9–histone demethylase fusion

Nicola A Kearns[1,2,4], Hannah Pham[1,2,4], Barbara Tabak[2,3], Ryan M Genga[1,2], Noah J Silverstein[3], Manuel Garber[2,3] & René Maehr[1,2]

Understanding of mammalian enhancers is limited by the lack of a technology to rapidly and thoroughly test the cell type–specific function. Here, we use a nuclease-deficient Cas9 (dCas9)–histone demethylase fusion to functionally characterize previously described and new enhancer elements for their roles in the embryonic stem cell state. Further, we distinguish the mechanism of action of dCas9-LSD1 at enhancers from previous dCas9-effectors.

Enhancers control development and cellular function, and intensive efforts are ongoing to elucidate cell fate–specific enhancer activity[1,2]. Indeed, a large number of genomic regions identified by genome-wide association studies of human disease fall within enhancer regions[3,4]. Thus, there is a pressing need for technologies to functionally annotate cell type–specific enhancer elements that control cellular function.

Engineered derivatives of clustered, regularly interspaced, short palindromic repeats (CRISPR) systems have enabled RNA-guided gene regulation through targeting of dCas9-coupled transcriptional regulators to promoter regions[5–10], and applicability of this approach in high-throughput contexts has been demonstrated[11,12]. Although a dCas9-KRAB repressor has previously been used to interfere with enhancer function, the study suggested steric rather than effector-mediated interference[13]. Targeted steric hindrance or nuclease-based disruption of an enhancer is possible, but such approaches require in-depth knowledge of the enhancer, may be inefficient and could have potential side effects such as perturbation of chromatin folding or DNA damage stress responses.

The histone demethylase LSD1 has been previously implicated in repression of enhancers[14,15], and transcription activator–like effector (TALE)-LSD1 can target histone modifications that correlate with active enhancers[16]. Although changes in gene expression were detected[16], the TALE-LSD1 approach allows for only low-throughput approaches, and it remains unclear whether the expression changes were a result of decommissioning an enhancer or due to an indirect effect on gene expression.

We generated mouse embryonic stem cells (mESCs) expressing versions of *Neisseria meningitidis* (Nm) dCas9 fused with LSD1, a non-effector BirA affinity tag (BAT) or a KRAB repressor (**Supplementary Figs. 1** and **2**) and used a viral delivery system for single guide RNAs (sgRNAs). We first targeted the well-characterized *cis*-regulatory region of *Oct4* (also known as *Pou5f1*; ref. 17) (**Fig. 1a**), which encodes a factor critical for the embryonic stem cell (ESC) state[18]. *Oct4* expression is regulated by a proximal enhancer (OPE) active in epiblast cells and by a distal enhancer (ODE) active in mESCs and cells of the inner cell mass[17,19]. Targeting of LSD1 to the ODE resulted in loss of *Oct4* expression and appearance of OCT4-negative colonies accompanied by phenotypic changes (**Fig. 1b** and **Supplementary Fig. 3a**,**b**) compared to control dCas9-BAT cells targeted to the same enhancer. No effects were observed when the OPE, the *Oct4* proximal promoter (OPP) or a control locus was targeted in dCas9-LSD1 or dCas9-BAT cells. In contrast, targeting dCas9-KRAB to the OPP led to downregulation of *Oct4*, demonstrating that dCas9-KRAB is not enhancer specific (**Fig. 1b** and **Supplementary Fig. 3a**,**b**). dCas9-effector–dependent and sgRNA position–specific changes in ESC morphology correlated with loss of OCT4 and SOX2 pluripotency factor expression (**Supplementary Fig. 3a**) and genome-wide transcriptomic changes (**Supplementary Fig. 3c** and **Supplementary Table 1**), indicating a more profound change in the cellular state after interfering with ODE activity. Notably, cluster analysis revealed high similarity between gene expression profiles from dCas9-KRAB with OPP sgRNA, dCas9-KRAB with ODE sgRNA and dCas9-LSD1 with ODE sgRNA cells compared to other dCas9-effector–sgRNA combinations. We conclude that the dCas9-LSD1 system can be used to delineate enhancers specifically, unlike the dCas9-KRAB system that more broadly affects the cellular state when targeted to promoters and enhancers.

We next targeted eight pluripotency-specific candidate enhancers in dCas9-LSD1–expressing ESCs (**Supplementary Figs. 4–6** and **Supplementary Table 2**), and four showed both morphological changes and loss of ESC-associated alkaline phosphatase (AP) activity (**Supplementary Fig. 7**). These four enhancers are thus critical for the appropriate expression of genes required for the ESC state. Absence of differentiation phenotypes upon enhancer targeting could occur if the enhancer is dispensable for maintenance of the pluripotent state. Alternatively, some sgRNAs might be nonfunctional or chromatin state could counteract dCas9-LSD1 function. Of note, no control sgRNAs (*Foxn1*, *N-Myc* (*Mycn*) or

[1]Diabetes Center of Excellence, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [2]Program in Molecular Medicine, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [3]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. [4]These authors contributed equally to this work. Correspondence should be addressed to R.M. (rene.maehr@umassmed.edu).

