



Genome editing demonstrates that the –5 kb Nanog enhancer regulates Nanog expression by modulating RNAPII initiation and/or recruitment

Received for publication, July 7, 2020, and in revised form, December 7, 2020. Published, Papers in Press, December 17, 2020.

<https://doi.org/10.1074/jbc.RA120.015152>

Puja Agrawal^{1,2,†}, Steven Blinka^{1,2,†}, Kirthi Pulakanti², Michael H. Reimer Jr², Cary Stelloh², Alison E. Meyer², and Sridhar Rao^{1,2,3,*}

From the ¹Department of Cell Biology, Neurobiology, and Anatomy, ²Department of Pediatrics, Medical College of Wisconsin, Milwaukee, Wisconsin, USA; and ³Blood Research Institute, Versiti, Milwaukee, Wisconsin, USA

Edited by Ronald Wek

Transcriptional enhancers have been defined by their ability to operate independent of distance and orientation in plasmid-based reporter assays of gene expression. At present, histone marks are used to identify and define enhancers but do not consider the endogenous role of an enhancer in the context of native chromatin. We employed a combination of genomic editing, single cell analyses, and sequencing approaches to investigate a *Nanog*-associated *cis*-regulatory element, which has been reported by others to be either an alternative promoter or a super-enhancer. We first demonstrate both distance and orientation independence in native chromatin, eliminating the issues raised with plasmid-based approaches. We next demonstrate that the dominant super-enhancer modulates *Nanog* globally and operates by recruiting and/or initiating RNA Polymerase II. Our studies have important implications to how transcriptional enhancers are defined and how they regulate gene expression.

Gene expression is regulated by two types of genetic elements: *Trans* elements typically encode proteins such as transcription factors (TFs), which subsequently bind *cis*-regulatory elements (CREs) that must be on the same DNA molecule as the gene they regulate. Different types of CREs have historically been classified based upon their behavior in plasmid-based reporter assays (1, 2). For almost 40 years it has been accepted that promoters are required to be in the correct orientation and immediately adjacent to the gene they regulate, whereas enhancers operate independent of both distance and orientation. The advent of enhancer-specific epigenetic signatures based on histone marks such as H3K27Ac or H3K4me1 permit genome-wide identification of enhancers, which then demonstrate enhancer activity in reporter assays (3, 4). However, plasmid-based assays are limited for multiple reasons. First, they do not fully recapitulate native chromatin structure, and therefore represent a highly artificial system.

Second, they typically are performed on smaller (<500 bp) DNA sequences rather than on the larger chromatin domains of many highly active enhancers. Third, they cannot precisely link a given enhancer sequence to the gene(s) they may regulate *in vivo*. As such, plasmid assays are far more effective at confirming DNA sequences with enhancer potential, rather than definitively identifying them as enhancers.

The advent of sequencing-based chromosomal conformation capture techniques has allowed the measurement of genome-wide enhancer-gene interactions, potentially identifying enhancer:gene regulatory interactions (5, 6). Of interest, this approach demonstrates that many enhancers interact with multiple genes and vice versa but in isolation are insufficient to properly determine if an enhancer is required for gene(s) expression (7, 8). The classic approach to address this question is through genetics, namely, deleting a putative enhancer and measuring the mRNA levels of nearby genes, a method made highly feasible through genomic editing approaches such as CRISPR-Cas9. One important point is that, although these approaches can identify which gene(s) are regulated by an enhancer, many of the mechanistic details of how the enhancer regulates transcription to modulate gene expression are not elucidated through solely this approach.

Multiple models of enhancer-mediated gene expression exist within the literature. Early theories postulated that enhancers looped in to interact with promoters and recruited RNA Polymerase II (RNAPII) to the gene's promoter (9). More recently, multiple mechanisms that focus on enhancers regulating transcriptional elongation have been proposed, including promoter-proximal pause release of RNAPII through various mechanisms (reviewed in Chen *et al.* 2018). It has also been proposed that enhancers modulate transcriptional bursting, or the periods of time during which transcription is active, which represents a combination of initiation and elongation (11, 12). New studies demonstrate that among enhancers there is a subclass of highly active enhancers called "super-enhancers" (SEs, (13, 14)), which may potentially form phase-separated droplets within the nucleus to concentrate transcriptional machinery around highly transcribed genes (15). It is important to note this current model remains to be definitively established. Collectively, this literature indicates

This article contains [supporting information](#).

[†] Equal contribution.

* For correspondence: Sridhar Rao, sridhar.rao@versiti.org.

Present address for Steven Blinka: Department of Internal Medicine, University of Washington, Seattle, WA 98105, USA.

Defining an enhancer that modulates recruitment

that there may be multiple mechanisms by which enhancers regulate gene expression.

The extended *Nanog* locus is a unique locus to study how super-enhancers regulate gene expression and pluripotency. The *Nanog* locus (150 kb) contains a number of different pluripotency-associated genes including *Dppa3*, *Gdf3*, and *Apobec1* (7, 16). It also contains three SEs (–5, –45, and +60, based upon distance in kilobases from *Nanog* TSS) that interact with *Nanog* and behave as enhancers in reporter assays (7). Recent work from our group and others has demonstrated that these three enhancers are super-enhancers based upon several criteria, including high levels of the epigenetic mark H3K27Ac, robust binding by the Mediator complex, and production of enhancer-transcribed RNAs (13, 14). One group has argued that the –5 SE/CRE is actually an alternative promoter, emphasizing that plasmid-based approaches are insufficient to determine if a DNA element is a promoter or enhancer (17). In this study we demonstrate that the –5 CRE is an enhancer by confirming that it operates in a distance- and orientation-independent fashion through genomic approaches and regulates *Nanog* by modulating RNAPII initiation or recruitment.

Results

The –5 *Nanog* CRE is required for embryonic stem cell pluripotency in a *Nanog*-dependent manner

Previously, we demonstrated that the –5 CRE physically interacts with the *Nanog* promoter by chromosomal conformation capture and could also activate *Nanog* expression in plasmid-based reporter assays (7). Monoallelic deletion of the –5 CRE causes a 50% reduction in *Nanog* expression (7). However, we were unable to recover mouse embryonic stem cells (ESCs) that exhibited biallelic deletion of the –5 CRE, leading us to hypothesize it is required for pluripotency. We refer to this element as a CRE rather than an SE because one group has previously reported that this element is an alternative promoter (17). To identify if this element is required for *Nanog* expression, we used genomic editing to insert a tamoxifen (4OHT)-inducible Cre-recombinase (CreER^{T2}) into the constitutively expressed *Rosa26* locus in ESCs to facilitate conditional deletions and then biallelically inserted *loxP* sites to flank a 2.5-kb region of the –5 CRE to encompass two *Nanog*, *Oct4*, and *Sox2* (NOS) binding sites (Fig. S1A, Fig. 1A, left). Insertion of *loxP* sites causes no change to *Nanog* expression (Fig. S1, B–C). Treatment with 4OHT induces complete biallelic deletion of the –5 CRE as compared with vehicle treated (ethanol; Fig. S1D, Fig. 1A, left). ESCs began to differentiate and became nonadherent, consistent with a loss of pluripotency, following 4OHT exposure. Staining for the pluripotency marker alkaline phosphatase was reduced in bulk cells treated with 6 days of 4OHT (Fig. S1E i–ii) compared with control (Fig. S1E iii–iv). Deletion of the –5 CRE resulted in a rapid loss of *Nanog* mRNA (Fig. 1A, right) and protein (Fig. S1F). By contrast, *Gdf3*, a nearby gene, showed little

change in expression following deletion of the –5 CRE (Fig. S1G). There was also a decrease in other pluripotency-associated TFs such as *Oct4*, *Esrrb*, and *Klf4*, demonstrating a progressive collapse of the transcriptional network regulating pluripotency (Fig. S1, H–J). Consistent with previous studies showing that *Nanog* represses endoderm specification, RT-qPCR for key differentiation genes (Fig. 1B) and endoderm-promoting TFs such as *Gata4*, *Gata6*, and *Hnf4a* (Fig. 1C) demonstrated an increased expression in specifically the endoderm-promoting genes following 4OHT treatment. These results demonstrate that the –5 CRE is required for ESC pluripotency, likely by regulating *Nanog* expression.

Next, we hypothesized that the –5 CRE maintains pluripotency solely by regulating *Nanog* expression rather than the expression of another gene on chromosome 6 (chr6). To test this, we made a stable cell line expressing murine *Nanog* with a ubiquitous promoter (CAG; Fig. 1D, left). Of importance, *Nanog*^{+/-} animals are viable and ESCs remain pluripotent (18, 19), indicating that 50% levels of *Nanog* do not compromise pluripotency in mice. Endogenous *Nanog* gene expression can be followed with RT-qPCR primers amplifying the *Nanog* 3' UTR, which is absent from the *mNanogV5* transgene. Prior to 4OHT treatment, bulk cells express *Nanog* mRNA levels approximately 50% higher than wildtype ESCs, which then falls following 4OHT treatment to 50% below wildtype (Fig. 1D, right). After 4OHT treatment, these cells show a profound (>90%) reduction in endogenous *Nanog* expression (Fig. 1D, right) and a small decrease in *Oct4* levels but no other significant change in other core pluripotency TFs such as *Esrrb* or *Klf4* (Fig. 1E). In addition, following 6 days of 4OHT treatment cells remain alkaline phosphatase positive (Fig. S2A), consistent with the *Nanog* transgene rescuing the loss of pluripotency seen following deletion of the –5 CRE (Fig. 1).

