LADL: light-activated dynamic looping for endogenous gene expression control

Ji Hun Kim[1,5], Mayuri Rege[1,4,5], Jacqueline Valeri[1], Margaret C. Dunagin[1], Aryeh Metzger[1], Katelyn R. Titus[1], Thomas G. Gilgenast[1], Wanfeng Gong[1], Jonathan A. Beagan[1], Arjun Raj [1,2,3] and Jennifer E. Phillips-Cremins [1,2,3]*

Mammalian genomes are folded into tens of thousands of long-range looping interactions. The cause-and-effect relationship between looping and genome function is poorly understood, and the extent to which loops are dynamic on short time scales remains an unanswered question. Here, we engineer a new class of synthetic architectural proteins for directed rearrangement of the three-dimensional genome using blue light. We target our light-activated-dynamic-looping (LADL) system to two genomic anchors with CRISPR guide RNAs and induce their spatial colocalization via light-induced heterodimerization of cryptochrome 2 and a dCas9-CIBN fusion protein. We apply LADL to redirect a stretch enhancer (SE) away from its endogenous *Klf4* target gene and to the *Zfp462* promoter. Using single-molecule RNA–FISH, we demonstrate that de novo formation of the *Zfp462*-SE loop correlates with a modest increase in *Zfp462* expression. LADL facilitates colocalization of genomic loci without exogenous chemical cofactors and will enable future efforts to engineer reversible and oscillatory loops on short time scales.

The development of tools to manipulate three-dimensional genome folding on demand with spatiotemporal precision is of critical importance for advancing studies in basic science, regenerative medicine, metabolic engineering and synthetic biology. Mouse and human genomes are folded into more than 10,000 loops[1,2], but the functional role for individual and combinations of long-range chromatin interactions in gene expression remains poorly understood. Published strategies for loop engineering involve synthetic transcription factors tethered to dCas9[3,4] or zinc fingers[5,6], and synthetic looping factors have thus far been constitutively expressed or induced over long time scales by the presence of small molecules[3–6]. The paucity of tools to engineer genome folding on short time scales has prohibited scientists' ability to understand the extent to which loops are dynamic and functionally contribute to the kinetics of transcriptional activation.

Results

Here, we engineer synthetic architectural proteins with the capability of forming long-range contacts between distal genomic loci on demand with blue light. We designed LADL in a modular manner with four key components (Fig. 1a, Supplementary Figs. 1 and 2 and Supplementary Tables 1–6). First, we designed a synthetic architectural protein consisting of enzymatically inactive Cas9 (dCas9) tethered to a truncated version of the CIB1 protein (CIBN) from *Arabidopsis thaliana*[7] (Fig. 1b). Second, we recruited the LADL Anchor (dCas9-CIBN) to two genomic target sites with sequence-specific CRISPR guide RNAs (gRNAs) (Fig. 1c). We designed two gRNAs per anchoring genomic target site. Third, we hypothesized that the CRY2 protein from *A. thaliana* could serve as an inducible bridging factor owing to its well-established ability to heterodimerize with CIBN in response to blue light on millisecond time scales in mammalian cells[8,9] (Fig. 1c). Finally, we used blue light of

wavelength 470 nm as the loop inducing agent[10]. Because it is well established that blue light illumination causes CIBN-CRY2 heterodimerization[8,9] and CRY2 oligomerization[8,11,12], we hypothesized that LADL would spatially connect the two anchoring genomic fragments via a light-induced dCas9-CIBN and CRY2 bridge (Fig. 1a). Thus, we designed LADL as a modular, four-component synthetic architectural protein system to spatially connect genomic loci in response to light via facile design of sequence-specific gRNAs.

To determine the conditions in which blue light would induce a spatial chromatin contact, we first employed 24 h of blue light exposure (Fig. 1d). We built a light box to achieve 5 mW cm^{-2} intensity and 1-s pulses at 0.067 Hz as previously reported for optimal CRY2-CIBN heterodimerization[10,13] (detailed in Methods). We confirmed that the light exposure conditions successfully induced CRY2 oligomerization (Supplementary Fig. 3). Mouse embryonic stem (ES) cells were cotransfected to achieve one of four conditions: (1) LADL (anchor (dCas9-CIBN) and bridge + target (CRY2 + gRNA) plasmids), (2) empty anchor control (empty anchor and bridge + target (CRY2 + gRNA) plasmids), (3) empty bridge control (anchor (dCas9-CIBN) and empty bridge (gRNA only) plasmids) or (4) empty target control (anchor (dCas9-CIBN) and (CRY2 only) plasmids). Overall plasmid mass and ratios were adjusted to optimize transfection efficiency (Supplementary Fig. 4).

We exposed transfected cells to 24 h of blue light or dark after puromycin selection (Fig. 1d). ES cell densities were similar across conditions and exhibited morphology characteristic of the v.6.5 feeder-dependent clone after passage onto gelatin (Supplementary Fig. 5). All conditions showed equivalently high expression of pluripotency markers *Oct4*, *Nanog* and *Sox2* and low levels of *Nestin*, suggesting that the pluripotent, self-renewing ES cell state was not compromised by transfection and light induction (Fig. 1e,f and Supplementary Table 7). dCas9-CIBN and CRY2 transcripts were

[1]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA. [2]Epigenetics Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [3]Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [4]Present address: DST-INSPIRE Faculty, Department of Microbiology, Ramanarain Ruia Autonomous College, Matunga, Mumbai, India. [5]These authors contributed equally: Ji Hun Kim, Mayuri Rege. *e-mail: jcremins@seas.upenn.edu
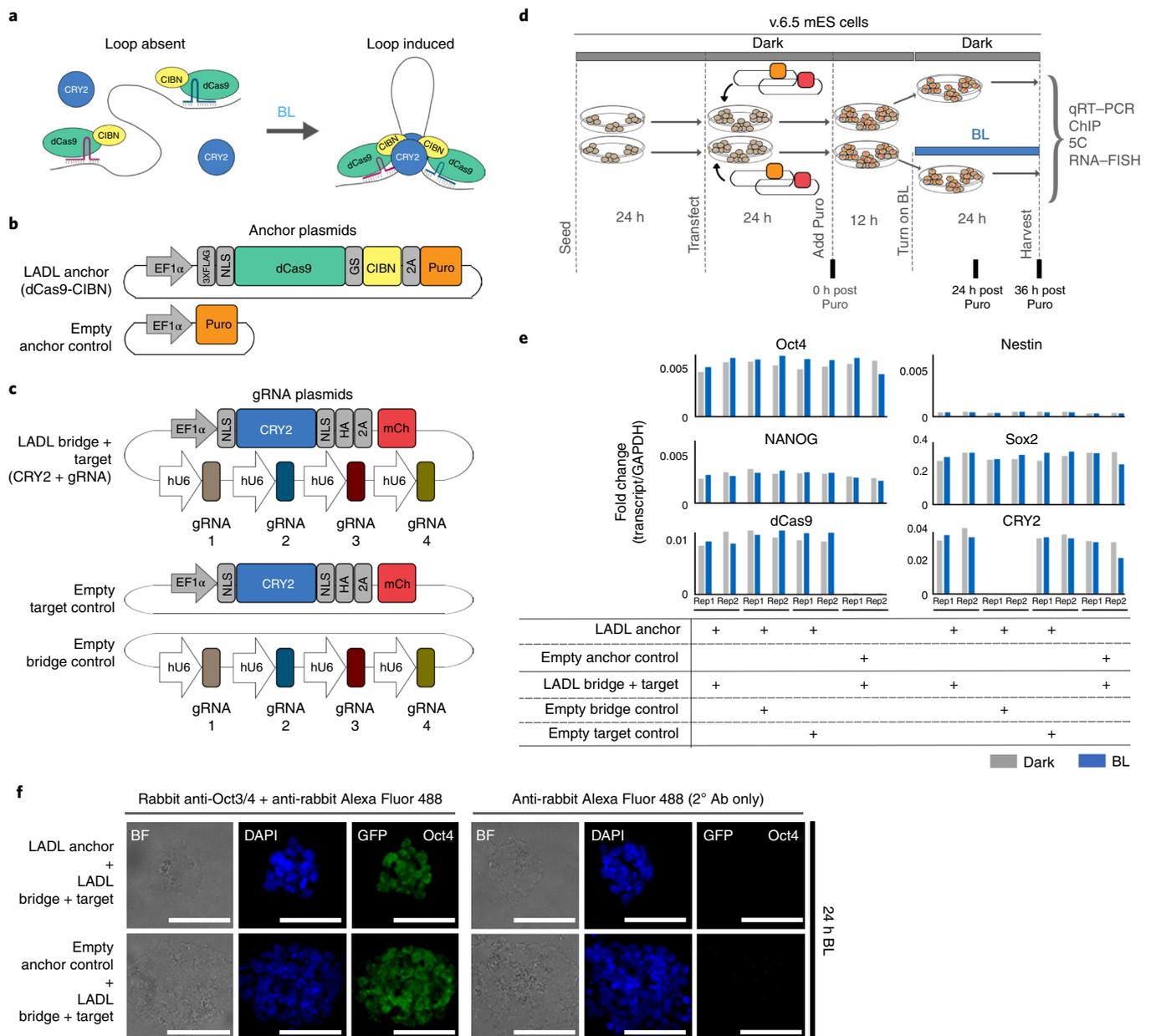
Fig. 1 | Concept, design and implementation of the LADL system. a, Schematic of the LADL system. **b,c**, Schematic of plasmid constructs encoding the puromycin-selectable LADL anchor and empty anchor control (**b**) and the LADL 'bridge + target', the empty target control and the empty bridge control (**c**). **d**, Schematic timeline of seeding, transfection, puromycin selection and blue light illumination of v.6.5 mouse ES cells. **e**, RT–qPCR analysis of *Oct4*, *Nestin*, *Nanog*, *Sox2*, dCas9 and CRY2 transcript levels in cotransfected mouse ES cells 36 h after puromycin selection. Data from two independent experiments are shown as replicates 1 (Rep1) and 2 (Rep2), respectively. BL, blue light. **f**, Immunofluorescence staining for *Oct4* in mouse ES cells cotransfected with the indicated plasmids. Scale bars, 50 μm. Images are representative of three independent experiments.

strongly expressed across all conditions transfected with vectors encoding the transgenes (Fig. 1e). Moreover, equivalent levels were seen in dark and blue light exposure, thus ruling out artifacts caused by differential transgene levels between conditions. Together, these results demonstrate that the two plasmids encoding our synthetic architectural protein system were equivalently expressed and have minimal negative impact on ES cell morphology, viability and pluripotent properties.

