A dynamic interplay of enhancer elements regulates *Klf4* expression in naïve pluripotency

Liangqi Xie,[1] Sharon E. Torigoe,[1,5] Jifang Xiao,[2] Daniel H. Mai,[2] Li Li,[3] Fred P. Davis,[4] Peng Dong,[3] Herve Marie-Nelly,[2] Jonathan Grimm,[3] Luke Lavis,[3] Xavier Darzacq,[1] Claudia Cattoglio,[1] Zhe Liu,[3] and Robert Tjian[1,2]

[1]Howard Hughes Medical Institute, Berkeley, California 94720, USA; [2]Department of Molecular and Cell Biology, Li Ka Shing Center for Biomedical and Health Sciences, California Institute for Regenerative Medicine Center of Excellence, University of California at Berkeley, Berkeley, California 94720, USA; [3]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA; [4]Molecular Immunology and Inflammation Branch, National Institute of Arthritis and Musculoskeletal and Skin Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

Transcription factor (TF)-directed enhanceosome assembly constitutes a fundamental regulatory mechanism driving spatiotemporal gene expression programs during animal development. Despite decades of study, we know little about the dynamics or order of events animating TF assembly at *cis*-regulatory elements in living cells and the long-range molecular "dialog" between enhancers and promoters. Here, combining genetic, genomic, and imaging approaches, we characterize a complex long-range enhancer cluster governing *Krüppel-like factor 4* (*Klf4*) expression in naïve pluripotency. Genome editing by CRISPR/Cas9 revealed that OCT4 and SOX2 safeguard an accessible chromatin neighborhood to assist the binding of other TFs/cofactors to the enhancer. Single-molecule live-cell imaging uncovered that two naïve pluripotency TFs, STAT3 and ESRRB, interrogate chromatin in a highly dynamic manner, in which SOX2 promotes ESRRB target search and chromatin-binding dynamics through a direct protein-tethering mechanism. Together, our results support a highly dynamic yet intrinsically ordered enhanceosome assembly to maintain the finely balanced transcription program underlying naïve pluripotency.

Precise spatiotemporal gene regulation is fundamental to organismal development, and disruption of this process often lies at the root of human diseases (Levine and Tjian 2003). In metazoans, cell type-specific transcription programs are initiated by the binding of site-specific transcription factors (TFs) to *cis*-regulatory elements and are reinforced by long-range communications between distal enhancers and proximal/core promoters (Levine et al. 2014). Understanding the function and mechanism of dynamic TF assembly at enhancers and long-distance enhancer–promoter cross-talk presents a formidable challenge in the post-genomic era.

Transcription regulation is seminal to lineage specification and tissue patterning during mammalian embryogenesis, in which the inner cell mass (ICM) segregates into the preimplantation epiblast representing the ground or naïve pluripotent state and the supporting hypoblast. Following uterine implantation, the epiblast can differentiate into the three germ layers but not blastocyst chimeras. The naïve epiblast-derived mouse embryonic stem cells (mESCs) and post-implantation epiblast-derived stem cells (EpiSCs) are believed to recapitulate the naïve and primed pluripotent states, respectively (Nichols and Smith 2009). Importantly, both human ESCs and induced pluripotent stem cells (PSCs) resemble the primed epiblast state and thus are limited in their use for various biomedical applications. Because the naïve pluripotent state holds promise to boost genome editing efficiency, allow industrial-scale culture growth, and efficiently contribute to humanized organogenesis (Mascetti and Pedersen 2014), there is great interest in deciphering the organizing principles of the transcription programs that establish, maintain, and terminate the naïve pluripotent state.

[5]**Present address: Department of Biology, Lewis and Clark College, Portland, OR 97219, USA.**
Corresponding authors: jmlim@berkeley.edu, liuz11@janelia.hhmi.org
Article is online at http://www.genesdev.org/cgi/doi/10.1101/gad.303321.117.

Xie et al.

The Yamanaka factor Krüppel-like factor 4 (KLF4) is a key TF regulating the gene expression program of naïve pluripotency (Takahashi and Yamanaka 2006). Ectopic *Klf4* expression is sufficient to reprogram primed mouse EpiSCs to naïve ESCs (Guo et al. 2009), while depletion of KLF4 alone is sufficient to collapse the human naïve pluripotent state (Takashima et al. 2014). Moreover, KLF4 is highly abundant in the ICM of human blastocysts and human naïve state-like PSCs but not in conventional human PSCs (hPSCs) resembling the primed state (Takashima et al. 2014). Despite this critical role in naïve pluripotency, the organizational structure, functional relationship, and assembly dynamics of TFs at the *Klf4* enhancer remain largely uncharted territory. This paucity of information is in part a consequence of inherent difficulties in accurately mapping and functionally characterizing complex enhancers in the context of an endogenous locus in native chromatin within a biologically relevant cell type. The task of identifying functional *cis*-elements that operate combinatorially is further exacerbated by metazoan enhancers that are typically dispersed over long stretches (10–100 kb) of DNA, with short functional elements interspersed between gene deserts. Moreover, the ability to discern mechanisms of TF activity and their highly dynamic binding properties in vivo have only recently become possible with the advent of imaging modalities capable of single-molecule measurements in individual live cells. Dissecting the architecture and dynamics of the TF circuitry governing *Klf4* expression in PSCs could shed new light on the fundamental mechanisms governing transcriptional control of naïve pluripotency, an essential step toward developing regenerative therapeutics.

Because of persistent uncertainties as well as inherent difficulties in reprogramming conventional hPSCs to naïve pluripotency (Rossant and Tam 2017), here we chose to study expression of *Klf4* and transcriptional programs directing naïve pluripotency in mESCs. By combining genetic and genomic approaches, we report the detailed functional mapping of a distal enhancer cluster driving *Klf4* gene expression. Using a CRISPR/dCas9 tiling screen and ChIP-exo (chromatin immunoprecipitation [ChIP] using λ exonuclease to digest TF-unbound DNA after ChIP) sequencing, we identified key TFs bound to *cis*-regulatory elements within the enhancer cluster. Targeted mutagenesis revealed that OCT4 and SOX2 serve as lead factors maintaining an open chromatin architecture and assisting the binding of other TFs/cofactors (e.g., STAT3 and ESRRB) to the *Klf4* enhancer cluster. Live-cell single-molecule imaging experiments revealed a highly dynamic binding of STAT3 and ESRRB to cognate sites in chromatin with much slower target site sampling and shorter temporal occupancy than lead factors OCT4 and SOX2. Our single-particle tracking (SPT) studies also found that SOX2 facilitates ESRRB target search in both live mESCs and terminally differentiated cells likely involving a protein tethering mechanism. Together, our results highlight a dynamic and ordered mechanism underlying TF assembly at enhancers and unmask key organizing principles governing distal enhancer-mediated regulation of the naïve pluripotency state.

Results

Identification of an enhancer cluster driving Klf4 expression in mESCs

The *Klf4* gene is located within a gene-poor region on mouse chromosome 4, on which there are three major DNase I-hypersensitive sites located 50–70 kb downstream from the *Klf4* transcription start site (TSS). As these sites are enriched for the active enhancer-associated histone modification H3K27ac but not the promoter-associated H3K4me3 mark, we designated these three putative enhancer domains as E1, E2, and E3 (Fig. 1A; Supplemental Fig. S1A). DNA fragments containing E1 and E2 specifically exhibited high transcriptional activity in naïve state mESCs relative to primed EpiSCs as determined by transgene-driven luciferase assays, consistent with the *Klf4* expression pattern in these two pluripotent cell types (Fig. 1B,C). To further evaluate the activities of these regions, we used the catalytically dead Cas9 (dCas9) (Jinek et al. 2012) with pools of guide RNAs (single-guide RNAs [sgRNAs]) to target and block the activities of each putative enhancer. Whereas blocking E3 had little effect, blocking either E1 or E2 significantly decreased *Klf4* expression (Fig. 1D). Additionally, biallelic deletion of either E1 or E2 by CRISPR/Cas9-mediated genome editing reduced *Klf4* expression by ∼70% and 85%, respectively, without significantly affecting the expression of an adjacent gene (*Rad23b*) or another pluripotency gene (*Pou5f1*) (Fig. 1E). Interestingly, although deletion of E3 alone had little effect, removing all three putative enhancers (ΔE123) reduced *Klf4* expression by 90% (Fig. 1E). It should be noted that the ΔE123 mESC clones did not display altered levels of other pluripotency factors or defects in colony formation, indicating that the observed reduction in *Klf4* expression is not due to cell differentiation (Supplemental Fig. S1B,C), underscoring the advantage of using mESCs to study *Klf4* regulation in naïve pluripotency.

To probe whether the putative enhancers E1 and E2 communicate with the *Klf4* promoter ∼55 kb away, we performed chromosome conformation capture (3C) assays to detect long-range interactions. For both E1 and E2, there was a significantly higher frequency of cross-linking to the *Klf4* promoter in mESCs compared with EpiSCs (Fig. 1F), indicating that E1 and E2 spatially juxtapose the promoter in mESCs. These 3C assays also revealed that E3 interacts with both the promoter and the other putative enhancers (Fig. 1F, Supplemental Fig. S1D), suggesting that E3 might have a role in *Klf4* transcription despite no apparent effects upon blocking or deleting this region. Nevertheless, for the remainder of this report, we primarily discuss our findings regarding E1 and E2, as these two enhancers clearly represent dominant elements that regulate *Klf4* expression in naïve state mESCs.