RECEIVED 20 OCTOBER 2014; ACCEPTED 26 JANUARY 2015; PUBLISHED ONLINE 16 MARCH 2015; DOI:10.1038/NMETH.3325

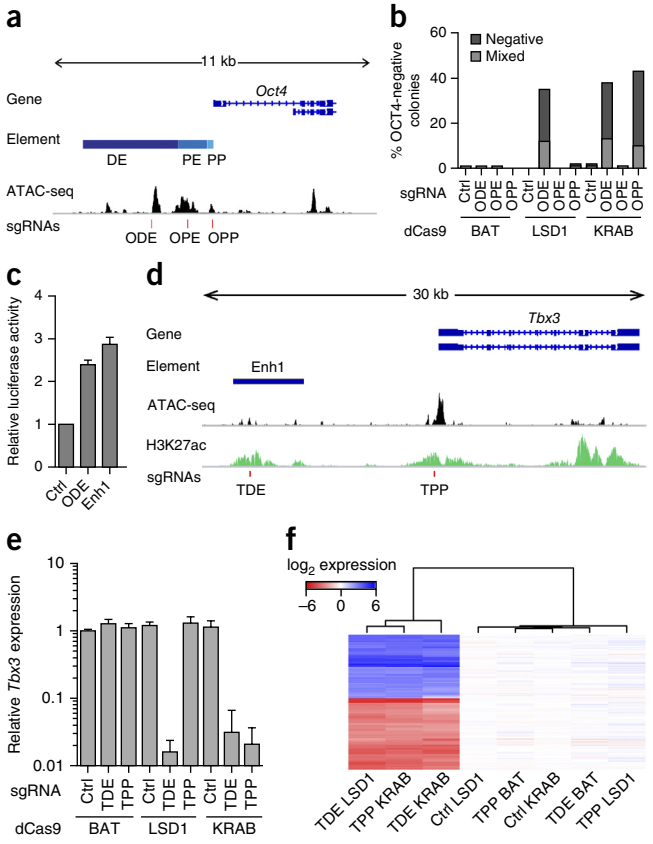

Figure 1 | dCas9-effector fusions regulate *cis*-regulatory elements in an effector-dependent manner. (**a**) Genomic organization of the *Oct4* locus: ODE, distal enhancer; OPE, proximal enhancer; OPP, proximal promoter; ATAC-seq signal, accessible genome; red lines indicate the binding sites of the sgRNAs used in this study. (**b**) Percentage of colonies that do not contain OCT4-expressing cells (negative) or that contain OCT4-negative cells among residual OCT4-expressing cells (mixed) after locus-specific sgRNA delivery. Ctrl, control. (**c**) Relative luciferase activity of reporter plasmids containing a fragment either of the ODE or of enhancer 1 (Enh1). $n = 3$ biological replicates; error bars, s.d. (**d**) Genomic organization of the *Tbx3* locus including H3K27ac. (**e**) Quantitative PCR analysis for *Tbx3* expression in Nm dCas9–effector mESCs treated with sgRNAs specific to an unrelated control genomic region (Ctrl), the putative *Tbx3* distal enhancer (TDE) or the *Tbx3* promoter (TPP). $n = 3$ biological replicates; error bars, s.d. (**f**) Heat map of gene expression microarray data from dCas9-effector mESCs with indicated sgRNAs displayed relative to the average expression levels of the dCas9-effector–Ctrl sgRNA samples. Unsupervised hierarchical clustering of 174 differentially expressed genes (listed in **Supplementary Table 3**) is displayed on the *y* axis. Hierarchical clustering of samples based on similarities in gene expression profiles is displayed on the *x* axis.

Oct6 (*Pou3f1*)) led to changes in morphology or AP activity, thereby indicating sgRNA-specific changes in cellular state.

For follow-up, we focused on the putative enhancer with the highest differential score, Enh1. Tests of Enh1 in a reporter assay confirmed its ability to enhance expression to levels comparable with those obtained with an ODE sequence (**Fig. 1c**). This previously unannotated ESC-specific enhancer is positioned ~10 kb upstream of transcription factor–encoding *Tbx3* (**Fig. 1d**), a gene previously implicated in the maintenance of pluripotency[20]. We therefore hypothesized that Enh1 may function in the ESC network by regulating *Tbx3* expression. Indeed, we detected a relative reduction of *Tbx3* mRNA and protein expression upon targeting Enh1 with dCas9-LSD1 (**Fig. 1e** and **Supplementary Fig. 8a**). This loss of *Tbx3* could be similarly obtained when dCas9-KRAB was used to target Enh1 or the *Tbx3* proximal promoter (TPP) but was not observed in control cells. Change in colony morphology and an increase of differentiation-associated markers were detected in a dCas9-effector– and sgRNA–dependent manner (**Fig. 1f**, **Supplementary Fig. 8** and **Supplementary Table 3**) concordant with the destabilization of the ESC state. Of note, none of the transcriptomic changes among any of the control conditions exceed twofold. Although this finding does not rule out minor off-target effects, these data support a compelling level of specificity of our system.