We chose an ~800-kb-sized locus around the *Klf4* and *Zfp462* genes as the genomic context for our LADL-engineered loop. The *Klf4* and *Zfp462* genes have high and low expression in pluripotent ES cells, respectively, and are under the control of distal enhancer elements (Fig. 2). As previously reported[14], *Zfp462* loops to at least

four independent putative enhancers (E1, E2, E3, E4) marked by positive enrichment of the histone modification H3K27ac (Fig. 2a and Supplementary Table 8). *Klf4* forms an ~70-kb-sized long-range interaction with a putative stretch enhancer (SE)[14,15]. We reasoned that we could test LADL's performance with a 'redirect and reinforce' strategy in which we spatially redirected the SE away from *Klf4* and reinforced its new interaction with the *Zfp462* promoter. To avoid disrupting endogenous transcription factor and architectural protein binding sites, we designed LADL gRNAs directly adjacent to, but not overlapping, H3K27ac and accessible chromatin at the *Klf4* SE and *Zfp462* promoter (Fig. 2a–c, blue and magenta gRNA markers, respectively, and Supplementary Table 8). We used chromatin immunoprecipitation (ChIP) followed by
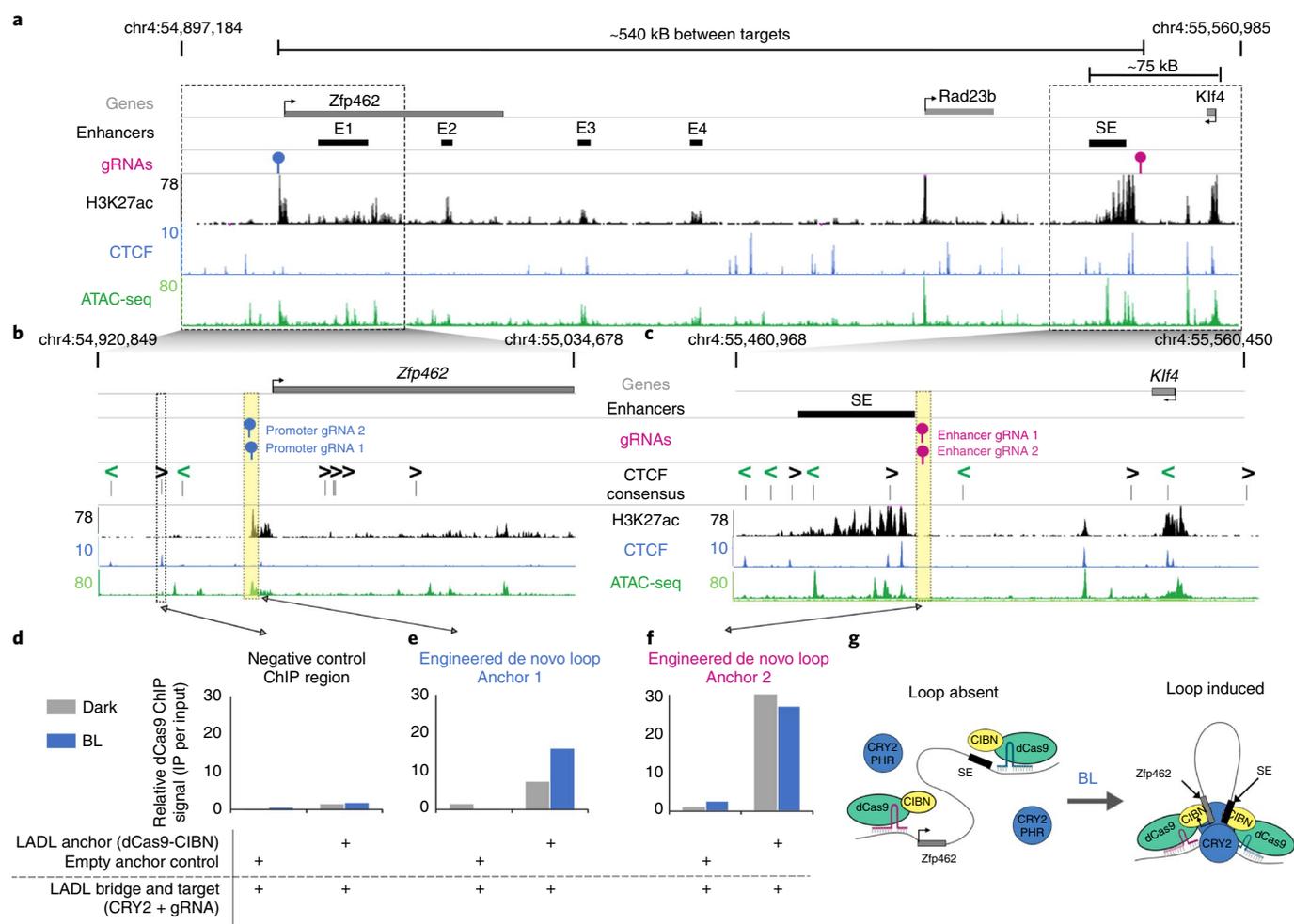
Fig. 2 | Chromatin binding of LADL anchor (dCas9-CIBN) at the engineered sites increases after blue light exposure. a, Genome browser tracks overlaid at the *Zfp462* and *Klf4* genes and their target enhancers (chr4:54,897,184–55,560,985; mm9 reference genome). SE, the *Klf4* SE. E1, E2, E3, E4, the *Zfp462* enhancers. **b,c**, Zoomed-in genome browser tracks at the regions surrounding the *Zfp462* promoter (**b**) and the SE (**c**). **d–f**, ChIP–qPCR data for the negative control chromatin site (IP, immunoprecipitation) (**d**), engineered gRNAs at the *Zfp462* promoter (**e**) and engineered gRNAs at the *Klf4* SE in cotransfected mouse ES cells after 24 h of dark or blue light exposure (**f**). Data were acquired from one ChIP–qPCR experiment. **g**, Model illustrating findings from the ChIP–qPCR data.

quantitative PCR (ChIP–qPCR) to confirm recruitment of the LADL system to the specifically targeted genomic locations (Fig. 2d–f and Supplementary Table 9). Using an anti-FLAG antibody, we demonstrated strong enrichment of FLAG-tagged dCas9-CIBN in the dark at both the *Zfp462* promoter (Fig. 2e) and the *Klf4* SE (Fig. 2f), but not a non-specific genomic region (Fig. 2d). This enrichment was not observed when the LADL anchor was absent (empty anchor control). Thus, the LADL anchor can be effectively targeted to genomic loci adjacent to accessible chromatin using two gRNAs.

We next set out to determine whether a spatial contact was induced by LADL in response to blue light. We hypothesized that an engineered long-range contact between our two targeted genomic fragments might alter dCas9-CIBN ChIP–qPCR signal due to indirect immunoprecipitation from the distal, spatially proximal fragment (Fig. 2g). We found that the intensity of dCas9-CIBN ChIP signal is altered after blue light illumination, increasing more than two fold at the *Zfp462* promoter and slightly decreasing at the *Klf4* SE (Fig. 2e,f). We then directly assessed higher-order chromatin architecture with chromosome-conformation-capture-carbon-copy (5C)[14,16–18] (Supplementary Tables 10–12). We generated a high-resolution map of long-range interactions for all genomic