Fine-mapping functional TF-binding sites in E1 and E2

Enhancers must be recognized and bound by sequence-specific TFs to coordinately drive transcription of target genes. To identify cognate TFs bound to E1 and E2, we devised a sgRNA/dCas9-based tiling strategy to
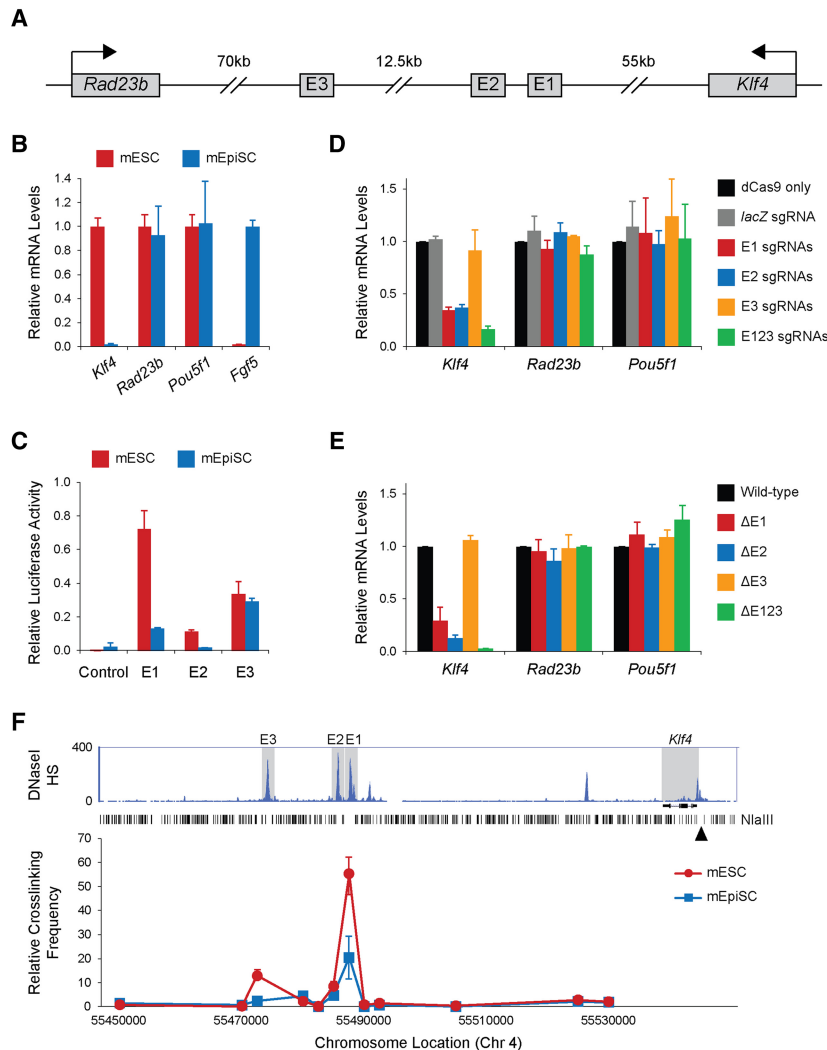
Figure 1. A long-range enhancer cluster controls *Klf4* gene expression in mESCs. (*A*) A schematic diagram of the putative regulatory elements E1, E2, and E3 relative to the *Klf4* gene and its neighbor gene, *Rad23b*, on chromosome 4. (*B*) *Klf4* is specifically expressed in mESCs and not in mouse EpiSCs (mEpiSCs) as evaluated by quantitative RT–PCR (qRT–PCR). Genes either with unchanged expression between mESCs and mEpiSCs (*Rad23b* and *Pou5f1*) or specifically expressed in mEpiSCs (*Fgf5*) were also evaluated as a control. (*C*) Dual-luciferase activity (firefly normalized to *Renilla*) for E1, E2, and E3 in mESCs and mEpiSCs. $n = 3$. (*D*) mESCs were transiently transfected with a pool of sgRNA–dCas9 constructs targeting each putative *cis*-regulatory region and analyzed by qRT–PCR. $n = 2$. (*E*) Deletion of E1 and E2 reduces *Klf4* expression. Individual putative regulatory elements or the entire region (ΔE123) were deleted using CRISPR/Cas9 in mESCs, and gene expression was analyzed by qRT–PCR from two independent clones. (*F*) All three putative regulatory elements interact with the *Klf4* promoter region. Chromosome conformation capture (3C)-qPCR was performed to assess long-range interactions using Nla III to produce restriction fragments. Relative cross-linking frequency or the relative level of ligation products is shown according to distance from the anchor point near the *Klf4* promoter region (arrowhead). DNase I hypersensitivity (HS) data are from ENCODE. All error bars represent SEM.

systematically block putative TF-binding sites (Fig. 2A,B). This unbiased "enhancer-bashing" experiment revealed a set of sgRNAs that significantly reduced *Klf4* expression but did not affect the adjacent gene *Rad23b* (Fig. 2C; Supplemental Fig. S2A). Using de novo TF motif prediction algorithms to scan the genomic sequences around these sgRNAs, we identified several candidate sequence-specific regulators of *Klf4* expression, including the pluripotency factors OCT4, SOX2, ESRRB, and STAT3 (see the Materials and Methods). To validate the direct binding of these candidate TFs to the *Klf4* enhancers, we performed genome-wide high-resolution ChIP-exo experiments for SOX2, STAT3, and ESRRB in mESCs (Rhee and Pugh 2011). Peaks identified by ChIP-exo were consistent with peaks from previous ChIP-seq (ChIP combined with high-throughput sequcing) protocols (Supplemental Fig. S2B,C). As expected, we observed significant enrichment of all three TFs at the *Klf4* enhancer cluster, particularly at E2 (Fig. 2D).

To test the functional contribution of individual TF-binding sites to *Klf4* expression in native chromatin, we next used CRISPR/Cas9 to biallelically delete OCT4/

SOX2-, STAT3-, or ESRRB-binding motifs within E2 validated by Sanger sequencing and found that removal of individual TF-binding sites significantly reduced *Klf4* gene expression (Fig. 2E). Strikingly, loss of the OCT4/SOX2 composite site in E2 profoundly reduced *Klf4* expression to a level comparable with deletion of the entire E2 enhancer (Fig. 1E). Consistent with these results, deletion or mutation of OCT4/SOX2, STAT3, and ESRRB motifs in E1 or E2 also reduced *Klf4* enhancer reporter activity in ESCs (Supplemental Fig. S2D). All together, these experiments strongly suggest that binding of OCT4, SOX2, ESRRB, and STAT3 at the native *Klf4* enhancer cluster drives the expression of *Klf4* in mESCs.

Enhancers must somehow respond to and integrate multiple extracellular signals that instruct cell fate maintenance and transitions. Among the key TFs that we identified as direct binders, STAT3 is activated by the LIF signaling pathway necessary for the derivation, maintenance, and acquisition of naïve pluripotency (Martello and Smith 2014). As expected, a *Stat3* knockout significantly decreased *Klf4* expression (which was largely rescued upon expression of full-length or Halo-tagged
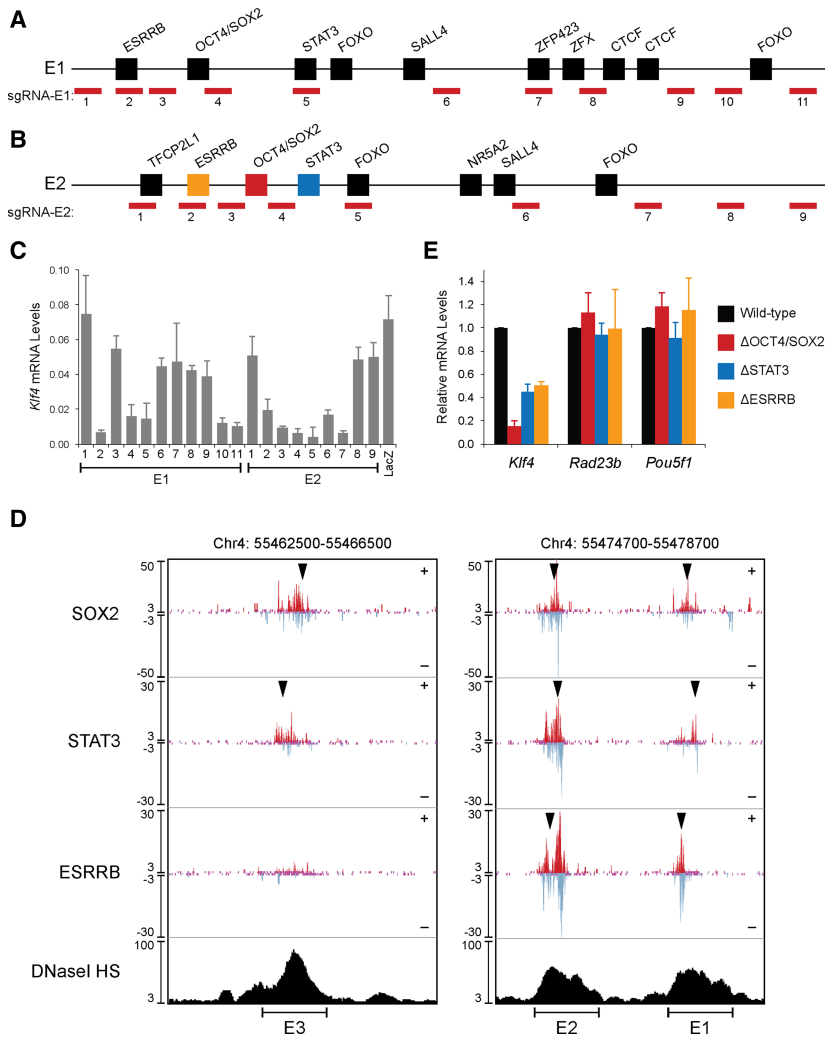
Figure 2. Fine-mapping of TFs functioning at *Klf4* active enhancers E1 and E2. (*A,B*) Schematic diagram of sgRNAs targeting to E1 (*A*) and E2 (*B*), with putative TF motifs shown. (*C*) mESC lines stably expressing individual sgRNAs and dCas9 were analyzed by qRT–PCR for *Klf4* expression. A sgRNA targeting *LacZ* was used as a negative control. $n = 2$. (*D*) SOX2, ESRRB, and STAT3 localize to E1, E2, and E3, as measured by ChIP-exo. DNase I HS data are from ENCODE. Arrowheads indicate predicted TF-binding sites. (*E*) qRT–PCR analysis of *Klf4*, *Rad23b*, and *Pou5f1* gene expression from mESC clones with biallelic deletion of the binding sites of OCT4/SOX2 ($n = 4$), ESRRB ($n = 2$), or STAT3 ($n = 2$) at E2 achieved by CRISPR/Cas9 genome editing. All error bars represent SEM.