Next we sought to dissect the mechanism by which targeting of Enh1, hereafter referred to as *Tbx3* distal enhancer (TDE), results in *Tbx3* downregulation. We examined a time point at which differential *Tbx3* expression levels were detected but changes in cell morphology were not yet observable (**Supplementary Fig. 9a**). Conducting chromosome conformation

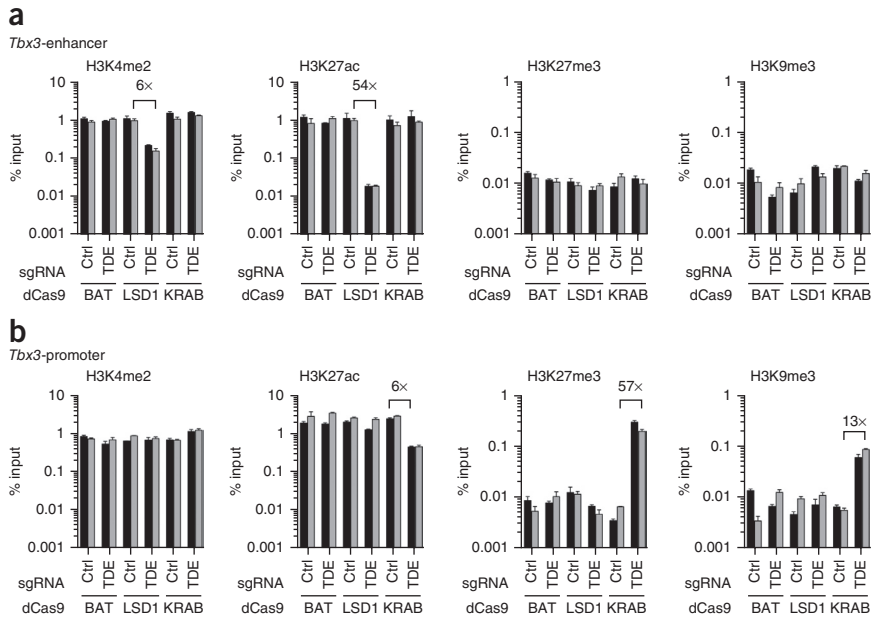

Figure 2 | Enhancer targeting by dCas9-LSD1 or dCas9-KRAB. (**a**,**b**) Chromatin immunoprecipitation followed by quantitative PCR (ChIP qPCR) analysis of H3K4me2, H3K27ac, H3K27me3 and H3K9me3 at the *Tbx3* enhancer 345 bases upstream of the sgRNA target site (**a**) and the *Tbx3* proximal promoter region (**b**) for dCas9-effector mESCs in the presence of control sgRNA (Ctrl) or an sgRNA specific to the putative *Tbx3* distal enhancer (TDE). Locations of qPCR amplicons are indicated in **Supplementary Figure 4**. The data are expressed as percent input values. Adjacent black and gray bars represent two independent experiments with mean ± s.d. of three technical replicates.

capture (3C), we observed a localized peak in interaction frequency between the TDE and the *Tbx3* promoter in dCas9-BAT ESCs with a control sgRNA supporting the presence of an enhancer-promoter loop (**Supplementary Fig. 9b**). Comparable interaction frequencies were observed for dCas9-LSD1 and dCas9-BAT with TDE sgRNA, indicating that the looping interaction between the *Tbx3* promoter and enhancer is not disrupted by dCas9-effector targeting *per se*.

Investigating local histone modifications in the presence of either dCas9-KRAB or dCas9-LSD1, we observed relative loss of the LSD1 substrate dimethylated histone 3 lysine 4 (H3K4me2) around the enhancer-sgRNA target site in dCas9-LSD1 cells (by 6- to 8-fold) (**Fig. 2a** and **Supplementary Fig. 10**). H3K27ac, indicative of active enhancers[2], was dramatically lost at the enhancer (by 33- to 54-fold) upon dCas9-LSD1 targeting compared to other tested conditions, a finding consistent with prior studies using TALE-LSD1 (ref. 16; **Fig. 2a** and **Supplementary Fig. 10**). Changes in H3 occupancy were not detected (**Supplementary Fig. 10**). Interestingly, although repressive marks have been previously associated with KRAB-mediated repression[21], we could not detect concomitant increases in repressive marks, H3K27me3 or H3K9me3, at the enhancer after targeting of any of our constructs (**Fig. 2a**). To further understand the differences between the LSD1 and KRAB effects, we surveyed chromatin state at the *Tbx3* promoter. At this promoter, we detected a loss of H3K27ac and an increase of H3K27me3 (~57-fold) and H3K9me3 (~13-fold) when the TDE was targeted with the KRAB effector, but we did not observe substantial changes with LSD1 (**Fig. 2b**). Taken together, these data offer strong evidence that LSD1-induced enhancer deactivation results in *Tbx3* downregulation, whereas KRAB-mediated *Tbx3* downregulation is due to promoter silencing. We therefore caution that the dCas9-KRAB effector, besides having potential side effects such as heterochromatin spreading[21], is silencing promoter activity rather than decommissioning the target enhancer.