fragments in an ~3.5-Mb region around the *Klf4* and *Zfp462* genes in the conditions of (1) LADL (anchor + bridge + target) after 24 h of blue light, (2) LADL (anchor + bridge + target) in dark and (3) empty target control (anchor + bridge only) in dark (Fig. 3a,b and Supplementary Fig. 6). On blue light illumination, a new long-range contact is gained between the SE and *Zfp462* in mouse ES cells transfected with LADL vectors (Fig. 3c–e and Supplementary Fig. 7a). The engineered loop is specific to the LADL + blue light condition and not present in LADL + dark or empty target + dark controls (Fig. 3c–e and Supplementary Fig. 7a). We reproduced the de novo *Zfp462-Klf4* SE loop in LADL-transfected ES cells after 24 h of blue light illumination at a lower intensity of 1.5 mW cm⁻² (Supplementary Fig. 7b), as well as in three more independent experiments at 5 mW cm⁻² (Supplementary Fig. 7c–f). An additional one-sided gRNA negative control (anchor + bridge + one-sided target (CRY2 + promoter-targeted sgRNA)) did not show looping signal (Supplementary Fig. 7c,f). Classic 4C looping efficiency plots from the viewpoint of both gRNA anchors across five independent experiments confirmed that the median strength of the *Zfp462–Klf4* SE interaction increased ~2.0–2.5-fold in the LADL + blue light versus the LADL + dark condition (Supplementary Fig. 8). Together, these results demonstrate that LADL can
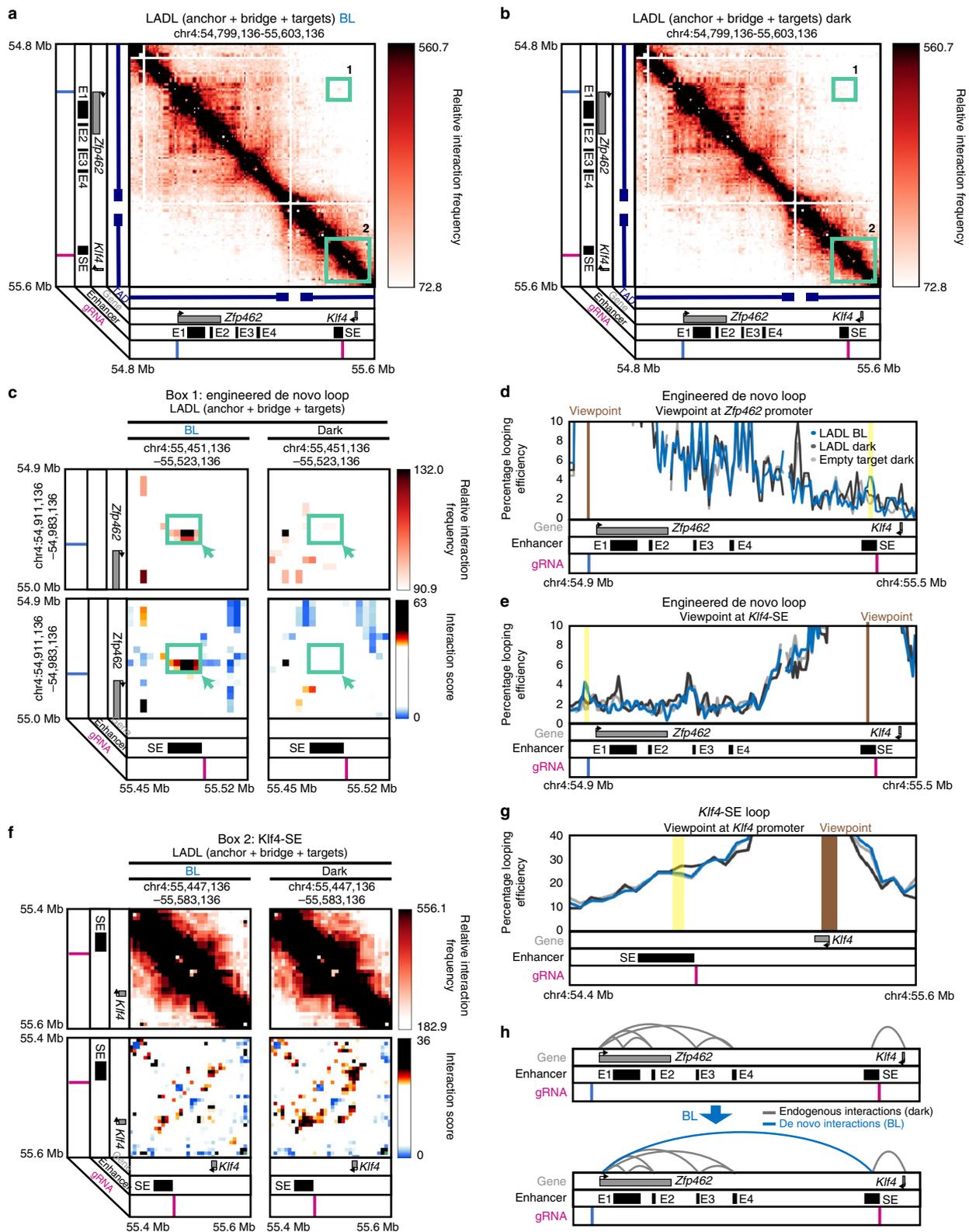
Fig. 3 | LADL redirects a long-range interaction between an SE and a new target gene on blue light illumination. a,b, Heat map of long-range interactions around an ~800-kb genomic region encompassing the *Klf4* and *Zfp462* genes. SE, the *Klf4* stretch enhancer. E1, E2, E3, E4, the *Zfp462* enhancers. Mouse ES cells were cotransfected with LADL (anchor + bridge + target) plasmids and then exposed to 24 h of blue light illumination (**a**), or in dark (**b**). Box 1, the target de novo engineered loop between the pluripotency-specific *Klf4* SE and the *Zfp462* promoter. Box 2, *Klf4* interaction with its upstream, pluripotency-specific SE. Additional controls are shown in Supplementary Fig. 6. **c,f,** Zoomed-in heat maps of box 1 (**c**) and box 2 (**f**). Top, relative interaction frequency 5C signal. Bottom, distance-corrected interaction score 5C signal. **d,e,g,** Classic 4C looping efficiency plots from the viewpoint of the *Zfp462* promoter-targeted gRNA (**d**), the SE targeted gRNA (**e**) and the *Klf4* promoter (**g**). Additional negative controls and replicates are shown in Supplementary Figs. 7–10. **h,** Model of looping interaction reconfiguration in response to LADL and blue light illumination.
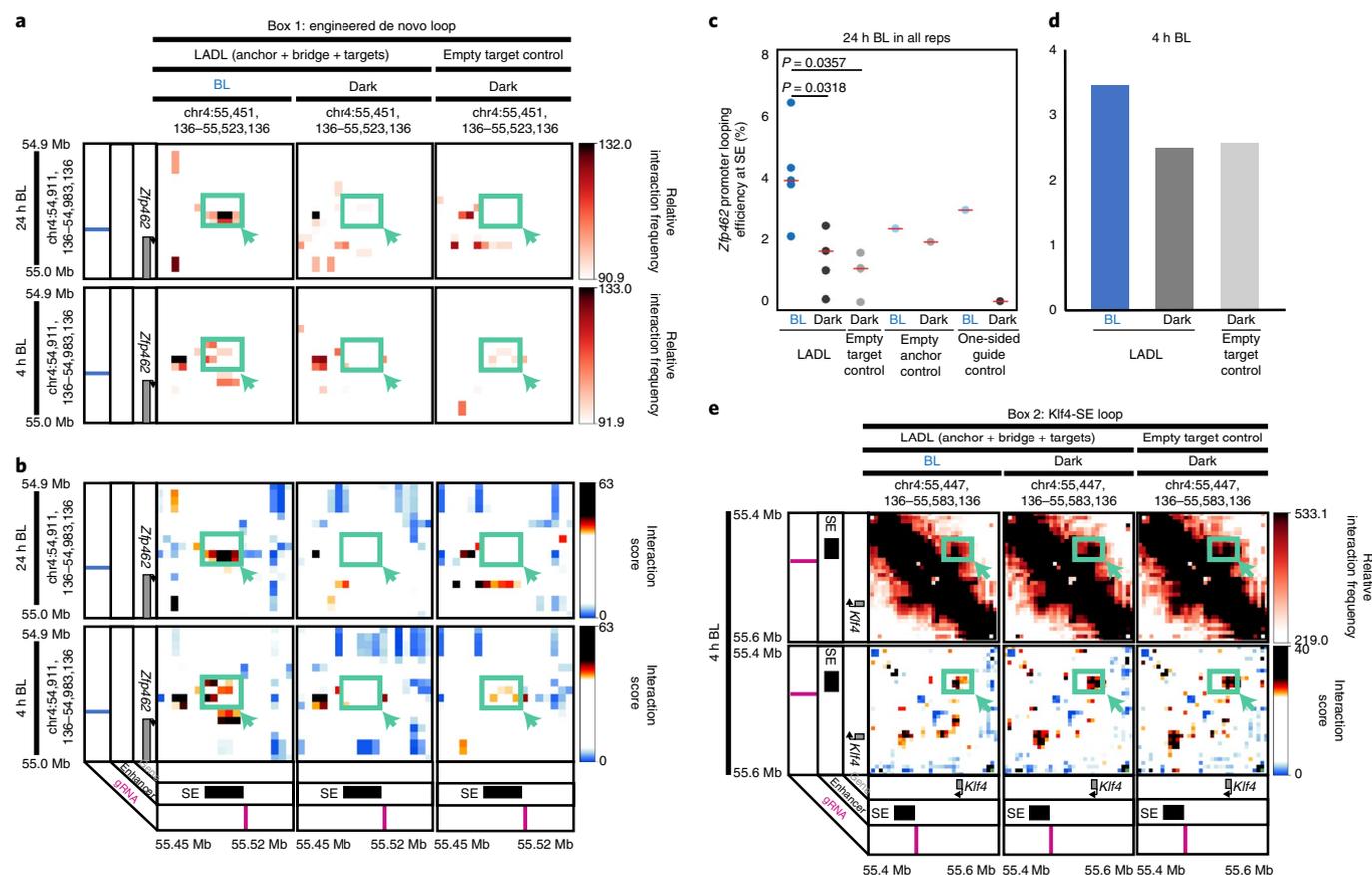
Fig. 4 | The LADL-engineered long-range interaction occurs as early as 4 h after blue light illumination. a,b, Zoomed-in heat maps corresponding to box 1 from Fig. 3a,b for (**a**) relative interaction frequency 5C signal and (**b**) distance-corrected interaction score 5C signal. **c,d**, Percentage looping efficiency for the *Zfp462* promoter and *Klf4*–SE interaction after (**c**) 24 h (*n* = 5 independent experiments) and (**d**) 4 h (*n* = 1 experiment) of blue light illumination. Red bars, medians of each condition. *P* values computed using the unpaired two-sided Mann–Whitney *U* test. **e**, Zoomed-in heat maps corresponding to box 2 from Fig. 3a,b for (top) relative interaction frequency 5C signal and (bottom) distance-corrected interaction score 5C signal.
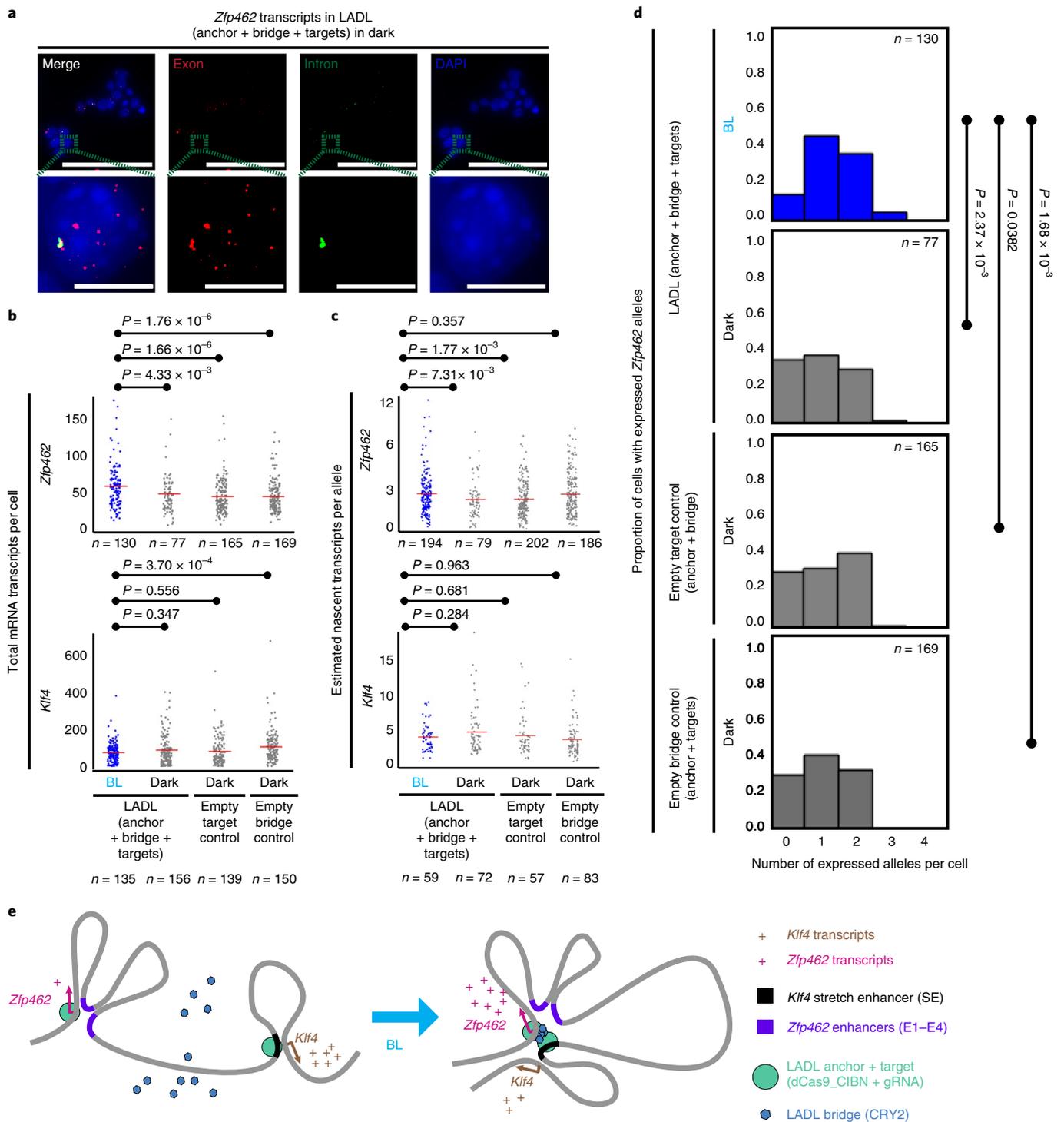
form a new long-range interaction between two genomic fragments in a blue-light-dependent manner.