Following 6 days of 4OHT treatment on the –5 CRE floxed cells expressing individual clones were selected, expanded, and analyzed (Fig. S2B). These cells remain pluripotent even after 6 days of leukemia inhibitory factor (LIF) withdrawal (Fig. S2C, right) because of the constitutive expression of *Nanog* off the ubiquitous CAG promoter of the transgene. Western blot analysis of total *Nanog* in isolated individual clones shows a decrease in *Nanog* expression of approximately 50% upon stable deletion of the –5 CRE (Fig. S2D). These data demonstrate that the loss of pluripotency in cells without the *Nanog* transgene is attributable to the loss of *Nanog* expression. It should be noted that pluripotency was determined by a combination of gene expression (Fig. 1E and see RNA-Seq below), morphology, and alkaline phosphatase staining (Fig. S2, A and C), but we were unable to perform the most rigorous test of either teratoma formation or tetraploid complementation owing to the presence of the *Nanog* transgene in these cells, which prevents differentiation.

To determine if the activity of other super-enhancers within the extended *Nanog* locus changed upon deletion of the –5 CRE, we measured enhancer-transcribed RNA levels, which

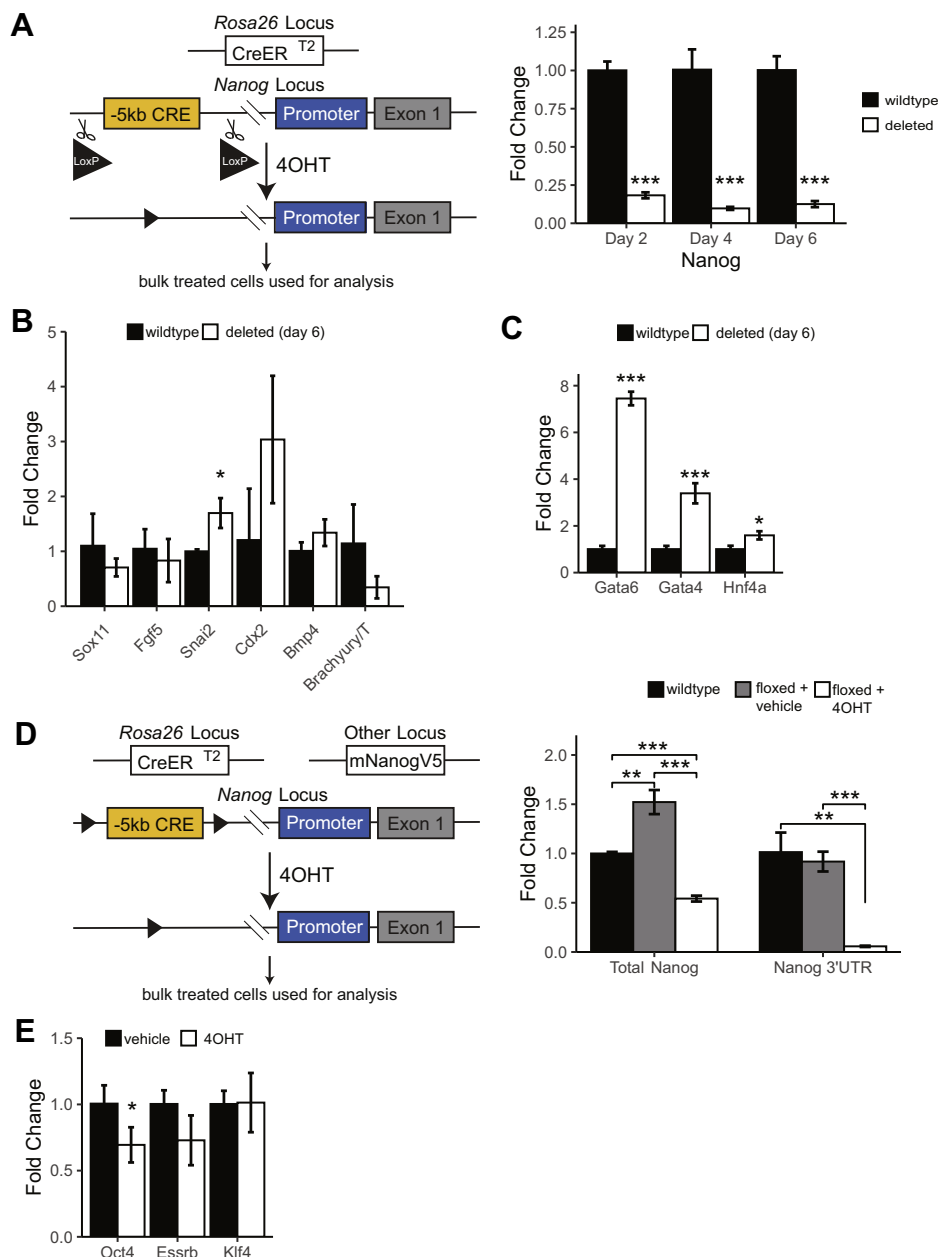


Figure 1. Deletion of the -5 CRE. *A*, biallelic deletion of -5 CRE was achieved by inserting two loxP sites around the enhancer and cells were treated with tamoxifen (4OHT) for 6 days. *Left panel*, schematic; *right panel*, Nanog mRNA levels in bulk vehicle or 4OHT-treated cells. $n = 3$. *B*, day 6 mRNA expression of mesodermal, ectodermal, and trophoctodermal differentiation markers. $n = 3$. *C*, day 6 mRNA expression of endodermal promoting transcription factors. $n = 3$. *D*, stable biallelic deletion of the -5 CRE was achieved by rescuing with a mouse Nanog cDNA. Endogenous Nanog expression is measured via the Nanog 3'UTR, while Total Nanog measured endogenous and the exogenous expression. *Left panel*, schematic; *right panel*, mRNA levels in bulk treated cells. $n = 3$. *E*, pluripotency markers in bulk treated cells in -5 CRE-deleted cells expressing Nanog *in trans* shown in (*D*). $n = 3$. All mRNA levels measured by RT-qPCR and shown as $2^{\Delta\Delta CT}$ compared with wildtype or vehicle treated. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Student's two sample *t* test. CRE, *cis*-regulatory element.

are widely used as a measure of enhancer activity (7, 13, 20). In individually expanded clones of the -5 CRE deletion with Nanog *in trans*, we observed a global decrease in enhancer activity (Fig. S2E), which may be due to the presence of the exogenous Nanog transgene at a decreased level (50%). However, we cannot rule out that the decreased enhancer activity is a direct result of -5 CRE deletion. From here forward, all experiments using a biallelically deleted -5 CRE deletion were done using clonal cell line(s) that supply Nanog *in trans* to prevent a loss of pluripotency.

To determine if the -5 CRE solely regulates Nanog we used RNA-Seq to identify other altered transcripts. We used the cell line described in Figure 1E, a stable line with the -5 CRE deleted that expresses exogenous mNanogV5 and compared it with the floxed -5 CRE expressing exogenous mNanogV5 cell line. First, we identified genes on chr6, which showed at least a 2-fold, statistically significant change (adj *p*-value < 0.05) between samples (Fig. 2A). As a control, we also compared our data with previously published RNA-Seq data where Nanog was depleted by RNAi (21). We further queried changes on

Defining an enhancer that modulates recruitment

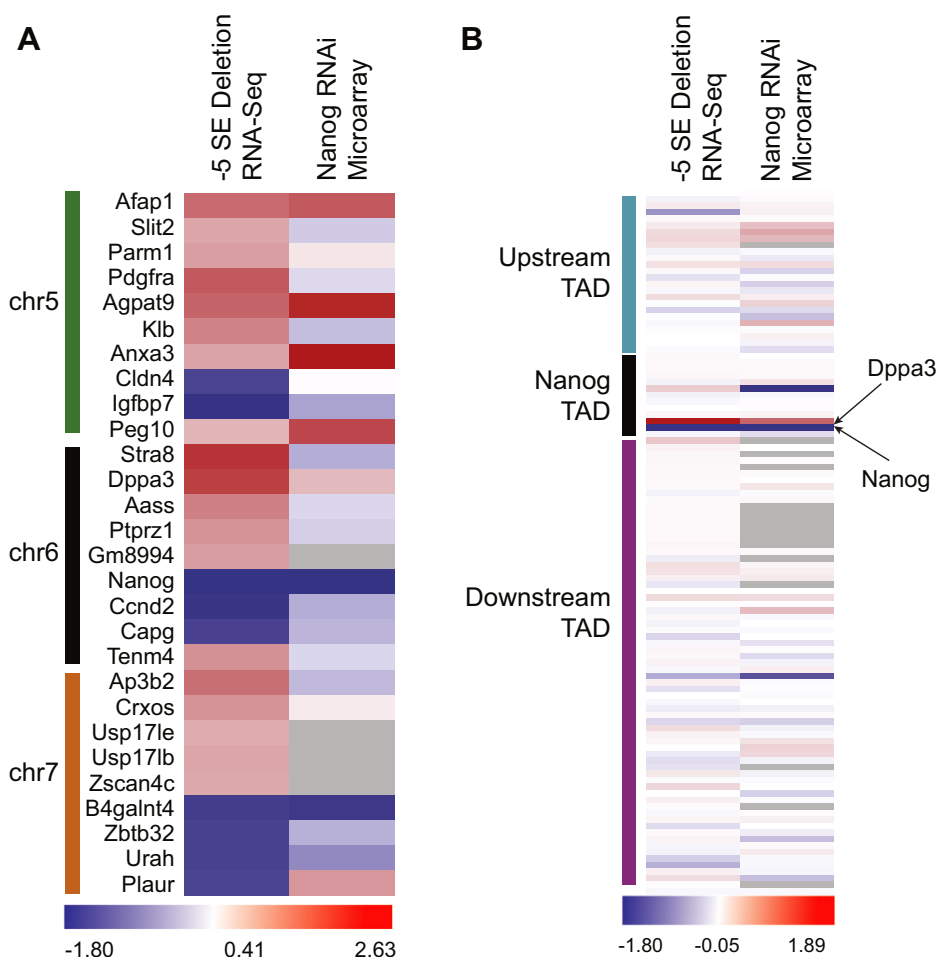


Figure 2. The -5 CRE regulates *Nanog* exclusively. A, differentially expressed genes, defined to be at least 2-fold, statistically significant ($\text{adj } p < 0.05$, $n = 3$) change, on chromosome 5, 6, and 7 via RNA-Seq. *Nanog* RNAi microarray data are shown for comparison (20). B, differentially expressed genes, irrespective of significance or fold-change, on chromosome 6 within the *Nanog* TAD, one TAD upstream and one downstream. The -5 SE-deleted cells express *Nanog* in *trans* to prevent a loss of pluripotency. CRE, *cis*-regulatory element; SE, super-enhancer; TAD, topologically associated domain.

