bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

TITLE: Live-Cell Imaging Reveals Enhancer-dependent *Sox2* Transcription in the Absence of Enhancer Proximity

AUTHOR NAMES AND AFFILIATIONS: Jeffrey M. Alexander[1], Juan Guan[2], Bo Huang[2,3,4], Stavros Lomvardas[5,6], Orion D. Weiner[1,3*]

[1]Cardiovascular Research Institute
[2]Department of Pharmaceutical Chemistry
[3]Department of Biochemistry and Biophysics, University of California, San Francisco;
[4]Chan Zuckerberg Biohub
[5]Department of Biochemistry and Molecular Biophysics
[6]Mortimer B. Zuckerman Mind Brain and Behavior Institute, Columbia University
*Corresponding Author

ABSTRACT

Enhancers are important regulatory elements that can control gene activity across vast genetic distances. However, the underlying nature of this regulation remains obscured because it has been difficult to observe in living cells. Here, we visualize the spatial organization and transcriptional output of the key pluripotency regulator *Sox2* and its essential enhancer *Sox2* Control Region (SCR) in living embryonic stem cells (ESCs). We find that *Sox2* and SCR show no evidence of enhanced spatial proximity and that spatial dynamics of this pair is limited over tens of minutes. *Sox2* transcription occurs in short, intermittent bursts in ESCs and, intriguingly, we find this activity demonstrates no association with enhancer proximity, suggesting that direct enhancer-promoter contacts do not drive contemporaneous *Sox2* transcription. Our study establishes a framework for interrogation of enhancer function in living cells and supports an unexpected mechanism for enhancer control of *Sox2* expression that uncouples transcription from enhancer proximity.

INTRODUCTION

Chromosomes are packaged and organized non-randomly within the mammalian nucleus. Emerging evidence suggests that 3D genome topology plays a fundamental role in genome control, including the regulation of gene expression programs (Bickmore, 2013; Krijger and de Laat, 2016; Schwarzer and Spitz, 2014). Within the nucleus, each chromosome occupies discrete chromosomal territories (Cremer et al., 2006). These territories are further structured into distinct compartments that separate active and repressive chromatin (Lieberman-Aiden et al., 2009; Sexton et al., 2012). At finer scales, chromosomes are partitioned into largely-invariant, sub-megabase sized topologically-associated domains (TADs), which break up the linear genome into interactive neighborhoods (Dixon et al., 2012; Nora et al., 2012). Chromosomal contacts are disfavored across TAD boundaries. Thus, most cell-type specific contacts occur within TAD boundaries, and disruption of TAD architecture leads to dysregulation of gene expression (Dowen et al., 2014; Gröschel et al., 2014; Guo et al., 2015; Lupiáñez et al., 2015; Narendra et al., 2015; Nora et al., 2017).

Within this 3D framework, gene expression programs are established by non-coding regulatory enhancer elements. First discovered within a metazoan genome over three decades ago (Banerji et al., 1983), it is now predicted that greater than 300,000 enhancers are encoded in the human genome (The ENCODE Project Consortium, 2012; Zhu et al., 2013). Enhancers demonstrate unique epigenetic markings, enriched for H3K4me1 and H3K27ac (Creyghton et al., 2010; Heintzman et al., 2007; Rada-Iglesias et al., 2010), and are highly accessible, as demonstrated by elevated DNase sensitivity and transposition susceptibility (Boyle et al., 2008; Buenrostro et al., 2013; Thurman et al., 2012). These features facilitate transcription factor occupancy, enrichment of co-activators such as p300 and Mediator, and transcription of non-coding enhancer RNAs (eRNAs), all of which play important roles in modulation of target gene expression (Kim et al., 2015; Long et al., 2016). Importantly, enhancer activity is highly specific across cell types (Heintzman et al., 2009; The ENCODE Project Consortium, 2012; Zhu et al.,

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2013) and dynamic during cellular differentiation (Blum et al., 2012; Buecker et al., 2014; Huang et al., 2016; Wamstad et al., 2012), and this activity correlates with nearby gene expression. Thus, enhancers are fundamental to achieving gene expression programs that orchestrate embryonic development and drive disease pathogenesis. Understanding the mechanism by which enhancers influence target genes is crucial to decode gene regulation.

The textbook model proposes that enhancers influence target gene promoters through protein-protein complexes and physical interaction mediated by a DNA loop (Alberts et al., 2014). Experimental support for this model comes primarily from numerous chromosome conformation capture (3C)-based studies that have identified enriched contacts between enhancer and promoter elements (Jin et al., 2013; Li et al., 2012; Rao et al., 2014; Sanyal et al., 2012; Weintraub et al., 2017) and recent observations that driving contacts between an enhancer-promoter pair is sufficient to augment gene expression (Bartman et al., 2016; Deng et al., 2012; 2014; Morgan et al., 2017). However, other observations fit this model poorly. For example, *sonic hedgehog* (*Shh*) enhancers that drive expression in the brain move further, rather than closer, to the *Shh* gene when activated (Benabdallah et al., 2017). Furthermore, in *Drosophila*, coupled reporter genes regulated by a shared enhancer nevertheless show coordinated transcriptional bursting, suggesting either that an enhancer can contact multiple genes at once or that contact can be decoupled from transcription (Fukaya et al., 2016; Lim et al., 2018). Super enhancers -- clusters of enhancers that are highly enriched for coactivators like Mediator and BRD4 (Lovén et al., 2013; Whyte et al., 2013) -- have been proposed to activate transcription through nucleation of activator droplets rather than stepwise assembly of transcription complexes (Hnisz et al., 2017), providing a possible mechanism for enhancer action at a distance, and recent imaging has provided support for this idea (Cho et al., 2018; Sabari et al., 2018). Thus, how distal elements communicate with and regulate gene promoters in living cells remains an open question.

Live-cell imaging represents a powerful approach to dissect chromatin architecture and gene regulation in the context of single cells to address these questions (Chen et al., 2013; 2018; Germier et al., 2017; Gu et al., 2018; Lucas et al., 2014). However, interrogation of both enhancer-gene spatial organization and real-time transcriptional activity of the regulated gene has not yet been realized in living mammalian cells. Here, we investigate the dynamic 3D organization and transcriptional activity of the *Sox2* gene and its distal enhancer *Sox2* Control Region (SCR) in mouse embryonic stem cells (ESCs) using live-cell microscopy.
We find that the *Sox2* promoter and SCR demonstrate similar spatial characteristics to non-regulatory regions in ESCs, while differentiation of ESCs leads to significant compaction throughout the *Sox2* region. Time-lapse microscopy revealed that individual loci explore only a fraction of their potential spatial range during the ~25 minute imaging window, driving high cell-to-cell variability in *Sox2* locus conformation and *Sox2*/SCR encounters. Incorporation of an MS2 transcriptional reporter into the *Sox2* gene demonstrated that transcription occurs in intermittent bursts in ESCs but, surprisingly, showed no correlation with spatial proximity between the enhancer-promoter pair. Together, our findings establish the spatial and transcriptional characteristics of an essential pluripotency gene and suggest an unconventional mechanism for enhancer control of *Sox2* expression that uncouples transcription from enhancer proximity.

RESULTS

Engineering the Endogenous *Sox2* Locus to Visualize Locus Organization in Living Embryonic Stem Cells

To visualize discrete loci within the mammalian genome, we turned to the well-established genetic labeling method of incorporating repetitive arrays of exogenous operator sequences, an approach that has been extensively used to visualize chromosomal loci (Belmont and Straight, 1998; Lucas et al., 2014; Marshall et al., 1997; Masui et al., 2011; Michaelis et al., 1997; Robinett et al., 1996; Roukos et al., 2013). To independently visualize two regions of interest, we utilized the tetO/TetR system to visualize one chromosomal location. For the other chromosomal location, because of the reported issues using lacO/lacI in ESCs (Lucas et al.,

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2014; Masui et al., 2011), we developed a new tool based on the cuO/CymR pair. This is a repressor system from the bacteria *Pseudomonas putida* that is involved in cumate metabolism and has been previously used as a tool for inducible gene expression (Mullick et al., 2006). We opted to target these arrays to the mouse genome using a two-step genetic engineering strategy with bacteriophage integrases for two reasons (**Figure 1A)**. First, repetitive sequences can be unstable during vector construction, making it advantageous to use generic targeting vectors portable between genomic loci. Second, we worried the repetitive arrays might recombine during genomic targeting using homologous recombination. To target the tetO/TetR and cuO/CymR labels to specific loci within the mouse genome, we first placed attP landing sites for the PhiC31 (Raymond and Soriano, 2007; Thyagarajan et al., 2001) and Bxb1(Xu et al., 2013) integrase systems using CRISPR/Cas9 homology directed repair. We then integrated generic PhiC31 or Bxb1 targeting vectors bearing either the tetO or cuO array, respectively, at the corresponding landing sites through transient expression of the PhiC31 and Bxb1 integrases. This strategy was both modular in design and portable between genomic loci. To target two regions on the same chromosome, we used 129/Cast F1 hybrid ESCs, derived from crossing the 129 mouse strain to the divergent subspecies *Mus musculus castaneus*. This allowed us to limit editing to the 129 allele by using genetic polymorphisms between the two parental genomes to design allele-specific CRISPR guide RNAs.

We chose the murine *Sox2* locus as our genetic model. *Sox2* encodes a high-mobility group (HMG) DNA-binding transcription factor with important roles in embryonic development (Kamachi and Kondoh, 2013; Lefebvre et al., 2007; Sarkar and Hochedlinger, 2013), embryonic and adult neural progenitors (Pevny and Nicolis, 2010), and the progression of many forms of cancer (Weina and Utikal, 2014; Wuebben and Rizzino, 2017). *Sox2* also functions as an essential regulator of pluripotency, where it cooperates with other transcriptional regulators to maintain the pluripotency transcriptional program and keep embryonic stem cells in the undifferentiated state (Chen et al., 2008a; Young, 2011). *Sox2* resides in an isolated neighborhood on chromosome 3, as the sole protein-coding gene in a ~1.6 Mb region. Numerous regulatory elements that modulate *Sox2* expression have been identified in this neighborhood across amniotes (Okamoto et al., 2014; Tomioka et al., 2002; Uchikawa et al., 2003; Zappone et al., 2000). However, *Sox2* expression in mouse ESCs is controlled by a single, strong distal enhancer called the *Sox2* Control Region (Li et al., 2014; Zhou et al., 2014), which is robustly enriched with H3K27ac, DNase hypersensitivity, RNA Polymerase II (RNAP), and transcription factor occupancy (herein referred to as SCR, **Figure 1B**). Genetic ablation of SCR in ESCs leads to loss of *Sox2* expression in *cis*. Moreover, SCR maintains *Sox2* expression levels in the context of compound deletion of alternative *Sox2* enhancers, suggesting SCR is sufficient for *Sox2* regulation in ESCs (Zhou et al., 2014). Publicly available circularized chromosome conformation capture (4C) and HiC datasets reveal enriched contacts between SCR and the *Sox2* promoter region, suggesting that these enhancer-promoter interactions may play an important role in SCR function (**Figure 1B**).

We generated three distinct modified cell lines in 129/Cast F1 hybrid ESCs (**Figure 1B**, bottom) First, we labeled the *Sox2* promoter region and SCR by integrating the cuO array 8 kb centromeric to the *Sox2* TSS (*Sox2*-8C) and the tetO array approximately 5 kb telomeric to the SCR boundary (i.e. 117 kb telomeric to *Sox2* TSS, *Sox2*-117T). We refer to this pair as *Sox2*-SCR. Secondly, we created two control ESC lines: one with two arbitrary loci labeled with cuO and tetO (*Sox2*-43T$^{tetO/+}$; *Sox2*-164T$^{cuO/+}$ or Control-Control) and a second where we labeled SCR along with a non-specific telomeric locus (*Sox2*-117T$^{tetO/+}$; *Sox2*-242T$^{cuO/+}$ or SCR-Control). In both cases, the genetic distance between labels was similar to that of *Sox2*-SCR. Both control pairs show low contact propensity in chromosome conformation capture data (**Figure 1B)**. We verified the correct placement of the cuO and tetO labels for each locus using PCR with primers that span the unique recombination arms generated after plasmid integration (**Figure 1—figure supplement 1**, **Supplementary file 1,2**). We detected a similar *Sox2* expression ratio (129 / CastEiJ) using an allele-specific qPCR assay for modified cell lines compared to the parental ESCs, suggesting *Sox2* regulation is intact despite genetic alteration

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

of the locus (Analysis of Variance, p = 0.215, **Figure 1—figure supplement 2**).