STAT3) without affecting the expression of the neighboring gene *Rad23b* or other pluripotency genes (Fig. 3A; Supplemental Fig. S3A). LIF binding to the heterodimeric LIFR/gp130 receptor activates Janus kinase (JAK) and triggers STAT3 tyrosine phosphorylation and nuclear entry for target gene activation. Inhibition of the LIF–STAT3 pathway (JAKi), but not of the BMP–Smad pathway (BMPi), specifically reduced the activities of E1 and E2 (Fig. 3B), and LIF readdition after LIF withdrawal for 24 h triggered rapid recruitment of activated STAT3 and coactivator p300 to both E1 and E2 (Fig. 3C), which corresponded to *Klf4* reactivation (Supplemental Fig. S3B). Together, these results indicate that E1 and E2 integrate LIF–STAT3 signaling into the regulation of *Klf4* expression in naïve pluripotency.

Ordered TF assembly at E2

A critical question generating ongoing debate in gene regulation is whether TF assembly at *cis*-regulatory elements is random or hierarchically ordered (Swinstead et al. 2016). Here, we probed whether the binding activities of OCT4, SOX2, ESRRB, and STAT at E2 are interdependent through ChIP-qPCR (ChIP combined with quantitative PCR) analysis following endogenous deletion of individual TF-binding sites. Sanger sequencing confirmed that deleting the OCT4/SOX2 composite site did not affect the *cis*-regulatory motifs recognized by ESRRB or STAT3 and vice versa (Supplemental Table S4). Interestingly, deletion of the OCT4/SOX2 composite site significantly reduced the occupancy of other factors, including ESRRB, STAT3, and the cofactor p300, at E2 but not at the Pou5f1 enhancer region, arguing against a global or secondary effect (Fig. 4B; Supplemental Fig. S4A). It has been reported that some active enhancers can be transcribed by RNA polymerase II (Pol II) (Heinz et al. 2015). As expected, recruitment of Pol II at E2 was also significantly reduced upon deletion of the OCT4/SOX2 composite site, consistent with the idea that deleting a composite OCT4/SOX2 motif is sufficient to collapse E2 enhancer activity. On the other hand, deletion of the STAT3-binding site markedly reduced STAT3 and p300 occupancy but did not significantly affect OCT4/SOX2 or ESRRB binding at E2 (Fig. 4C). These results suggest that TF assembly at E2 is initiated by OCT4/SOX2 and followed by STAT3, which could play an important role in
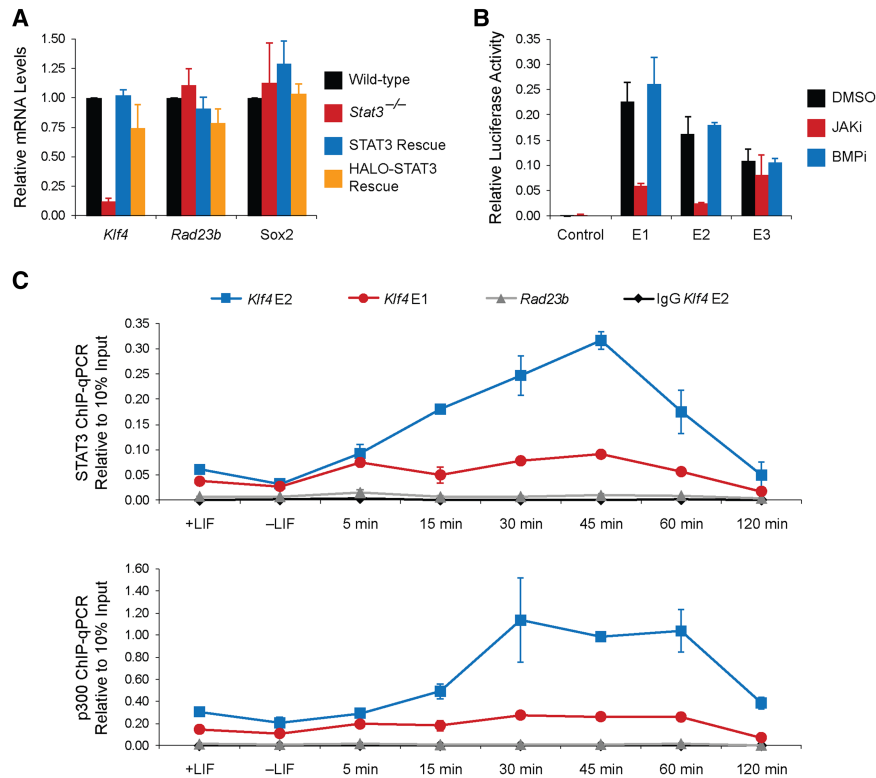
Figure 3. The *Klf4* enhancer cluster integrates naïve pluripotency signaling pathways. (*A*) Loss of *Stat3* (*Stat3*$^{-/-}$) reduces expression of *Klf4*, which can be rescued upon exogenous expression of STAT3. Gene expression was measured by qRT–PCR in wild type, *Stat3*$^{-/-}$, and *Stat3*$^{-/-}$ transfected with constructs to express either full-length STAT3 or HaloTag-STAT3. For comparison, STAT3 loss does not affect the *Klf4* neighbor gene *Rad23b* or the pluripotency gene *Sox2*. (*B*) Relative luciferase activity of different *Klf4* enhancer elements in response to inhibitors of the LIF–STAT3 (JAKi) or BMP–SMADs (BMPi) pathway. $n = 3$. (*C*) STAT3 and p300 are recruited rapidly to enhancer E2 upon LIF readdition following 24 h of LIF withdrawal. mESCs were collected at the indicated time points and analyzed by ChIP-qPCR (ChIP combined with quantitative PCR). $n = 2$. All error bars represent SEM.

recruiting chromatin-modifying activities like p300 (Hou et al. 2008). Likewise, ESRRB site deletion or even ESRRB knockout did not affect OCT4/SOX2 or STAT3/p300 binding at *Klf4* enhancers, suggesting that ESRRB binding is also hierarchically downstream from OCT4/SOX2 binding (Fig. 4D; Supplemental Fig. S4B).

OCT4 and SOX2 often heterodimerize on DNA to initiate gene expression. To further dissect the mechanisms by which OCT4 and SOX2 nucleate the TF ensemble at E2, we performed binding assays using purified proteins and E2-derived DNA probes. As expected, we observed efficient binding of OCT4 and SOX2 to the wild-type OCT4/SOX2 composite site. However, 1- or 2-base-pair (bp) insertions between the OCT4 and SOX2 sites markedly reduced this codependent binding without significant changes in individual protein binding (Supplemental Fig. S4C). Furthermore, these insertions dramatically reduced E2-driven enhancer reporter activity (Supplemental Fig. S4D). Taken together, these results support a model in which coordinate binding of OCT4 and SOX2 with precise spatial positioning nucleates the assembly of a higher-order TF ensemble that includes STAT3 and ESRRB at E2 and regulates enhancer activity.

SOX2 and OCT4 maintain accessible chromatin architecture

As shown above, deletion of a single OCT4/SOX2-binding site in the *Klf4* enhancer is sufficient to collapse enhanceosome assembly, thereby compromising expression of *Klf4*. To gain further insight into the function of OCT4

and SOX2, we analyzed the chromatin architecture in mESCs upon deletion of the OCT4/SOX2 site in E2 (ΔOCT4/SOX2) by genome-wide ATAC-seq (assay for transposase-accessible chromatin [ATAC] using sequencing), a method that faithfully recapitulates the chromatin accessibility pattern identified by traditional DNase sequencing (DNase-seq) approaches (Buenrostro et al. 2013). We found selective loss of chromatin accessibility at E2 in the ΔOCT4/SOX2 cells, with no significant changes in ATAC-seq signals at regulatory regions for other genes such as *Pou5f1* (Fig. 4E). Consistent with this finding, in the OCT4/SOX2 site-deleted ESCs, we observed increased local nucleosome density (histone H3 occupancy) and decreased acetylation of H3K27 at E2 (Supplemental Fig. S4E,F). These findings underscore the essential role of OCT4 and SOX2 in generating and maintaining an accessible chromatin neighborhood that is permissive for higher-order assembly of TF complexes necessary for *Klf4* expression in mESCs.