chr5&7 to estimate gene expression changes secondary to *trans* effects from changes in *Nanog* protein levels. We observed that gene expression changes in the -5 CRE deleted line mimics the *Nanog* RNAi data, implying that all changes are due to alteration in the levels of *Nanog* protein operating in *trans*. On chr6, none of the altered genes were within 1 MB of *Nanog* except *Dppa3*. The observed increase in *Dppa3* expression was expected, since it is directly repressed by *Nanog* protein binding to its promoter (7). These data suggest that the -5 CRE only regulates *Nanog*. To further clarify if loss of the -5 CRE affected the expression of other genes, *in cis* we queried expression changes of all genes within the *Nanog* topologically associated domain (TAD) as well as the two adjacent topologically associated domains (22), irrespective of statistical significance or fold-change (Fig. 2B, Table S1). With the exception of *Nanog* and *Dppa3*, most genes showed minimal gene expression changes that were comparable with the *Nanog* RNAi data, implying this was due to reduced *Nanog* protein operating in *trans*. We therefore conclude that the -5 CRE exclusively regulates *Nanog* expression in ESCs, with no evidence that it regulates other genes on chr6 *in cis*. Collectively, these experiments demonstrate that the *Nanog* -5 CRE

is required for pluripotency through its direct regulation of *Nanog* expression.

The *Nanog* -5 CRE operates in a distance- and orientation-independent fashion

The *Nanog* -5 CRE in plasmid assays acts independent of distance and orientation and has been extensively referred to as an enhancer within the literature (23, 24). By contrast, in at least one report the -5 CRE was considered an alternative promoter that played a critical role in regulating pluripotency through an alternative *Nanog* isoform (17). Given this ambiguity we chose to definitively establish if this element had enhancer activity within the context of normal chromatin with our 4OHT-inducible Cre-LoxP system by inserting one of the *LoxP* sites in the opposite orientation (Fig. 3A, left). In this configuration, Cre activation by 4OHT treatment will induce biallelic “flipping” of the -5 CRE continuously between the two orientations. Following treatment with 4OHT, individual clones were isolated and expanded and biallelic inversion of the -5 CRE was verified by PCR (Fig. S3A). On comparison of the biallelic inversion with the wild-type orientation of the -5

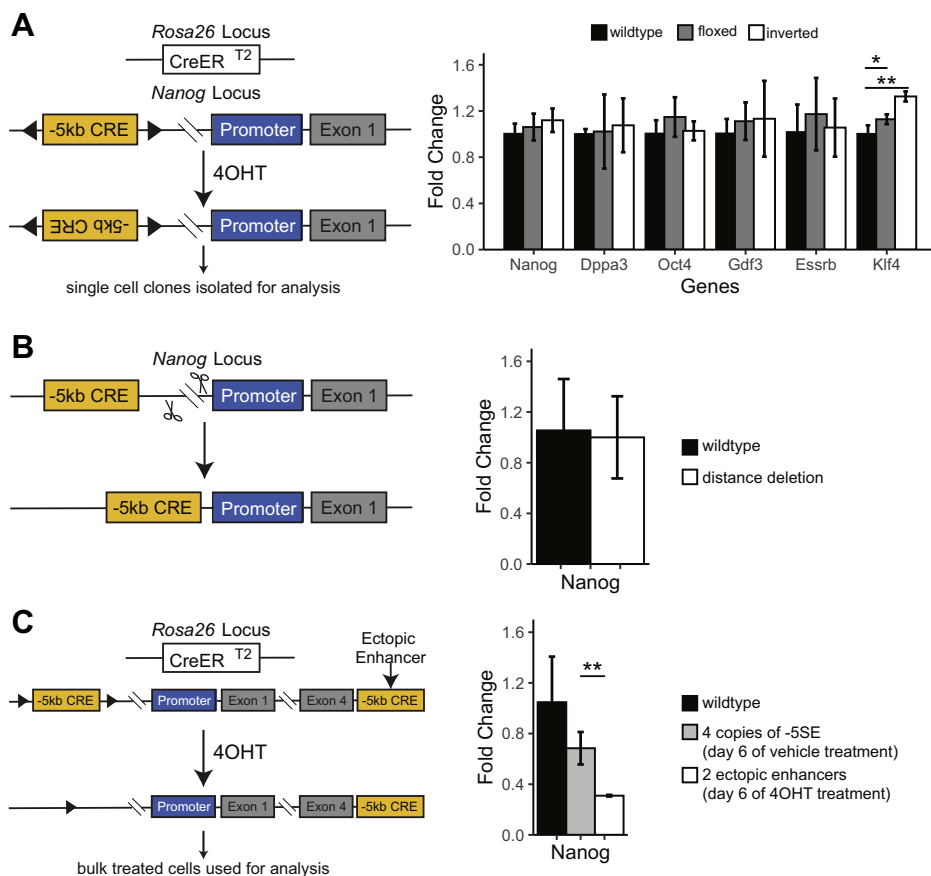


Figure 3. Manipulation of the -5 CRE. A, the -5 CRE was flipped by inserting opposing loxP sites followed by tamoxifen (4OHT) treatment. *Left panel*, schematic; *right panel*, mRNA levels of *Nanog* and relevant pluripotency genes in individual isolated clones. $n = 3$. B, an approximately 2-kb region between the -5 CRE and *Nanog* TSS was deleted. *Left panel*, schematic; *right panel*, *Nanog* mRNA levels. $n = 5$. C, two copies of the -5 CRE were inserted downstream of *Nanog* in cells where the endogenous enhancer is floxed. Endogenous enhancers were deleted *via* treatment with tamoxifen. *Left panel*, schematic; *right panel*, mRNA levels of *Nanog* in bulk cells treated with vehicle or 4OHT. $n = 3$. All mRNA levels measured by RT-qPCR and shown as $2^{-\Delta\Delta CT}$ compared with wildtype. * $p < 0.05$, ** $p < 0.01$ Student's two sample t test. None of the cell lines shown express *Nanog* in *trans*. CRE, *cis*-regulatory element.

CRE, we observed no statistically significant changes in the expression of *Nanog*, *Dppa3*, *Oct4*, or *Esrrb* and only minor changes in *Klf4* (Fig. 3A, right). To verify there was no change in the *Nanog* protein banding pattern to indicate a potential change in protein isoforms, we performed Western blots and did not observe any changes in the banding pattern (Fig. S3B). These data are consistent with the -5 CRE regulating *Nanog* expression in an orientation-independent manner, a classic property of an enhancer but not a promoter. Although we cannot rule out that this CRE can simultaneously act as both an enhancer and an alternative promoter, our data demonstrate that any alternative isoforms produced *via* the -5 CRE as a promoter (17) are dispensable for pluripotency.

We next hypothesized that the -5 CRE would also operate independent of distance from the *Nanog* transcriptional start site (TSS). To determine this, we first deleted the intervening ≈ 2 kb between the -5 CRE and promoter (Fig. S3C, Fig. 3B, left) and found no significant change in *Nanog* expression (Fig. 3B, right). Next, we biallelically inserted an additional copy of the -5 CRE between the *Nanog* transcriptional end site (TES) and the nearest CCCTC binding factor site to ensure it remained within the same insulated neighborhood (Fig. S3, D-E, Fig. 3C, left, 25). Insertion of the additional -5 CRE caused

no change in *Nanog* mRNA levels (Fig. 3C, right). Treatment with 4OHT for 6 days resulted in deletion of the endogenous -5 CRE in bulk cells and caused a reduction in *Nanog* mRNA by approximately 50% (Fig. 3C, right). Expression of other key pluripotency markers such as *Oct4*, *Esrrb*, and *Klf4* were unchanged, indicating that pluripotency was maintained (data not shown). The partial recapitulation of *Nanog* expression and pluripotency by the ectopic enhancers is consistent with the -5 CRE operating independent of distance, albeit less effectively than its native chromatin position (Fig. 3C, right). *Nanog* expression in the cells only containing the ectopic enhancer is $\approx 35\%$ compared with wildtype, which is higher than the $\approx 10\%$ *Nanog* seen on the conditional deletion of the -5 CRE in bulk cells (Fig. 1A) and near 0% *Nanog* we see on true complete deletion in individual clones (Fig. 1D, right). We do observe an insignificant decrease in *Nanog* in cells with ectopic and endogenous enhancers, treated with vehicle; however, this could be due to minor changes in chromatin architecture from the insertion of the ectopic enhancers. One reasonable explanation for the reduced *Nanog* expression with the ectopic enhancer is that the -5 CRE includes a larger chromatin domain, whereas we inserted only the core ≈ 2.5 kb containing two NOS sites into an alternative

Defining an enhancer that modulates recruitment

location. This may imply that additional sequences surrounding the core are required for full activity. Alternatively, moving the enhancer further away may prevent it from fully activating Nanog. Nonetheless, this demonstration of enhancer function represents a highly feasible, native chromatin approach to confirm that a DNA element is an enhancer. Collectively, these data demonstrate that the -5 CRE is an enhancer, and we will hereafter refer to it as the -5 SE (14).