We next queried whether endogenous chromatin interactions were disrupted during the process of redirecting the *Klf4* SE to *Zfp462*. In wild-type mouse ES cells, the *Klf4* gene forms a strong long-range interaction with its target SE[14,15]. We detected high-frequency *Klf4*–SE interactions in both LADL (anchor + bridge + target) and the empty target control (anchor + bridge only) in the dark (Fig. 3f,g and Supplementary Fig. 9a). The *Klf4*-SE loop remained largely intact, with slightly reduced contact frequency, in LADL-transfected ES cells after 24 h of 5 mW cm⁻² blue light illumination (Fig. 3f,g and Supplementary Fig. 9a). We observed a slight reduction in *Klf4*-SE interaction strength in the LADL + blue light condition compared with that for negative controls across *n* = 4 total replicates at 5 mW cm⁻² and one additional replicate at 1.5 mW cm⁻² blue light exposure (Supplementary Fig. 9b–f and Supplementary Fig. 10). Our data indicate that the endogenous hub of enhancer–enhancer and enhancer–*Zfp462* interactions[14,15] present in pluripotent ES cells was largely undisturbed across all conditions (Supplementary Fig. 11). Together, these data demonstrate that the *Klf4* SE can be redirected across a population of cells to *Zfp462*, with a slight disruption in endogenous *Klf4*–SE interactions and a negligible effect on endogenous *Zfp462*–enhancer interactions (Fig. 3h).

To gain insight into the time scale on which the LADL-induced loops are formed, we mapped chromatin architecture in LADL-engineered ES cells with 5C after varying the time scale of blue light exposure (Supplementary Fig. 12). We observed the de novo

engineered interaction between *Zfp462* and the *Klf4* SE after as little as 4 h of blue light illumination, but it was slightly shifted spatially compared to the LADL-induced contact observed at 24 h of blue light (Fig. 4a,b and Supplementary Fig. 8a). Classic 4C looping efficiency plots confirmed that the engineered *Zfp462*-*Klf4* SE contact showed an increase in interaction frequency between LADL blue light and dark conditions after 4 h of light exposure (Fig. 4c,d). Consistent with our observations at 24 h, the *Klf4*–SE interaction was only slightly reduced in LADL-engineered ES cells after 4 h of 5 mW cm⁻² blue light compared to the LADL + dark and empty target + dark conditions (Fig. 4e). Together, these results indicate that LADL can enable the formation of long-range interactions on demand in as little as 4 h after application of the induction stimulus. Chemical induction of looping is reported to occur on the time scale of 24 h or more[3]; thus LADL may provide an advance in shortening the time scale of loop induction.

To understand the possible functional role of the de novo engineered loop, we next measured the effects of the LADL-engineered interactions on gene expression. We performed single-molecule RNA–fluorescence in situ hybridization (FISH) to assess *Zfp462* and *Klf4* expression changes on a single-cell basis after 24 h of blue light illumination (Fig. 5a and Supplementary Table 13). The mean number of total *Zfp462* messenger RNA transcripts per cell (53.47; 95% confidence interval (CI): 48.54 < μ$_{Zfp462\_LADL + blue\ light}$ < 58.40) was significantly higher in LADL + blue light compared to LADL + dark (43.68; 95% CI: 38.24 < μ$_{Zfp462\_LADL + dark}$ < 49.11), empty target control + dark (40.14; 95% CI: 36.69 < μ$_{Zfp462\_Empty\ target\ control + dark}$

Fig. 5 | Functional effect of the LADL-engineered de novo interaction on endogenous gene expression. a, Full field (top row) and zoomed microscopy images (bottom row) from single-molecule RNA–FISH analysis of LADL-engineered mouse ES cells in dark. Quantitative image analysis of *Zfp462* and *Klf4* transcripts from LADL-engineered mouse ES cells after 24 h of blue light (detailed in Methods). Scale bars, 100 μm (top) or 10 μm (bottom). Images are representative of *n* = 3 independent experiments. **b,c**, Strip charts representing the total number of mRNA transcripts per cell (**b**) and the estimated number of nascent transcripts per allele for *Zfp462* (upper row) and *Klf4* (lower row) (**c**). Red bars, means of each condition. **d**, Histograms represent the proportion of cells with a specific number of actively expressing *Zfp462* alleles. *P* values computed using the unpaired one-tailed Mann–Whitney *U* test. Sample sizes represent number of cells (**b,d**) or number of active transcription alleles (**c**). Additional replicates from independent experiments are shown in Supplementary Fig. 13. **e**, Schematic diagram illustrating the link between LADL-engineered de novo loop formation and changes in endogenous gene expression.

< 43.59), or empty bridge control + dark (40.33; 95% CI: 37.24 < $\mu_{Zfp462\_Empty\ bridge\ control + dark}$ < 43.41) (Fig. 5b). Moreover, the mean of estimated nascent *Zfp462* transcripts per allele (2.62; 95%

CI: 2.43 < $\mu_{Zfp462\_LADL + blue\ light}$ < 2.81) was also significantly higher in LADL + blue light compared to LADL + dark (2.25; 95% CI: 1.99 < $\mu_{Zfp462\_LADL + dark}$ < 2.51) and empty target control + dark

$(2.26; 95\% \text{ CI: } 2.1 < \mu_{Zfp462\_Empty\ target\ control + dark} < 2.42)$ (Fig. 5c). Total *Zfp462* mRNA transcripts per cell were reproducibly upregulated upon LADL-induced interaction formation in two out of the three RNA–FISH experiments (Supplementary Fig. 13a,c). Moreover, estimated nascent *Zfp462* transcripts per allele were reproducibly upregulated upon LADL-induced interaction formation in three out of three experiments (Supplementary Fig. 13b,d). We also queried *Klf4* expression and observed that total mRNA and estimated nascent transcript levels were highly variable across experiments, consistent with the slight but non-significant reduction in loop formation in LADL + blue light versus negative controls (Fig. 5b,c and Supplementary Fig. 13a–d). Overall, our data are consistent with our working model that forced spatial interactions between the *Klf4* SE and the *Zfp462* promoter correlate with a modest increase in total mRNA and nascent transcripts of *Zfp462* per cell.

We also hypothesized that an increase in loop frequency might lead to an increase in the number of transcriptionally active alleles within a population that express the target gene. Using single-molecule RNA–FISH, we observed that the number of alleles per cell actively transcribing *Zfp462* was significantly increased in LADL + blue light $(1.32; 95\% \text{ CI: } 1.19 < \mu_{Zfp462\_LADL + dark} < 1.46)$ compared to LADL + dark $(0.97; 95\% \text{ CI: } 0.79 < \mu_{Zfp462\_LADL + dark} < 1.16)$, empty target control + dark $(1.12; 95\% \text{ CI: } 0.99 < \mu_{Zfp462\_Empty\ target\ control + blue\ light} < 1.24)$ or empty bridge control + dark $(1.02; 95\% \text{ CI: } 0.91 < \mu_{Zfp462\_Empty\ bridge\ control + blue\ light} < 1.14)$ (Fig. 5d). The increase in *Zfp462*-expressing alleles in LADL + blue light was reproducible in two out of three RNA–FISH experiments (Supplementary Fig. 13e,f). Our data suggest that LADL-induced formation of the de novo *Zfp462-Klf4* SE interaction can result in an increase in the proportion of alleles expressing *Zfp462* (Fig. 5e and Supplementary Fig. 13).

Discussion

Overall, we present LADL as a new synthetic architectural protein system that is capable of forming inducible long-range interactions in response to light. Our new 'three-dimensional optoepigenetic tools' to engineer chromatin topology will be useful in the future to (1) facilitate loop engagement and reversibility on rapid time scales, (2) enable the previously unachievable ability to oscillate spatial contacts, and (3) overcome signal-to-noise issues in population-based genomics assays by synchronizing chromatin topology across a large population of cells via blue light illumination. Although the interaction strength achieved in this first variation of LADL was modest, we can further optimize LADL-induced contacts in the future by adjusting light intensity, gRNA numbers, or CRY2 bridge size, or by building other light-inducible dimerization systems. At the single locus investigated here, we see that an ~2–2.5-fold increase in interaction frequency correlated with a modest ~1.2–1.3-fold increase in gene expression. It will be important to use LADL and other three-dimensional genome engineering tools in the future to obtain a truly quantitative understanding of the relationships among loop strength, enhancer activity and gene expression levels. We also see opportunities to use LADL to form possible phase separated nuclear bodies or hubs of multi-way chromatin contacts. Should three-dimensional genome engineering prove useful for correcting chromatin misfolding patterns in disease, LADL will open up future opportunities for spatial targeting of specific cell types in vivo for dynamic looping and control of gene expression on short time scales.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41592-019-0436-5.

References

1. Phanstiel, D. H. et al. Static and dynamic DNA loops form AP-1-bound activation hubs during macrophage development. *Mol. Cell* **67**, 1037–1048 e1036 (2017).
2. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
3. Morgan, S. L. et al. Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* **8**, 15993 (2017).
4. Hao, N., Shearwin, K. E. & Dodd, I. B. Programmable DNA looping using engineered bivalent dCas9 complexes. *Nat. Commun.* **8**, 1628 (2017).
5. Deng, W. et al. Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* **149**, 1233–1244 (2012).
6. Deng, W. et al. Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* **158**, 849–860 (2014).
7. Liu, H. et al. Photoexcited CRY2 interacts with CIB1 to regulate transcription and floral initiation in *Arabidopsis*. *Science* **322**, 1535–1539 (2008).
8. Bugaj, L. J., Choksi, A. T., Mesuda, C. K., Kane, R. S. & Schaffer, D. V. Optogenetic protein clustering and signaling activation in mammalian cells. *Nat. Methods* **10**, 249 (2013).
9. Kennedy, M. J. et al. Rapid blue-light-mediated induction of protein interactions in living cells. *Nat. Methods* **7**, 973–975 (2010).
10. Konermann, S. et al. Optical control of mammalian endogenous transcription and epigenetic states. *Nature* **500**, 472–476 (2013).
11. Ozkan-Dagliyan, I. et al. Formation of *Arabidopsis* cryptochrome 2 photobodies in mammalian nuclei: application as an optogenetic DNA damage checkpoint switch. *J. Biol. Chem.* **288**, 23244–23251 (2013).
12. Che, D. L., Duan, L., Zhang, K. & Cui, B. The dual characteristics of light-induced cryptochrome 2, homo-oligomerization and heterodimerization, for optogenetic manipulation in mammalian cells. *ACS Synth. Biol.* **4**, 1124–1135 (2015).
13. Polstein, L. R. & Gersbach, C. A. A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nat. Chem. Biol.* **11**, 198–200 (2015).
14. Beagan, J. A. et al. Local genome topology can exhibit an incompletely rewired 3D-folding state during somatic cell reprogramming. *Cell Stem Cell* **18**, 611–624 (2016).
15. Beagan, J. A. et al. YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. *Genome Res.* **27**, 1139–1152 (2017).
16. Dostie, J. et al. Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* **16**, 1299–1309 (2006).
17. Phillips-Cremins, J. E. et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* **153**, 1281–1295 (2013).
18. Kim, J. H. et al. 5C-ID: increased resolution chromosome-conformation-capture-carbon-copy with in situ 3C and double alternating primer design. *Methods* **142**, 39–46 (2018).