SOX2 interaction with ESRRB potentiates enhancer activation

In addition to signal integration, enhancers must also communicate with core promoter elements to activate gene transcription. ESRRB is an orphan nuclear receptor implicated in ESC self-renewal, somatic reprogramming, and instating naïve pluripotency (Feng et al. 2009; Festuccia et al. 2012; Martello et al. 2012). Previous immunoprecipitation-mass spectrometry experiments have revealed extensive interactions between ESRRB and the basal
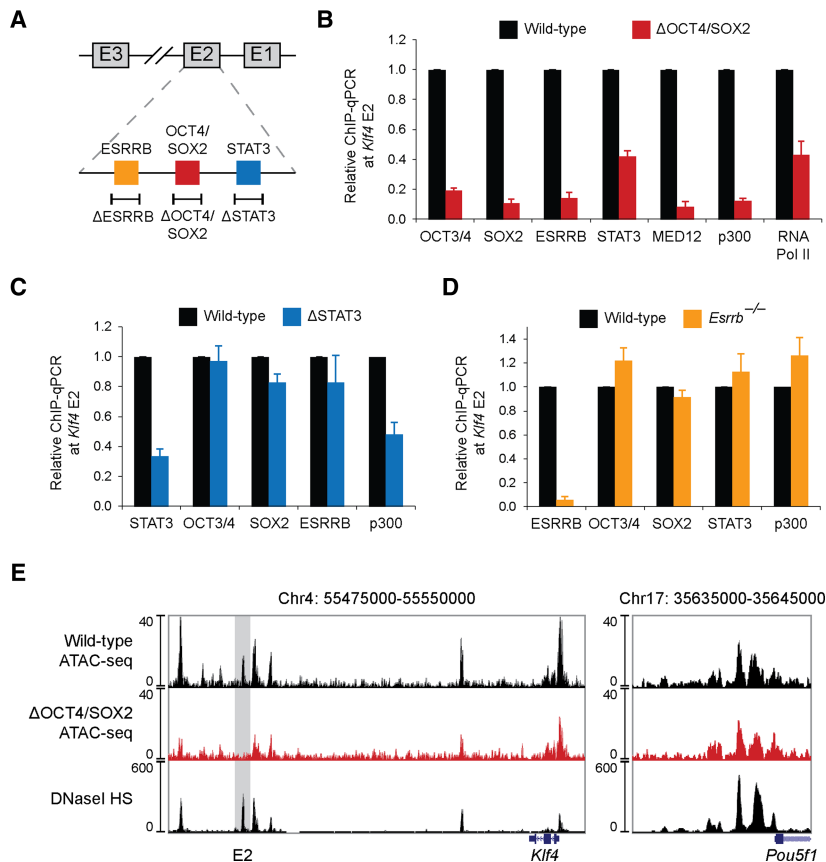
Figure 4. Lead factors SOX2 and OCT4 hierarchically regulate a higher-order protein ensemble at the *Klf4* enhancer. (*A*) Schematic illustration of binding sites for OCT4/SOX2, ESRRB, and STAT3 in the *Klf4* enhancer E2. Sanger sequencing validated the individual biallelic site deletion with no impact on nearby predicted sites. (*B*) Reduced occupancy of TFs, cofactors, and RNA Pol II upon deletion of the OCT4/SOX2-binding site in enhancer E2 (ΔOCT4/SOX2). Wild-type and ΔOCT4/SOX2 mESCs were analyzed by ChIP-qPCR. $n = 4$. (*C*) Deletion of the STAT3-binding site in enhancer E2 (ΔSTAT3) affects occupancy of p300 but not of OCT4 and SOX2. Wild-type and ΔSTAT3 mESCs were analyzed by ChIP-qPCR. $n = 4$. (*D*) Knockout of *Esrrb* does not affect occupancy of other TFs in enhancer E2. Wild-type and $Esrrb^{-/-}$ mESCs were analyzed by ChIP-qPCR. $n = 4$. (*E*) In ΔOCT4/SOX2 mESCs, chromatin accessibility decreases at enhancer E2 but not at other regions in the genome (e.g., the *Pou5f1* locus), as measured by ATAC-seq (assay for transposase-accessible chromatin [ATAC] using sequencing). The ENCODE DNase I HS data track at the *bottom* validates the ATAC-seq results. All error bars represent SEM.

transcription machinery, including Mediator, TFIID, and RNA Pol II (van den Berg et al. 2010), suggesting that ESRRB is likely important for mediating enhancer–promoter communication, although a direct role for ESRRB specifically at the *Klf4* enhancer had not been established. Our ChIP-exo analysis demonstrated that ESRRB binds predominantly to E1 and E2 but not appreciably to E3 (Fig. 2D). Moreover, deletion of ESRRB-binding sites in E2 significantly reduced *Klf4* expression to an extent similar to an ESRRB knockout (Fig. 2E; Supplemental Fig. S5A).

The above results suggest that ESRRB binding to the *Klf4* enhancer occurs hierarchically downstream from OCT4 and SOX2. To gain further mechanistic insights into the role of ESRRB in *Klf4* gene activation, we analyzed the spacing between pairs of TF motifs, as defined by our ChIP-exo data sets (see the Materials and Methods). We found a statistically preferred spacing between ESRRB and SOX2 motifs (P value = 6.31^{-7}, distance = 50, 53, 56 nucleotides [nt]) that is suggestive of a potential proximity-dependent interaction between ESRRB and SOX2 (Supplemental Fig. S5B). Coimmunoprecipitation (co-IP) experiments showed that ESRRB can interact directly with SOX2 (Fig. 5A). A reciprocal co-IP with ESRRB and SOX2 deletion mutants demonstrated that the DNA-binding domains (DBDs) of both ESRRB and SOX2 mediate this interaction (Fig. 5A; Supplemental Fig. S5C). Treatment with benzonase or RNase did not abrogate

this interaction (data not shown), suggesting that the interaction occurs via protein:protein contact.

To examine the functional significance of this SOX2–ESRRB interaction, we measured *Klf4* enhancer reporter activity in EpiSCs in which *Oct4* and *Sox2* expression levels are comparable with those in naïve ESCs, but *Esrrb* is largely inactive (Martello and Smith 2014). As expected, ectopic ESRRB expression alone was sufficient to activate the composite E1 + E2 enhancer reporter in EpiSCs (Fig. 5B). Notably, coexpression of ESRRB with SOX2, but not with OCT4, was able to synergistically increase the *Klf4* enhancer reporter activity. Addition of the SOX2 DBD but not the noninteracting transactivation domain (TAD) mimicked the effect of the full-length SOX2, emphasizing the importance of the SOX2–ESRRB direct interaction for cooperative gene activation (Fig. 5B). Furthermore, point mutations in SOX2 that compromise DNA-binding activity (SOX2M) or its OCT4-interacting surface (SOX2O) (Reményi et al. 2003) also abolished the synergistic effect with ESRRB (Supplemental Fig. S5D). It should be noted that this synergism between SOX2 and ESRRB is observed for enhancer E2 but not enhancer E1 and corresponds to a shorter spacing between SOX2 and ESRRB (Supplemental Fig. S5E), suggesting that different enhancer modules within the same spatial cluster can harbor distinct gene activation mechanisms (Fig. 5C). These findings indicate that the activity of ESRRB in the *Klf4* enhancer is
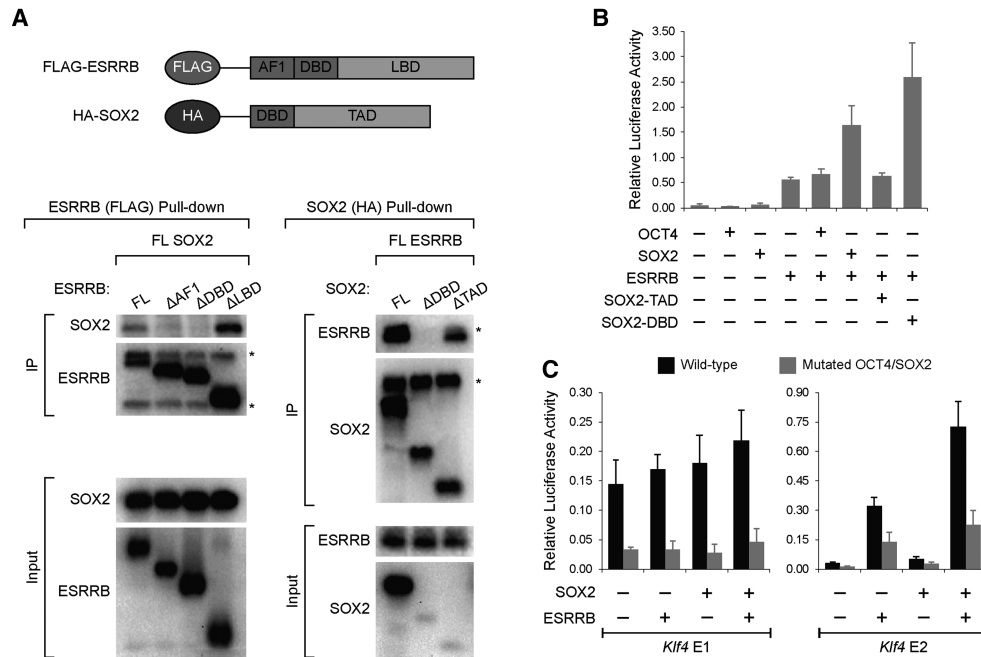
Figure 5. ESRRB and SOX2 synergistically activate the *Klf4* enhancer. (*A*) Co-IP was performed on nuclear lysates from HEK293T cells transfected with plasmids to express Flag-tagged ESRRB and HA-tagged SOX2. (AF1) Activation function 1 domain; (LBD) ligand-binding domain; (TAD) transactivation domain; (*) IgG heavy or light chains. (*B*) ESRRB and SOX2 synergistically activate expression through enhancers E1 and E2. mEpiSCs expressing exogenous OCT4, SOX2, and/or ESRRB were cotransfected with the composite E1 and E2 enhancer reporter and assayed for dual-luciferase activity. n = 3. (*C*) Luciferase assay for ESRRB- and SOX2-mediated gene activation on individual *Klf4* enhancer reporters driven by E1 or E2 with wild type or OCT4/SOX2 composite motif mutations. n = 3.

dependent on physical interaction with SOX2 as well as the proper formation of the SOX2–OCT4–DNA ternary complex.

Single-molecule dynamics of STAT3 and ESRRB in mESCs

Thus far, we used ensemble biochemistry, genomics, and genetics to probe the structure and function of TFs driving transcription of *Klf4*, a key regulator of naïve pluripotency. These conventional techniques, although instructive for mapping *cis*-control elements and identifying *trans*-acting factors, are unable to provide other critically important mechanistic insights such as the dynamics of TF binding, the mode of TF exploration of the nucleus in living cells, and the residence times required for activation at target enhancers. SPT analysis can provide such insights into TF dynamics, as we showed previously for OCT4 and SOX2 in ESCs (Chen et al. 2014). Having uncovered the important role of STAT3 and ESRRB in directly regulating *Klf4*, we carried out SPT analysis on HaloTag-STAT3 and HaloTag-ESRRB fusion proteins to more deeply dissect the mechanisms by which these TFs search for and decode regulatory elements in native chromatin under physiological conditions. First, we confirmed that these HaloTag fusion proteins are functional, can rescue the *Stat3* knockout phenotype, and can activate the *Klf4* enhancer (Fig. 3A; Supplemental Figs. S3A, S5F). To determine the fraction of STAT3 and ESRRB mol-

ecules bound to chromatin, we used 100-Hz image acquisitions, which allow tracking of both bound and diffusing single molecules. Model fitting of single-molecule displacements revealed that ∼35% of STAT3 and ∼44% of ESRRB molecules are likely chromatin-bound (Fig. 6A,B) and showed displacements comparable with "immobile" histone H2B molecules on chromatin (Hansen et al. 2017). For STAT3 or ESRRB, deletion of the DBD significantly reduced the "target-bound" fraction, as did specific mutation of Y705 in STAT3, which abolishes tyrosine phosphorylation and nuclear entry (Supplemental Fig. S6A,B).