Constituent enhancers within the -5 SE are additive in regulating Nanog expression

Several groups have demonstrated that, within a super-enhancer, a single, smaller constituent enhancer is required for proper SE function, with the remaining constituent enhancers being dispensable for regulating gene expression (26). To determine if the -5 SE has a dominant constituent enhancer, we first reviewed published chromatin immunoprecipitation coupled with next-generation sequencing (ChIP-seq) datasets from other groups to determine if there were constituent enhancers within the larger -5 SE. We observed that there were two distinct regions occupied by the classic pluripotency transcription factors Nanog, Oct4, and Sox2 (NOS, Fig. 4A). To determine if one or both constituent enhancers were critical to pluripotency, we deleted each individually with CRISPR-Cas9 using a pair of distinct guide RNAs (gRNAs) (Fig. 4, A–B, Fig. S3F). It is surprising that we were able to recover biallelically deleted clones of the individual 5' or 3' constituent enhancers without difficulty. Deletion of

either the 5' or 3' constituent enhancer results in an approximately 50% reduction in *Nanog* mRNA, but these reductions were insufficient to alter pluripotency as measured by *Oct4*, *Esrrb*, or *Klf4* expression (Fig. 4C). This demonstrates that neither the 5' nor the 3' constituent enhancer is required for pluripotency, even though together they promote normal *Nanog* expression. Thus, the two constituent enhancers function in an additive fashion and are required for proper *Nanog* expression.

The -5 SE regulates expression in all cells

Previous studies from our laboratory demonstrated that monoallelic deletion of the -5 SE, biallelic deletion of the -45 SE, and biallelic deletion of the $+60$ SE have different effects on *Nanog* expression despite each enhancer physically interacting with the gene, as shown by chromosomal conformation capture (7). Specifically, deletion of the -45 SE causes a 50% decrease in *Nanog* expression, whereas deletion of the $+60$ SE had no change in *Nanog* expression. Work from this study has further shown that the -5 SE is critical to *Nanog* expression, as there is a 90% decrease in *Nanog* expression upon biallelic deletion (Fig. 1E, right). We hypothesized that each enhancer may operate on distinct subpopulations of cells, which we could not distinguish using a bulk population. Specifically, the -5 SE could be regulating a larger proportion of high-*Nanog*-expressing cells (Fig. S4, A–C).

To investigate this possibility, we performed single-cell RT-qPCR on the -5 SE biallelically deleted cells with *Nanog* in

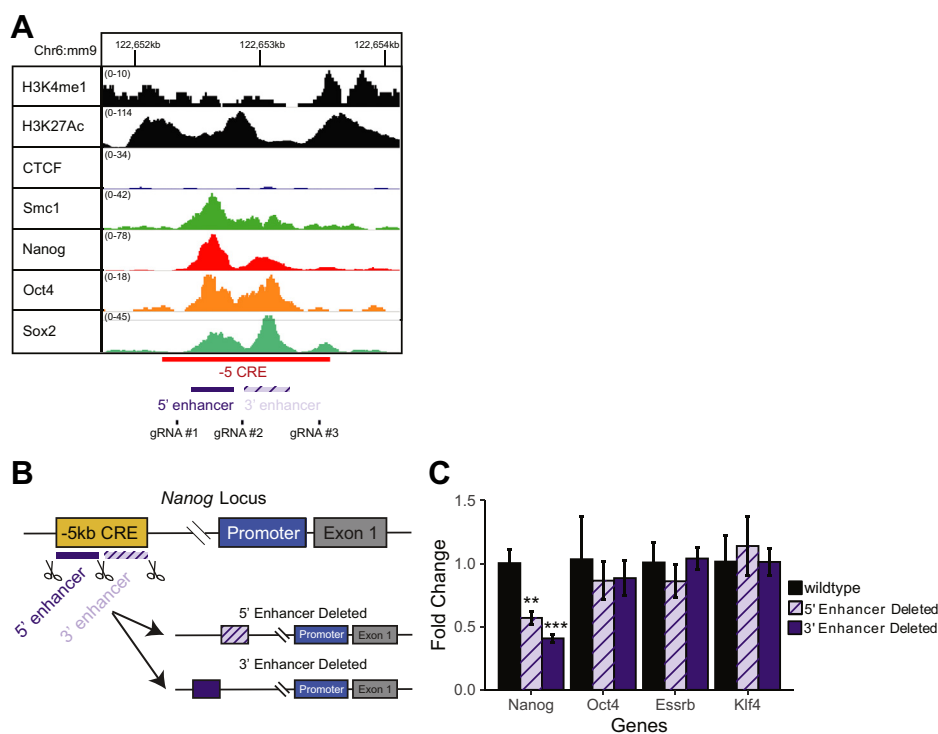


Figure 4. -5 SE constituent enhancers are both required. A, Integrated Genome Viewer snapshot of the -5 SE showing two constituent enhancers and locations of guide RNAs. The x-axis is genomic position; y-axis is normalized tag count. B, schematic. C, mRNA levels of *Nanog* and relevant pluripotency genes when two constituent enhancers within the -5 SE were deleted. $n = 3$. All mRNA levels measured by RT-qPCR and shown as $2^{\Delta\Delta\Delta C_T}$ compared with wildtype. ** $p < 0.01$, *** $p < 0.001$ Student's two sample t test. SE, super-enhancer.

trans (Fig. 1E). As described above, these are stably deleted clones that remain pluripotent in the presence of the exogenously supplied *Nanog* allowing us to examine changes in endogenous *Nanog* due to the loss of the -5 SE. If the -5 SE operates on *Nanog* in all cells, we should observe a uniform reduction in *Nanog* mRNA levels (Fig. S4B). By contrast, if the -5 SE regulates a different population of cells, we should observe a bimodal distribution of *Nanog* expression following enhancer deletion (Fig. S4B). It should be noted that, although some groups have shown that *Nanog* has bimodal expression in single cells (27), the presence of truly bimodal *Nanog* expression is debated (28, 29). Of interest, we found a uniform reduction in both total *Nanog* and endogenous *Nanog* expression as measured by the 3' UTR when the -5 SE is deleted (Fig. 5). Calculation of the bimodal coefficient (30) shows that both the wildtype and deleted lines demonstrate a nonbimodal distribution as both have a coefficient that is less than 5/9 (Fig. S4D). It should be noted that single-cell RT-qPCR data are more sensitive than bulk qPCR and the data are represented as a Log2 of expression rather than $\Delta\Delta CT$. All analyzed data points are present within Table S2. Thus, the difference in the mean of the single cell expression is a 2.6-fold reduction between the control and deleted lines, which is equivalent to an $\approx 84\%$ decrease in expression, consistent with our bulk qPCR data (Fig. 1E, right). Two controls, *Oct4* and *ERCC3*, showed minimal changes in expression. Collectively, these single cell experiments support a model that the -5 SE actively regulates *Nanog* expression in all ESCs grown in serum/LIF.

The -5 SE regulates *Nanog* by regulating transcriptional initiation/recruitment

Previous studies have argued that SEs regulate gene expression through promoter-proximal pause release (hereafter referred to as pause release) of RNAPII (31, 32). Briefly, transcription begins with the recruitment of RNAPII to the promoter, which is immediately phosphorylated on Ser 5 (Ser5P) of its C-terminal domain, resulting in bidirectional transcription around the TSS and is referred to as “paused” RNAPII because it cannot elongate further into the gene body.

The “pausing” of RNAPII occurs 20 to 120 bp downstream of the TSS, which must be relieved for productive gene transcription (10). “Pause release” is mediated by phosphorylation of Ser2, releasing RNAPII to transcribe the gene body and is referred to as elongating RNAPII. To measure changes in RNAPII dynamics, we performed CUT&Tag (33) with antibodies specific to total and paused (RNAPII-Ser5P) RNAPII in WT and -5 SE deleted cells with *Nanog* supplied *in trans* (0-copy cell line, Fig. 1E). We chose to use the 0-copy cell line to directly compare two clonal pluripotent cell lines where changes to *Nanog* are solely due to the enhancer deletion without the confounding issue of 4OHT treatment effects on the transcriptome. We note that we cannot judge if RNAPII dynamics or transcription itself may regulate the interaction between the -5 SE and the *Nanog* promoter.