Next, we showed that the LSD1-specific inhibitor TCP (*trans*-2-phenylcyclopropylamine hydrochloride), previously shown to overcome strong differentiation cues in mESCs, such as loss of *Oct4* expression[14], inhibited the loss of *Tbx3* mRNA and protein expression in dCas9-LSD1 cells but not in dCas9-KRAB cells after TDE targeting (**Supplementary Fig. 11**). This indicates that the loss of *Tbx3* expression is directly dependent on the enzymatic activity of LSD1 and that KRAB can repress *Tbx3* independently.

We conclude that the dCas9-LSD1 fusion protein allows for an effector-dependent definition of functional, native enhancer elements that help to maintain a given cellular state. Accordingly, dCas9-LSD1 provides a rapid and powerful approach to understanding distal *cis*-regulatory regions such as enhancers without major disruption of the local genomic architecture. Although we used the Nm dCas9, we expect that similar systems could be built on orthogonal CRISPR-Cas9 systems including the widely used *Streptococcus pyogenes* dCas9. Further, the use of chromatin modifiers with differential functionality and orthogonal dCas9 systems promises to enable even more complex epigenome engineering. This level of genome control has previously not been achieved with other RNA-guided regulatory technologies such as RNAi. Combined with rapid advances in cell type–specific annotation of regulatory elements and genome-wide association studies, this technology will be instrumental in dissecting the contribution of distal *cis*-regulatory elements to development and disease.

METHODS

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

Accession codes. NCBI Gene Expression Omnibus: ATAC-seq and microarray data are available at GSE64059.

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

ACKNOWLEDGMENTS

We thank T. Fazzio, J. Dekker and S. Wolfe for helpful discussions and E. Sontheimer and M. Ziller for critical reading of the manuscript. We thank K. Eggan (Harvard University) and A. McMahon (University of Southern California) for providing the V6.5 ESCs and the *Rosa26* targeting vector backbone, respectively. We are grateful to S. Hainer, H. Belaghzal, K. Morrison and the University of Massachusetts Medical School morphology core for technical help and discussions. R.M. is supported by The Leona M. and Harry B. Helmsley Charitable Trust (2015PG-T1D057 and 2015PG-T1D035), a Charles H. Hood Foundation Child Health Research Award, the Glass Family Charitable Foundation and the US National Institutes of Health (NIH; UC4 DK104218). M.G. is supported by a grant from the NIH (1R01HD080224-01A1).

AUTHOR CONTRIBUTIONS

N.A.K., H.P. and R.M. were responsible for the conception, design and interpretation of experiments. M.G. conceived the bioinformatic approach for novel enhancer identification. N.A.K., H.P. and R.M.G. conducted experiments. B.T. and N.J.S. performed bioinformatics analyses. All authors wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Ernst, J. *et al. Nature* **473**, 43–49 (2011).
2. Rada-Iglesias, A. *et al. Nature* **470**, 279–283 (2011).
3. Hnisz, D. *et al. Cell* **155**, 934–947 (2013).
4. Parker, S.C.J. *et al. Proc. Natl. Acad. Sci. USA* **110**, 17921–17926 (2013).
5. Gilbert, L.A. *et al. Cell* **154**, 442–451 (2013).
6. Konermann, S. *et al. Nature* **500**, 472–476 (2013).
7. Mali, P. *et al. Science* **339**, 823–826 (2013).
8. Esvelt, K.M. *et al. Nat. Methods* **10**, 1116–1121 (2013).
9. Kearns, N.A. *et al. Development* **141**, 219–223 (2014).
10. Cheng, A.W. *et al. Cell Res.* **23**, 1163–1171 (2013).
11. Gilbert, L.A. *et al. Cell* **159**, 647–661 (2014).
12. Konermann, S. *et al. Nature* **517**, 583–588 (2015).
13. Gao, X. *et al. Nucleic Acids Res.* **42**, e155 (2014).
14. Whyte, W.A. *et al. Nature* **482**, 221–225 (2012).
15. Gray, S. & Levine, M. *Genes Dev.* **10**, 700–710 (1996).
16. Mendenhall, E.M. *et al. Nat. Biotechnol.* **31**, 1133–1136 (2013).
17. Yeom, Y.I. *et al. Development* **122**, 881–894 (1996).
18. Niwa, H., Miyazaki, J. & Smith, A.G. *Nat. Genet.* **24**, 372–376 (2000).
19. Yoshimizu, T. *et al. Dev. Growth Differ.* **41**, 675–684 (1999).
20. Ivanova, N. *et al. Nature* **442**, 533–538 (2006).
21. Groner, A.C. *et al. PLoS Genet.* **6**, e1000869 (2010).