By exploiting a motion-blurring strategy using a longer acquisition time (2 Hz) and low excitation power, we recorded the dwell times of TF binding to chromatin in live cells and reconstructed a residence time histogram. Consistent with previous results (Chen et al. 2014), a two-component exponential decay model fit well with our data, generating average lifetimes of 0.9 sec and ∼8 sec and percentages of bound molecules of ∼87% and ∼13% for the short-lived and long-lived populations of STAT3, respectively (Fig. 6C; Supplemental Fig. S6C). Similarly, the long-lived lifetime and fraction for ESRRB were found to be ∼10 sec and 23%, with the short-lived population being ∼0.9 sec and 77% (Fig. 6D; Supplemental Fig. S6D). Deletion or mutation of the DBD in STAT3 or ESRRB largely obliterated the long-lived populations and resulted in a single-component fit with average dwell times of 0.8–0.9 sec (Supplemental Fig. S6C,D),
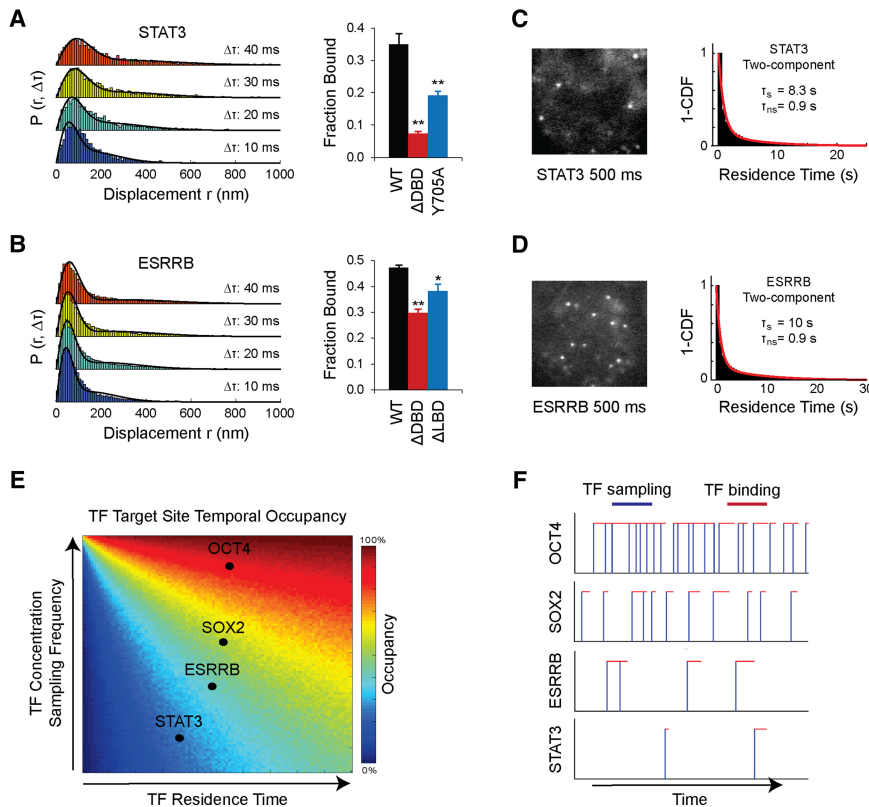
Xie et al.

Figure 6. Single-molecule dynamics analysis of key naïve pluripotency TFs in live ESCs. (*A,B*) Analysis of trajectories of single molecules of STAT3 (*A*) and ESRRB (*B*). Single-molecule displacement was analyzed from 100-Hz camera integration time over multiple time scales (*left* panel), and the bound fraction of molecules was determined by fitting a two-state model incorporating a bound and a diffusive state (*right* panel). n = 12–16. The model fit is overlaid on the raw displacements. (*C,D*) Analysis of bound STAT3 (*C*) and ESRRB (*D*) molecules. (*Left* panel) Immobile molecules were detected as diffraction-limited spots with a camera integration time of 500 msec. (*Right* panel) A two-component (long-lived and short-lived components) exponential decay model was fitted to a one-cumulative distribution function (1-CDF) of the fluorescence dwell time. (*E*) Simulation heat map to illustrate the temporal sampling frequency, target site temporal occupancy, and TF residence time for the naïve pluripotency enhanceosome TFs OCT4, SOX2, ESRRB, and STAT3. One-thousand continuous binding events for each TF were simulated, and the temporal occupancy was calculated based on the binding on–off durations and is presented in the Jet color map. (*F*) Illustration of simulation tracks for individual TFs with distinct sampling intervals and temporal occupancy (see the Supplemental Material for details).

confirming that the long-lived populations of STAT3 and ESRRB very likely represent stable on-target chromatin-binding events.

The majority of STAT3 (~87%) and ESRRB (~77%) molecules in the bound fraction has short-lived lifetimes (~0.9 sec), similar to those of SOX2 and OCT4 (Chen et al. 2014), and likely engages in nonspecific chromatin interactions (see the Supplemental Material). These nonspecific chromatin interactions might include a one-dimensional (1D) sliding process along short stretches of open chromatin regions to sample TF-binding sites in vivo (Chen et al. 2014). Moreover, we calculated that, on average, only ~5% of STAT3 and ~10% of ESRRB molecules in the ESC nucleus at any given time are binding at cognate sites. We showed previously that TF sampling frequency is highly sensitive to TF concentrations in the nucleus (Chen et al. 2014). Using a series of dye dilution experiments, we found that the nuclear concentrations of STAT3 (~0.15–0.3 μM) and ESRRB (~0.5–0.75 μM) were significantly lower than those of SOX2 (~1–1.3μM) and OCT4 (~2–3 μM) (Supplemental Fig. S7). Simulations based on SPT data suggested that different TFs displayed distinct cognate site-sampling frequencies and temporal occupancies (Fig. 6E,F). Thus, these differences in TF concentrations and target site-sampling dynamics might contribute to their cooperative interplay with lead factors such as SOX2 in directing gene regulatory programs. Taken together, these SPT analyses reveal that both STAT3 and ESRRB display highly dynamic three-dimensional

(3D) target search and chromatin-binding modes in live ESCs while participating in enhanceosome formation that helps drive the naïve pluripotency gene expression program.

Imaging dynamic and hierarchical enhanceosome assembly in live cells

Having established a critical role of OCT4/SOX2 as lead factors to nucleate the higher-order TF ensemble at the *Klf4* enhancer cluster, we wondered whether this ordered and hierarchical TF assembly mechanism might hold true globally for OCT4 and/or SOX2 to dictate the naïve pluripotency program. Using drug-inducible knockout ESC lines, we examined whether loss of SOX2 or OCT4 could influence ESRRB dynamics throughout the nucleus of individual cells, which in turn could potentially regulate enhanceosome communication with the basal transcriptional machinery. We first confirmed that doxycycline treatment for 24 h almost completely depleted SOX2 or OCT4 protein levels without altering the nuclear concentrations of other TFs (Supplemental Fig. S8A). Next, we performed SPT analysis of HaloTag-ESRRB stably expressed in SOX2-inducible knockout ESCs with or without 24 h of doxycycline treatment (Supplemental Fig. S8B). Whereas the loss of OCT4 had no appreciable effects on ESRRB dynamics (Supplemental Fig. S8D), depletion of SOX2 reduced the ESRRB residence time and the long-lived ESRRB fraction (Fig. 7A), leaving the short-lived
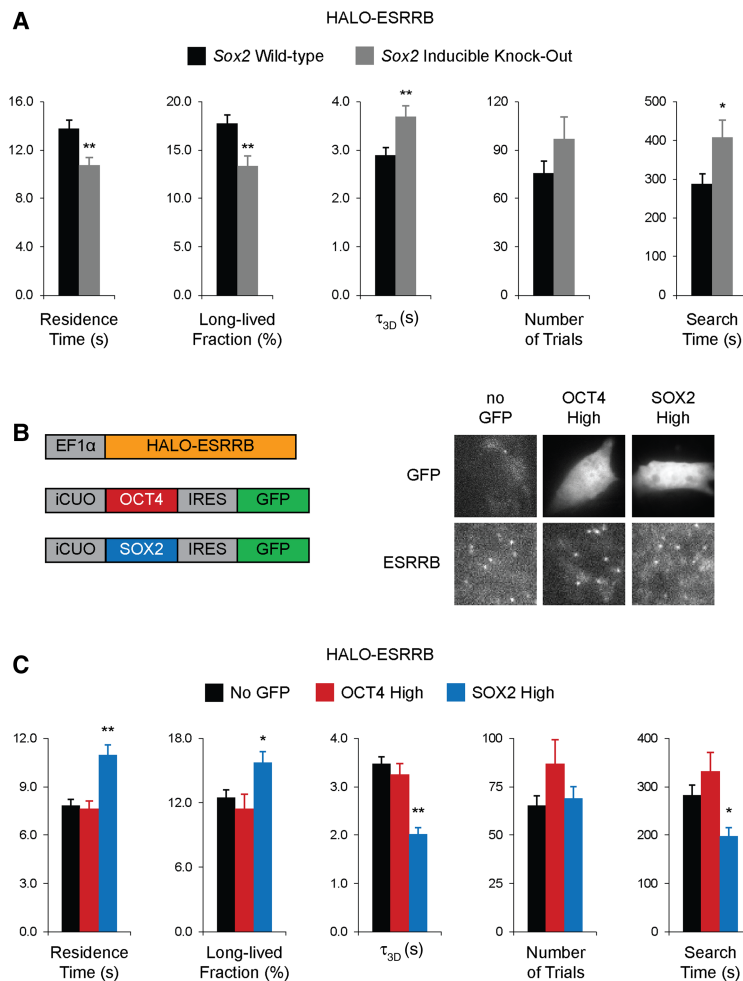
Figure 7. Dynamics of the hierarchical assembly of the enhanceosome in live cells. (*A*) Analysis of residence time, long-lived fraction, τ_{3D}, number of trials (N_{trials}), and search time for ESRRB after inducible knockout of SOX2. $n = 23$ cells. (*B*) Exogenous expression of OCT4, SOX2, and ESRRB in NIH/3T3 cells. (*Left* panel) Halo-Tag-ESRRB was stably expressed under the *EF1α* promoter, and OCT4 and SOX2 expression was driven by a cumate-inducible promoter (iCUO) and monitored by IRES-linked GFP expression. (*Right* panel) Representative images of two-dimensional SPT for HaloTag-ESRRB in GFP-negative and GFP-positive cells upon OCT4 or SOX2 induction. (*C*) Kinetic analysis of residence time, long-lived fraction, τ_{3D}, number of trials (N_{trials}), and search time for ESRRB upon *Sox2* or *Oct4* induction in 3T3 cells. $n = 16$ cells. (*) $P < 0.05$; (**) $P < 0.01$ by Student's *t*-test. Error bars represent SEM.