Depending on which phase of transcription an enhancer is regulating, RNAPII's genomic location will change as shown in Fig. S5. As described above, RNAPII is phosphorylated on the Ser5 position of its C-terminal domain after recruitment, at which point it is paused. If recruitment is regulated by the enhancer, loss of the enhancer will cause a loss RNAPII-Ser5P at the TSS (Fig. S5-i). If pause release is being regulated, loss of an enhancer will cause a build-up of RNAPII-Ser5P that cannot be released (Fig. S5-ii). If neither of these are the steps being regulated by the enhancer, RNAPII-Ser5P enrichment will remain unchanged (Fig. S5-iii). A confounding issue of this system is that RNAPII binding at *Nanog* exons in the 0-copy cell line is obscured by the presence of the *Nanog* cDNA *in trans*, because the exogenous *Nanog* cDNA is identical to the endogenous coding regions. Therefore, we are unable to distinguish between the binding to the coding exons within endogenous *Nanog* versus *Nanog* cDNA supplied *in trans*. Thus, we limited our analysis to regions of the endogenous *Nanog* transcript that do not overlap with the exogenous transcript, which are the intronic and noncoding regions (blue and yellow areas in Fig. 6, A–B) and not the coding regions (grayed areas in Fig. 6, A–B). We observed a complete loss of paused RNAPII at both the -5 SE and *Nanog* in the 0-copy cell line (Fig. 6A), consistent with the -5 SE playing a critical role in RNAPII recruitment and/or phosphorylation on Ser5. These

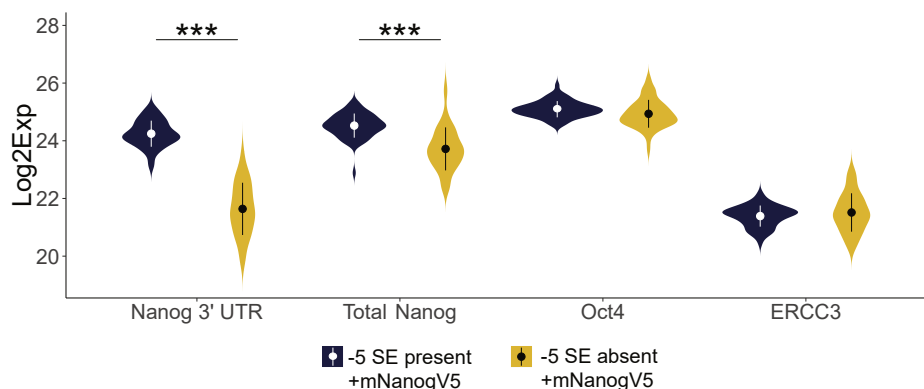


Figure 5. -5 SE operates in all cells. Single-cell RT-qPCR for endogenous *Nanog* via *Nanog* 3'UTR, *Total Nanog*, *Oct4*, and *ERCC3* in -5 SE-deleted cells. Expression is depicted as Log2exp relative to the Limit of Detection, as described by Fluidigm. $n = 51$ to 77. $***p < 0.001$ by Mann-Whitney test. The -5 SE-deleted cells express *Nanog* *in trans* to prevent a loss of pluripotency. SE, super-enhancer.

Defining an enhancer that modulates recruitment

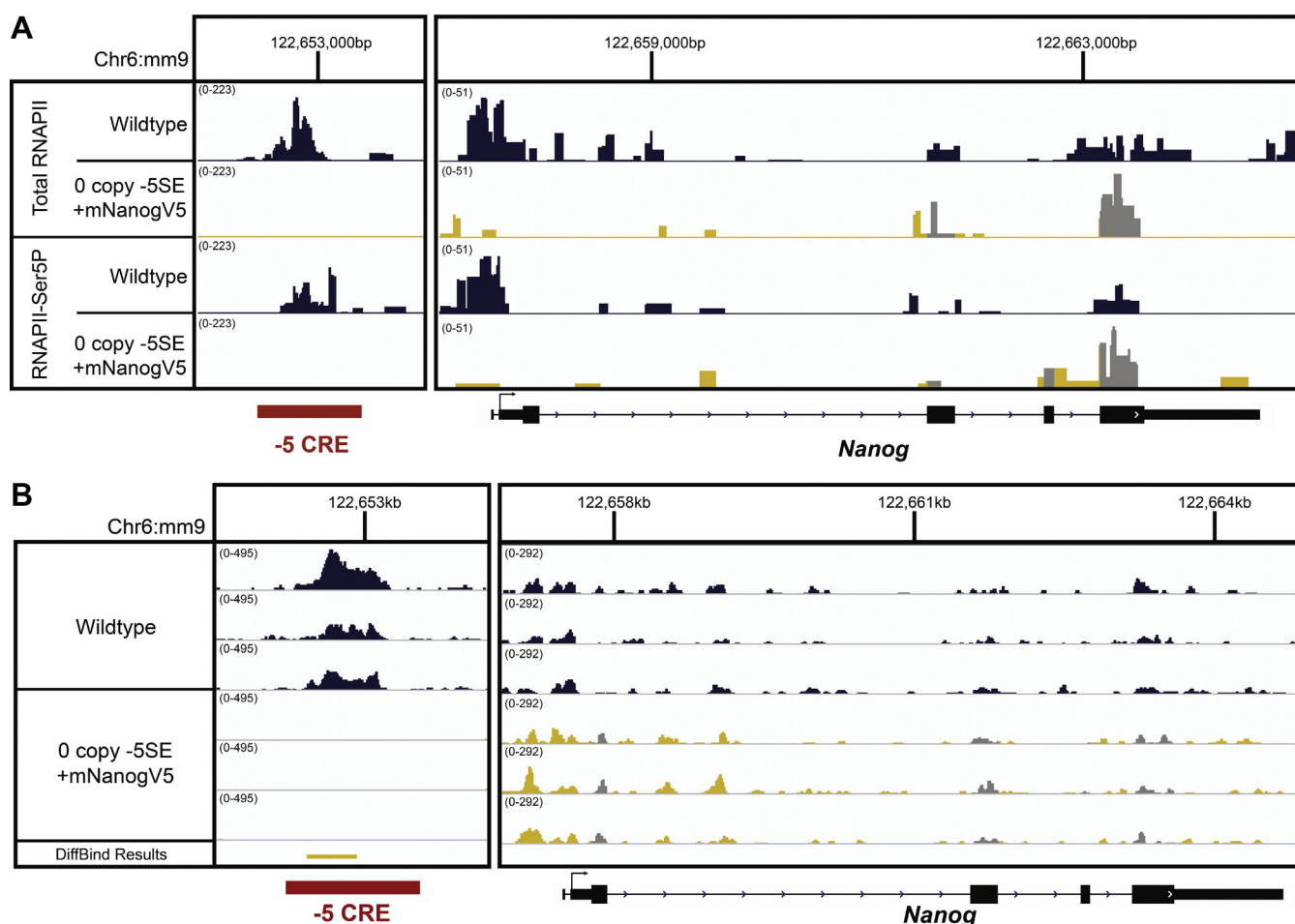


Figure 6. -5 SE operates prior to RNAPII pause release. *A*, Integrated Genome Viewer snapshot of CUT&Tag for Total RNAPII ($n = 1$) and RNAPII-Ser5P ($n = 3$) in wildtype and 0 copy of -5 SE cells. The *left panel* shows the -5 SE (y -axis = 0–233), and the *right panel* shows Nanog (y -axis = 0–51). Note that the exonic regions of the 0 copy cell line are confounded by the exogenous Nanog cDNA. *B*, Integrated Genome Viewer snapshot of ATAC-Seq data, separated by each sample of each wildtype cells where the -5 SE is floxed and cells with the -5 SE completely deleted. Differential peaks identified using DiffBind are shown in the last track. The *left panel* shows the -5 SE (y -axis = 0–495) and the *right panel* shows Nanog (y -axis = 0–292). Genes and SEs are shown below. The x -axis is genomic position; y -axis is normalized read count. The -5 SE-deleted cells express Nanog in *trans* (mNanogV5) to maintain pluripotency, and overlapping regions are shown in grayscale and are not included in the analysis. SE, super-enhancer.

data indicate that the -5 SE regulates *Nanog* not through RNAPII pause-release, but rather through modulating transcriptional initiation/recruitment.

These data led us to question whether the -5 SE modulates the initial binding of RNAPII to *Nanog* by changing chromatin accessibility, rather than recruiting RNAPII. To investigate the chromatin landscape, we performed assay for transposase-accessible chromatin using sequencing (ATAC-Seq) on wildtype cells and -5 SE-deleted cells with Nanog supplied in *trans* (Fig. 1D). As with the CUT&Tag data, chromatin accessibility changes within the *Nanog* exons will be confounded by the presence of the exogenous mNanogV5, and thus we focused only on changes within the regions of the *Nanog* transcript that do not overlap with the mNanogV5 (intronic and noncoding regions). We observed no significant differences in chromatin accessibility within the *Nanog* locus, although there is an expected change at the -5 SE given its deletion (Fig. 6B). DiffBind analysis (bottom track, Fig. 6B) concluded that the only major change in accessibility was at the -5 SE. Thus, we conclude that the mechanism by which the -5 SE regulates *Nanog* is by

modulating RNAPII recruitment/initiation but not through a change in chromatin accessibility. Although we cannot distinguish between transcriptional initiation *versus* recruitment, given the complete loss of RNAPII at *Nanog* upon deletion of the -5 SE, recruitment of RNAPII to the promoter is likely the rate-limiting step modulated by the enhancer, since generation of the initiating form of RNAPII (Ser5P) is not rate limiting. Together, these data show that the -5 SE is an enhancer critical to *Nanog* expression prior to pause-release of RNAPII.