ONLINE METHODS

Effector and sgRNA plasmids. A human codon–optimized, nuclease-inactive version of Nm Cas9 was gene synthesized with two SV40 nuclear localization signals and a triple Flag tag (GenScript Technologies) and cloned in-frame with either a BirA affinity tag, a KRAB repressor or LSD1 (for amino acid sequences, see **Supplementary Fig. 1**). A *Rosa26* CAGS *loxP* STOP *loxP* DEST vector was generated by inserting a gateway acceptor cassette into a *Rosa26* CAGS *loxP* STOP *loxP* targeting vector (a kind gift from A. McMahon). The various dCas9-effector fusions were inserted into the *Rosa26* targeting vector by standard gateway cloning.

An sgRNA expression vector was generated by replacing the H1 shRNA and tetO IRES RFP-Puro expression cassettes in a pLV-H1TetO-RFP-Puro lentiviral backbone (Addgene plasmid number 36297) with a U6 Nm sgRNA expression cassette harboring two BsmBI sites for gRNA cloning and EF1α Puro selection cassettes (see **Supplementary Fig. 1** for cloning sites). The Nm sgRNA acceptor and expression cassette was generated as a gene block containing a short form of the tracrRNA (ref. 8 and **Supplementary Fig. 1**).

Note: *N. meningitidis* (Nm) dCas9 is an orthogonal and substantially smaller Cas9 variant to the commonly used Cas9 from *S. pyogenes* (Sp) that is about 25% longer in its nucleotide sequence. Although not tested in our study, we expect an orthogonal Sp dCas9-LSD1 to perform similarly to the Nm dCas9-LSD1 and to work on alternative PAM sequences. The ability to use either Sp Cas9 or Nm Cas9 broadens the spectrum of potential target sites in the genome[8,22], although a thorough test in mammalian cells has to be conducted. Further, off-target effects are currently a major concern of Cas9 systems, and appropriate software and multiple independent sgRNAs should be used to minimize the off-target effects that might influence interpretation of a given study.

sgRNA design. sgRNA target sites were obtained by using a custom script (available upon request) that searches for Nm Cas9 PAM sequences NNNNGATT, NNNNGCTT or NNNNGGTT within the enhancer or promoter regions of interest. Bowtie2 was used to map candidate targets to the mouse genome build GRCm38 with sensitive parameters (--local -f -k 10 --very-sensitive-local -L 9 -N 1) to detect potential off-target sites. Only sgRNAs were used that had two or more mismatches to other sites in the genome than the selected target site. All our sgRNAs had no other genomic matches at the alignment stringency used. Oligonucleotides for sgRNA cloning were obtained from Integrated DNA Technologies or Life Technologies and are listed in **Supplementary Table 4**.

ESC culture. V6.5 (a kind gift from K. Eggan) and *Rosa26*-targeted mouse ESC lines were maintained on mouse embryonic fibroblast feeder layers in ESC-Media (KnockOUT DMEM (GIBCO, 10829) supplemented with 15% Hyclone FBS (Thermo Scientific, SH30070.03), 1% GlutaMAX (GIBCO, 35050079), 1% NEAA (Cellgro, 25-025-CI), 100 μM β-mercaptoethanol (Invitrogen, 21985023) and 10^3 U/mL LIF) according to standard methodology[23]. ESC lines were routinely tested to confirm the absence of mycoplasma.

Generation of stable Nm dCas9–effector mESCs. V6.5 mESCs were electroporated (230 V, 500 mF) with a *Rosa26* targeting construct and selected from day 2–8 with 300 μg/mL Geneticin (Life Technologies, 10131035). Picked clones were screened by PCR analysis for integration (**Supplementary Fig. 2**; knock-in primers were GCCGCCTAAAGAAGAGGCTGTGCTTTGG (forward) and TACCGTAAGTTATGTAACGCG (reverse)). The following PCR conditions were used: 2-min initial denaturation at 95 °C, 30 cycles with 30-s denaturation at 95 °C, annealing at 50 °C for 30 s, extension at 72 °C for 2 min and a final extension at 72 °C for 5 min. After transient transfection of a Cre recombinase expression plasmid (Addgene plasmid 13775), individual clones were picked, expanded and screened for removal of the *loxP*-flanked STOP cassette (**Supplementary Fig. 2**; STOP removal primers were TGCTCCTGCCGAGAAAGTATCCATCA TGGC (forward) and CGCCAAGCTCTTCAGCAATATCACGG GTAG (reverse)). The following PCR conditions were used: 1-min initial denaturation at 93 °C and 30 cycles with 20-s denaturation at 95 °C, annealing/extension at 68 °C for 3 min.