ESRRB fraction unaffected (Supplemental Fig. S8E), suggesting that SOX2 may affect the off rate (k_{off}) of ESRRB by stabilizing its specific binding to chromatin. Additionally, in the absence of SOX2, we also observed increases in the average time between two binding events (τ_{3D}) and in the search time for ESRRB, which suggest that SOX2 may also modulate the on rate (k_{on}) of ESRRB searching for cognate binding sites (Fig. 7A). This SOX2-dependent regulation of ESRRB kinetics is consistent with our observation that SOX2 physically interacts with ESRRB and synergistically activates the *Klf4* enhancer (Fig. 5). In contrast, the target search parameters and chromatin binding for SOX2 did not noticeably change in the absence of ESRRB (Supplemental Fig. S8C), consistent with the hierarchical nature of the codependent interaction of TFs.

To complement the analysis performed in the inducible knockout cell lines, we sought to reconstitute a similar TF assembly in NIH/3T3 cells, which are differentiated cells that do not express the pluripotency factors OCT4, SOX2, or ESRRB. We generated stable cells that constitutively express HaloTag-ESRRB together with cumate-inducible OCT4 or SOX2 linked to IRES-GFP. OCT4 or SOX2 expression could thus be monitored by GFP, while the dynamics of ESRRB were measured by SPT (Fig. 7B). Consistent with data from the inducible knockout cell

lines, OCT4 induction did not appreciably impact the diffusion dynamics or chromatin binding of ESRRB, whereas SOX2 induction led to an increase in ESRRB residence time and a decrease in the τ_{3D} (Fig. 7C).

These single-molecule studies suggest a general strategy for an ordered and hierarchical enhanceosome assembly that could operate in the maintenance of naïve pluripotency. Specifically, SOX2 is kinetically favored over OCT4 to initiate chromatin binding to its cognate site first, while ESRRB binding is hierarchically downstream and can be assisted by SOX2, suggesting a leading role for SOX2 in an ordered enhanceosome assembly process.

Discussion

Division of labor within the Klf4 enhancer cluster

Deciphering the elaborate "molecular dialog" between enhancers and promoters presents a formidable challenge in our quest to better understand the remarkable decoding of the mammalian genome with its evolving size and complexity. Our studies revealed a 3D architectural arrangement of long-distance *cis*-regulatory elements and provided new insights into dynamic TF interactions within the enhancer cluster that accommodate a functional

"division of labor" between key TFs working in combination to integrate key signal transduction pathways. Specifically, E1 and E2 are nonredundant cooperative active enhancers engaging the LIF/STAT3 signaling pathway. Our analysis also revealed a strong positive correlation between the spatial clustering of enhancer elements as determined by 3C experiments and high levels of gene activity in ESCs. Presumably, the enhancer clustering via looping of distal *cis*-elements may increase local concentrations of binding sites and facilitate the TF target search process (Liu et al. 2014). Strikingly, the functional clustering between the three enhancer elements appears to dismantle in primed EpiSCs where the *Klf4* gene becomes inactive, suggesting an intricate interplay instigated by TFs between chromatin organization and gene activity. Although deletion of E3 had no detectable effect on *Klf4* expression, 3C assays revealed significantly higher cross-linking frequency between E3 and the *Klf4* promoter, E1, and E2, suggesting that E3 might yet play some currently unknown role in *Klf4* transcription. The apparent functional redundancy of E3 with E1 and E2 is reminiscent of "shadow enhancers" in *Drosophila* that may fine-tune transcriptional precision and robustness during cell fate determination (Hong et al. 2008).

Using a combination of dCas9/sgRNA-directed tiling screen and site-specific deletions, we identified different classes of TF-binding sites within the *Klf4* enhancers. A more detailed functional analysis of the interdependence between TFs that recognize and bind these cognate elements revealed an ordered assembly process at the *Klf4* enhancer cluster. Specifically, deletion of the OCT4/SOX2 site in E2 reduces chromatin accessibility and prevents STAT3, ESRRB, and p300 from binding to E2. On the other hand, deletion of STAT3- or ESRRB-binding sites in E2 did not significantly affect SOX2 and OCT4 binding but still impacted *Klf4* expression. STAT3 mediates LIF signaling and helps recruit p300 to target genes, while ESRRB is thought to interact with the general transcription apparatus and MLL family chromatin modifiers (van den Berg et al. 2010). Thus, the emerging model from these results is that, upon SOX2 and OCT4 dynamic binding to E2, STAT3 and ESRRB are recruited to build up functional components that link signal transduction and chromatin modification to gene activation. Taken together, our data suggest that modular and ordered TF assembly empowers a functional division of labor carried out by multiple TFs assembled at the *Klf4* enhancer cluster to control precise programs of gene expression.

Previous pioneering studies of the human interferon-β (IFN-β) enhanceosome established a prototype of signal-dependent higher-order nucleoprotein complex assembly driving gene activation in eukaryotes (Maniatis et al. 1998). It involved a coordinated assembly of at least eight proteins with tightly arranged overlapping binding sites spanning a mere ~60-bp region directly upstream of the IFN-β TSS. In stark contrast, the *Klf4* enhanceosome described here is characterized by a complex 3D enhancer cluster architecture with highly dispersed, nonoverlapping multifactor-binding elements located some 55–70 kb downstream from the TSS. Unlike the IFN-β case in

which the enhanceosome proteins could form a stable complex in vitro, key TFs bound to *Klf4* enhancer elements exhibit dynamic binding and unbinding from cognate sites in chromatin on the time scale of seconds as measured in living cells. Thus, it is possible that these two well-characterized cases represent distinct classes of enhancers that may regulate transcription via different mechanisms. The IFN-β enhancer functions as a compact unit closely linked to the core promoter, while the *Klf4* cluster involves *cis*-elements located at widely separated sites distal to the core promoter. Although not many enhancer clusters have been functionally dissected at this detailed level, genomic analyses suggest that these highly dispersed long-distance enhancer clusters are more likely the rule rather than the exception in animal genomes (Levine et al. 2014). Recent studies suggest that only specific enhancer/promoter pairs can work synergistically to activate gene expression (Zabidi et al. 2014). It is tempting to speculate that at least part of the molecular basis for selective gene expression arises from distinct paths of communication between proximal and distal enhancers. It will also be interesting to examine whether some of the features uncovered from dissection of the *Klf4* enhanceosome (such as the highly dynamic interplay within long-range enhancer clusters) represent a pervasive mode of action controlling transcription during metazoan development.

Mechanisms of TF target search on enhancer DNA

Since the discovery of the first human sequence-specific TF almost three decades ago (Dynan and Tjian 1983), how a TF molecule navigates through complex 3D nuclear space to locate its targets has been a subject of intense theoretical discussions. However, direct experimental observations, especially in live mammalian cells, had largely been beyond reach because of technical limitations associated with live-cell single-molecule imaging. By taking advantage of recent developments in bright and cell-permeable dyes and live-cell imaging platforms, we recently demonstrated the possibility of tracking TF dynamics in single live ESCs (Chen et al. 2014). Single-molecule tracking charts a global cartography of molecular dynamics in living cells with high spatial (<50 nm) and temporal (approximate milliseconds) resolution, providing important kinetic insights into the behavior and process by which TFs find their cognate sites in vivo (3D diffusion rate, nonspecific binding, residence time, etc.). Here, we analyzed the single-molecule dynamics of STAT3 and ESRRB, two key naïve pluripotency TFs in mESCs, and found that both TFs undergo a 3D diffusion-dominant process to find their targets via multiple rounds of "trial and error" nonspecific collisions interspersed by relatively long 3D diffusion events. Previous in vitro single-molecule experiments suggested that these nonspecific events likely include 1D sliding/hopping to efficiently search for target sites within an "accessible" chromatin domain (Chen et al. 2014). We surmise that a 1D searching mode might be especially important for finding the nonoverlapping highly dispersed TF-binding sites within scattered enhancer clusters across the mammalian genome.

Single-molecule imaging combined with loss-of-function and reconstitution experiments revealed that SOX2 but not OCT4 regulates the target search and chromatin-binding dynamics of ESRRB. The SOX2-dependent modulation of ESRRB dynamics is also consistent with our finding of a direct physical interaction between SOX2 and ESRRB and the ability of SOX2 to synergize with ESRRB to trigger *Klf4* enhancer reporter activation. These findings also concur with our previous studies in which SOX2 was observed to be kinetically favored to bind chromatin first and then assist OCT4 loading, which in turn stabilizes SOX2 (Chen et al. 2014). Thus, SOX2 likely behaves as a recruiting "antenna" via protein–protein interactions with various partner TFs. It had been suggested previously that proteins searching for targets at near-optimal efficiency have a 1D sliding length of ~50–100 bp (Halford 2009). Genome-wide high-resolution ChIP-exo analysis found an enrichment of ESRRB-binding motifs roughly 50 bp from SOX2 motifs. Thus, we envision that SOX2 may facilitate the ESRRB search process by positioning ESRRB adjacent to SOX2-binding sites and, accordingly, increasing its chances of colliding with a cognate ESRRB target. This facilitated target search mechanism mediated by protein–protein tethering provides a kinetically optimized pathway for ESRRB to load into the enhanceosome that has already been seeded by SOX2/OCT4.