Author contributions

J.E.P.C., M.R., J.V. and A.M. conceptualized the system. M.R., J.H.K., J.V. and W.G. designed and performed the experiments. M.C.D. and A.R. designed and conducted RNA–FISH experiments. M.C.D., A.R. and J.E.P.C. analyzed FISH data. J.H.K., J.A.B., K.R.T., T.G.G. and J.E.P.C. performed the 5C data analysis. J.E.P.C. wrote the manuscript with help from all other authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-019-0436-5.

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

Correspondence and requests for materials should be addressed to J.E.P.

Peer review information: Nicole Rusk was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

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

Methods

Construction of anchor plasmids. *Anchor plasmid backbone assembly.* We cloned the LADL anchor and related control plasmids into a minimally sized backbone for optimal transfections. First, we created a minimal backbone containing the transcriptional terminator bGH along with appropriate restriction enzyme sites. Second, we cloned the individual anchor plasmids into this backbone. To create the minimal backbone, we digested pUC19 plasmid with ZraI and PciI, gel extracted the 1,809-kb fragment containing the ampicillin promoter and open reading frame, and treated with alkaline phosphatase at 37 °C for 60 min. The bGH polyA signal was PCR amplified using primers MRP175, 176 (Supplementary Table 1) and the template Addgene no. 62987 plasmid[19]. MRP175 and MRP176 primers also incorporated requisite restriction for PciI, SnaBI, EcoRI and ZraI sites for use in downstream applications. Thus, the 282-bp PCR product was digested with ZraI and PciI and ligated with the pUC19-derived vector backbone from above to get the anchor backbone plasmid (Plasmid S13.1; see Supplementary Fig. 2a). The anchor plasmids (Supplementary Fig. 2b,c) were cloned into this minimal backbone derived from pUC19 as described below.

LADL anchor (dCas9-CIBN) plasmid. We next built the LADL anchor plasmid (Supplementary Fig. 2b) by ligating multiple fragments in a single-step Gibson assembly. We first had to ensure we were using the appropriate PCR templates for the individual fragments. We had access to Cas9n (Addgene no. 62987 plasmid)[19] as a PCR template, which would need to be mutated at H840A before use as a template for the dCas9 construct. We performed site-directed mutagenesis using the Quikchange II XL mutagenesis kit (Agilent, no. 200521) and mutated the H840A amino acid in Cas9n (Addgene no. 62987 plasmid)[19]. We verified the resultant dCas9 sequence with Sanger sequencing and used it as a PCR template for further cloning. Primers MRP171 and MRP172 were used for the site-directed mutagenesis to mutate the H840A in the Cas9n construct (Supplementary Table 1).

Next, we PCR amplified the individual inserts (EF1a, 3XFLAG-dCas9, GS-CIBN and 2A-Puro) using the high-fidelity Q5 polymerase (NEB) (Supplementary Table 2). We verified a single band of expected size on an agarose gel, treated with DpnII, purified with a Qiagen PCR clean-up kit and quantified using the Nanodrop. The inserts were ligated into the SnaBI, EcoRI digested anchor backbone plasmid S13.1 (Supplementary Fig. 2a) using the NEB Gibson assembly mix (100 ng vector backbone, 0.3 pmol of total DNA fragments for four inserts ligated for 60 min at 50 °C in a thermocycler). Finally, we added the Kozak sequence upstream of the start site of 3XFLAG by digesting an intermediate plasmid with BamHI, ClaI and ligating the annealed and phosphorylated double-stranded DNA oligo (MRP207 and MRP208; Supplementary Table 1) to get the 'LADL anchor' plasmid (Supplementary Fig. 2b).

Empty anchor control plasmid. As a negative control for the LADL anchor, we created a vector containing the EF1a promoter-puromycin using the same backbone plasmid S13.1 (see Supplementary Fig. 2a) using a Gibson assembly. The individual inserts (EF1a and puromycin) were PCR amplified (primers detailed in Supplementary Table 2) using the high-fidelity Q5 polymerase (NEB), verified to give a single band of the expected size on an agarose gel, DpnI treated, cleaned-up using the Qiagen PCR clean-up kit and quantified using Nanodrop. These were then cloned into the SnaBI, EcoRI digested Addgene no. 58771 plasmid[20] used as a backbone with a Gibson assembly. Positive clones were screened using diagnostic digests and verified by Sanger sequencing to give the 'empty anchor control plasmid' (Supplementary Fig. 2c).

Construction of gRNA plasmids. *Overview.* To achieve multiplexing of four gRNAs in a single plasmid, we adopted and modified the system developed by the Yamamoto laboratory, where single gRNAs are cloned into individual plasmids first and then combined together using the Golden Gate assembly[20]. Our multiplexed four gRNA plasmids have two versions: without or with soluble CRY2 (plasmids in Supplementary Fig. 2i,l). The multiplexed gRNA plasmid without soluble CRY2 was created first and sequence verified to contain the multiplexed gRNAs in the expected order (Supplementary Fig. 2i). Subsequently, the soluble CRY2 expression construct was inserted into this multiplexed gRNA plasmids (Supplementary Fig. 2l). In the current study, we designed two gRNAs per engineered loop anchor (two gRNA × two loop anchors). Published CRISPRa and CRISPRi studies[13,21] oftentimes use multiple guides and, although these studies have a completely different goal, this did influence our decision. We have only tried two guides on each loop anchor and cannot predict whether loop efficiency would be altered with more or less guides. All gRNA primer sequences are provided in Supplementary Table 3.

Individual gRNA plasmids without soluble CRY2. The Addgene no. 58768 plasmid was digested with SnaBI and EcoRI to excise the Cas9 open reading frame and ligated with the annealed and phosphorylated dsDNA oligo (MRP173, MRP174; see Supplementary Table 1) to get the ampicillin-resistant S12.1 gRNA multiplex backbone plasmid (Supplementary Fig. 2d). For step 1 of multiplexing, sgRNAs were cloned into one of the following plasmids: S12.1 (ampicillin resistant) (Supplementary Fig. 2e), B1 (Addgene no. 58778)[20] (Supplementary Fig. 2f), B2 (Addgene no. 58779)[20] (Supplementary Fig. 2g) or B3 (Addgene no. 58780)[20]

(Supplementary Fig. 2h). The gRNA sequences and the plasmids they were cloned into are listed in Supplementary Tables 3 and 4, respectively. Positive clones were screened with a diagnostic digest and verified by Sanger sequencing using the U6 promoter primer (GAGGGCCTATTTCCCATGATTCC).

Multiplexed gRNA plasmids without soluble CRY2. For step 2 of multiplexing, plasmids containing single gRNAs were mixed together and multiplexed using the NEB Golden Gate assembly mix[20]. Specifically, we used 75 ng of the gRNA 129 plasmid clone (Supplementary Fig. 2e) and 114 ng of gRNA 135 plasmid, gRNA 115 plasmid, gRNA 117 plasmid each (Supplementary Fig. 2f–h, respectively). For optimal efficiency, we performed the Golden Gate assembly using the following cycling parameters: (37 °C, 5 min → 16 °C, 5 min) × 30 cycles → 55 °C, 5 min. This procedure consistently gave us >90% efficiency in multiplexing four gRNAs at a time. Sanger sequencing for the multiplexed plasmids was performed using each individual gRNA as the sequencing primer (therefore, not present itself in the Sanger trace) and checking for the presence of the adjacent gRNA sequence. The sequence verified multiplexed gRNA plasmid without soluble CRY2 was named 'empty bridge control' (Supplementary Fig. 2i) and the individual gRNAs present in this plasmid are listed in Supplementary Table 5. All gRNA plasmids were transformed into NEB Stable Competent Cells (NEB, C3040I) to minimize recombination between repetitive U6 promoters present in the multiplexed plasmid.

Multiplexed gRNA plasmids with soluble CRY2. For greater modularity, we built a separate plasmid as the source of soluble CRY2 that could be inserted into any gRNA expressing plasmid. We created the soluble CRY2 cassette in the plasmid S13.1 (Supplementary Fig. 2a). The individual inserts (EF1a, CRY2PHR and 2A_mCherry) were PCR amplified from the templates listed in Supplementary Table 2 using the high-fidelity Q5 polymerase (NEB), verified to give a single band of the expected size on an agarose gel, DpnI treated, cleaned-up using the Qiagen PCR clean-up kit and quantified using Nanodrop. The three PCR products were then cloned into the EcoRI + SnaBI digested S13.1 plasmid (Supplementary Fig. 2a) using a Gibson assembly. Positive clones were screened using diagnostic digests and verified by Sanger sequencing to give the plasmid called the 'empty target control plasmid' (Supplementary Fig. 2k).

To demonstrate the modularity above, we used the empty bridge control (Supplementary Fig. 2i) as a backbone to create a multiplexed plasmid that also contains the soluble CRY2 transgene (termed the LADL bridge + target plasmid) (Supplementary Fig. 2l). To build this vector, we digested the fragment containing the EF1alpha promoter and the CRY2-HA-2A-mCherry transgene from the empty target control plasmid (Supplementary Fig. 2k) with SnaBI + EcoRI and gel extracted the band. The fragment was then ligated into the multiplexed gRNA empty bridge control plasmid (Supplementary Fig. 2i) digested with SnaBI + EcoRI using NEB Quick Ligase. Positive clones were screened using diagnostic digests and verified by Sanger sequencing to give the plasmid called 'LADL bridge + target' (Supplementary Fig. 2l). The individual gRNAs present in the LADL bridge + target plasmid are listed in Supplementary Table 6.

One-sided guide control plasmid. We included an additional control containing two gRNAs that target the *Zfp462* promoter, but without the two gRNAs that target the *Klf4* SE. Initially, CRY2 was cloned into S12.1 using the EcoRI and SnaBI sites, and then gRNA 115 was cloned into S12.1. gRNA 117 was cloned into B1 (Addgene no. 58778)[20]. We multiplexed gRNA 115 plasmid and gRNA 117 plasmid as well as B2 (Addgene no. 58779)[20] and B3 (Addgene no. 58780)[20] together to give the 'one-sided guide control' plasmid (Supplementary Fig. 2m and Supplementary Table 6).

CRY2olig and derived plasmids. The CRY2olig plasmid (Addgene no. 60032) used in the functional validation of the light box was a gift from C. Tucker[22]. We used this plasmid as a template to amplify mCherry. First, we mutated the two BbsI sites with two synonymous point mutations to ensure they would not be cut by BbsI during gRNA cloning. Thus, we derived our 'CRY2olig mut 2–1 plasmid' (Supplementary Fig. 2j) from the Addgene no. 60032 plasmid by sequentially mutating the two BbsI sites using the NEB Q5 Site-Directed Mutagenesis kit. At site 1 nucleotide 729 was changed from a C to an A using primers AM_43 and AM_44. At site 2, nucleotide 2574 was changed from a G to an A using primers AM_45 and AM_46. The sequences of all primers used for cloning this vector are given in Supplementary Table 1.