Spontaneous differentiation capability of Nm dCas9-effector cells was assessed by injecting ~10^7 undifferentiated cells subcutaneously into 8- to 13-week-old male and female NOD.Cg-*Rag1*^{tm1Mom}*Il2rg*^{tmWjl}/SzJ mice according to institutional guidelines. Teratomas were retrieved and analyzed 3–4 weeks after injection. Hematoxylin and eosin images were acquired on a Nikon Eclipse 90i microscope using a 20× objective (Nikon Plan Fluor 20x/0.50 Ph1 DLL).

sgRNA delivery. *Rosa26*-targeted mESC lines were transduced with lentivirus containing an sgRNA specific to a genomic region of interest. HEK293T cells were used for lentiviral production as described[9]. HEK293T cells were routinely tested to confirm the absence of mycoplasma. ESCs were washed once with PBS and split with 0.25% trypsin (GIBCO, 25200114). Cells were MEF depleted on 0.1% gelatinized plates for 30 min and then collected and counted. Cells were incubated with sgRNA lentivirus on low-attachment plates for 3 h in 2i medium (ESC-Media supplemented with 1 μM PD0325901 (Tocris, 4192) and 3 μM CHIR99021 (Tocris, 4423)) and were then plated onto 0.1% gelatin–coated plates at 12.5×10^4 cells/cm^2. After 48 h, cells were split using 0.25% trypsin and plated in 2i medium supplemented with 1 μM puromycin to select for cells expressing the sgRNA. Transduced cells were maintained in 2i medium supplemented with 1 μM puromycin and split every 3 d. Cells were analyzed at the time of observed morphological changes (8–14 d after infection with sgRNA carrying lentivirus). To examine the effect of dCas9-LSD1 on local histone modifications, we analyzed cells 3 d before morphological changes were observed. To assess the requirement for LSD1 enzymatic activity, we carried out experiments in 2i medium supplemented with 500 nM *trans*-2-phenylcyclopropylamine hydrochloride (TCP) (Tocris 3852). After two passages, TCP was withdrawn where indicated, and cells were analyzed 14 d after TCP withdrawal.

Immunofluorescence analysis. Cells were fixed and processed as described[9]. Phase-contrast images were acquired on a Nikon Eclipse TS100 microscope using a 20× objective (Nikon LWD 20x/0.40 Ph1 ADL). Fluorescent images were acquired on a Nikon

doi:10.1038/nmeth.3325

Eclipse Ti microscope using a 20× objective (Nikon S Plan Fluor ELWD 20x/0.45). We used filter cubes ET CY3 (Chroma, 49004), ET FITC (Chroma, 49002) and ET DAPI (Chroma, 49000) to isolate fluorescence from Alexa 594, Alexa 488 and Hoechst, respectively. Microscope and camera were controlled through NIS-Elements software (Nikon). Adobe Photoshop was used to apply linear contrast adjustments equally across all images. Further details are listed in the **Supplementary Note**. Antibodies were diluted as indicated in **Supplementary Table 5**. OCT4-positive, OCT4-negative and mixed colonies were quantified using NIS-Elements Analysis Software. Between 520 and 1,300 colonies were identified for each condition by Hoechst staining of clusters comprising more than five cells. Colonies were scored as mixed if <50% of cells in the colony expressed OCT4.

Quantitative PCR analysis. Cells were trypsinized and washed once with PBS, and total RNA was isolated with Trizol reagent (Invitrogen, 15596-018) following the manufacturer's instructions. 1 µg of total RNA was reverse transcribed with the Superscript III First Strand Synthesis System (Invitrogen, 18080-051). 15 ng cDNA were used in quantitative PCR analyses with iTAQ Universal Sybr Green Supermix (Bio-Rad, 172-5124) with specific primers listed in **Supplementary Table 6**. Relative gene expression was calculated using the ΔΔCt method. All genes were normalized to *Gapdh*.

RNA-seq processing. RNA-seq data for ESCs and epiblast-like cells (EpiLCs) from a prior study[24] were downloaded from the Short Read Archive (SRP040451). The raw FASTQ files were processed using the UMass Medical School Bioinformatics Core Galaxy pipeline. Briefly, reads were aligned to the mouse ribosomal sequence and discarded from further analysis. Remaining reads were processed with RSEM[25] (version 1.2.7) using the RefSeq[26] gene annotation set for the mouse build GRCm38. The inferred transcripts per million (TPMs) values were used for further analysis.