Discussion

Although enhancers have been well known regulators of gene expression for 40 years, it has become apparent with new technologies that they are far more numerous than classical protein-coding genes (>4-fold) and unlike promoters far more variable across tissues, implying that enhancers play a central role in regulating tissue-specific expression (34–36). Given their importance, the breadth of questions that remain within the field is profound. In particular, the reliance on plasmid-based approaches have been *de rigueur* for the formal

definition of whether a CRE has enhancer activity. For the -5 SE, given its relative proximity to the *Nanog* TSS and the literature suggesting it may have a promoter-like activity (17), plasmid-based assays alone were unlikely to definitively address if the -5 CRE operates as an enhancer. Of importance, this DNA element may function as an alternative promoter in other contexts, such as alternative pluripotent states or in primordial germ cells, where *Nanog* is also expressed (37). Given the ease of genetically engineering ESCs using CRISPR-Cas9 technology, our initial goal was to move beyond plasmid-based approaches and utilize native cells/chromatin to determine whether the -5 SE truly displayed enhancer potential. We utilized a combination of classical genetic approaches to draw several conclusions about the -5 SE. First, the -5 SE is required for pluripotency, which can be genetically rescued by supplying *Nanog* *in trans*. Second, reversing the orientation of the -5 SE had no effect on pluripotency or *Nanog* expression. Third, moving the enhancer either closer to the *Nanog* TSS or within the insulated neighborhood but downstream of the TES permitted sufficient *Nanog* expression to maintain pluripotency. Of importance, these data conclusively demonstrate that the -5 SE is an enhancer. However, three caveats remain. First, although the -5 SE is an enhancer, our studies do not rule out that it can also act simultaneously as an alternative promoter, as has been suggested by others (17). Reversing the orientation of the -5 SE had no effect on *Nanog* expression and pluripotency, however, which formally demonstrates that any potential promoter activity of the -5 SE is not required for pluripotency. The second caveat is that insertion of the -5 SE downstream of the *Nanog* TES did not completely recapitulate normal expression. One reasonable explanation is that the -5 SE includes a larger chromatin domain, whereas we inserted only the core ≈ 2.5 kb into an alternative location. This may imply that additional sequences surrounding the core are required for full activation when it is moved into a new location. Nonetheless, this demonstration of enhancer function represents a highly feasible, native chromatin approach to assess enhancer function. Finally, given the fact that *Nanog* can bind to its own promoter to regulate its own expression, we cannot rule out that subtle changes in autoregulation are being disrupted through our various genetic alterations.

By the end of our studies we realized the *Nanog* locus has essentially been converted to a reporter gene, with minimal alteration beyond those described and the insertion of *LoxP* sites. This minimizes confounding variables inherent to artificial reporters such as the extensive presence of bacterial DNA sequences and/or insertion into a heterologous region of the genome. In addition, because *Nanog* is critical to ESCs, this permitted us to perform additional experiments to understand how the enhancer may modulate pluripotency and/or differentiation. One caveat is that many of our experiments were performed by supplying *Nanog* *in trans*, thereby permitting the cells to maintain pluripotency even when native *Nanog* expression was significantly reduced by deleting the -5 SE. Given our use of a heterologous, constitutively active promoter (CAG) and a *Nanog* cDNA that lacked the 5' UTR, we cannot ensure that subtle changes in the temporal regulation of *Nanog*

expression or *Nanog* protein levels were preserved. Nonetheless, supplying *Nanog* *in trans* was able to restore approximately heterozygous levels of *Nanog*, which is sufficient to maintain pluripotency and substantially suppress the spontaneous differentiation of ESCs (18, 19).

The recent literature has shown that there are many mechanisms by which enhancers regulate gene expression. It was previously thought that enhancers promoted the recruitment of RNAPII and other transcriptional machinery to the promoter to "promote" transcription (9). Current studies focusing on highly active enhancers have centered on their role in RNAPII pause release, which constitutes the first steps of RNAPII converting from the initiating to the elongating form (31, 38). Multiple studies have also shown that enhancers regulate transcriptional bursting (11, 12), the observation of oscillating transcriptional activity over time, in which one transcriptional burst is a period of time during which there is active transcription. Enhancers have been shown to regulate the frequency of these bursts of activity, which represent a combination of initiation and elongation. More recently, enhancers have been implicated in forming phase-separated condensates that concentrate transcriptional machinery for actively transcribed genes (15, 39, 40). Critically, these studies have not demonstrated how multiple enhancers could simultaneously regulate expression of a single gene, and if enhancers uniformly operate through the same or different mechanisms on the same gene. Since the -5 SE is indispensable for *Nanog* expression, and cannot be compensated for by another enhancer, the obvious question is whether this is because the other enhancers function through different mechanisms, or perhaps function in other pluripotent states. Deletion of each enhancer at the locus causes variable changes in *Nanog* expression (7), leading us to ask if each enhancer may play a unique role in regulating *Nanog* expression through different phases of transcription. For example, it may be that, while the -5 SE does not regulate *Nanog* through pause-release, another enhancer plays this more traditional role.

Given the broad role of enhancers in regulating tissue-specific gene expression, our work has implications for how other gene:enhancer pairs are studied. In the absence of genetic confirmation, it is difficult to confirm an enhancer:gene functional dyad based solely on plasmid-based approaches. In addition, further attention needs to be paid to the other enhancers in the region to understand how multiple enhancers work together to regulate a gene. Understanding the interplay of the three SEs around *Nanog* will further drive changes in how gene:enhancer pairs are studied, especially since they may operate through different phases of transcription to regulate expression.

Experimental procedures

For further information and requests for reagent and resources, please contact the Lead Contact, Sridhar Rao (Sridhar.rao@versiti.org; 414-937-3841).

Cell culture

Gelatin-adapted ESCs were utilized for all experiments. This cell line is a male, in-house generated, ICM-derived 129SVJ-

Defining an enhancer that modulates recruitment

derived murine ESC line, similar to the one we have used previously and cultured under similar Serum/LIF conditions (41, 42). Briefly, cells were propagated under feeder-free conditions in Dulbecco's modified Eagle's medium (Corning #10-017-CV) with the following supplements (FBS, Gibco #100-106; Penicillin/Streptomycin, Corning #30-002-Cl; MEM Nonessential Amino Acids, Corning #25-025-Cl; L-glutamine, Corning #25-005-Cl; Nucleosides, Sigma #ES-008-D; LIF, β -mercaptoethanol at the appropriate concentration). Two micromoles of 4-hydroxytamoxifen (4OHT) in 70% ethanol (EtOH) was used for all experiments and diluted 1:1000 for drug treatments with EtOH as a control.

CRISPR-Cas9-mediated genomic editing

To generate biallelic *loxP* ESC clones, single gRNAs targeting specific regions flanking CREs were designed using the CRISPR design tool (<http://crispr.mit.edu/>). gRNAs were cloned into the Cas9 expressing vector px459 v2.0 (Addgene #62988, 43, 44). Single-strand DNA oligos were designed with ~60-bp homology directed repair (HDR) arms flanking each side of the 34-bp *loxP* sequence and a restriction enzyme palindromic sequence (BamHI) for restriction digest genotyping of genomic PCR products. The *loxP* and restriction enzyme sequence was inserted between the PAM recognition sequence and the gRNA genomic targeting sequence. A single gRNA and single-strand HDR oligo were cotransfected along with the gRNA (1–2 μ g of each plasmid) into 1 to 2 $\times 10^6$ WT ESCs using Lipofectamine 2000 (Invitrogen #1168-019) in a single well of a 6-well plate. Transfected ESCs were selected with puromycin (2 μ g/ml for 2 days only) and then passaged onto 10-cm dishes at various dilutions and grown until single colonies appeared. Individual clones that were resistant to puromycin were isolated and expanded for genotyping. Primers designed outside of the HDR arms were used to genotype for enhancer deletion. Following genomic PCR, products were digested with XbaI (NEB, R0145s) for genotyping. Clones that demonstrated biallelic cutting were cloned into TOPO TA (Thermo Fisher #45-0641) for sequencing to confirm correct integration. *loxP* sequences (upstream or downstream of the targeted region) were inserted one at a time.

Cell lines described in Table S3 were generated using the following CRISPR strategy. gRNAs were cloned into px459 v2. The plasmid was digested using BbsI (NEB, R0539) and purified. gRNA oligos were phosphorylated and annealed using T4 PNK (NEB, M0201). The cut vector and annealed oligos were ligated overnight at 16 °C. Ligated plasmids were transformed into NEB High Efficiency (NEB, C2987) bacteria, plated on LB+Amp plates and incubated overnight at 37 °C. Colonies were picked and mini-prepped for sequencing, followed by maxi-preps once gRNA presence was verified. The same transfection protocol described above was followed, and at least two single cell clones were picked and analyzed for the following cell lines.

To generate a floxed –5 *Nanog* CRE ESC line for conditional deletion the 3' *loxP* was inserted first using a gRNA and a HDR

arm (Table S4). The 5' *loxP* sequence was subsequently inserted using a gRNA and a HDR arm. The PAM sequences adjacent to the *loxP* sequences were mutated to prevent cutting of Cas9 following HDR. The gRNAs were used to constitutively delete the –5 CRE in (7). ESCs were treated with 4OHT for 4 days at 2 μ M to delete the –5 *Nanog* CRE.

To generate –5 *Nanog* CRE inverted clones we inserted the 5' HDR oligo containing *loxP* in the opposite orientation into the ESC clone that contains the 3' *loxP* above. The same 5' gRNA above was used with a single-strand HDR oligo. An ESC clone with *loxP* sequences in the opposite orientations was treated with 4-OHT at 2 μ M for 3 days and cells were subcloned as described above. We confirmed biallelic inversion by genomic PCR with primers inside and outside of the *loxP* sequences. Clones that demonstrated wildtype, monoallelic, and biallelic orientation were cloned into TOPO TA plasmids. At least four individual clones were then isolated and sequenced to confirm correct integration.