Visualization of the *Sox2* Region in ESCs Reveals Minimal Evidence for *Sox2*/SCR Interactions

We were first interested in measuring the 3D distance between *Sox2* and the SCR enhancer in living ESCs. To this end, we stably coexpressed CymR-GFP and TetR-tdTomato (TetR-tdTom) fusion proteins in *Sox2*-SCR ESCs using ePiggyBac transposon-based gene delivery (Lacoste et al., 2009). This allowed for visualization of both the *Sox2* promoter (cuO) and SCR (tetO) within the nucleus using live-cell fluorescence confocal microscopy. We confirmed that coexpression of CymR-GFP and TetR-tdTom did not significantly alter *Sox2* expression from the modified 129 allele by qPCR (**Figure 1—figure supplement 2**). 3D time series of proliferating ESCs showed the majority of cells demonstrated a single, bright focus of

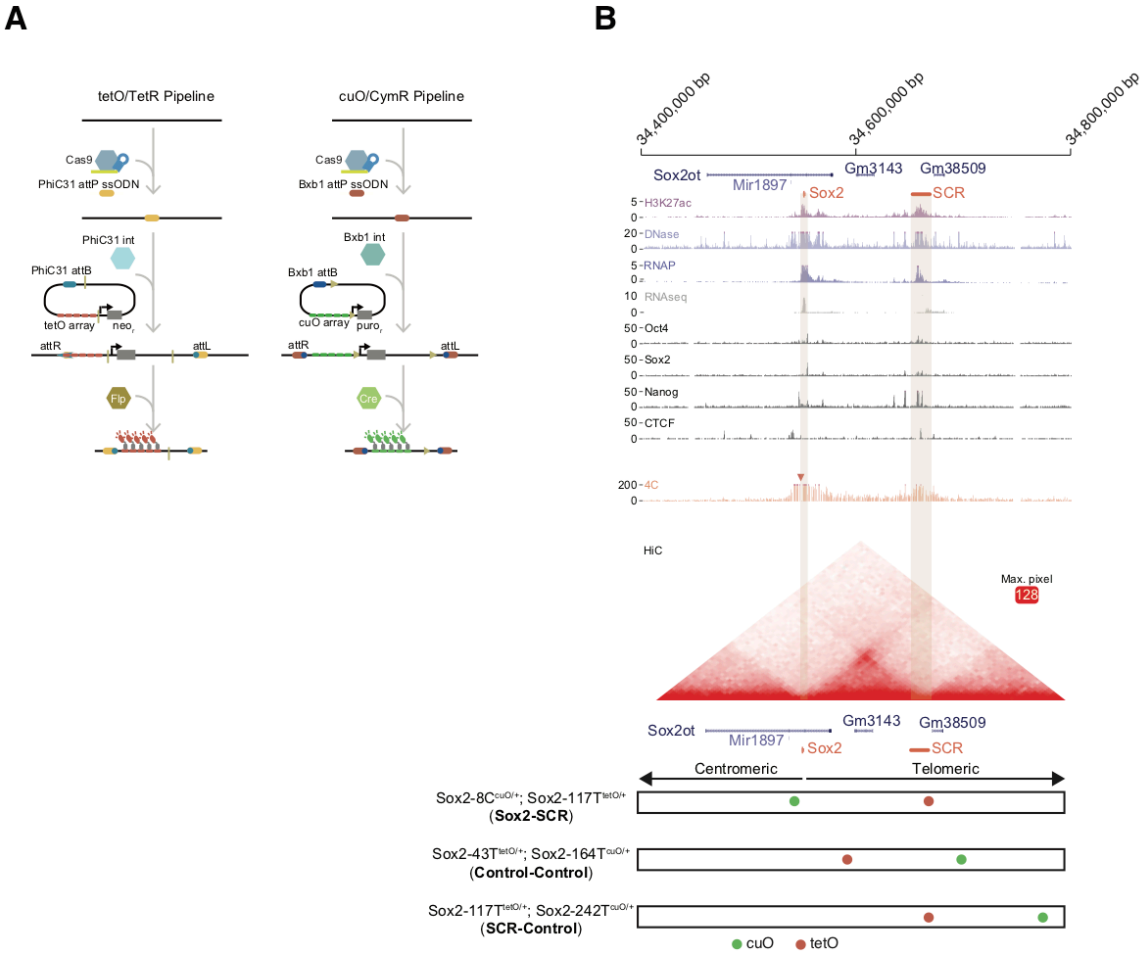

Figure 1. The *Sox2* Locus As a Model for Visualization of Enhancer-Promoter Regulation in Mouse Embryonic Stem Cells. A) To visualize chromosome loci in living cells, we have used tetO/TetR and cuO/CymR genetic labels. Our pipeline for insertion of these labels into the mouse genome is shown. First, CRISPR-Cas9 is used to place an attP intergrase landing site. Second, a targeting plasmid bearing the compatible attB sequence, the tetO or cuO array, and a selection cassette is introduced along the integrase (Int) to mediate site-specific integration. The selection cassette can then be subsequently removed by Cre/Flp recombinase. B) The *Sox2* locus in mouse ESCs. Genomic browser tracks of epigenomic and expression data demonstrate high levels of histone acetylation, RNA polymerase II, and transcription factor (OCT4, SOX2, NANOG, CTCF) occupancy at *Sox2* and the distal *Sox2* Control Region enhancer (tan boxes). Data from 4C and HiC experiments demonstrate chromosomal contacts at the *Sox2* locus. For 4C data, read density indicates contact frequency with a fixed position near the *Sox2* promoter (red triangle). Y-axis for browser tracks is reads per million. For HiC, all pairwise contact frequencies are mapped using a heatmap. The intensity of each pixel represents the normalized number of contacts detected between a pair of loci. The maximum intensity is indicated in red square. At bottom, locations of the cuO- and tetO-arrays for the three cell lines utilized for this study. Sox2-8C$^{cuO/+}$; Sox2-117T$^{tetO/+}$ (Sox2-SCR) ESCs were used to track Sox2/SCR location. Two control lines, Sox2-43T$^{tetO/+}$; Sox2-164T$^{cuO/+}$ (Control-Control) and Sox2-117T$^{tetO/+}$; Sox2-242T$^{cuO/+}$ (SCR-Control) were analyzed for comparison. H3K27ac, RNA polymerase II (RNAP), and RNAseq data from GSE47949 (Wamstad et al., 2012); DNase data from GSE51336 (Vierstra et al., 2014); SOX2, OCT4, NANOG, and CTCF data from GSE11431 (Chen et al., 2008b); 4C data from GSE72539 (de Wit et al., 2015); and HiChIP data from GSE96107 (Bonev et al., 2017).

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

CymR-GFP and TetR-tdTom in the ESC nucleus in close proximity. Many of these foci revealed the presence of two juxtaposed sister chromatids (**Video 1**). Because the overlapping signal from adjacent, identical arrays would degrade the resolution of our localization, we excluded these loci from our analysis and focused on cells demonstrating single, diffraction-limited spots for cuO and tetO, likely representing cells in the G1/early S phase of the cell cycle.

3C data demonstrate enriched contacts between *Sox2* and SCR (Beagan et al., 2017; Bonev et al., 2017; de Wit et al., 2015; Kieffer-Kwon et al., 2013; Mumbach et al., 2016; Phillips-Cremins et al., 2013; Zhou et al., 2014), supporting the possibility of a looped locus configuration with *Sox2* and SCR juxtaposed in 3D space. A mixture of looped and unlooped configurations across the population might be expected to produce a multimodal distance distribution with short and large distance peaks representing looped and unlooped states, respectively, as was recently observed for an enhancer system in *Drosophila* (Chen et al., 2018). To investigate the distribution of *Sox2*/SCR distances, we determined the 3D position of cuO and tetO for each locus, assembled 3D tracks, and calculated 3D separation distances between the labels across time (**Figure 2A, Supplementary file 3**). By localization of fluorescent beads at a comparable signal-to-noise ratio, we estimate our localization precision in the X, Y, and Z dimensions to be 14.7 $\pm$ 4.4 nm, 11.8 $\pm$ 3.4 nm, and 38.4 $\pm$ 12 nm, respectively, for cuO/CymR and 20.4 $\pm$ 7.6 nm, 19 $\pm$ 7.8 nm, 59.5 $\pm$ 21 nm for tetO/TetR (**Figure 2—figure supplement 1**). 84% and 62% of our assembled tracks span the full time series (> 75 frames) for cuO and tetO, respectively (**Figure 2—figure supplement 2**). Visualization of these data as a histogram revealed a unimodal distribution with positive skew (Hartigan's Dip Test for multimodality, p = 1). On average, *Sox2*/SCR labels are separated by a few hundred nanometers in the ESC nucleus (mean = 339 nm, **Figure 2B**). Infrequently, we observed the *Sox2* region to adopt an extended conformation, leading to considerable *Sox2*/SCR separation distance (2.1% of measurements > 750 nm, 0.35% of measurements > 1 um).

One possible interpretation of a unimodal distance distribution is that the *Sox2*/SCR pair exists predominantly in an interacting state. To investigate this possibility, we repeated our analysis with our two control locus pairs. We found that, while one control pair (Control-Control) did show increased separation distance as compared to Sox2/SCR, our other control set (SCR-Control), consisting of the SCR paired with a non-specific partner, showed a similar distribution to *Sox2*/SCR (**Figure 2C**). Indeed, no significant differences between *Sox2*-SCR and SCR – Control were found when comparing the mean distance per cell, while Control-Control demonstrated significantly increased distances (**Figure 2D**). Reinspection of chromosomal contact maps revealed evidence for a topological boundary, potentially established by the SCR element, separating the two labeled regions in the Control-Control configuration (**Figure 1A**), which could account for the elevated 3D distances measured for Control-Control, as has been observed for TAD boundaries (Dixon et al., 2012; Nora et al., 2012). These results demonstrate that SCR does not show greater proximity to the *Sox2* gene than to a non-specific control.

To further exclude the possibility that our measurements reflected a constitutive interaction state, we sought to estimate the distance profile for a static *Sox2*/SCR interaction. To this end, we used CRISPR/Cas9 to delete a ~111 kb fragment between the cuO and tetO labels in the *Sox2*-SCR configuration, leaving a 14 kb tether between the labels (**Figure1—figure supplement 1**). This is similar in length to the effective tether (~17 kb) between labels expected during a direct interaction between the *Sox2* TSS and the center of the SCR. Visualization of