mESC-specific enhancer definition. Enhancers specific to ESC and EpiLC cellular states were identified on the basis of chromatin immunoprecipitation–sequencing (ChIP-seq) for the H3K27ac histone mark for each cell type[24]. Reads were aligned with Bowtie version 1.0.0 using "-v 0 -a --strata --best -m 1" parameters, and peaks were identified with MACS version 1.4.2 using "--tsize=35 --bw=300 -g 186550000." Each peak called by MACS was initially considered a candidate enhancer. We then rescored the candidate peaks by way of a 300-nt moving window, advancing 10 nt for sequential windows as described[27]. An enrichment score, depth for H3K27ac as compared to the null, was associated to each window, and the maximum enrichment score across the windows was associated to the peaks. Any peaks with scores below a twofold enrichment relative to the null were filtered out. Low-quality peaks, with scores below the 60% quantile, were also filtered out. Each remaining peak was then trimmed according to the read depth associated to 50-nt moving windows. Ends were trimmed until a window score exceeding 20% of scale for the peak at hand was located. Each high-quality peak for one stage, EpiLC or ESC, was then compared to the other by scoring the trimmed region for the alternate stage. After adding a pseudocount of 10 to the score for each stage, those peaks exhibiting at least twofold enrichment were considered to be stage specific.

For each cell type, any stage-specific peak with the nearest gene transcription start site more than 5,000 bases away was identified as a potential distal enhancer. Peaks for which the nearest transcription start site was associated to a transcription factor were selected. To determine the strongest enhancer candidates for the ESC state, we further filtered the list to include only those peaks whose closest transcription factor exhibited at least a fourfold enrichment in TPM for the ESC state relative to the EpiLC state, after adding a pseudocount of 5 TPM for each stage. The resulting list contained 24 candidates (**Supplementary Table 2**). We excluded the 13 intronic enhancers from further analysis. We also excluded enhancer chr8:72326162–72328532 as it was within 5,000 bases of a gene body and enhancer chr8:72332459–72335859 as it was closer to the gene body of a non-transcription factor than to the transcription factor KLF2. We chose to test two of the three candidates near KLF5 and all other remaining candidates, leaving eight putative distal enhancers that are listed in **Supplementary Table 2a**.

ATAC-seq. Assay for transposase-accessible chromatin using sequencing (ATAC-seq) was performed on V6.5 mouse ESCs according to the protocol as described[28]. As starting material for the protocol, V6.5 mouse ESCs were grown on 0.1% gelatinized plates and maintained in 2i medium. Cells were harvested with 0.25% trypsin for 5 min, collected and counted. 50,000 cells were spun down at 500*g* for 5 min at 4 °C and washed once with 50 µL of cold PBS. ATAC-seq library generation and analysis occurred according to standard methodology[28]. Sequencing reads were mapped to the genome using Bowtie2 version 2.1.0 using the parameters "--phred33 -5 9 -3 2." All ATAC-seq data are available through GEO under accession number GSE64059.

Microarray analysis. Total RNA was isolated with Trizol reagent (Invitrogen, 15596-018) following the manufacturer's instructions either directly (*Oct4* experiments) or after trypsinization (*Tbx3* experiments). 250 ng of total RNA were amplified to generate labeled cRNA using the Illumina Total Prep RNA Amplification Kit (Ambion, AMIL1791). 750 ng of each labeled cRNA sample were hybridized to MouseRef-8 v2 Expression BeadChip (Illumina, BD-202-0202) and the chip scanned on an Illumina BeadArray Reader. Raw probe level data were exported from Illumina Bead Studio for further processing in R (http://www.r-project.org/). Raw probe level intensity values were adjusted using the R package limma[29] to apply quantile normalization and background subtraction and to transform to $\log_2$ expression values. To identify differentially expressed genes, we ranked probes by highest difference in $\log_2$ expression value between any two populations. Genes with log fold changes greater than 3 were used for the generation of hierarchically clustered heat maps. Data are displayed relative to the average $\log_2$ expression signal in dCas9-effector–Ctrl sgRNA samples. In instances where multiple probes mapped to the same gene, only expression values for the highest ranked probe were retained. All gene expression data are available through GEO under accession number GSE64059.

Luciferase assay. Reporter plasmids containing a fragment of the *Oct4* distal enhancer or putative *Tbx3* distal enhancer were cloned into the pGL3-Promoter plasmid (Promega, E176A) upstream of an SV40 minimal promoter and firefly luciferase–encoding gene.

For the luciferase assay, 5×10^4 V6.5 mouse ESCs were plated in 12-well plates and transfected with 900 ng of the firefly reporter plasmid, and 100 ng of pRL-SV0 *Renilla* luciferase plasmid (Promega, E223A), using Lipofectamine 2000 (Invitrogen, 11668019). Cells were harvested 48 h after transfection, and the luciferase activities were measured with the Dual-Luciferase Reporter Assay System (Promega, E1910) according to the manufacturer's protocol. The ratio of firefly luciferase to *Renilla* luciferase was calculated, and the data are expressed relative to an empty vector control.