To insert the –5 *Nanog* CRE downstream of the *Nanog* gene, a single gRNA was used to stimulate HDR of a modified version of pL451 (*loxP* sequence removed). The HDR vector contains a *Neomycin resistance (Neo)* cassette flanked *frt* sites and by homology regions (left arm chr6:122667133–122668329, 1197 bp, mm9; right arm chr6:122668389–122669450, 1062 bp, mm9). The left arm was cloned using KpnI and Sall, and the right arm was cloned using BamHI and NotI. The enhancer (same sequence used in reporter assays in Blinka *et al.*, 2016) was inserted adjacent to the left arm using Sall and EcoRI sites. The HDR plasmid was cotransfected along with a gRNA 5'-TGGCTTGCATCCAATCTCTT-3' chr6: 122668369, mm9 (2–3 μ g of gRNA and 6 μ g of HDR plasmid) into 10 $\times 10^6$ WT ESCs using Lipofectamine 2000. HDR vector arms and the enhancer were amplified off of a BAC (7) and fully sequenced in pBlueScript II SK(+) and matched the genomic reference sequence. Transfected ESCs were selected with puromycin (2 μ g/ml first 2 days only) and G418 (350 μ g/ml days 2–14) until single colonies appeared. Individual clones that were resistant to both puromycin and G418 were isolated and expanded for genotyping. *Neo* was removed by transfecting cells with a FLPe expressing plasmid driven by the CAG promoter. Primers designed outside of the HDR arms were used to genotype for enhancer insertion. Following genomic PCR to genotype, homozygous clones containing *Neo* were amplified and cloned into TOPO TA for sequencing to confirm correct integration.

Constituent enhancer deletions were generated using three gRNAs. Clones were genotyped using PCR primers designed around the constituent enhancers. Distance deletion clones were generated using four gRNAs. Clones were genotyped using PCR primers that surrounded the deleted portion.

All gRNAs and genotyping primers are listed in Table S4.

Generation of murine *Nanogv5* rescue cell line

The mouse *Nanog* sequence was synthesized by GeneArt Strings DNA Fragment (ThermoFisher). A C-terminal v5 tag was added to distinguish from endogenous *Nanog* protein. The synthesized DNA fragment was A-tailed and cloned into

TOPO TA (Thermo Fisher #45-0641) to confirm the sequence. XhoI and NotI sites were designed at the 5' and 3' end of the DNA fragment so that it could be cloned into and the pPyCAG iH vector (hygromycin resistance, gifted from Austin Smith) for expression under a ubiquitous (CAG) promoter (13, 42). ESCs were electroporated with the linearized plasmid (FspI NEB R0135) in the presence of hygromycin, and individual clones were isolated as we have done previously and expanded for further experiments.

Total RNA RT-qPCR

Total RNA was harvested from cells following manufacturer's protocol (TRIzol Reagent, Invitrogen #15596018). Genomic DNA was removed from the total RNA samples using a DNA eliminator column step and passing RNA over a column following manufacturer's protocol (RNeasy Plus Mini Kit, Qiagen #74134). Equal amounts of DNA-free total RNA were converted to cDNA using the iScript cDNA synthesis kit (Bio-Rad #1708891). Quantitative PCR (qPCR) was performed on a QuantStudio 6 Flex Real-Time PCR System (Thermo-Fisher). Quantifications were normalized to an internal control (Actin) for reverse transcriptase-qPCR (RT-qPCR) using the $\Delta\Delta C_t$ method as we have done previously (7). Primers used for RT-qPCR are described in Table S4.

Alkaline phosphatase staining

Bright-field images and alkaline phosphatase staining were performed as previously described (Rao *et al.*, 2010, Sigma 86R-1kt). Briefly, cells were plated in 10-cm dishes and treated for up to 6 days with vehicle or tamoxifen. Plates were rinsed 1x with PBS, fixed using citrate-acetone-formaldehyde fixative for 30s, and rinsed with deionized water for 45s. An alkaline-dye mixture (diazonium salt solution + deionized water + Naphthol As-BI Alkaline Solution) was added to the plate and incubated for 15 min at room temperature in the dark. The dye mixture was removed from the plates, and the plates were rinsed for 2 min with deionized water and then air dried. For LIF withdrawal experiments, 5000 cells were plated in 6-well dishes and provided new media daily with no LIF for 6 days. Cells were then stained with alkaline phosphatase as described above.

Western blots

Proteins were extracted in radioimmunoprecipitation buffer and quantified as described in (45). A total of 10 μ g of protein was loaded in each well of a gel (Bio-Rad # 567-1094, 567-1095, 456-1036). Blots were blocked in 5% milk/tris buffered saline with tween (TBST) for 20 min to 2 h at room temperature (RT) or overnight at 4 °C. Primary antibodies to Nanog (Millipore; Cat # 5731) was used at 1:1000, beta-Actin (Sigma; Cat # a5441) was used at 1:5000 in 5% milk/TBST or GAPDH-HRP (Cell Signaling Tech; 51332S) at 1:3000 in 5% milk/TBST for 90 min at RT or overnight at 4 °C. Blots were then washed with TBST and secondary antibody donkey anti-rabbit IgG-HRP (Santa Cruz; Cat # sc2313) was used at 1:5000 for 30 to 60 min at RT for Nanog. For beta-Actin, a secondary antibody

Defining an enhancer that modulates recruitment

(Santa Cruz; Cat # sc2064) goat anti-mouse IgM-HRP was used at 1:5000 for 30 min at RT. Blots were washed with TBST and then antibody labeled proteins were detected using Amersham ECL Prime Western Blotting Detection Reagent (Cat # RPN2232).

RNA-Seq

RNA-Seq libraries were generated using the NEBNext Ultra RNA Library Prep Kit for Illumina (NEB #E7530). Libraries were quantified using the NEBNext Quant Kit (NEB #E7630) and Agilent TapeStation 2200 (D1000 tapes) and were sequenced on a NextSeq 500(36 × 36 PE). Library preparation and sequencing were performed following manufacturer's protocol. Data were analyzed using STAR (mm9, (46)), Cufflinks (47), and DESeq (48) using default parameters through Basepair (www.basepairtech.com). Genes altered on chromosome 6 are provided in Table S1.

Single-cell analysis

Single-cell analysis was performed using the Fluidigm C1 and BiomarkHD system following manufacturer's protocol. Data were analyzed first using the Fluidigm Real-Time PCR Analysis software to remove any data point with a poor melt curve or no amplification and were further analyzed using R. Cts were normalized to *ACTB* measurements and any cell with an *ACTB* measurement above 8 was excluded as the quality of those samples could not be ensured (Table S2). Data are represented as a difference from the Limit of Detection (as described by Fluidigm, SINGuLAR Analysis Toolset) and expressed as Log₂Expression. Statistical difference was tested using a Mann-Whitney test with a *p*-value of 0.001. Bimodal distributions were analyzed by calculating the bimodal coefficient (30).

CUT&Tag

A total of 100,000 cells were collected and processed through the method described in Kaya-Okur *et al.*, 2019, for Total RNA Polymerase II and RNA Polymerase II Ser5P (Cell Signaling Technologies, #54020). Libraries were quantified using the KAPA Quant Kit (#07960140001) and Agilent TapeStation and were sequenced on a NextSeq 500(36 × 36 PE). Data were processed as described in Kaya-Okur *et al.*, 2019.

ATAC-Seq

ATAC-Seq libraries were generated as described previously on cells with the -5 SE floxed and cells with the -5 SE deleted, with Nanog expressed exogenously (41). ESCs were plated 24 h prior to the experiment, collected, and transposed for 30 min. Data were analyzed using bowtie2 (49) using default parameters through Basepair (www.basepairtech.com). Differential peaks were identified using DiffBind (50).

Data set reanalyses

All Chromatin Immunoprecipitation Sequencing (ChIP-Seq) and Global Run-on Sequencing (GRO-Seq) data sets were

Defining an enhancer that modulates recruitment

displayed using the Integrated Genome Viewer (data.broadinstitute.org). These data sets were previously downloaded and analyzed from the GEO omnibus (13). Data sets are listed in Table S5.

Statistical analyses

Statistical analyses were done using Microsoft Excel and R. Statistical details of experiments can be found in the figure legends. Two sample two-tailed Student's *t* test comparisons were performed and *p*-values < 0.05 were considered significantly different. Statistical significance was not shown for values within 20% of the control or between experimental values for RT-qPCR experiments as that is within the error of the assay. All error bars shown in figures are standard deviation between independent experimental replicates. For single-cell RT-qPCR, the Mann-Whitney test was performed and *p* values < 0.001 were considered significantly different. Error bars are shown as SD.

Data availability

All NGS-derived data are available on the GEO (GSE143993). All other data are available upon request to Dr Sridhar Rao (sridhar.rao@versiti.org).

Acknowledgments—The authors would like to thank Drs Emery Bresnick (University of Wisconsin); Erin M. Wissink and John T. Lis (Cornell University) for their assistance.

Author contributions—P. A., S. B., C. S., A. E. M., and M. H. R. performed all experiments. K. P. and P. A. analyzed NGS data. P. A., S. B., M. H. R., and S. R. designed and interpreted experiments collectively. P. A., S. B., and S. R. wrote the manuscript with input from all authors.

Funding and additional information—This work was supported in part by funding from NIDDK (DK120152, DK10350) to P. A. and S. B., respectively, and the MCW MSTP T32 (NIGMS, GM080202) to P. A. and S. B. Additional support came from NIH (CA204231) to S. R. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflicts of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: 4OHT, 4-hydroxytamoxifen; CRE, *cis* regulatory element; ESC, embryonic stem cells; gRNA, guide RNA; HDR, homology directed repair; LIF, leukemia inhibitory factor; NOS, Nanog, Oct4, and Sox2; RNAPII, RNA Polymerase II; SE, super-enhancer; TBST, tris buffered saline with tween; TES, transcriptional end site; TF, transcription factors; TSS, transcriptional start site.