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

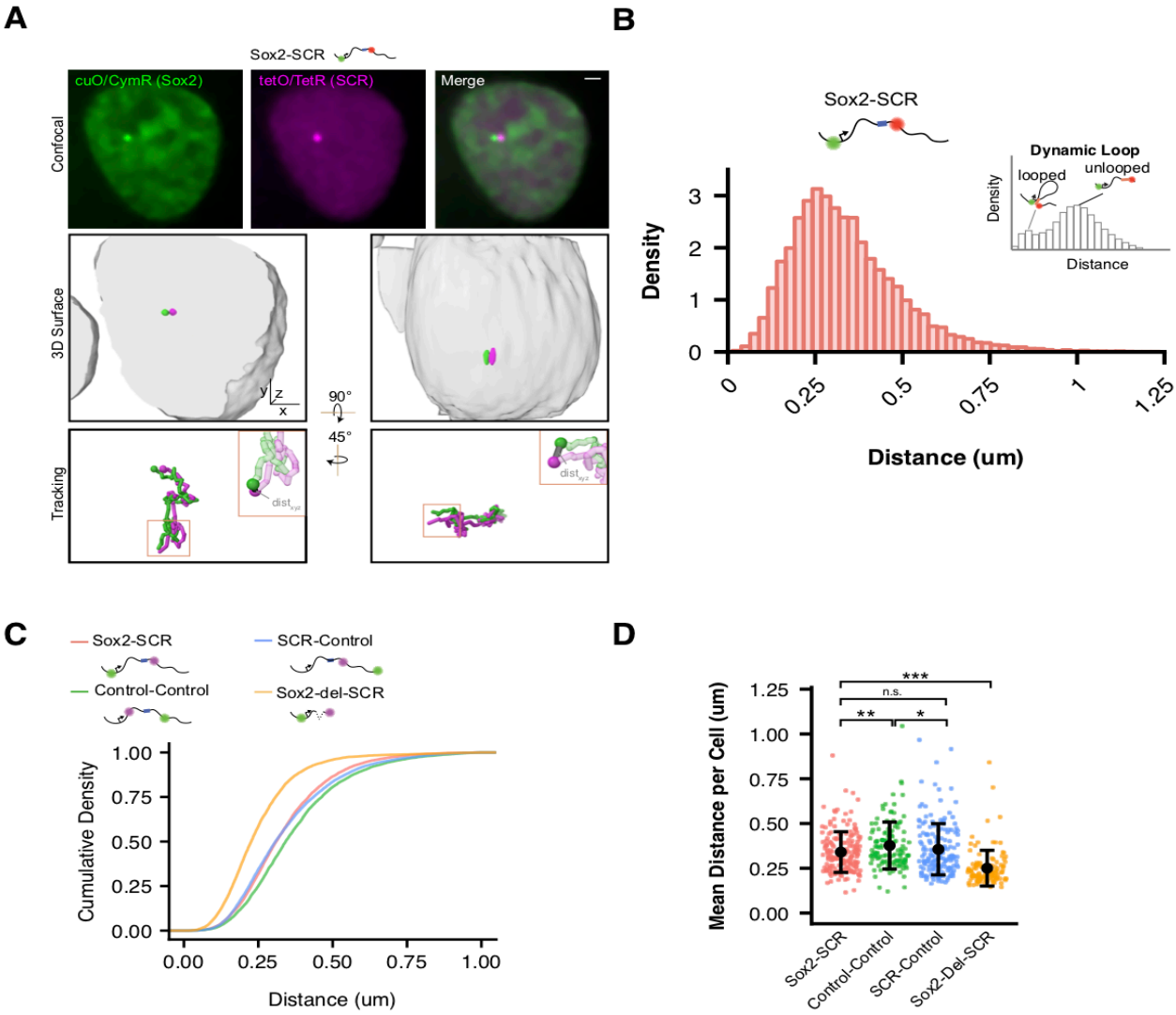

Figure 2. Visualization of the *Sox2* Region in ESCs Reveals Minimal Evidence for *Sox2*/SCR Interactions
A) Top, confocal Z slices of CymR-GFP and TetR-tdTom in Sox2-SCR ESCs, labeling the Sox2 promoter and SCR region with bright puncta, respectively. Middle, 3D surface rendering of the ESC nucleus shown above. A single fluorescence channel was rendered white and transparent to outline the nucleus, and GFP and tdTom surfaces were rendered with high threshold to highlight the cuO and tetO arrays, respectively. Bottom, tracking data is rendered for the nucleus shown above. Inset shows example of calculated 3D separation distance between the two labels. Scale bar is 1 um. B) Normalized histogram of 3D separation distance for *Sox2*-SCR ESCs demonstrates a single peak (Hartigan's Dip Test for multimodality, p = 1). Schematic for an hypothetical looping enhancer-promoter pair is shown as an inset, with two peaks. C) Cumulative density of 3D separation distance for *Sox2*-SCR versus control comparisons. D) Mean 3D separation distance per cell for each label pair. Population means and standard deviations are shown for each sample. Mann-Whitney, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

this label configuration in living ESCs demonstrated a significant shift to more proximal distance values (**Figure 2C,D**). These results are consistent with our expectation that a direct Sox2/SCR interaction would be confined shorter 3D distances than those observed for the *Sox2*-SCR pair and validate our experimental capacity to measure these differences. Taken together, these data demonstrate no unique characteristics for Sox2 and SCR distances in ESCs. While these observations could suggest very infrequent interaction events, they also may allude to fundamental differences between spatial proximity and the features captured by proximity ligation using 3C approaches (see DISCUSSION).

Differentiation of ESCs to Diverse Lineages Correlates with *Sox2* Locus Compaction

We next differentiated our modified cell lines in order to determine how *Sox2* locus

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

organization is altered upon cellular differentiation (**Figure 3A**). To this end, we derived neural precursor cells (NPCs), a cell-type that maintains *Sox2* expression despite inactivation of the SCR and reduced of *Sox2*/SCR contacts by chromosome conformation capture carbon copy (5C) (**Figure 3B**)(Beagan et al., 2017). We validated that our NPC lines expressed NPC marker genes and demonstrated their ability to differentiate into both neurons and astrocytes (**Figure 3—figure supplement 1**). As an additional comparison, we differentiated our ESC lines into FLK1$^+$/PDGFRα$^+$ mesodermal precursors (MES), a cell type which downregulates *Sox2* expression and inactivates the SCR element (**Figure 3B**). Interestingly, we observed that all label pairs embedded in the *Sox2* locus showed greater proximity in differentiated cells compared to ESCs (**Figure 3C**). These changes were significant when comparing mean distances per cell between label pairs in NPCs or MES with ESCs (**Figure 3D**). These data suggest the entire *Sox2* locus adopts a more compact conformation upon ESC differentiation, regardless of transcriptional status of *Sox2*.

To explore if compaction of the *Sox2* locus conformation might be driven by inactivation of the SCR element (which occurs in both NPCs and MES) or could be driven by other factors related to cellular differentiation, we generated a heterozygous genetic deletion of the SCR element on the 129 allele in ESCs using CRISPR/Cas9 (**Figure 1—figure supplement 1, Figure 3—figure supplement 2**). These cells show no signs of differentiation and maintained naive ESC morphology, consistent with previous studies (Zhou et al., 2014). Moreover, SCR deletion led to reduction of *Sox2* expression from the *cis* allele to undetectable levels by qPCR (**Figure 1—figure supplement 2**). Live-cell visualization of the cuO and tetO labels in these cells demonstrated a slight shift in 3D distances towards greater proximity; however, this shift was small compared to that seen after differentiation to NPCs or MES (**Figure 3—figure**

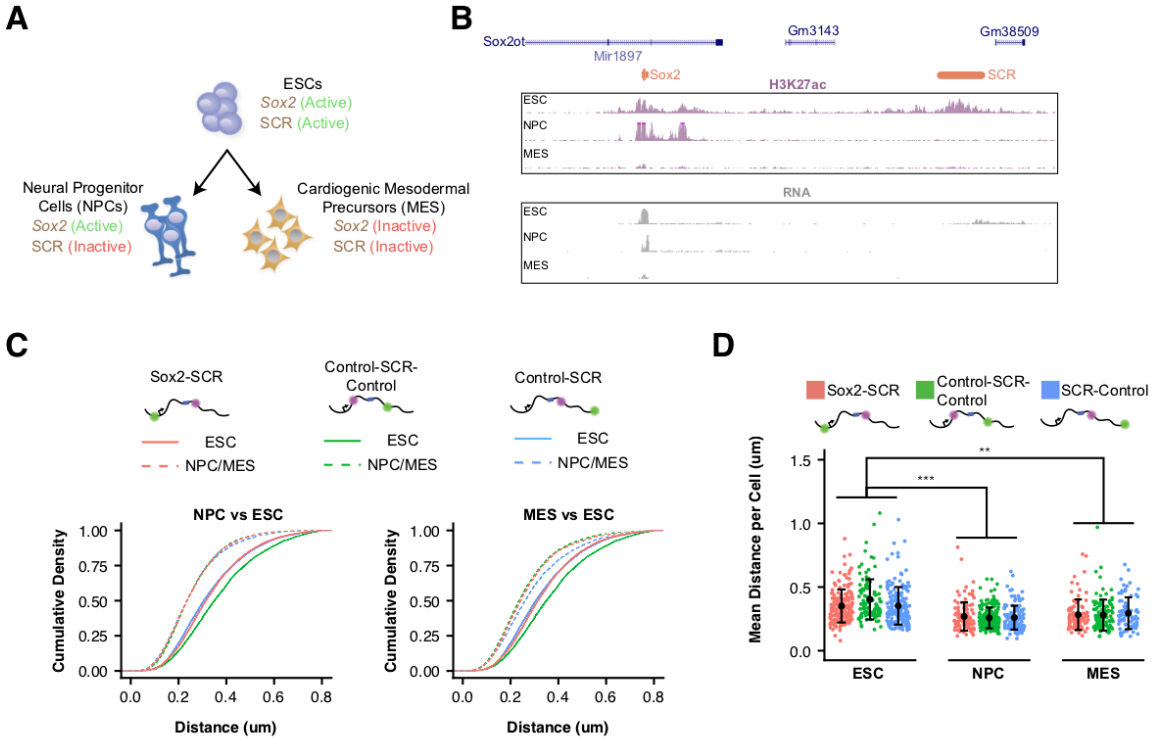

Figure 3. *Sox2* Locus Compacts upon ESC Differentiation. A) ESCs were differentiated into neural progenitor cells (NPCs), which maintain expression of *Sox2* but inactivate the SCR, and cardiogenic mesodermal precursors (MES), which inactivate both *Sox2* and the SCR. B) Browser tracks of H3K27ac and RNA-seq data from ESCs, NPCs, and MES demonstrate the activation status of *Sox2* and SCR in each cell type. Y-axis is 0-5 reads per million for H3K27ac data and 0-10 reads per million for RNA-seq data. C) Cumulative density of 3D separation distance for *Sox2*-SCR and two control pairs for NPCs (left) and MES (right). ESC data is shown for comparison as solid lines on each graph and reproduced from Figure 2C. D) Mean 3D separation distance per cell for each label pair, organized by cell type. Statistical analysis is for each matched pair-wise comparison between cell types. All p-values are below reported value. Mann-Whitney (** $p < 0.01$, *** $p < 0.001$). H3K27ac data from GSE47949 (Wamstad et al., 2012) and GSE24164 (Creyghton et al., 2010). RNAseq data from GSE47949 and GSE44067 (Zhang et al., 2013).

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

supplement 2). Hierarchical clustering analysis of the similarity between distance histograms revealed that SCR-deleted ESCs were most similar to other ESC lines (**Figure 3—figure supplement 2**). These observations suggest that *Sox2* locus organization is significantly altered with ESC differentiation but largely robust to changes in *Sox2* or SCR activity.

Slow *Sox2* Locus Conformation Dynamics Lead to Limited Exploration and Variable Enhancer Encounters

We next investigated the dynamics of *Sox2* spatial organization and focused our analysis of the ESC state. While all three label pairs showed comparable distance profiles across the cell population, we observed striking variation in locus organization between individual cells (**Figure 4A,B, Video 2**). We observed label pairs in prolonged compact or extended conformations as well as gradual or sharp transitions between the two (**Figure 4A**). However, few label pairs explored their entire range – the distance spread observed across our cell population -- during our imaging window (~25 minutes), demonstrating that *Sox2* locus conformation dynamics are slow over tens of minutes.

To better understand this phenomenon, we investigated the dynamic properties of our *Sox2*-SCR label pair, as well as both control pairs. Both relative step sizes (defined as the 3D displacement of the cuO label between frames if the tetO location is fixed) and the change in 3D separation distance between frames were significant (e.g.180 nm and 79 nm, respectively, for the *Sox2*-SCR pair, 20 sec per frame, **Figure 4—figure supplement 1**). However, the relative displacement between two adjacent frames was rarely additive. Instead, we found a strong bias in the XY plane angle between two successive relative displacements toward 180°, demonstrating anti-correlation in the directionality of two adjacent steps in time (**Figure 4—figure supplement 1**, Kuiper's Test $p < 0.01$). These data point to extensive oversampling of the local environment by individual loci within the *Sox2* region. These findings are consistent with the polymer nature of chromatin (Dekker and Mirny, 2016) and the viscoelastic nature of the nucleoplasm (Lucas et al., 2014).