Tissue culture and cell preparations. *Mouse ES cell culture.* Murine v.6.5 ES cells (v.6.5; genotype 129SvJae × C57BL/6; male) were purchased from Novus Biologicals. Mouse ES cells were cultured in the following medium: DMEM (Corning, 10013CV) supplemented with 15% Hyclone fetal bovine serum (FBS) (Thermo Fisher, SH3007003E), 1× MEM non-essential amino acid (Thermo Fisher, 11140076), 2 mM L-glutamine (Thermo Fisher, 25030164), 100 U ml^{-1} penicillin-streptomycin (Thermo Fisher, 15140163), 1× 2-mercaptoethanol (EMD, Millipore, ES-007-E), 10^3 U ml^{-1} Leukemia Inhibitory Factor (EMD Millipore, ESG1107) and maintained on Mitomycin-C (Fisher Scientific, BP2531-2) inactivated mouse embryonic fibroblast (pMEF) feeders at 37 °C and 5% CO_2, as previously described[14,15]. Before transfection and puromycin selection, mouse ES cells were passaged once on gelatin-coated feeder-free plates to minimize pMEF contamination.

Mouse embryonic fibroblast (MEF) culture. MEFs were cultured at 37 °C and 5% CO_2 in DMEM (Corning, 10013CV) supplemented with 10% FBS (Atlanta Biologicals, S11550), 1× MEM non-essential amino acid (Thermo Fisher, 11140076), 2 mM L-glutamine (Thermo Fisher, 25030164), 100 U ml⁻¹ penicillin-streptomycin (Thermo Fisher, 15140163). At ~90% confluency, MEFs were inactivated in 10 µg ml⁻¹ Mitomycin-C (Fisher Scientific, BP2531-2) in culture media at 37 °C and 5% CO_2 for 2 h. Then, 1.5×10^6 inactivated MEFs were plated on a 10 cm gelatin-coated plate to be used as a feeder layer for mouse ES cell culture.

Gelatin-coating plates. All plates for mouse ES cells and MEF cultures were coated with EmbryoMax 0.1% Gelatin Solution (EMD Millipore, ES-006-B) for ~20 min at room temperature and dried before plating cells.

Transfection conditions. We seeded v.6.5 mouse ES cells at 2.4×10^4 cells cm⁻² on gelatin-coated feeder-free plates. At 24 h post-seeding, we cotransfected with 1.5 fM cm⁻² of the puro-resistant LADL anchor plasmid and LADL bridge + target (CRY2 + gRNAs) plasmid for 24 h in dark using Lipofectamine2000 (Thermo Fisher, 11-668-019) according to the manufacturer's protocol. All plasmids to be transfected were maxi-prepped with Qiagen Endofree Maxiprep kit (Qiagen, 12362) before transfection. At 24 h post-transfection, cells were selected in puromycin-selection media (3.5 µg ml⁻¹ of puromycin in mouse ES cell culture media) for 36 h. Mouse ES cells were either exposed to blue light or cultured in the dark during puromycin selection before collection, as outlined in Fig. 1d. Transfection efficiency of the two plasmids was evaluated by visually assessing the number of mCherry positive cells that survive puromycin selection. The optimal DNA mass and ratio of the two plasmids to be cotransfected were determined (Supplementary Fig. 4).

Blue light illumination to cells. LADL-engineered cells were stimulated using blue light (470 nm) with an intensity of ~1.5 or ~5 mW cm⁻² at 1 s pulse every 14.925 s or 0.067 Hz (ref. [13]).

Fixation for ChIP and 5C. We crosslinked the LADL-engineered mouse ES cells after puromycin selection for ChIP and 5C experiments as previously described[14,15,17]. In brief, we washed puromycin-selected ES cells three times with 1× PBS to get rid of dead, un-transfected cells. The transfected cells that were still adhered to the plates were crosslinked with 1% (v/v) formaldehyde in DMEM (Corning, 10013CV) at room temperature for 10 min, followed by quenching in 125 mM glycine at room temperature for 5 min and at 4 °C for an additional 15 min before cell collection. The mouse ES cells were ~50–70% confluent at the time of fixation[14].

Construction of the light box. *Overall design.* We constructed a light box with a large enough footprint to conduct experiments on the cell numbers required for ChIP and 5C. A significant design change from previously published methods was required to illuminate six-well plates[13,21]. The main challenge we overcame was getting the same light intensity and blinking parameters across all the light emitting diodes (LEDs) in the circuit (Supplementary Fig. 3b). First, a light box was built to illuminate cells at ~1.5 mW cm⁻². One 5-m blue LED strip was cut into six smaller strips of 24 LEDs each (12 V DC weatherproof IP66 LED Tape Light 226 lumens per foot with 5050SMD 470 nm LEDs, WFLS-X3, superbrightleds.com). The six smaller LED strips were connected to each other in series with interconnects (2 Contact 10 mm Flexible Light Strip Interconnects, WFLS10-2CH, www.superbrightleds.com) and aligned parallel on the base of an acrylic box ~1 inch apart from each other. The box is a quarter-inch custom built black acrylic laser cut box measuring 36 × 48 × 8 cm. A quarter-inch clear acrylic lid was laser cut to the same dimensions. The LED strip was powered via a 12 V power supply (Mean Well LED Switching Power Supply LPV Series Single Output LED Power Supply 60 W 12 V DC, LPV-60-12, superbrightleds.com). The LED blinking at 0.067 Hz (1 s on, 13.925 s off) was controlled by a metal–oxide–semiconductor field-effect transistor (MOSFET) (IRF520 Power MOSFET, SiHF520, Vishay Siliconix) and an Arduino (Arduino Uno Rev3, A000066, Arduino). We inserted a 10 kΩ resistor between the MOSFET gate and the ground to prevent gate breakdown. Next, a light box was built to illuminate cells at ~5 mW cm⁻². Twelve blue LEDs were arranged in four parallel lanes of three LEDs in series (470 nm Rebel LED on a SinkPAD-II 10 mm Square Base −74 lm at 700 mA, SP-05-B6, Luxeon Star LEDs). We soldered the LEDs together and secured them in an 8 × 12 × 8 cm plastic box. The LED strip was powered via a 12 V 2,000 mA power supply (3–12 V Selectable Output Variable DC Supply, 9902 PS, MPJA). The LED blinking was controlled in the same way as in the lower intensity light box.

Arduino code. void setup() {// put your setup code here, to run once: pinMode(13, OUTPUT); digitalWrite(13, HIGH); // Turn on the LED delay(60000); // Wait for one minute digitalWrite(13, LOW); // Turn off the LED delay(1000); // Wait for one second} void loop() {// put your main code here, to run repeatedly: digitalWrite(13, HIGH); // Turn on the LED delay(1000); // Wait for one second digitalWrite(13, LOW); // Turn off the LED delay(13925); // Wait for 14.925 s}

Functional validation. We tested the light box functionality using transfected cells with 115 fM of mCherry-conjugated CRY2 (CRY2olig_mCherry; see Supplementary Fig. 3a) in six-well plates[22]. The CRY2 oligomers that were assembled in response to blue light illumination were readily visualized as punctate signals with a fluorescence microscope[22]. Mouse ES cells were seeded for 24 h and transfected for 24 h. Transfected mouse ES cells with CRY2olig_mCherry were first focused using white light passing through a red film (intensity 10% laser power) to ensure minimal exposure to ambient light. Then, the cells were imaged before and after blue light exposure using the Texas Red filter (excitation: 560/40; emission: 630/75) to observe association and dissociation kinetics of the clusters, respectively. Images of the same ES cell colony were taken at the following time points: before exposure to 470 nm blue light, after 4 min of blue light exposure (CRY2 clustering observed) and 26 min after blue light was turned off (CRY2 clusters dissociate). (Supplementary Fig. 3c and d).

RT–qPCR. *RNA extraction.* We collected ~100,000 puromycin-selected mouse ES cells for RNA extraction using *mir*Vana miRNA Isolation kit (Thermo Fisher, AM1560) according to the manufacturer's instructions.

Reverse transcription. We treated the extracted RNA with TURBO DNase I (Thermo Fisher, AM2239) and quantified it using Qubit RNA BR assay (Thermo Fisher, Q10210). We used 100 ng of RNA to prepare complementary DNA using SuperScript First-Strand Synthesis System for qPCR with reverse transcription (RT–qPCR) (Thermo Fisher, 11904018) according to the manufacturer's instructions.

qPCR. We mixed 1 µl of cDNA with 10 mM forward and 10 mM reverse primers in 1× Power SYBR Green PCR Master Mix (Thermo Fisher, 4368706) and ran on qPCR using SYBR Green standard curve method of StepOnePlus Real-Time PCR System (Thermo Fisher, 4376600). PCR cycles start with 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 65 °C for 45 s. We validated the primer pair specificity by looking at single peaks from melting curve analysis at the end of each qPCR run.

Standard curve preparation. We designed RT–qPCR primers using Primer III (https://primer3plus.com/cgi-bin/dev/primer3plus.cgi) and computationally validated their specificity using BLAT (https://genome.ucsc.edu/cgi-bin/hgBlat) and NCBI Primer blast (https://www.ncbi.nlm.nih.gov/tools/primer-blast/primertool.cgi). We mixed 10 mM of forward and reverse primers (Supplementary Table 7) with 1 µl of a cDNA in 1× Taq Master mix (NEB, M0270) and then amplified by 40 PCR cycles according to the manufacturer's instructions. After confirming unique primer amplicons from each primer pair, PCR products were purified using DNA Clean & Concentrator−5 (Zymo Research, D4013) according to the manufacturer's protocol, followed by measuring their concentrations using the Qubit dsDNA HS assay (Thermo Fisher, Q32851). We prepared serial dilutions of the purified products from 1 to 0.00001 fmol µl⁻¹.

ChIP–qPCR. *Antibody-bead binding.* We performed ChIP for the LADL anchor (dCas9-CIBN) on the crosslinked cell pellets, as previously described[23]. To immunoprecipitate the LADL anchor, we used 5 µg anti-FLAG antibody (Sigma, F1804-200UG) and for pre-clearing step in ChIP, we used 100 µg IgG (Sigma, I8140-10MG). Both antibodies were pre-bound to 20 µl protein A (Thermo Fisher, 15918014) and 20 µl protein G (Thermo Fisher, 15920010) agarose beads in PBS at 4 °C for overnight with a rotation at 10 r.p.m. Next day, we washed both antibody-beads in 1 ml ice-chilled PBS twice before use in ChIP.