References

1. Banerji, J., Rusconi, S., and Schaffner, W. (1981) Expression of a β -globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**, 299–308
2. Moreau, P., Hen, R., Wasyluk, B., Everett, R., Gaub, M. P., and Chambon, P. (1981) The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res.* **9**, 6047–6068
3. Catarino, R. R., and Stark, A. (2018) Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev.* **32**, 202–223
4. Rickels, R., and Shilatifard, A. (2016) Enhancer logic and mechanics in development and disease. *Trends Cell Biol.* **28**, 608–630
5. Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002) Capturing chromosome conformation. *Science* **295**, 1306–1311
6. Yu, M., and Ren, B. (2017) The three-dimensional organization of mammalian genomes. *Annu. Rev. Cell Dev. Biol.* **33**, 265–289
7. Blinka, S., Reimer, M. H., Pulakanti, K., and Rao, S. (2016) Super-enhancers at the Nanog locus differentially regulate neighboring pluripotency-associated genes. *Cell Rep.* **17**, 19–28
8. Blinka, S., and Rao, S. (2017) Nanog expression in embryonic stem cells - an ideal model system to dissect enhancer function. *BioEssays* **39**, 1700086
9. Rippe, K., von Hippel, P. H., and Langowski, J. (1995) Action at a distance: DNA-looping and initiation of transcription. *Trends Biochem. Sci.* **20**, 500–506
10. Chen, F. X., Smith, E. R., and Shilatifard, A. (2018) Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **19**, 464–478
11. Bartman, C. R., Hsu, S. C., Hsiung, C. C. S., Raj, A., and Blobel, G. A. (2016) Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol. Cell* **62**, 237–247
12. Fukaya, T., Lim, B., and Levine, M. (2016) Enhancer control of transcriptional bursting. *Cell* **166**, 358–368
13. Pulakanti, K., Pinello, L., Stelloh, C., Blinka, S., Allred, J., Milanovich, S., Kiblawi, S., Peterson, J., Wang, A., Yuan, G. C., and Rao, S. (2013) Enhancer transcribed RNAs arise from hypomethylated, Tet-occupied genomic regions. *Epigenetics* **8**, 1303–1320
14. Whyte, W. A., Orlando, D. A., Hnisz, D., Abraham, B. J., Lin, C. Y., Kagey, M. H., Rahl, P. B., Lee, T. I., and Young, R. A. (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319
15. Sabari, B. R., Dall'Agnese, A., Boija, A., Klein, I. A., Coffey, E. L., Shrinivas, K., Abraham, B. J., Hannett, N. M., Zamudio, A. V., Manteiga, J. C., Li, C. H., Guo, Y. E., Day, D. S., Schuijers, J., Vasile, E., *et al.* (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958
16. Levasseur, D. N., Wang, J., Dorschner, M. O., Stamatoyannopoulos, J. A., and Orkin, S. H. (2008) Oct4 dependence of chromatin structure within the extended Nanog locus in ES cells. *Genes Dev.* **22**, 575–580
17. Das, S., Jena, S., and Levasseur, D. N. (2011) Alternative splicing produces nanog protein variants with different capacities for self-renewal and pluripotency in embryonic stem cells. *J. Biol. Chem.* **286**, 42690–42703
18. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655
19. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003) The homeo-protein nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642
20. Mikhaylichenko, O., Bondarenko, V., Harnett, D., Schor, I. E., Males, M., Viales, R. R., and Furlong, E. E. M. (2018) The degree of enhancer or promoter activity is reflected by the levels and directionality of eRNA transcription. *Genes Dev.* **32**, 42–57
21. Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K. Y., Sung, K. W., Lee, C. W. H., Zhao, X. D., Chiu, K. P., *et al.* (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38**, 431–440
22. Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., and Ren, B. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380

23. Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S. H. (2008) An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049–1061
24. Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y. H., Yeo, H. C., Yeo, Z. X., Narang, V., Govindarajan, K. R., *et al.* (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117
25. Downen, J. M., Fan, Z. P., Hnisz, D., Ren, G., Abraham, B. J., Zhang, L. N., Weintraub, A. S., Schuijers, J., Lee, T. I., Zhao, K., and Young, R. A. (2014) Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* **159**, 374–387
26. Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., Trompouki, E., Bowman, T. V., Zon, L. I., Yuan, G. C., Orkin, S. H., and Xu, J. (2016) Dynamic control of enhancer repertoires drives lineage and stage-specific transcription during hematopoiesis. *Dev. Cell* **36**, 9–23
27. Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007) Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234
28. Faddah, D. A., Wang, H., Cheng, A. W., Katz, Y., Buganim, Y., and Jaenisch, R. (2013) Single-cell analysis reveals that expression of nanog is biallelic and equally variable as that of other pluripotency factors in mouse escs. *Cell Stem Cell* **13**, 23–29
29. Filipczyk, A., Gkatzis, K., Fu, J., Hoppe, P. S., Lickert, H., Anastassiadis, K., and Schroeder, T. (2013) Biallelic expression of nanog protein in mouse embryonic stem cells. *Cell Stem Cell* **13**, 12–13
30. Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., and Freeman, J. B. (2013) Good things peak in pairs: a note on the bimodality coefficient. *Front. Psychol.* **4**, 700
31. Chen, F. X., Xie, P., Collings, C. K., Cao, K., Aoi, Y., Marshall, S. A., Rendleman, E. J., Ugarenko, M., Ozark, P. A., Zhang, A., Shiekhattar, R., Smith, E. R., Zhang, M. Q., and Shilatifard, A. (2017) PAF1 regulation of promoter-proximal pause release via enhancer activation. *Science* **357**, 1294–1298
32. Henriques, T., Scruggs, B. S., Inouye, M. O., Muse, G. W., Williams, L. H., Burkholder, A. B., Lavender, C. A., Fargo, D. C., and Adelman, K. (2018) Widespread transcriptional pausing and elongation control at enhancers. *Genes Dev.* **32**, 1–16
33. Kaya-Okur, H. S., Wu, S. J., Codomo, C. A., Pledger, E. S., Bryson, T. D., Henikoff, J. G., Ahmad, K., and Henikoff, S. (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1930
34. Heintzman, N. D., and Ren, B. (2009) Finding distal regulatory elements in the human genome. *Curr. Opin. Genet. Dev.* **19**, 541–549
35. Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311–318
36. Heintzman, N. D., Hon, G. C., Hawkins, R. D., Kheradpour, P., Stark, A., Harp, L. F., Ye, Z., Lee, L. K., Stuart, R. K., Ching, C. W., Ching, K. A., Antosiewicz-Bourget, J. E., Liu, H., Zhang, X., Green, R. D., *et al.* (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* **459**, 108–112
37. Murakami, K., Günesdogan, U., Zyllicz, J. J., Tang, W. W. C., Sengupta, R., Kobayashi, T., Kim, S., Butler, R., Dietmann, S., and Surani, M. A. (2016) NANOG alone induces germ cells in primed epiblast *in vitro* by activation of enhancers. *Nature* **529**, 403–407
38. Schaukowitz, K., Joo, J. Y., Liu, X., Watts, J. K., Martinez, C., and Kim, T. K. (2014) Enhancer RNA facilitates NELF release from immediate early genes. *Mol. Cell* **56**, 29–42
39. Cho, W. K., Spille, J. H., Hecht, M., Lee, C., Li, C., Grube, V., and Cisse, I. I. (2018) Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415
40. Guo, Y. E., Manteiga, J. C., Henninger, J. E., Sabari, B. R., Agnese, A. D., Shrinivas, K., Abraham, B. J., Hannett, N. M., Spille, J., Afeyan, L. K., Alicia, V., Bojja, A., Decker, T., Rimel, J. K., Fant, C. B., *et al.* (2019) Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature* **572**, 543–548
41. Volk, A., Liang, K., Suraneni, P., Li, X., Zhao, J., Bulic, M., Marshall, S., Pulakanti, K., Malinge, S., Taub, J., Ge, Y., Rao, S., Bartom, E., Shilatifard, A., and Crispino, J. D. (2018) A CHAF1B-dependent molecular switch in hematopoiesis and leukemia pathogenesis. *Cancer Cell* **34**, 707–723.e7
42. Rao, S., Zhen, S., Roumiantsev, S., McDonald, L. T., Yuan, G.-C., and Orkin, S. H. (2010) Differential roles of Sall4 isoforms in embryonic stem cell pluripotency. *Mol. Cell Biol.* **30**, 5364–5380
43. Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013) Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308
44. Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823
45. Minella, A. C., Loeb, K. R., Knecht, A., Welcker, M., Varnum-Finney, B. J., Bernstein, I. D., Roberts, J. M., and Clurman, B. E. (2008) Cyclin E phosphorylation regulates cell proliferation in hematopoietic and epithelial lineages *in vivo*. *Genes Dev.* **22**, 1677–1689
46. Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21
47. Trapnell, C., Williams, B. A., Pertea, G., Mortazavi, A., Kwan, G., Van Baren, M. J., Salzberg, S. L., Wold, B. J., and Pachter, L. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515
48. Anders, S., and Huber, W. (2010) Differential expression analysis for sequence count data. *Genome Biol* **11**, R106
49. Langmead, B., and Salzberg, S. L. (2012) Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359
50. Ross-Innes, C. S., Stark, R., Teschendorff, A. E., Holmes, K. A., Ali, H. R., Dunning, M. J., Brown, G. D., Gojis, O., Ellis, I. O., Green, A. R., Ali, S., Chin, S. F., Palmieri, C., Caldas, C., and Carroll, J. S. (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393