To further capture the dynamics of *Sox2* organization, we computed the autocorrelation function. The autocorrelation function describes the correlation between measurements separated by various lag times and can be utilized to quantify memory or inertia in single cell quantities (e.g. protein levels) compared to the population average (Sigal et al., 2006)(**Figure 4C**). Autocorrelation values near one are expected between closely spaced measurements, decaying towards zero for larger lag times. An autocorrelation coefficient of zero indicates that the underlying process has randomized during the time lag between the relevant measurements. Computation of the autocorrelation function for each label pair revealed a monotonic autocorrelation decay with increasing lag times (**Figure 4D**). We observe an initial rapid reduction in autocorrelation in the small time lag regime, driven by a period of effective local exploration. As our probes begin to oversample the local environment (1-2 mins), the autocorrelation decay slows, reflecting the constraint on locus diffusion within the nuclear environment. Interestingly, at long time lags (> 10 mins), the autocorrelation function for both control pairs appears to flatten to a slope of zero, suggesting that conformational memory for some loci may be quite long-lived.

An important implication of this behavior of chromatin is that first encounter times between loci is highly dependent on the initial configuration of the genomic region (**Figure 4E**, Fisher's Exact Test, $p = 1.42 \times 10^{-3}$). Indeed, while 73% of *Sox2*-SCR pairs that start within 200 nm of each other are observed to have at least one encounter (defined as closer than 100 nm) over the 25 minute imaging window, this drops to 18% for pairs that start greater than 600 nm away. These data suggest that enhancer proximity, and therefore the capacity for direct enhancer-promoter contact, is likely to be highly variable across time within a cell and between cells within a population.

Visualization of *Sox2* Transcriptional Bursts in Living ESCs

We next explored the temporal relation between 3D organization of the *Sox2* locus and transcription. To this end, we utilized the well-established MS2 reporter system to directly visualize nascent transcription in single living ESCs (Bertrand et al., 1998). Using CRISPR/Cas9

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

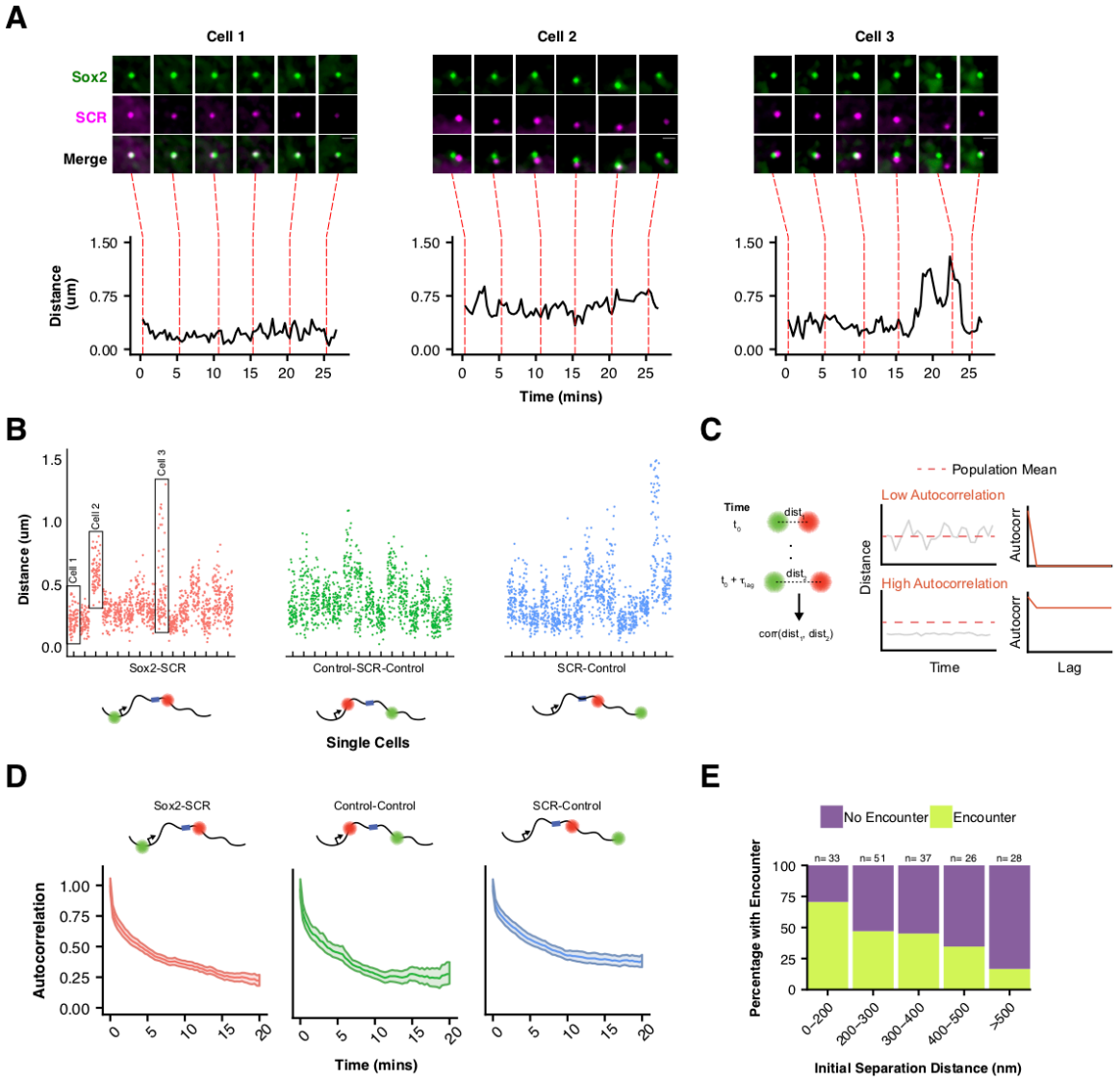

Figure 4. Slow *Sox2* Locus Conformation Dynamics Lead to Limited Exploration and Variable Enhancer Encounters
A) Maximum-intensity projection images (top) centered on the *Sox2* locus and associated 3D distance measurements (bottom) highlight distinct conformations and dynamics of the *Sox2* locus across cells. Scale bar is 1 um. B) 3D separation distance measurements for individual cells for *Sox2*-SCR, Control-Control, and SCR-Control highlight the heterogeneity of *Sox2* locus organization across the cell population. The three cells depicted in A are boxed. C) Cartoon description of autocorrelation analysis. Distance measurement between two time points are correlated using population statistics, revealing the time scale over which local measurements diverge from the population mean. A cell with low autocorrelation will randomly fluctuate around the population mean, leading the autocorrelation function to quickly decay to zero. A cell with high autocorrelation will deviate substantially from the expected value, only slowly relaxing back to the population mean. In this case, the autocorrelation function will stay significantly above zero for large lag times. D) Autocorrelation function for *Sox2*-SCR, Control-Control, and SCR-Control pairs demonstrates significant autocorrelation at large lag times, indicating significant memory in 3D conformation across a 20 minute window. The plotted values are mean ± 95% CI. E) Percent of cells with an encounter between *Sox2* and SCR (defined as within 100 nm) shown as a function of the initial separation distance measured for the cell. Likelihood of an encounter is sharply dependent on the initial conformation of the locus.

genome engineering, we replaced the endogenous 129 *Sox2* allele with a modified version that includes a P2A-puromycin resistance gene fusion and 24 MS2 stem loops inserted into the 3' UTR of the *Sox2* gene (**Figure 5—figure supplement 1**). We generated this MS2 reporter allele in our *Sox2*-SCR labeled cell line to generate *Sox2*-8C$^{cuO/+}$, *Sox2*-117T$^{tetO/+}$, *Sox2*$^{MS2/WT}$ ESCs (or simply *Sox2*-MS2 ESCs). Transcription levels derived from the *Sox2*-MS2 reporter allele were lower than those from the untargeted CastEiJ allele (**Figure 1—figure supplement**

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

2), potentially due to reduced stability of transcripts labeled with MS2 stem loops (Ochiai et al., 2014). Western blotting of Sox2-MS2 lysate revealed a SOX2 doublet as expected, suggesting proper expression of both wild-type SOX2 and the SOX-P2A fusion (**Figure 5—figure supplement 1**).

We first characterized the transcriptional activity of *Sox2*-MS2 reporter allele. We co-expressed a tandem-dimer of the MS2 coat protein fused with 2 copies of tagRFP-T (tdMS2cp-tagRFP-Tx2), TetR fused with 2 copies of GFP (TetR-GFPx2), and CymR fused with 2 copies of Halo tag (CymR-Halox2) in *Sox2*-MS2 ESCs. These ESCs enabled simultaneous visualization of the *Sox2* promoter, SCR, and nascent *Sox2* transcription in living ESCs when imaged in the presence of the Halo-tag ligand JF646 (Grimm et al., 2015) (**Figure 5A**). Time-lapse confocal microscopy revealed bright flashes of MS2cp signal in the ESC nucleus, which occurred in spatial proximity to the cuO and tetO labels, and were similar to the MS2 transcriptional bursts observed elsewhere (Bothma et al., 2014; Chubb et al., 2006; Lionnet et al., 2011; Martin et al., 2013; Ochiai et al., 2014). These results suggested the *Sox2* MS2 reporter allele enables visualization of *Sox2* transcription.

We developed a computational pipeline for identification and quantitation of *Sox2* transcriptional activity using the MS2 reporter (see MATERIALS AND METHODS). Briefly, we identified an ROI surrounding the *Sox2* promoter (cuO signal) for each nucleus in each frame, performed a 2D-gaussian fit on the maximum Z projection of the MS2cp fluorescence signal, and classified a successful Gaussian fit as transcriptional signal using a k-nearest neighbor classification scheme. After training, our pipeline classifications matched a manually scored testing set with 98% accuracy, correctly identifying 70% of curated MS2 signal while maintaining a false positive rate of 0.1%. Thus, we consider our identified bursts to be a conservative estimate of *Sox2* transcriptional activity.

Using our pipeline, we identified a total of 603 individual bursts across 1,208 cells (**Figure 5B**, **Supplementary Files 4,5**, **Video 3**). We found *Sox2* transcriptional activity to be sporadic both between cells and within individual cells across time (**Figure 5C**). Nearly two-thirds (66.1%) of nuclei lacked detectable *Sox2* transcription during our 30 minute imaging window, with the majority of remaining cells demonstrating transcriptional activity in less than 20% of frames (29.3%, **Figure 5—figure supplement 1**). However, we did observe rare cells that demonstrated robust transcriptional activity in greater than half the measured frames (0.25% of cells, **Video 4**). We also found substantial variability in the intensity of transcriptional bursts and their duration (**Figure 5D**). As a population, we found Sox2-MS2 ESCs spent 4% of their time with a detectable MS2 burst (**Figure 5E**). Thus, our live-cell measurements of *Sox2* transcription suggest short, intermittent transcriptional activity in ESCs.

To ensure that our MS2 analysis identified bona fide transcriptional activity, we repeated our analysis in a number of control contexts. First, we measured bursting frequency in ESCs that expressed the MS2 coat protein but lacked the *Sox2*-MS2 reporter allele(*Sox2*-8C$^{cuO/+}$, *Sox2*-117T$^{tetO/+}$, *Sox2*$^{WT/WT}$). Second, we measured bursting frequency in *Sox2*-MS2 ESCs that harbored an SCR deletion in *cis* (*Sox2*-8C$^{cuO/+}$, *Sox2*-117T$^{tetO/+}$, *Sox2*$^{MS/WT}$, SCR$^{del/+}$). Third, we measured bursting frequency in *Sox2*-MS2 ESCs that were treated with the transcriptional inhibitor 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB). In each case, we observed a significant drop in *Sox2* burst frequency (**Figure 5E**). Taken together, these data demonstrate our ability to accurately identify *Sox2* transcriptional events using our MS2 reporter cell line.