Recently, a combination of OCT4, SOX2, and ESRRB was reported to be sufficient to reprogram somatic cells to induced PSCs cultured in LIF (Feng et al. 2009). ESRRB was also shown recently to be important for selecting and activating pluripotency gene enhancers during somatic reprogramming by cooperating with a SOX2-containing Yamanaka TF cocktail (Chronis et al. 2017). Thus, the increased k_{on} and decreased k_{off} of ESRRB binding to targets facilitated by SOX2 could contribute to both the maintenance of and the induction toward the naïve pluripotent state.

The role of dynamic TF transactions in establishing naïve pluripotency

Our kinetic analysis uncovered a highly dynamic enhanceosome assembly process in live PSCs. Combined with measurements of nuclear TF concentrations and ChIP-exo analysis in mESCs, we were able to estimate the average temporal-sampling intervals of individual TFs at specific binding sites. Assuming that OCT4, SOX2, STAT3, and ESRRB sample similar numbers of nonspecific sites in the nucleus, we found, surprisingly, that OCT4 and SOX2 displayed shorter target site sampling intervals (~15 sec for OCT4 and ~30 sec for SOX2) than ESRRB (~64 sec) and STAT3 (~159 sec) (see the Supplemental Material). Combined with the measured specific residence times for STAT3 (~8 sec) and ESRRB (~10 sec), we estimated that the temporal occupancy for STAT3/ESRRB at their cognate binding sites is ~5%–16%, which is significantly lower than OCT4/SOX2 (40%–80%). The much faster sampling frequency and longer temporal occupancy of OCT4/SOX2 at naïve pluripotency enhancers may underlie their kinetic advantage and intrinsic capability to access and maintain a chromatin structure that helps nucleate the assembly of other critical TFs, such as STAT3 and ESRRB. Intriguingly, STAT3 and ESRRB displayed much slower target site sampling and shorter occupancy. Such inefficient target site sampling might in part explain the observed transcriptional bursting with long latency times (Coulon et al. 2013). In sum, these numerical simulations and modeling suggest a rather dynamic enhanceosome assembly that is highly sensitive to fluctuations in TF concentration and are likely quite vulnerable to TF disengagement and enhanceosome collapse. This model is consistent with the observation that activation of the FGF/MAPK and GSK3β pathways can readily counteract the LIF–STAT3 axis and cause mosaic ESRRB down-regulation, thereby facilitating exit from the naïve pluripotency program (Kalkan et al. 2017). This unexpectedly dynamic and transient nature of enhanceosome assembly in vivo underscores the developmental plasticity of the naïve pluripotent state poised to exit and transition toward primed pluripotency followed by lineage specification.

Materials and methods

Cell culture

Mouse JM8.N4 ESCs were cultured on 0.1% gelatin-coated plates without feeders. The ESC medium was prepared by supplementing knockout DMEM with 15% FBS (HyClone), 1 mM GlutaMAX, 0.1 mM nonessential amino acids, 0.1 mM 2-mercaptoethanol, and 1000 U of LIF (Millipore). STAT3 or ESRRB knockout ESCs were maintained in ESC medium supplemented with 2i (1 µM MEKi PD0325901, 3 µM GSK3i CHIR99021). Mouse EpiSCs (OEC2 cell line; a kind gift from Dr. Austin Smith) were cultured in N2B27 medium supplemented with 20 ng/mL Activin A and 12 ng/mL Fgf2 as described in Guo et al. (2009). 293T or NIH/3T3 cells were grown in DMEM high glucose with GlutaMAX supplemented with 10% FBS. Small molecule inhibitors are used as follows: 20 µM TGFβRI inhibitor SB 431542 (Sigma) and 10 µM JAK inhibitor InSolution JAK inhibitor I (EMD).

ChIP-qPCR

ChIP experiments were performed as described previously (Li et al. 2016) with minor modifications. Briefly, cells were cross-linked for 10 min at room temperature with 1% paraformaldehyde (methanol-free) in PBS and quenched with 0.125 M glycine. Cells were scraped and resuspended in cold cell lysis buffer (5 mM PIPES at pH 8.0, 85 mM KCl, 0.5% NP-40) and incubated for 10 min on ice. After centrifuging, nuclear pellets were resuspended in at least 6 vol of sonication buffer (50 mM Tris-HCl at pH 8.1, 10 mM EDTA, 0.1% SDS), incubated for 10 min on ice, and sonicated using the Covaris S220 sonicator to obtain an average fragment length of ~500 bp examined by electrophoresis. Sonicated chromatin was diluted in RIPA buffer, aliquoted, and incubated with Protein G Dynabeads that were prebound with individual antibodies for at least 1 h at room temperature. Immunoprecipitation was performed overnight at 4°C with 2–4 µg of antibodies. Ten percent of the chromatin was saved as input. Immunoprecipitated DNA was washed extensively with RIPA buffer, reverse-cross-linked, and digested with proteinase K and RNase A followed by purification with the Qiagen QIAquick PCR

purification kit. Eluted ChIP product was analyzed by qPCR using the SYBR Select master mix for CFX. The antibodies and primers used for ChIP assay are listed in Supplemental Tables S1 and S2.

ChIP-exo library preparation and sequencing analysis

ChIP-exo libraries for ESRRB and STAT3 were prepared as described previously (Rhee and Pugh 2011; Li et al. 2016) by adapting the SoLid sequencer adaptors/primers compatible with the Illumina sequencing on a HiSeq 2500 in 50-bp single-end format. After trimming the 3′-most 14-bp error-prone regions, we aligned ChIP-exo reads to the mouse genome (University of California at Santa Cruz build mm10) using Bowtie (version 0.12.7; options -S -t -m 1), accepting only uniquely aligning reads. The resulting SAM files were converted to BAM format, sorted (SAMtools version 0.1.19), and converted to bigWig files scaled to 10 million total mapped reads. We called peaks in all samples using several methods: MACS version 2 (options call peak -t $BAM_FN –f BAM -g mm -n $NAME -B -q 0.01 --outdir $MACS2_OUTDIR) or GEM (options –t $NUMCPU –g $GENOMESIZE_FN –out $GEM_OUTBASEDIR/$GEM_OUTDIR --f SAM --genome $GENOME FASTA_DIR --k_min 6 --k_max 15 --expt $BAM_FN). We used 1% false discovery rate cutoffs for peak calling. We converted the GEM output to 50-nt peaks centered around the GEM peak summit (custom Perl script). For GEM analysis of the STAT3 ChIP-exo data, we "seeded" the joint peak/motif calling by specifying seed motifs of TTCCTGGAA. Raw sequencing data were deposited to NCBI Gene Expression Omnibus (GEO) with accession number GSE97304.

To enable side-by-side comparison, STAT3 and ESRRB ChIP-exo were compared with STAT3 (GSM288353) and ESRRB (GSM288355) ChIP-seq data sets. SOX2 and SP1 ChIP-exo libraries were reported previously (Chen et al. 2014; Li et al. 2016).

Co-IP

Full-length *Esrrb/Sox2* were PCR-amplified from the mESC cDNA and cloned into a PiggyBac construct backbone with the N-terminal HA or 3XFlag tag. The domain deletions for Esrrb or Sox2 were further generated by PCR or site-directed mutagenesis and verified by Sanger sequencing. HEK293T cells grown on 10-cm tissue culture plates were transfected with 5 μg of each construct by Lipofectamine 2000 following the manufacturer's instructions. Forty-eight hours after transfection, cells were scraped off the plate with ice-cold PBS and centrifuged, and the cell pellet was resuspended in ice-cold cell lysis buffer (5 mM PIPES at pH 8.0, 85 mM KCl, 0.5% NP-40, protease inhibitors) and incubated for 10 min on ice. Nuclei were pelleted by centrifuge at 4°C and resuspended in 500 μL of low-salt cell lysis buffer. After preclearing with protein G sepharose beads for 1 h at 4°C, ~1 mg of proteins was diluted in 1 mL of co-IP buffer (0.2 M NaCl, 25 mM HEPES, 1 mM MgCl$_2$, 0.2 mM EDTA, 0.5% NP-40, protease inhibitors) and incubated with 4 μg of IgGs or specific antibodies overnight in a cold room. Fifty microliters of precleared lysate was kept overnight at 4°C as input. Fifty microliters of Protein G sepharose beads in co-IP buffer supplemented with 0.5% BSA was added to the samples and incubated for 2 h in a cold room. After extensive washes in co-IP buffer, proteins were eluted from the beads by boiling for 5 min in 1× SDS loading buffer and analyzed by SDS-PAGE and Western blot with the indicated antibodies (Supplemental Table S1). Western blot was exposed to Western Lightning Plus-ECL (PerkinElmer) and imaged in a ChemiDoc MP (Bio-Rad) detection system.

3C

3C experiments were performed as described previously (Hagège et al. 2007). ESCs or EpiSCs in single-cell suspension were cross-linked with 1% paraformaldehyde for 10 min at room temperature. Reactions were quenched by 250 mM glycine. Chromatin from 10^7 cells was digested with Nla III twice in CutSmart buffer overnight at 37°C. Samples with digestion efficiency >90% as assessed by qPCR were used for the following treatments. After inactivation of the Nla III for 30min at 65°C, chromatin fragments were ligated with T4 DNA ligase (New England Biolabs) overnight at 16°C followed by reverse-cross-linking overnight at 65°C in the presence of 300 μg/mL Proteinase K and 0.5 M NaCl. After RNase A treatment, samples were purified with UltraPure phenol:chloroform:isoamyl alcohol three times, and DNA was harvested by ethanol precipitation. One-hundred nanograms of DNA was used for each qPCR reaction in the triplicate. BAC control DNA template (RP23 111D23) was prepared similarly, and 100 ng of ligated BAC DNA was used for each qPCR reaction as the BAC control. 3C data were first corrected for bias of PCR amplification by using the BAC control template. To compare between samples, data from ESCs and EpiSCs were normalized to each other using the interaction frequencies between fragments in control regions (XRCC RP23-148C24 or Pdhb RP23-366F9). 3C primer information is listed in Supplemental Table S2.