Alkaline phosphatase staining. Cells were harvested 14 d after infection with sgRNA-carrying lentivirus. Cells were fixed with formalin solution (Sigma, HT5014) for 20 min at room temperature and washed three times with 100 mM Tris, pH 8.5. Alkaline phosphatase (AP) expression was detected using the BCIP/NBT Alkaline Phosphatase Substrate Kit IV (Vector, SK-5400) according to the manufacturer's protocol. Overview images were acquired on a Nikon Eclipse SMZ1500 microscope with a 1× objective (Nikon HR Plan Apo 1x WD54).

Chromatin immunoprecipitation. Nm dCas9–effector ESCs transduced with lentivirus containing an sgRNA were plated onto 0.1% gelatin–coated plates at 5×10^4 cells/cm^2 in 2i medium. After 48 h, cells were passaged and plated in 2i medium supplemented with 1 μM puromycin. Cells were maintained in 2i medium with 1 μM puromycin and harvested 9 d later. Cells were cross-linked with fixing solution (11% formaldehyde, 100 mM NaCl, 1 mM EDTA, 50 mM HEPES-KOH (pH 7.6)) and incubated for 10 min at room temperature. The cross-linking reaction was stopped by the addition of glycine to 125 mM. Cells were washed once with PBS, pelleted, frozen and stored at −80 °C. Chromatin immunoprecipitation (ChIP) was performed as described[30]. 10^7 cells and 50 μL of antibody-coupled protein A magnetic beads (NEB) were used for each ChIP. Protein A magnetic beads were coupled to either H3K4me2 (Active Motif, 39141), H3 (Abcam, ab1791), H3K27me3 (Millipore, 07-449), H3K9me3 (Active Motif, 39161) or H3K27ac (Abcam, ab4729) antibodies. ChIP and input samples were used as a template for qPCR analysis using Sybr Fast (KAPA Biosystems, KK4602) with the primers listed in **Supplementary Table 7** or control primers (Active Motif, 71017 and 103727). Relative enrichment for each primer set was expressed as percent input.

Chromosome conformation capture (3C). 3C templates were generated as described previously[31] with minor differences. Briefly, 30×10^6 cells were harvested and cross-linked per condition. Cross-linked cells were digested with HindIII (NEB, 400U) overnight at 37 °C. Purified 3C templates were desalted and concentrated using Amicon Ultra Centrifugal Filter Units (Millipore, UFC503096). Control 3C template was generated from purified BAC DNA covering the *Tbx3* locus (RP23-406M3, CHORI BACPAC). BAC DNA was purified using a Qiagen Large-Construct Kit (12462), and 10 μg of purified BAC DNA were digested overnight with HindIII at 37 °C. 3C primers (**Supplementary Table 8**) were designed around HindIII sites both upstream and downstream of the *Tbx3* transcriptional start site using an online web module (http://3DG.umassmed.edu/)[32]. PCR analysis was performed in triplicate with each primer paired with the anchor for each 3C template. Interaction frequencies were determined by quantifying PCR amplicons on a gel using Bio-Rad Quantity One analysis software. Primer efficiencies were normalized by dividing the average value of each 3C template by the average value of the BAC 3C control template. 3C template generation efficiencies were normalized by calculating interaction frequencies through a control genomic region using gene desert primer pairs. 3C templates were normalized to each other by calculating the log ratio of the average interaction frequencies for each 3C template through the gene desert region. Final normalized interaction frequencies were calculated by multiplying the average BAC-normalized interaction frequencies with the determined normalization factor for each 3C template.

Reproducibility. No sample size estimates were performed to ensure adequate power to detect a prespecified effect size. Experiments were not randomized, and experimenters were not blinded to experimental conditions. No data were excluded from analysis.

22. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. *Nat. Biotechnol.* **31**, 233–239 (2013).
23. Kearns, N.A. *et al. Stem Cell Res.* **11**, 1003–1012 (2013).
24. Buecker, C. *et al. Cell Stem Cell* **14**, 838–853 (2014).
25. Li, B. & Dewey, C.N. *BMC Bioinformatics* **12**, 323 (2011).
26. Pruitt, K.D., Tatusova, T. & Maglott, D.R. *Nucleic Acids Res.* **35**, D61–D65 (2007).
27. Garber, M. *et al. Mol. Cell* **47**, 810–822 (2012).
28. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. *Nat. Methods* **10**, 1213–1218 (2013).
29. Smyth, G.K. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds. Gentleman, R., Carey, V., Huber, W., Irizarry, R. & Dudoit, S.) Ch. 23, 397–420 (Springer, 2005).
30. Chen, P.B. *et al. eLife* **2**, e01557 (2013).
31. Miele, A., Gheldof, N., Tabuchi, T.M., Dostie, J. & Dekker, J. *Curr. Protoc. Mol. Biol.* **74**, 21.11 (2006).
32. Lajoie, B.R., van Berkum, N.L., Sanyal, A. & Dekker, J. *Nat. Methods* **6**, 690–691 (2009).

doi:10.1038/nmeth.3325