Sox2 Transcription Is Not Associated with SCR Proximity

Assuming SCR regulates *Sox2* transcription via the conventional enhancer looping model, we would expect *Sox2* transcriptional activity to occur during interactions or periods of *Sox2*/SCR proximity (**Figure 6A**), given that *Sox2* depends of SCR for its ESC expression. To investigate this prediction, we restricted our analysis to nuclei with single, diffraction-limited spots for the cuO and tetO labels in our *Sox2*-MS2 ESC dataset. We calculated 3D distances between *Sox2* and SCR and compared single cell distance traces with matched MS2 signal traces. We identified some transcriptionally active cells that showed prolonged proximity of the *Sox2*/SCR labels. However, we also observed cells which showed robust transcriptional bursting despite a prolonged extended conformation of the *Sox2* region, driving *Sox2*/SCR

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

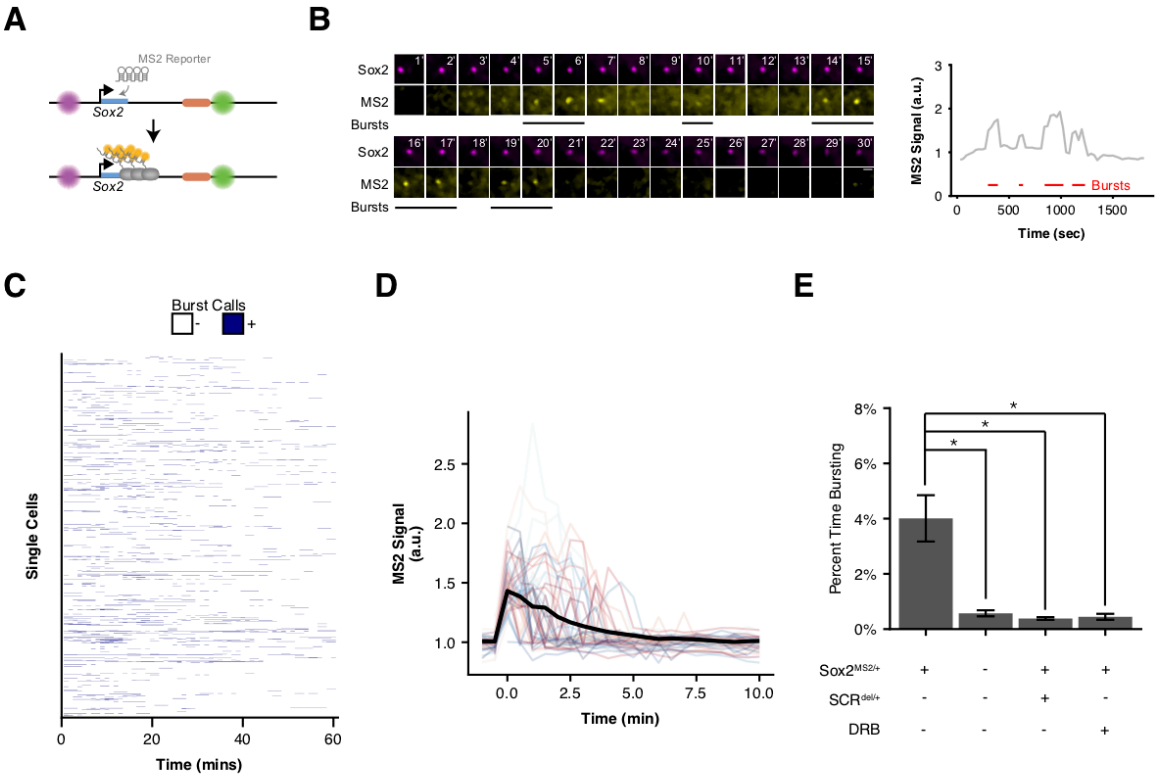

Figure 5. Visualizing *Sox2* Expression in Single Living ESCs Reveals Intermittent Bursts of Transcription. A) *Sox2* locus with cuO-labeled *Sox2* promoter and tetO-labeled SCR was further modified to introduce an MS2 transcriptional reporter cassette into the *Sox2* gene. Transcription of *Sox2* leads to visible spot at the *Sox2* gene due to binding and clustering of MS2 coat protein to the MS2 hairpin sequence. B) Maximum-intensity projection images centered on the *Sox2* promoter (cuO) show intermittent bursts of MS2 signal, which are quantified on the right. Scale bar is 1 um. C) Single cell trajectories of *Sox2* transcriptional bursts as representatively shown in B. D) Aligned *Sox2* transcriptional bursts. Randomly selected *Sox2* bursts are shown as color traces (n = 50). Black line is mean MS2 signal for all annotated bursts. E) Percent time *Sox2* transcriptional bursting for various experimental conditions. Bars are mean ± standard error of ≥ 3 independent experiments. *Sox2*$^{MS2/+}$ indicates cell line harbors the *Sox2*-MS2 reporter allele. SCR$^{del/+}$ indicates presence of an SCR deletion *in cis* with the *Sox2*-MS2 reporter. DRB indicates treatment with the transcriptional inhibitor 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB).

distance above the population average for the duration of our 30 minute imaging window (**Figure 6B, Video 5**). We binned time points according to the measured distance between *Sox2* and SCR and calculated the percent time spent bursting for each bin and found that all bins showed similar transcriptional activity (**Figure 6C**). Furthermore, segregating time points into bursting and non-bursting frames for each cell demonstrated no significant differences between the two groups (**Figure 6D**, Mann-Whitney, p = 0.72).

We next considered the possibility that *Sox2*/SCR proximity might precede transcriptional bursting by a characteristic time (e.g. due to lags required for transcription complex assembly or elongation through the gene body). To this end, we identified the initiation point for all bursts in our dataset and considered a 25-minute window centered at each burst initiation event. Alignment and meta-analysis of these bursts showed little change in *Sox2*/SCR distance across the time window. To determine if *Sox2*/SCR distance significantly deviated from expected values across transcriptional bursts, we compared aligned bursts to a randomly shuffled control dataset and found no significant differences between the burst-centered and random-centered analysis (**Figure 6E**, **Supplementary File 6**). This analysis suggests *Sox2*/SCR proximity and *Sox2* transcription is not separated by a characteristic lag within the time frame considered.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

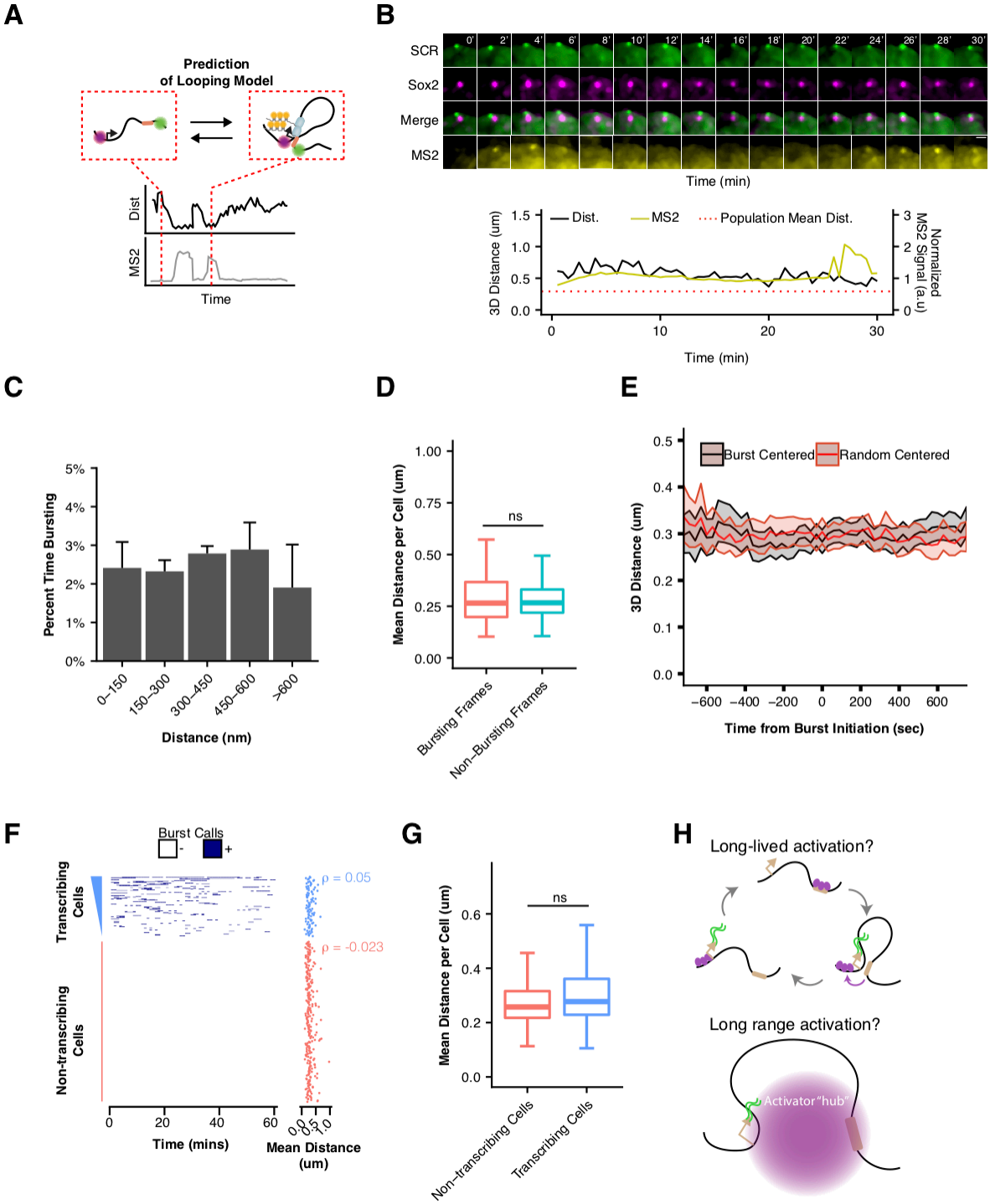

Figure 6. *Sox2* Transcription Is Not Associated with SCR Proximity. A) Schematic illustrating the expected relation between *Sox2*/SCR distance and MS2 transcription for a looping enhancer model. B) Maximum-intensity projection images centered on the *Sox2* promoter (cuO) show transcriptional activity without correlation to *Sox2*/SCR distance changes. The measured distance and MS2 signal are shown at bottom. The mean separation distance across the cell population is shown as a dotted red line. Scale bar is 1 um. C) Percent time with *Sox2* transcriptional burst as a function of *Sox2*/SCR distance. Weighted mean + SE for 7 experiments are shown. Weights were determined based on the proportion of frames in each bin contributed by individual experiments. D) Mean separation distance per cell, separated into bursting and non-bursting frames. (Mann-Whitney, p = 0.72). E) Mean separation distance across a 25 minute window for all transcriptional bursts (black) or randomly select time points (red), aligned according the burst initiation frame. Values plotted are mean ± 95% CI. F) Single cell trajectories of *Sox2* transcriptional bursts ranked by number of bursting frames per cell. At right, matched mean separation distances for each cell shown at left. Spearman's correlation coefficient for each is shown. G) Mean separation distance per cell for transcribing and non-transcribing cells. (Mann-Whitney, p = 0.12). H) Potential models of SCR regulation of *Sox2* that would uncouple *Sox2*/SCR proximity from transcriptional activity. Above, SCR leads to long-lived activation of the *Sox2* promoter that can persist long after *Sox2*/SCR contact is disassembled. Below, SCR nucleates a large hub of activator proteins that can modify the *Sox2* promoter environment despite large distances between *Sox2* and SCR.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Finally, given the high degree of cell-to-cell variability in *Sox2* locus organization, we investigated whether cells with greater average *Sox2*-SCR proximity, which would enable more frequent *Sox2*/SCR encounters, demonstrated higher transcriptional activity. We rank ordered cells based on cumulative transcriptional activity (i.e. number of transcriptionally active frames) and compared mean *Sox2*/SCR distance per cell (**Figure 6F**). As expected, non-transcribing cells showed no correlation between order and distance, given the ordering within this group was essentially random (Spearman's $\rho = -0.02$). However, transcribing cells also showed no correlation between transcriptional activity and distance (Spearman's $\rho = 0.05$). As a group, transcribing cells demonstrated no significant difference in mean *Sox2*/SCR separation distance compared to non-transcribing cells (**Figure 6G**, Mann-Whitney, $p = 0.12$). These data reveal little relation between the 3D conformation of *Sox2* relative to the SCR enhancer and its transcriptional output. These data suggest SCR is unlikely to modulate *Sox2* expression through a conventional enhancer looping mechanism.