ATAC-seq and analysis

ATAC-seq was performed according to Buenrostro et al. (2013) using the Nextera DNA library preparation kit (Illumina). Briefly, we performed ATAC-seq on 50,000 cells from wild-type and Klf4_E2_OCT4/SOX2 site deletion ESCs (delOS), which were trypsinized, washed with PBS on ice, and resuspended in the reaction mix (25 μL of 2× TD buffer, 2.5 μL of transposase, 22.5 μL of nuclease-free H$_2$O) for 30 min at 37°C. Following transposition, genomic DNA was purified and amplified by 14 cycles of PCR. Both ATAC-seq samples were sequenced in one lane of an Illumina HiSeq 2500 (Janelia Functional Genomics Core) with 50-bp paired-end sequencing.

To analyze ATAC-seq libraries, paired-end reads were first adapter-removed by Cutadapt and mapped to the mm9 genome build using Bowtie2 with the following parameters: --no-discordant --no-mixed --phred33 -X2000. Reads mapped to mitochondria and PCR duplicates were removed by SAMtools. We obtained 159,467,372 and 191,574,252 QC-passed reads for wild-type and delOS ESCs, of which 77.31% and 79.63%, respectively, corresponded to paired-end mapped reads. To compare the two ATAC-seq libraries, sequencing reads were normalized to 1× sequence depth, defined by total number of mapped reads × fragment length/effective genome size (2,150,570,000). Raw sequencing data were deposited to NCBI GEO with accession number GSE97304.

Single-molecule imaging

Single-molecule imaging experiments were carried out primarily as described in Chen et al. (2014). Cells were seeded on a 25-mm #1.5 coverglass precleaned with KOH and ethanol and coated with Matrigel according to the manufacturer's instructions. All live-cell imaging experiments were conducted using respective culture medium without Pheno-red (FluoroBrite DMEM, ThermoFisher). After testing multiple concentrations, 1 nM JF549 HaloTag ligand was added to cells for 10 min, and then cells were washed three times with imaging medium. The coverglasses were then transferred to a metal holder and mounted

onto a custom-built Nikon Eclipse Ti microscope with a 100× oil objective lens with NA 1.49, perfect focusing system, EMCCD camera (iXon3, Andor), and a humidified incubation chamber maintained at 37°C with 5% CO_2. To excite the JF549 HaloTag ligand, we used a 561-nm laser (MPB Lasertech) adjusted to the illumination intensity of ~50 W cm^{-2} and a 500-msec acquisition time to track stable binding molecules and ~0.5 kW cm^{-2} and 10 msec for fast tracking. The excitation laser was controlled by an AOTF (acousto-optic tunable filter) and reflected into the objective by a multiband dichroic (405/488/561/633 BrightLine quad-band bandpass filter, Semrock). The emission light was filtered by a single-band filter centered at 593 nm (FF01 593/40, Semrock) placed in front of the camera. Nikon NIS-Elements software was used to control the microscope, laser, and camera. Imaging was performed after a minimum of 2–10 sec of prebleaching. We tracked at least 5000 frames for 10-msec acquisition and 500 frames for 500-msec acquisition for eight to 10 cells per condition for at least three independent biological replicates.

Acknowledgments

References

Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. 2013. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10:** 1213–1218.

Chen J, Zhang Z, Li L, Chen BC, Revyakin A, Hajj B, Legant W, Dahan M, Lionnet T, Betzig E, et al. 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156:** 1274–1285.

Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K. 2017. Cooperative binding of transcription factors orchestrates reprogramming. *Cell* **168:** 442–459.

Coulon A, Chow CC, Singer RH, Larson DR. 2013. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nat Rev Genet* **14:** 572–584.

Dynan WS, Tjian R. 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35:** 79–87.

Feng B, Jiang J, Kraus P, Ng J-H, Heng J-CD, Chan Y-S, Yaw L-P, Zhang W, Loh Y-H, Han J, et al. 2009. Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nat Cell Biol* **11:** 197–203.

Festuccia N, Osorno R, Halbritter F, Karwacki-Neisius V, Navarro P, Colby D, Wong F, Yates A, Tomlinson SR, Chambers I. 2012. Esrrb is a direct Nanog target gene that can substitute for Nanog function in pluripotent cells. *Cell Stem Cell* **11:** 477–490.

Guo G, Yang J, Nichols J, Hall JS, Eyres I, Mansfield W, Smith A. 2009. Klf4 reverts developmentally programmed restriction of ground state pluripotency. *Development* **136:** 1063–1069.

Hagège H, Klous P, Braem C, Splinter E, Dekker J, Cathala G, de Laat W, Forné T. 2007. Quantitative analysis of chromosome conformation capture assays (3C-qPCR). *Nat Protoc* **2:** 1722–1733.

Halford SE. 2009. An end to 40 years of mistakes in DNA–protein association kinetics? *Biochem Soc Trans* **37:** 343–348.

Hansen AS, Pustova I, Cattoglio C, Tjian R, Darzacq X. 2017. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* **6:** e25776.

Heinz S, Romanoski CE, Benner C, Glass CK. 2015. The selection and function of cell type-specific enhancers. *Nat Rev Mol Cell Biol* **16:** 144–54.

Hong J-W, Hendrix DA, Levine MS. 2008. Shadow enhancers as a source of evolutionary novelty. *Science* **321:** 1314.

Hou T, Ray S, Lee C, Brasier AR. 2008. The STAT3 NH2-terminal domain stabilizes enhanceosome assembly by interacting with the p300 bromodomain. *J Biol Chem* **283:** 30725–30734.

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. 2012. A programmable dual-RNA-guided DNA endonuclease in adaptice bacterial immunity. *Science* **337:** 816–822.

Kalkan T, Olova N, Roode M, Mulas C, Lee HJ, Nett I, Marks H, Walker R, Stunnenberg H, Lilley KS, et al. 2017. Tracking the embryonic stem cell transition from ground state pluripotency. *Development* **5:** 733–736.

Levine M, Tjian R. 2003. Transcription regulation and animal diversity. *Nature* **424:** 147–151.

Levine M, Cattoglio C, Tjian R. 2014. Looping back to leap forward: transcription enters a new era. *Cell* **157:** 13–25.

Li L, Liu H, Dong P, Li D, Legant WR, Grimm JB, Lavis LD, Betzig E, Tjian R, Liu Z. 2016. Real-time imaging of Huntingtin aggregates diverting target search and gene transcription. *Elife* **5:** e17056.

Liu Z, Legant WR, Chen B-C, Li L, Grimm JB, Lavis LD, Betzig E, Tjian R. 2014. 3D imaging of Sox2 enhancer clusters in embryonic stem cells. *Elife* **3:** e04236.

Maniatis T, Falvo JV, Kim TH, Kim TK, Lin CH, Parekh BS, Wathelet MG. 1998. Structure and function of the interferon-β enhanceosome. *Cold Spring Harb Symp Quant Biol* **63:** 609–620.

Martello G, Smith A. 2014. The nature of embryonic stem cells. *Annu Rev Cell Dev Biol* **30:** 647–675.

Martello G, Sugimoto T, Diamanti E, Joshi A, Hannah R, Ohtsuka S, Göttgens B, Niwa H, Smith A. 2012. Esrrb is a pivotal target of the Gsk3/Tcf3 axis regulating embryonic stem cell self-renewal. *Cell Stem Cell* **11:** 491–504.

Mascetti VL, Pedersen RA. 2014. Naiveté of the human pluripotent stem cell. *Nat Biotechnol* **32:** 68–70.

Nichols J, Smith A. 2009. Naive and primed pluripotent states. *Cell Stem Cell* **4:** 487–492.

Reményi A, Lins K, Nissen LJ, Reinbold R, Schöler HR, Wilmanns M. 2003. Crystal structure of a POU/HMG/DNA ternary complex suggests differential assembly of Oct4 and Sox2 on two enhancers. *Genes Dev* **17:** 2048–2059.

Rhee HS, Pugh BF. 2011. Comprehensive genome-wide protein–DNA interactions detected at single-nucleotide resolution. *Cell* **147:** 1408–1419.

Rossant J, Tam PPL. 2017. New insights into early human development: lessons for stem cell derivation and differentiation. *Cell Stem Cell* **20:** 18–28.

Swinstead EE, Miranda TB, Paakinaho V, Baek S, Goldstein I, Hawkins M, Karpova TS, Ball D, Mazza D, Lavis LD, et al. 2016. Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell* **165:** 593–605.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:** 663–676.

Takashima Y, Guo G, Loos R, Nichols J, Ficz G, Krueger F, Oxley D, Santos F, Clarke J, Mansfield W, et al. 2014. Resetting transcription factor control circuitry toward ground-state pluripotency in human. *Cell* **158:** 1254–1269.

van den Berg DLC, Snoek T, Mullin NP, Yates A, Bezstarosti K, Demmers J, Chambers I, Poot RA. 2010. An Oct4-centered protein interaction network in embryonic stem cells. *Cell Stem Cell* **6:** 369–381.

Zabidi MA, Arnold CD, Schernhuber K, Pagani M, Rath M, Frank O, Stark A. 2014. Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518:** 556–559.

Supplemental Material	http://genesdev.cshlp.org/content/suppl/2017/10/05/31.17.1795.DC1
References	This article cites 33 articles, 8 of which can be accessed free at: http://genesdev.cshlp.org/content/31/17/1795.full.html#ref-list-1
Creative Commons License	This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genesdev.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**