ChIP. First, we lysed the crosslinked cell pellets in cell lysis buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.2% (v/v) Nonidet-P40, 10% (v/v) Protease Inhibitor Cocktail (Sigma, P8340-5ML) and 0.1 mM PMSF (Roche, 8553 S)), and dounce homogenized 30× using pestle A at room temperature. The nuclei fraction of the lysate was spun down at 2,500*g* and 4 °C for 5 min, and solubilized in 900 µl sonication buffer (50 mM Tris, pH 8.0, 10 mM EDTA, 0.5% (w/v) SDS, 0.1 mM PMSF). To shear the DNA into 300–500 bp, we sonicated the samples using QSonica (Qsonica, Q800R2) at 100% amplitude for 30 min at 4 °C with a cycle of 30 s on and 30 s off, and diluted in immunoprecipitation dilution buffer (20 mM Tris, pH 8.0, 2 mM EDTA, 150 mM NaCl, 1% (v/v) Triton X-100, 0.1 mM PMSF). We precleared non-specific DNA fragments from the sheared lysates using pre-bound IgG-beads at 4 °C for 2 h with rotation at 10 r.p.m., and centrifuged at 2,000 r.p.m. for 5 min at 4 °C. We aliquoted 200 µl of the supernatant to a separate tube to reserve the DNA as input. To immunoprecipitate dCas9-CIBN, we mixed the remaining supernatant with pre-bound anti-FLAG antibody-beads at 4 °C overnight with rotation at 10 r.p.m.

After spinning the lysate with anti-FLAG antibody-beads at 2,000 r.p.m. at 4 °C for 5 min, we washed the beads in different washing buffers in the following order: one wash with immunoprecipitation wash buffer I (20 mM Tris, pH 8.0, 2 mM EDTA, 50 mM NaCl, 1% (v/v) Triton X-100, 0.1% (w/v) SDS, 0.1 mM PMSF), two washes with High Salt buffer (20 mM Tris, pH 8.0, 2 mM EDTA, 500 mM NaCl, 1% (v/v) Triton X-100, 0.01% (w/v) SDS, 0.1 mM PMSF), one wash with immunoprecipitation wash buffer II (10 mM Tris, pH 8.0, 1 mM EDTA, 250 mM LiCl, 1% (v/v) NP40, 1% (w/v) sodium deoxycholate, 0.1 mM PMSF) and last, two washes in 1× Tris-EDTA. Each of the above washes was performed at 4 °C

with a rotation at 10 r.p.m. for 5 min each. Finally, protein-DNA complexes were eluted from beads by vortexing in elution buffer (100 mM sodium bicarbonate, 1% (w/v) SDS) at room temperature for 1 min. We reverse crosslinked the eluent and the input DNA aliquots at 65 °C overnight, and added 1× Tris-EDTA to bring the final volume up to 400 µl before digesting proteins using 20 U of Proteinase K (NEB, P8107S) at 65 °C for 2 h. We extracted and purified the DNA from the samples using phenol:chloroform:isoamyl alcohol (Fisher Scientific, BP1752I100) and ethanol precipitation method using 30 µg glycogen (Ambion, AM9510) and 80 mM NaCl. We resolved the DNA precipitates in 20 µl of 1× Tris-EDTA before subsequent analysis.

qPCR. To compare the LADL anchor enrichment in the ChIP DNA, it is essential to use equal masses of DNA for all the samples for qPCR. We measured the ChIP and input DNA concentrations using Qubit dsDNA HS assay (Thermo Fisher, Q32851). Then, 20 pg of input and eluent DNAs were mixed with 10 mM of forward and reverse primers, The 1× Power SYBR Green PCR Master Mix (Thermo Fisher, 4368706) and qPCR was performed using StepOne Real-Time PCR System (Thermo Fisher, 4376357) according to Standard SYBR Green protocol[14].

For PCR cycles, the PCR reaction was melted at 95 °C for 10 min, followed by 40 cycles of 95 °C for 15 s and 65 °C for 45 s using the SYBR Green standard curve method of the StepOnePlus Real-Time PCR System (Thermo Fisher, 4376600). We confirmed the primer pair specificity by looking at the single peaks of the melting curves in the end of each PCR run. Primer sequences used in ChIP–qPCR are described in Supplementary Table 9.

In situ 3C and 5C. *3C.* We created 3C libraries using the in situ 3C method with minor modifications[2,18]. We lysed the crosslinked cell pellets in 250 µl of cell lysis buffer (10 mM Tris-HCl, pH 8.0, 10 mM NaCl, 0.2% (v/v) Nonidet-P40) supplemented with 50 µl of Protease Inhibitor Cocktail (Sigma, P8340-5ML)) and incubated in ice for 15 min. The nuclei in lysates were spun down at 2,500g at 4 °C for 5 min, and washed in 500 µl of cell lysis buffer. We permeabilized the nuclei in 0.5% (w/v) SDS at 62 °C for 10 min, followed by quenching in 1.13% (v/v) Triton X-100 (final concentration) at 37 °C for 15 min. We digested genomic DNA with 100 U HindIII (NEB, R0104s) in 1× NEBuffer2 (NEB, B7002S) at 37 °C overnight. Next day, after HindIII inactivation at 62 °C for 20 min, we ligated the digested genomic DNA fragments in the nuclei with 2,000 U T4 DNA ligase (NEB, M0202S) in ligase buffer (0.83% Triton X-100, 0.1 mg ml⁻¹ bovine serum albumin in 1× T4 DNA ligase buffer (NEB, B0202S)) at 16 °C for 2 h. We spun down the nuclei at 2,500g at 4 °C for 5 min, and lysed in nuclear lysis buffer (10 mM Tris-HCl, pH 8.0, 500 mM NaCl, 1% (w/v) SDS). We reverse crosslinked the DNA in the lysates at 65 °C initially for 4 h in 20 U Proteinase K (NEB, P8107S) and subsequently overnight with additional 20 U of Proteinase K (NEB, P8107S). To purify DNA from residue proteins and RNA, we treated the samples with 50 mg RNaseA at 37 °C for 30 min (Roche, 10109169001) and performed phenol:chloroform extraction (Fisher Scientific, BP1752I100) and ethanol precipitation methods. After dissolving the DNA pellets in 500 µl TE, we centrifuged the samples on Amicon column filters (Millipore, MFC5030BKS) at 14,000g for 10 min at room temperature. To wash out the salts in the samples, we washed the column filters with 500 µl TE at 14,000g for 10 min at room temperature twice and inverted the column filters and centrifuged at 1,000g for 4 min at room temperature to elute the DNA. The 3C libraries were kept at −20 °C until 5C was performed.

5C primer design. All 5C primers were designed according to a double alternating design with the My5C primer design software (http://my5c.umassmed.edu/my5Cprimers/5C.php)[18,24,25]. Details of 5C primer sequences are described in Supplementary Tables 10 and 11 (ref. [18]).

5C library preparation. The 5C was performed as previously described[15]. We mixed 370 ng of replicate 1 3C library (Figs. 3 and 4 and Supplementary Figs. 6–11), or 500 ng from replicate 2 3C libraries (Supplementary Figs. 7–10) or 200 ng from replicate 5 3C libraries (Supplementary Figs. 7–10) or 590 ng from replicate 3–4 3C libraries (Supplementary Figs. 7–10) with salmon sperm DNA (Thermo Fisher, 15632–011) to ensure a final DNA mass of 1.5 µg. We mixed the DNA with 1 fmol of each 5C primer (Supplementary Tables 10 and 11) in 1× NEBuffer4 (NEB, B7004S). We denatured the DNA in the 5C reaction at 95 °C for 5 min and annealed 5C primers at 55 °C for 16 h. We then nick ligated annealed 5C primers using 10 U Taq ligase (NEB, M0208L) for 1 h at 55 °C, followed by inactivation at 75 °C for 10 min. We used 30 PCR cycles to amplify the 5C ligation product at an annealing temperature of 55 °C using 0.5 U of Phusion High-Fidelity DNA polymerase (NEB, M0530L), 120 µM universal T3/T7 primers, and 25 mM dNTP (Promega, U1330) in 1× Phusion HF Buffer (NEB, M0530L) according to the manufacturer's protocol. Before Illumina adaptor ligation, we purified 5C libraries using AgenCourt Ampure XP beads (Beckman Coulter, A63881) as described by the manufacturer.

To prepare sequencing libraries, we A-tailed 100 ng of purified 5C library before ligating Illumina sequencing adaptors using the NEBNext Ultra DNA library prep kit (NEB, E7370S) for replicates 1–3 and using the NEBNext Ultra II DNA Library Prep Kit (NEB, E7645S) for replicates 4–5. We used NEBNext

Multiplex oligos I and II (NEB, E7335S and E7550S) for all replicates according to the manufacturer's instructions. The linkered 5C libraries were size-selected at 230 bp using AgenCourt Ampure XP beads (Beckman Coulter, A63881) before amplification by nine cycles of PCR for replicates 1–3, and five cycles for replicates 4–5 according to the manufacturer's protocol. We further purified sequencing libraries using AgenCourt Ampure XP beads (Beckman Coulter, A63881) and assessed their quality using the Agilent DNA 1000 reagent kit (Agilent, 5067–1504) on the Agilent Bioanalyzer 2100 (Agilent, 5067–4626). Each library was quantitated using Library Quantification Kit—Illumina/ABI Prism (Kapa Biosystems, KK4835) before pooling and pair-ended Illumina sequencing on the NextSeq500 instrument (Illumina). Replicates 1–2 were sequenced with 37-bp paired end reads, and replicates 3–5 were sequenced with 75-bp paired end reads.

5C data analysis. We analyzed 5C data as detailed in ref. [15] with minor modifications. The 37-bp pair-ended sequencing reads for replicates 1 and 2 were directly mapped to a pseudo-genome consisting of 5C primer sequences with Bowtie using parameters --tryhard and -m 2 and --trim5 6 (Supplementary Tables 10 and 11). A summary of the mapped reads for replicates 1 and 2 is described in Supplementary Table 12. The 5C primer pairs were counted as previously described[14,15,17]. Outlier values were removed if they were greater than the sum of 30 surrounding pixels by 32-fold. Raw counts were quantile normalized, binned into 4-kb resolution matrices and balanced using the ICED algorithm[26]. We evaluated looping interactions by modeling the chromatin domains and distance-dependence signal using the upper half of the donut filter for short-range interactions under 100-kb distance, and the full donut filter for longer range interaction greater than 100 kb using parameters $P = 10$ and $w = 40$ (ref. [2,27]). We modeled the distance- and TAD-corrected interaction frequency data (that is, observed/expected) with a parameterized log-logistic distribution as described in refs. [15,27]. Interaction scores were computed as $-10\log_2(P\ \text{value})$.

Parsing 5C monomers and mapping. The 75-bp ends of pair-ended sequencing reads from replicates 3, 4 and 5 were independently mapped to the pseudo-genome consisting of 5C primer sequences. The length of these reads required additional processing. First, only the reads containing the HindIII recognition sequence (AAGCTT) were split into two subreads corresponding to the 20 bp of sequence immediately 3′ and immediately 5′ of the HindIII cut site, respectively. To assign specific 5C primer–primer ligations to each 75-bp end, subreads were mapped to the pseudo-genome consisting of 5C primers using Bowtie with parameters --tryhard and -m 2. To identify the 5C 'monomers' that have only one 5C ligation junction, we compared the primer–primer ligations between the paired reads. If both paired reads had the same primer–primer ligation junction, we classified them as a 'monomer' and constructed counts files for downstream analysis as described above. A summary of the mapped reads for replicates 3–5 is described in Supplementary Table 12.