DISCUSSION

We have investigated the dynamic 3D organization and underlying transcriptional activity of the established enhancer-gene pair *Sox2* and SCR. Interestingly, we observe few unique spatial characteristics for Sox2/SCR in ESCs; observed distance distributions and their spatial dynamics for SCR and the *Sox2* promoter region are similar to those observed between SCR and an equally-spaced non-specific region. In contrast, 3C-based assays have identified enriched contacts between Sox2/SCR as compared to the surrounding neighborhood. We note, however, that these results need not be incompatible. Proximity ligation (3C) and separation distance (microscopy) are distinct measures of chromatin structure with unique biases, assumptions, and limitations, and thus provide snapshots of chromatin architecture that may differ (Dekker, 2016; Fudenberg and Imakaev, 2017; Giorgetti and Heard, 2016). 3C-based assays often utilize millions of cells and so may capture rare conformations in the cell population; these rare conformations would have minimal impact on overall distance distributions constructed using microscopy. Moreover, it remains unclear what spatial proximity is required to enable ligation events during 3C, and this property may differ for distinct genomic regions. Indeed, enrichment of *Sox2*/SCR contacts in 3C assays may reflect only subtle differences in very proximal conformations (e.g. < 50 nm), conformations unlikely to be captured by our microscopy measurements due to technical limitations in localization resolution. Alternatively, large macromolecular bridges or hubs may enable crosslinking and ligation over larger distances that need not demonstrate pronounced spatial proximity, as recently demonstrated (Quinodoz et al., 2018). Moreover, chromatin composition and accessibility are likely to influence crosslinkability, distances permissive for proximity ligation, as well as spatial distances between probes. These characteristics raise important sources of uncertainty in how these measures translate to underlying chromatin structure. Thus, while a comprehensive picture of *Sox2* locus organization remains out of view, our study provides guidance as to what structures are unlikely. For instance, the absence of enhanced proximity between the *Sox2* and SCR pair suggests a prolonged, proximal conformation established by stable, direct pairing of the *Sox2* promoter with SCR is unlikely to be the predominant structure in ESCs.

Surprisingly, we also observe no association between *Sox2*/SCR proximity and *Sox2* transcription in real time. Indeed, we detect no correlation between transcriptional activity and instantaneous *Sox2*/SCR distances, no reduction in *Sox2*/SCR distances prior to transcriptional bursts, and no tendency for transcriptionally active cells to display reduced *Sox2*/SCR distance. These results strongly suggest that direct interactions between *Sox2* and SCR do not lead to contemporaneous transcriptional activity of the *Sox2* gene, a central feature of enhancer looping models.

The *Sox2* locus displays distinct behavior from an enhancer reporter recently used to explore the regulatory logic of the even-skipped (*eve*) enhancers in *Drosophila* embryos. In this study, the authors integrated an enhancer reporter ~142 kb upstream of *eve* locus and promoted pairing between the two loci by including an ectopic insulator sequence, which pairs with a similar sequence embedded near the *eve* enhancers. In this system, the authors observe both bimodality in distance measurements as well as clear correlation between enhancer-

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

reporter proximity and reporter transcription. While it is not yet clear why these systems behave so differently, we note the considerable differences in the 3D distances we report for *Sox2* (339 nm for *Sox2*/SCR) and those reported for the even-skipped reporter (709 nm for unpaired and 353 nm for paired). It seems plausible that the more extended conformation of the *Drosophila* chromosome necessitates pairing in order to bring the *eve* enhancer sufficiently close the reporter, particularly for enhancers evolved to function within 10 kb of their target gene. Our analysis suggests that most *Sox2*/SCR loci sample this distance range, perhaps lowering the importance of locus conformation for SCR function. Indeed, SCR transcriptional control does demonstrate proximity dependence on some scale, as SCR ablation is not compensated for by a normal copy located on the homologous chromosome (Li et al., 2014; Zhou et al., 2014). In other contexts, such as during olfactory receptor gene choice or transvection in *Drosophila* , regulation can occur over very large distances in *cis* (~ 80 Mb) or in *trans,* and transcriptional activity may be more closely tied to pairing events that promote spatial proximity, as recently demonstrated for the latter (Horta et al., 2018; Lim et al., 2018; Markenscoff-Papadimitriou et al., 2014). Hence, genomic interactions and other features of genome topology may differ in importance depending of the spatial distances navigated by enhancer-gene pairs.

What essential function SCR plays in *Sox2* transcription remains unclear, but we propose two mutually non-exclusive models that can account for our observations: 1) long-lived activation of the *Sox2* locus and promoter by SCR and/or 2) *Sox2*/SCR communication across measurable 3D distances. For example, SCR might induce long-lived *Sox2* activation through direct contacts with the promoter region, which could persist after disassembly of these interactions (**Figure 6H, top**). This mechanism might be achieved through delivery of durable factors (e.g. chromatin modifiers) to the *Sox2* promoter during contact. Indeed, this mechanism could explain why disruption of DNA loops genome-wide through acute RAD21 degradation leads to only modest changes in nascent transcription after 6 hours (Rao et al., 2017). In addition, numerous mechanisms for long-range communication between enhancers and promoters have been proposed (Bulger and Groudine, 2010). For example, SCR may play a critical role in the nucleation and spreading of important epigenetic activators and chromatin accessibility, establishing a permissive environment of *Sox2* transcription. An intriguing mechanism for action at a distance comes from recent observations that super-enhancers are capable of nucleating large (>300 nm), phase-separated condensates of coactivators, chromatin regulators, and transcription complexes (Cho et al., 2018; Sabari et al., 2018). SCR is a bona fide super-enhancer in ESCs (Whyte et al., 2013). Thus, SCR may deliver activation factors over hundreds of nanometers through inclusion of the *Sox2* promoter into an activator hub or condensate (**Figure 6H, bottom**). Future studies that couple visualization of the *Sox2* locus with that of important molecular components of transcriptional activation will provide critical insights regarding how, when, and over what distances these factors function at this critical pluripotency gene.

MATERIALS AND METHODS

ESC Culture

129/CastEiJ F1 hybrid mouse embryonic stem cells were maintained in 2i + Lif media, composed of a 1:1 mixture of DMEM/F12 (Thermo Fisher Waltham, MA, #11320-033) and Neurobasal (Thermo Fisher #21103-049) supplemented with N2 supplement (Thermo Fisher #17502-048), B27 with retinoid acid (Thermo Fisher #17504-044), 0.05% BSA (Thermo Fisher #15260-037), 2 mM GlutaMax (Thermo Fisher #35050-061), 150 uM 1-thioglycerol (Sigma St. Louis, MO, M6145), 1 uM PD03259010 (Selleckchem Houston, TX, #1036), 3 uM CHIR99021 (Selleckchem #S2924) and 10^6 U/L leukemia inhibitory factor (Peprotech Rocky Hill, NJ, #250-02). Media was changed daily and cells were passaged every 2 days.

ESC Genome Modification

For insertion of PhiC31 and Bxb1 attP sequences, 150,000 cells were electroporated with 1 uM of single-stranded oligonucleotide donor containing the attP sequence and 400 ng of the sgRNA/Cas9 dual expression plasmid pX330 (a gift from Feng Zhang, Addgene Plasmid #42230) using the Neon Transfection System (Thermo Fisher). Neon settings for the

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

electroporation were as follows: 1400V, 10ms pulse width, 3 pulses. Electroporated ESCs were given 3 days to recover, followed by seeding approximately 5000 cells on a 10cm dish for clone isolation (see Clone Isolation).

For integration of the tetO and cuO array, 150,000 cells were electroporated with 300ng each of 1) a tetOx224 repeat plasmid bearing a PhiC31 attB sequence and a FRT-flanked neomycin resistance cassette, 2) a cuOx144 repeat (approximate) plasmid bearing a Bxb1 attB sequence and a floxed puromycin or blasticidin resistance cassette, 3) an expression plasmid for the PhiC31 integrase (a gift from Philippe Soriano, Addgene Plasmid #13795), and 4) an expression plasmid for the Bxb1 integrase using the Neon Transfection System. Electroporated ESCs were allowed to recover for 3 days, followed by 7 days of drug selection using 500 ug/mL G418 and either 1 ug/mL puromycin or 8 ug/mL blasticidin in antibiotic-free media. After drug selection, cells were electroporated again with 400 ng each of Cre and Flpo expression plasmids to remove the resistance cassettes. 3 days after electroporation, approximately 5000 cells were seeded on a 10cm plate for clone isolation (see Clone Isolation).

For targeting of the MS2 reporter construct into the endogenous *Sox2* allele, we generated a targeting plasmid that inserted a P2A sequence followed by the puromycin resistance gene upstream of the endogenous *Sox2* stop codon with 1kb homolog arms on either side. We next mutated the PAM sequence for our sgRNA in the 3' homology arm by site-directed mutagenesis. 24 repeats of the MS2 hairpin sequence were inserted into an EcoRI restriction site located just 3' of the puromycin stop codon. 150,000 cells were electroporated with 400ng of targeting plasmid and 400ng of pX330 expressing the appropriate sgRNA. Electroporated ESCs were given 3 days to recover, followed by 5 days of puromycin selection. Approximately 5000 cells were subsequently seeded on a 10cm dish for clone isolation (see Clone Isolation). A positive clone was identified by PCR. DNA sequencing confirmed no mutations in the *Sox2*-P2A-puro$_r$ cassette and identified a single bp deletion in the 3' UTR of the non-targeted CastEiJ allele due to residual targeting of a non-canonical NAG PAM.

For deletion of the *Sox2* Control Region or the Sox2-1-112T fragment, 150,000 cells were electroporated with 400ng each of pX330 expressing sgRNAs targeting genomic regions centromeric and telomeric to the deletion fragment. 3 days after electroporation, approximately 5000 cells were seeded on a 10cm plate for clone isolation (see Clone Isolation).

ESC Clone Isolation

After 5-6 days of growth at low density (~5000 cells per 10 cm dish), individual colonies were picked and transferred to a 96-well plate. Briefly, colonies were aspirated and transferred to a well with trypsin, followed by quenching and dissociation with 2i+Lif+5%FBS. Once the 96-well plate had grown to confluency, we split the clones into 2 identical 96-well plates. One plate was frozen at -80°C by resuspending the clones in 80% FBS/20% DMSO freezing media. The second plate was used for DNA extraction. All wells were washed once with PBS and subsequently lysed overnight at 55°C in a humidified chamber with 50 uL lysis buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 1.5 mM MgCl2, 0.45% NP40, 0.45% Tween 20, 100 ug/mL Proteinase K). Genomic DNA was concentrated by ethanol precipitation and resuspended in 100 uL of double distilled water. 1 uL of suspension was used for subsequent PCR screening reactions using GoTaq Master Mix (Promega Madison, WI, #M7123).

Stable Expression of Fluorescent Transgenes

To generate stable lines expressing CymR, TetR, and MS2cp fluorescent protein fusions, 150,000 cells were electroporated with 400ng of an ePiggyBac Transposase expression plasmid (a gift from Ali Brivanlou) and 50ng of expression plasmid bearing PiggyBac terminal repeats. 7 days after electroporation, fluorescent cells were resuspended in fluorescence-activated cell sorting (FACS) buffer (5% FBS in PBS) and purified via FACS using a FACSAria II (BD). To enrich cells expressing the CymR-Halox2 fusion protein, ESCs were incubated in 100 nM of Janeila Fluor 646 (a gift from Luke Lavis) for 30 minutes at room temperature, washed once in FACS Buffer, incubated for 30 minutes at room temperature in

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

FACS Buffer, washed again, and sorted using a FACSAria II.

Isolation of Neural Progenitor Cells from ESCs

ESCs were passaged onto gelatinized 6 wells at 50,000-100,000 cells. The following day, these cultures were switched to N2B27 media (1:1 composition of DMEM/F12 and Neurobasal, N2 supplement, B27 with retinoic acid, 0.05% BSA, 2 mM GlutaMax, 150 uM 1-thioglycerol, 25 ug/mL insulin (Sigma #I6634)). After 4 days, we dissociated the cultures and seeded 1 million cells in an ungelatinized 10 cm dish in N2B27 with 10 ng/mL FGF basic (R&D Systems Minneapolis, MN, #233-FB) and 10 ng/mL EGF (Peprotech #315-09) to form neurospheres. After 3-4 days of outgrowth, neurospheres were collected by gentle centrifugation (180xg, 3 minutes) and plated onto a pre-gelatinized 6 well. Neural progenitor cell (NPCs) lines were established by passaging (4-6 passages). For maintainance of NPCs, cells were cultured on wells pre-treated with poly-D-lysine and 4 ug/mL natural mouse laminin (Invitrogen #23017015) in N2B27 with 10ng/mL FGF basic and 10 ng/mL EGF and passaged every 4-5 days.

Differentiation of NPCs to neurons and astrocytes

To differentiate NPCs to astrocytes, 30,000 cells were plated onto coverglass within a 24 well pre-treated with poly-D-lysine and laminin. The following day, cells were switched to N2B27 with 10 ng/mL BMP4 (R&D Systems #314-BP) and allowed to differentiate for 12 days.

To differentiate NPCs to neurons, 30,000 cells were plated onto coverglass within a 24 well pre-treated with poly-D-lysine and laminin. The following day, cells were switched to N2B27 with 10 ng/mL FGF basic and allowed to differentiate for 6 days. Cells were then switched to N2B27 without additional factors and grown for 6 days.

Differentiation of Cardiogenic Mesodermal Precursors from ESCs

ESCs were dissociated and seeded to form embryoid bodies at 1 million cells per dish in SFD media (3:1 composition of IMDM (Thermo Fisher #12440-053) and Ham's F12 (Thermo Fisher #11765-054), N2 supplement, B27 without retinoic acid (Thermo Fisher #12587-010), 0.05% BSA, 2 mM GlutaMax, 50 ug/mL ascorbic acid (Sigma #A-4544), 450 uM 1-thioglycerol). After 2 days, EBs were dissociated and reaggregated at 1 million cells per dish in SFD media with 5ng/mL VEGF (R&D Systems #293-VE), 5 ng/mL Activin A (R&D Systems #338-AC), and 0.75 ng/mL BMP4 to induce cardiogenic mesoderm. 40hrs after induction, cells were dissociated and stained for Flk1 and PDGFRα. Briefly, cells were wash four times in FACS Buffer, followed by incubation for 30 minutes with a biotinylated anti-FLK-1 antibody (Hybridoma Clone D218, 1:100). Cells were then washed three times with FACS Buffer and incubated with a PE-conjugated anti-PDGFRα (Thermo Fisher #12-1401-81, 1:400) and APC-Streptavidin (BD Biosciences Franklin Lakes, NJ, #554067, 1:200) for 30 minutes at room temperature. Cells were then washed two times with FACS Buffer and sorted for FLK1$^+$/PDGFRα^+ cells.

Immunofluorescence

NPCs or differentiated astrocytes/neurons on coverglass were fixed for 10 minutes at room temperature with 4% paraformaldehyde in PBS. After fixing, the coverglass were washed twice with PBS, permeabilized in PBS with 0.5% Triton X-10 for 10 minutes, and washed once in PBS with 0.1% Triton. Cells were then blocked for 1 hour at room temperature in PBS/0.1% Triton/4% goat serum. After blocking, coverglass were incubated in primary antibody in PBS/0.1% Triton/4% goat serum overnight at 4°C in a humidified chamber. Coverglass were subsequently washed three times with PBS/0.1% Triton and incubated in secondary antibody in PBS/0.1% Triton/4% goat serum at room temperature for 1 hour. After secondary incubation, coverglass were washed three times with PBS/0.1% Triton, stained with DAPI in PBS (1 ug/mL), and mounted on a slide for imaging in mounting medium (1x PBS, pH7.4, 90% glycerol, 5 mg/mL propyl gallate). Antibodies used were anti-SOX2 (Santa Cruz Biotechnology Dallas, TX, #sc-365823, Lot# K1414), anti-PAX6 (Biolegend San Diego, CA, #901301, Lot# B235967), anti-TUBB3 (Biolegend #801201, Lot# B199846), and anti-GFAP (Sigma #G3893, Lot#

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

105M4784V).

Western Blotting

3 million cells were collected, washed once with PBS, and lysed in 4x Laemmli Buffer. Protein lysate was passed through a 30 gauge needle twenty times to shear the genomic DNA and the lysate was cleared by centrifugation at 13,000 RPM for 10 minutes at 4°C. Subsequently, lysate was supplemented with 100 mM DTT and boiled at 95°C for 10 minutes. 200,000 cells of protein lysate were loaded onto a Bis-Tris 4-12% polyacrylamide gels (ThermoFisher #NW04120BOX) and electrophoresis was carried out using the Bolt system (ThermoFisher). Protein was transferred to a PVDF membrane. Membranes were blocked for 1 hour at room temperature with 4% milk PBS Tween (PBST). Membrane was subsequently incubated in primary antibody overnight in 4% milk PBST at 4°C. Membranes were then washed four times 15 minutes at room temperature in PBST and incubated in secondary antibody in 4% milk PBST for 1 hour at room temperature. After secondary incubation, membranes were washed four times 15 minutes at room temperature in PBST, incubated in SuperSignal chemiluminescence HRP substrate (ThermoFisher #34075), and visualized by film exposure. Antibodies used were anti-SOX2 (Santa Cruz #sc-365823, Lot# K1414) and anti-β-actin (Abcam Cambridge, UK, ab8227, Lot# GR92448-1).

Quantitative PCR

RNA was extracted from 500,000-1,000,000 millions using TRIzol and 200ng of RNA was reversed transcribed using the QuantiTect Reverse Transcription kit (Qiagen Hilden, Germany). Quantitative PCR was perform on 8ng cDNA in technical triplicates using TaqMan Gene Expression Master Mix (ThermoFisher #4369016) on a 790HT Fast Real-Time PCR System (ThermoFisher). The primer and probe sets used are as follows:

Sox2 Forward primer – 5'CTACGCGCACATGAACGG3',
Sox2 Reverse primer – 5'CGAGCTGGTCATGGAGTTGT3',
Sox2 129 allele probe – /56-FAM/CAACCGATG/ZEN/CACCGCTACGA/3IABkFQ/,
Sox2 CastEiJ allele probe – /56-FAM/CAGCCGATG/ZEN/CACCGATACGA/3IABkFQ/,
Tbp Forward primer – 5'ACACTCAGTTACAGGTGGCA3',
Tbp Reverse primer – 5'AGTAGTGCTGCAGGGTGATT3',
Tbp probe - /56-FAM/ACACTGTGT/ZEN/GTCCTACTGCA/3IABkFQ.
56-FAM = Fluorescein
ZEN = internal quencher (IDT)
3IABkFQ = 3' Iowa Black quencher

Live-Cell Microscopy

We imaged all live-cell experiments on a Nikon Ti-E microscope and the following setup for live, spinning disk confocal microscopy: Yokogawa CSU-22 spinning disk, 150 mW Coherent OBIS 488 nm laser, 100 mW Coherent OBIS 561 nm laser, 100 mW Coherent OBIS 640 nm laser, a Yokogawa 405/491/561/640 dichroic, zET405/488/561/635m quad pass emission filter, Piezo Z-drive, Okolab enclosure allowing for heating to 37°C, humidity control, and CO_2 control, and a Plan Apo VC 100x/1.4 oil immersion objective. Image acquisition utilized either a Photometric Evolve Delta EMCCD or an Andor iXon Ultra EMCCD camera.

ESCs were plated one day prior to imaging on a 8-chambered coverglass (VWR Radnor, PA, #155409) pretreated for at least 2 hours with 3.1 ug/mL Laminin-511 (iWaichem Tokyo, Japan #N-892011) at 120,000 cells per chamber. Just prior to imaging, 2i+Lif media was pre-mixed with 50 ug/mL ascorbic acid and a 1:100 dilution of Prolong Live Antifade Reagent (ThermoFisher P36975). If the cells to be imaged also expressed CymRHalox2, 100 nM of JF646 was also added to the media. After a one hour incubation, we added this media to the ESCs to be imaged.

NPCs were plated at least 8 hours prior to imaging on a 8-chambered coverglass pre-treated with poly-D-lysine and laminin at 120,000 cells per chamber. Prior to imaging, N2B27 with FGF basic and EGF was pre-mixed with 50 ug/mL ascorbic acid and a 1:100 dilution of Prolong Live Antifade Reagent. After a one hour incubation, we added this media to the NPCs

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

to be imaged.

Cardiogenic mesodermal cells enriched by FACS for FLK1 and PDGFRα were plated on 8-chambered coverglass precoated with 0.1% gelatin in StemPro-34 (Thermo Fisher #10639-011) supplemented with 2 mM GlutaMax, 50 ug/mL ascorbic acid, 5 ng/mL VEGF, 10 ng/mL FGF basic, and 25 ng/mL FGF10 (R&D Systems #345-FG) and cultured for 24 hours. Just prior to imaging, StemPro-34 media (with the additives listed above) was supplemented with a 1:100 dilution of Prolong Live Antifade Reagent, incubated for one hour, and subsequently added to the cultures for imaging.

For imaging dual color experiments (CymRGFP and TetRtdTom), we captured alternating green and red images for each plane by toggling the 488 nm and 561 nm lasers, enabling fast acquisition of both colors and minimal time between images (30ms exposure). A z-series of two color planes was acquired by moving between planes quickly using the Piezo Z-drive. For imaging tri-color experiments (CymRHalox2-JF646, TetRGFPx2, tdMS2cp-tagRFP-Tx2), we imaged the green and far red channels as above (toggling the 488 nm and 640 nm lasers), followed by a second pass through all z planes with the 561 nm after a ET525/50m emission filter was inserted in the light path. This eliminated bleed-through signal from the JF646 dye during 561 nm excitation allowed by the quad pass emission filter. All images were acquired using μManager (Edelstein et al., 2010).

Imaging data for each condition is composed of a minimum of three imaging sessions, except for cardiogenic mesodermal cultures, in which duplicate differentiations were performed.

Image Processing

Images were background subtracted using a dark image, converted to 32-bit, and denoised using NDSafir (Carlton et al., 2010; Kervrann and Boulanger, 2006). Denoised images were reverted back to 16-bit, fluorescence bleach corrected using exponential fitting, and despeckled to remove high-frequency noise using ImageJ (Schindelin et al., 2012; Schneider et al., 2012).

Image Analysis

Tracking Loci

Maximum Z-projections of 3D time series were manually analyzed to identify cuO/CymR and tetO/TetR spots in nuclei and annotate individual loci as doublets (likely two sister chromatids) or singlets. Loci that showed any frames with doublet spots for either channel were not included in downstream analysis. For each *Sox2* locus with well-behaved singlets, an ROI was drawn that included the locus location throughout the timecourse (or if the locus became untrackable from leaving the field of view, the duration of its visibility). In some cases (e.g. NPCs), multiple ROIs were needed to track a single loci because of large-scale movements of the cell nucleus. In these cases, location data was merged together after tracking. For each locus, the 3D location for the cuO/CymR spot and the tetO/TetR spot was tracked within the delimited ROI using TrackMate (Tinevez et al., 2017). TrackMate tracks for each spot were manually inspected, and if multiple tracks existed (due to gaps in the tracking), these were merged through manual curation. Spot positions were shifted by 0.5 pixels and converted to physical distances using a 0.091 um pixel size and a 0.3um z-step. We corrected for chromatic aberration by shifting position based on displacements observed using TetraSpeck fluorescent beads (ThermoFisher #T7279).

3D positions of cuO/CymR and tetO/TetR were associated with each *Sox2* locus.

Localization Error Estimation

Tetraspeck (Thermo Fisher T7279) multicolor fluorescent beads were embedded in 2% agarose and a one hundred frame Z-stack time series was constructed at various laser intensities. The max spot intensity as well as the mean and standard deviation of the nuclear background was estimated from ten nuclei for both cuO/CymR and tetO/TetR using our raw time-lapse data. Bead time series were modified to add background noise using ImageJ to approximate the nuclear background and then denoised as described above. 9-10 beads that showed signal within one standard deviation of that observed for either the cuO/CymR or

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

tetO/TetR spots were tracked using TrackMate and the standard deviation of position in the X, Y, and Z dimensions was computed using a 10 frame sliding window.

Euclidean Distance

1D, 2D, and 3D euclidean distances were calculated using the formula:

$$Dist_{ij} = \sqrt{\sum_{v=1}^{n} \left(X_{vi} - X_{vj}\right)^2}$$

where i and j represent the cuO/CymR and tetO/TetR spot, respectively, and n the number of dimensions.