Public data analysis. A list of all publicly available sequencing datasets that were used in this study is described in Supplementary Table 8. Sequencing reads were downloaded from the Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) and mapped to NCBI Build 37 (UCSC mm9) using the Bowtie with parameters --tryhard and -m 2 for ChIP-seq, and Bowtie2 with parameters -X2000 --no-mixed --no-discordant for ATAC-seq. Only the sequencing reads that were uniquely mapped to the genome were analyzed in this study. A summary of mapped reads corresponding to the publicly available data is also described in Supplementary Table 8.

Immunofluorescence staining. We fixed the LADL-engineered mouse ES cells in 4% PFA in PBS for 15–20 min at room temperature and washed three times with 1× PBS. The fixed cells were stored at 4 °C until immunofluorescence was performed. We incubated the fixed cells in blocking solution (10% (v/v) Normal Donkey Serum (Jackson ImmunoResearch, 017-000-121), 0.1% (v/v) Triton X-100 in PBS) with gentle nutation for 1 h at room temperature. We next probed the cells with Rabbit α-Oct3/4 (Thermo Fisher, SC-9081) at 1:200 dilution in blocking solution with gentle nutation overnight at 4 °C. The next day, we washed cells three times in 0.1% (v/v) Tween 20 in PBS for 10 min each to remove excess primary antibodies, and probed with Goat Anti-rabbit Alexa Fluor 488 (Thermo Fisher, A-11006) at 1:500 dilution in blocking solution for 2 h at room temperature in the dark. To remove excess secondary antibodies, we washed cells twice in 0.1% Tween 20 in PBS for 10 min each and twice in PBS for 10 min each. Finally, we mounted cells onto slide glass with ProLong Gold antifade reagent with 4,6-diamidino-2-phenylindole (Thermo Fisher, P36935) before imaging on a Leica DMi8/LAS X microscope.

RNA–FISH. We designed oligonucleotides for RNA–FISH using the Stellaris probe design software available online (https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer). Pools of 32 oligonucleotides were labeled with Atto674N (atto-tec) for *Klf4* and *Zfp462* exons and Atto700 for *Klf4* and *Zfp462* introns. We trypsinized cells and fixed in 3.7% formaldehyde and performed RNA–FISH as previously described[28]. After blue light illumination at 5 mW cm⁻² for 24 h, *Zfp462* or *Klf4* transcripts in LADL-engineered mouse

ES cells and in three other controls (LADL + dark, empty target control + dark, empty bridge control + dark) were hybridized with 32 exon- and 32 intron-specific fluorescently labeled oligonucleotides before acquiring images for quantitative analysis. For each field of view, 40 z-section images spaced at 0.3 μm were acquired on a Nikon Ti-E widefield microscope using a ×100 1.4 numerical aperture objective and a cooled charge-coupled device (CCD) camera. We used custom image processing scripts written in MATLAB to count mRNA and identify transcription sites. This software is available for download at https://bitbucket. org/arjunrajlaboratory/rajlabimagetools/wiki/Home. The estimates of nascent transcript numbers in Fig. 5c and Supplementary Fig. 13b,d were calculated by dividing the intensity of exon probe signal at the transcription site by the median intensity of all exon probe signals (primarily from mRNA) in the dataset. Fluorescence-labeled oligonucleotide sequences for RNA–FISH are given in Supplementary Table 13.

Statistics. The sample numbers corresponding to the individual experiments are included in the figures. Figure 1e shows two independent experiments. Figure 1f is a representative image of three independent experiments. Supplementary Fig. 3d shows representative images of two independent experiments. Supplementary Fig. 4 shows representative images of two independent experiments. Supplementary Fig. 5 shows representative images of more than ten independent experiments. Figure 2d–f shows one experiment. Supplementary Figs. 7f and 9f include box plots showing central tendency = median, box minima = 25th percentile, box maxima = 75th percentile, notches = 95% confidence interval, whiskers = 1.5× interquartile range. Figure 4c and Supplementary Figs. 7–10 show five independent experiments performed with LADL + blue light and LADL + dark ($n = 5$), empty target + dark ($n = 3$), empty anchor + blue light ($n = 1$), empty anchor dark ($n = 1$), one-sided guide control blue light ($n = 1$), one-sided guide control dark ($n = 1$) where n = number of independent experiments. The strip charts in Fig. 4c and Supplementary Figs. 8d and 10c show the median (red line). Figure 4c and Supplementary Fig. 8d show P values that were computed using an unpaired, two-sided Mann–Whitney U-test. Figure 5a shows representative images of three independent experiments. Figure 5b–d and Supplementary Fig. 13 show three independent experiments and P values were computed using an unpaired one-tailed Mann–Whitney U-test with a null hypothesis that *Zfp462* levels in the LADL + blue light condition are equal to the negative control conditions and an alternative hypothesis that *Zfp462* levels in the LADL + blue light condition are greater than the negative control conditions. Figure 5b,c and Supplementary Fig. 13a–d show P values that were computed using a one-tailed Mann–Whitney U-test, with a null hypothesis that *Klf4* levels in the LADL + blue light condition are equal to the negative control conditions and an alternative hypothesis that

Klf4 levels in the LADL + blue light condition are lower than the negative control conditions. Figure 5b,c and Supplementary Fig. 13a–d contain strip charts showing the mean (red line). Sample sizes (n) represent (Fig. 5b,d and Supplementary Fig. 13a,c,e,f) the number of cells or (Fig. 5c and Supplementary Fig. 13b,d) the number of active transcription alleles.

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

Data availability
The 5C data from this study have been submitted to the NCBI Gene Expression Omnibus under accession number GSE115963. Custom code for full reproducibility of all analyses is available upon request.

References
19. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* **517**, 583–588 (2015).
20. Sakuma, T., Nishikawa, A., Kume, S., Chayama, K. & Yamamoto, T. Multiplex genome engineering in human cells using all-in-one CRISPR/Cas9 vector system. *Sci. Rep.* **4**, 5400 (2014).
21. Cheng, A. W. et al. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. *Cell Res.* **23**, 1163 (2013).
22. Taslimi, A. et al. An optimized optogenetic clustering tool for probing protein interaction and function. *Nat. Commun.* **5**, 4925 (2014).
23. Hsu, S. C. et al. The BET protein BRD2 cooperates with CTCF to enforce transcriptional and architectural boundaries. *Mol. Cell* **66**, 102–116 (2017).
24. Lajoie, B. R., van Berkum, N. L., Sanyal, A. & Dekker, J. My5C: web tools for chromosome conformation capture studies. *Nat. Methods* **6**, 690–691 (2009).
25. Hnisz, D. et al. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science* **351**, 1454–1458 (2016).
26. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
27. Gilgenast, T. G. & Phillips-Cremins, J. E. Systematic evaluation of statistical methods for identifying looping interactions in 5C data. *Cell Syst.* **8**, 197–211 (2019).
28. Raj, A., van den Bogaard, P., Rifkin, S. A., van Oudenaarden, A. & Tyagi, S. Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* **5**, 877–879 (2008).

naturereasearch

Corresponding author(s): Jennifer E. Phillips-Cremins

Last updated by author(s): Mar 29, 2019

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.

Statistics

For all statistical analyses, confirm that the following items are present in the 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
☐	☒	A statement on 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 statistical parameters 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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used Python 2.7.5, R 3.0.1, R3.4.3, R3.5.1
Data analysis	We used Python 2.7.5, R 3.0.1, R3.4.3, R3.5.1, Numpy 1.7.1, Scipy 0.12.0., Bowtie ver0.12.7, Bowtie2 ver2.2.5 Custom codes used for RNA FISH analysis are available in the following links https://bitbucket.org/arjunrajlaboratory/rajlabimagetools/wiki/Home Custom codes used for 5C analysis is available upon the request to the corresponding author.

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. 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

The raw and processed data is deposited to GEO with the accession number GSE115963. It is publicly available now.

Field-specific reporting

Please select the one below that is 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/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	No statistical method were used to predetermine the sample sizes. Overall, 5 replicates of 5C and 3 replicates of single molecule RNA FISH were performed. We selected the number of replicates for 5C and single molecule RNA FISH as the highest number of samples we could perform under reasonable financial and logistical constraints to provide precise and accurate estimates of the data's central tendency and variance and allow for the computational of confidence intervals around estimates.
Data exclusions	No exclusion
Replication	All attempts at replication were successful. To convincingly test if the LADL system worked, 5 independent replicate experiments were performed. These 5 replicates include two different light intensities and loop formation correlated to the intensity of the inducing blue light signal. We also tested loop formation at a couple of different time points (4 hours and 24 hours) after blue light exposure. We also sequenced the replicates at different reading lengths (37bp PE and 75bp PE).
Randomization	For gene expression and 5C experiments, multiple controls had to be processed in addition to the experimental samples, at specific time-points. At each time point, one replicate of each condition was collected together, and this was done in succession for the rest of the biological replicates.
Blinding	Investigators were not blinded to group allocation. Blinding was not relevant to our study because no human subjective qualitative metrics were reported.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

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

Methods

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

Antibodies

Antibodies used	anti-FLAG antibody (Sigma, F1804-200UG, LOT SLBQ6349V, LOT SLBS3530V) : 5ug per sample was used for ChIP as described in Methods IgG (Sigma, I8140-10MG, LOT SLBK4078V) : 100ug per sample was used for ChIP as described in Methods Rabbit anti-Oct3/4 antibody (Santa Cruz, SC-9081) : 1:200 dilution ratio for IF as described in Methods Goat Anti-rabbit Alexa Fluor 488 (Thermo Fisher, A-11006) : 1:500 dilution ratio for IF as described in Methods Note. LOT numbers of some antibodies above could not be provided because the tubes that was used in the experiment is not available anymore.
Validation	Full information of each antibody is stated in the manufactorer's website/antibody bulletins as followings: anti-FLAG : https://www.sigmaaldrich.com/catalog/product/sigma/f1804?lang=en&region=US IgG : https://www.sigmaaldrich.com/catalog/product/sigma/i8140?lang=en&region=US anti-Oct3/4 antibody : https://www.scbt.com/scbt/product/oct-3-4-antibody-h-134?productCanUrl=oct-3-4-antibody-h-134&_requestid=3965428

Goat Anti-rabbit Alexa Fluor 488 : https://www.thermofisher.com/antibody/product/Goat-anti-Rat-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/A-11006

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Murine v6.5 Embryonic Stem (mES) cells (v6.5; genotype 129SvJae x C57BL/6; male) purchased from Novus Biologicals, mouse induced Pluripotent Stem cells (iPSC) reprogrammed on pNPC derived from a mouse with Sox2-eGFP (Eminli et al 2008, Ellis et al 2004), Primary Mouse Embryonic Fibroblast (pMEF) was derived from bodies and limbs of mouse embryos Day 13 or 14.
Authentication	None of the cell lines have been authenticated.
Mycoplasma contamination	The cell lines were not tested for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified lines were used.