Relative Displacement

The relative position of spot1 (CymRGFP) with respect to spot2 (TetRtTom) for the vth dimension was calculated as follows:

$$X_{v\hat{i}} = (X_{vi} - X_{vj})$$

The relative displacement was then calculated as the change is the relative position of spot 1.

$$Disp_t = \sqrt{\sum_{v=1}^{n} \left(X_{vi}(t) - X_{vi}(t-1)\right)^2}$$

where t is the current frame and n is the number of dimensions.

Angle between Displacement Vectors in the XY Plane

We calculated the displacement vectors of two adjacent frames.

$$u = \sqrt{\sum_{v=1}^{2} \left(X_{vi}(t) - X_{vi}(t-1)\right)^2} \qquad v = \sqrt{\sum_{v=1}^{2} \left(X_{vi}(t-1) - X_{vi}(t-2)\right)^2}$$

where t is the current frame and v reflects the dimension. The angle between these vectors was then calculated by the following:

$$\theta_{xy} = atan2(det_{uv}, dot_{uv})$$

where atan2(y, x) is the arctan(y, x) function that uses the sign of y and x to determine the appropriate quadrant for the result and

$$det_{uv} = \begin{vmatrix} u_x & u_y \\ v_x & v_y \end{vmatrix} = u_x v_y - u_y v_x$$

$$dot_{uv} = u \cdot v = u_x v_x - u_y v_y$$

Autocorrelaton Analysis

Autocorrelation values were calculated according to the formula

$$A(\tau) = \frac{E[(D_t - \mu)(D_{t+\tau} - \mu)]}{\sigma^2}$$

where D_t represents distance at time t, τ is the time lag, μ and σ^2 are the average and variance of 3D distance measured across the cell population, and E is the expected (i.e. average) value. Confidence intervals were computed by bootstrapping and recalculating $A(\tau)$ across 1000 simulations to estimate 95% confidence.

Distribution Distances and Clustering

The distance between 3D distance probability distributions from two cell lines or cell types was computed using earth mover's distance (EMD). Briefly, the earth mover's distance is the minimum cost to convert one probability distribution to the other over a defined region. We

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

calculated pairwise EMD for each 3D probability distribution using the R package *earthmovdist*. Complete-linkage hierachical clustering was then performed to generate a dendrogram.

MS2 Signal Identification and Quantification

3D time-lapse images of tdMS2cp-tagRFP-Tx2 were converted into 2D images by maximum Z projection. For each *Sox2* locus considered for analysis, a 20x20 pixel region centered on the XY tracking position of the cuO/CymR spot, reflecting the position of the *Sox2* promoter region, was analyzed for each frame. If tracking information was missing for a given frame, the position coordinates from the nearest frame were used. This 20x20 region was used for parameter estimation for 2D Gaussian fitting using the equation:

$$f(x,y) = Ae^{-\left(\frac{(x-x_o)^2}{2\sigma_x^2}+\frac{(y-y_o)^2}{2\sigma_y^2}\right)} + C$$

where A (Gaussian height), x_o, y_o (location of Gaussian peak), σ_x^2, σ_y^2 (Gaussian variance), and C (offset) are all estimated parameters. Initial estimate of the offset was defined as the median pixel value in the ROI, A was estimated as the maximum pixel value minus the estimated offset, σ_x^2 and σ_y^2 were estimated as 1, and x_o, y_o was estimated as the location of the brightest pixel in the ROI. These initial estimates were used attempt a Gaussian fit on a 10x10 pixel region centered on the estimated Gaussian position. We constrained the potential Gaussian fit to have a minimum height of 10% above background fluorescence, a fit position of no more than 3 pixels from the estimated position, and a width of no more than 4. Successful Gaussian fits were filtered for likelihood to reflect true MS2 signal using a k-nearest neighbor model trained on manually classified data and 4 parameters of the fit (A, σ_x^2, σ_y^2, and an R^2 value). Furthermore, frames were also required to have at least one neighboring frame ($\pm$ 3 frames) also demonstrate MS2 signal, eliminating high frequency noise. Time points which passed these filter steps were assigned a relative MS2 Signal based on:

$$Signal = \frac{A + C}{Normalization\ Factor}$$

were the normalization factor was the median pixel value for the 20x20 pixel ROI across all timepoints. For time points that did not pass filter, MS2 signal was taken as the median value of the 20x20 ROI for the current frame normalized as above.

Sox2 Burst Classification

Sox2 burst initiation events were classified as frames positive for MS2 signal (see above) that lack MS2 positive classifications for the preceding three frames. All frames spanning the burst initiation and the last positive MS2 classification prior to the next burst initiation were labeled as one burst event.

Aligned Sox2 Burst Analysis

To align our MS2 data across all *Sox2* bursts, a defined window was sampled for each burst centered on the burst initiation event. We subsequently generated a randomly sampled control comparison for this analysis by randomly shuffling the frames labeled as burst initiation events and repeating the analysis. Mean distances or MS2 signal were then calculated based on relative frame compared to the burst initiation event. Confidence intervals were computed by bootstrapping and recalculating the mean value for each relative frame across 1000 simulations to estimate 95% confidence.

Browser tracks

Unless wiggle files were available as part of the accession, sequencing read archives (SRA) were downloaded from NCBI and reads were aligned to the mm9 mouse genome using Bowtie (Langmead et al., 2009) as part of the Galaxy platform (Afgan et al., 2018). Sequences were extended by 200 bp and allocated into 25 bp bins to generate wiggle files. HiC data was visualized using JuiceBox (Durand et al., 2016). Browser tracks were visualized on the UCSC

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Genome browser (Kent et al., 2002).

FIGURE LEGENDS

Figure 1. The *Sox2* Locus As a Model for Visualization of Enhancer-Promoter Regulation in Mouse Embryonic Stem Cells. A) To visualize chromosome loci in living cells, we have used tetO/TetR and cuO/CymR genetic labels. Our pipeline for insertion of these labels into the mouse genome is shown. First, CRISPR-Cas9 is used to place an attP intergrase landing site. Second, a targeting plasmid bearing the compatible attB sequence, the tetO or cuO array, and a selection cassette is introduced along the integrase (Int) to mediate site-specific integration. The selection cassette can then be subsequently removed by Cre/Flp recombinase. B) The *Sox2* locus in mouse ESCs. Genomic browser tracks of epigenomic and expression data demonstrate high levels of histone acetylation, RNA polymerase II, and transcription factor (OCT4, SOX2, NANOG, CTCF) occupancy at *Sox2* and the distal *Sox2* Control Region enhancer (tan boxes). Data from 4C and HiC experiments demonstrate chromosomal contacts at the *Sox2* locus. For 4C data, read density indicates contact frequency with a fixed position near the *Sox2* promoter (red triangle). Y-axis for browser tracks is reads per million. For HiC, all pairwise contact frequencies are mapped using a heatmap. The intensity of each pixel represents the normalized number of contacts detected between a pair of loci. The maximum intensity is indicated in red square. At bottom, locations of the cuO- and tetO-arrays for the three cell lines utilized for this study. Sox2-8C$^{cuO/+}$; Sox2-117T$^{tetO/+}$ (Sox2-SCR) ESCs were used to track Sox2/SCR location. Two control lines, Sox2-43T$^{tetO/+}$; Sox2-164T$^{cuO/+}$ (Control-Control) and Sox2-117T$^{tetO/+}$; Sox2-242T$^{cuO/+}$ (SCR-Control) were analyzed for comparison. H3K27ac, RNA polymerase II (RNAP), and RNAseq data from GSE47949 (Wamstad et al., 2012); DNase data from GSE51336 (Vierstra et al., 2014); SOX2, OCT4, NANOG, and CTCF data from GSE11431 (Chen et al., 2008b); 4C data from GSE72539 (de Wit et al., 2015); and HiChIP data from GSE96107 (Bonev et al., 2017).

Figure 2. Visualization of the *Sox2* Region in ESCs Reveals Minimal Evidence for *Sox2*/SCR Interactions

A) Top, confocal Z slices of CymR-GFP and TetR-tdTom in Sox2-SCR ESCs, labeling the Sox2 promoter and SCR region with bright puncta, respectively. Middle, 3D surface rendering of the ESC nucleus shown above. A single fluorescence channel was rendered white and transparent to outline the nucleus, and GFP and tdTom surfaces were rendered with high threshold to highlight the cuO and tetO arrays, respectively. Bottom, tracking data is rendered for the nucleus shown above. Inset shows example of calculated 3D separation distance between the two labels. Scale bar is 1 um. B) Normalized histogram of 3D separation distance for *Sox2*-SCR ESCs demonstrates a single peak (Hartigan's Dip Test for multimodality, $p = 1$). Schematic for an hypothetical looping enhancer-promoter pair is shown as an inset, with two peaks. C) Cumulative density of 3D separation distance for *Sox2*-SCR versus control comparisons. D) Mean 3D separation distance per cell for each label pair. Population means and standard deviations are shown for each sample. Mann-Whitney, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure 3. *Sox2* Locus Compacts upon ESC Differentiation. A) ESCs were differentiated into neural progenitor cells (NPCs), which maintain expression of *Sox2* but inactivate the SCR, and cardiogenic mesodermal precursors (MES), which inactivate both *Sox2* and the SCR. B) Browser tracks of H3K27ac and RNA-seq data from ESCs, NPCs, and MES demonstrate the activation status of *Sox2* and SCR in each cell type. Y-axis is 0-5 reads per million for H3K27ac data and 0-10 reads per million for RNA-seq data. C) Cumulative density of 3D separation distance for *Sox2*-SCR and two control pairs for NPCs (left) and MES (right). ESC data is shown for comparison as solid lines on each graph and reproduced from Figure 2C. D) Mean 3D separation distance per cell for each label pair, organized by cell type. Statistical analysis is for each matched pair-wise comparison between cell types. All p-values are below reported value. Mann-Whitney (** $p < 0.01$, *** $p < 0.001$). H3K27ac data from GSE47949 (Wamstad et al., 2012) and GSE24164 (Creyghton et al., 2010). RNAseq data from GSE47949 and

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

GSE44067 (Zhang et al., 2013).

Figure 4. Slow *Sox2* Locus Conformation Dynamics Lead to Limited Exploration and Variable Enhancer Encounters
A) Maximum-intensity projection images (top) centered on the *Sox2* locus and associated 3D distance measurements (bottom) highlight distinct conformations and dynamics of the *Sox2* locus across cells. Scale bar is 1 um. B) 3D separation distance measurements for individual cells for *Sox2*-SCR, Control-Control, and SCR-Control highlight the heterogeneity of *Sox2* locus organization across the cell population. The three cells depicted in A are boxed. C) Cartoon description of autocorrelation analysis. Distance measurement between two time points are correlated using population statistics, revealing the time scale over which local measurements diverge from the population mean. A cell with low autocorrelation will randomly fluctuate around the population mean, leading the autocorrelation function to quickly decay to zero. A cell with high autocorrelation will deviate substantially from the expected value, only slowly relaxing back to the population mean. In this case, the autocorrelation function will stay significantly above zero for large lag times. D) Autocorrelation function for *Sox2*-SCR, Control-Control, and SCR-Control pairs demonstrates significant autocorrelation at large lag times, indicating significant memory in 3D conformation across a 20 minute window. The plotted values are mean ± 95% CI. E) Percent of cells with an encounter between *Sox2* and SCR (defined as within 100 nm) shown as a function of the initial separation distance measured for the cell. Likelihood of an encounter is sharply dependent on the initial conformation of the locus.

Figure 5. Visualizing *Sox2* Expression in Single Living ESCs Reveals Intermittent Bursts of Transcription. A) *Sox2* locus with cuO-labeled *Sox2* promoter and tetO-labeled SCR was further modified to introduce an MS2 transcriptional reporter cassette into the *Sox2* gene. Transcription of *Sox2* leads to visible spot at the *Sox2* gene due to binding and clustering of MS2 coat protein to the MS2 hairpin sequence. B) Maximum-intensity projection images centered on the *Sox2* promoter (cuO) show intermittent bursts of MS2 signal, which are quantified on the right. Scale bar is 1 um. C) Single cell trajectories of *Sox2* transcriptional bursts as representatively shown in B. D) Aligned *Sox2* transcriptional bursts. Randomly selected *Sox2* bursts are shown as color traces (n = 50). Black line is mean MS2 signal for all annotated bursts. E) Percent time *Sox2* transcriptional bursting for various experimental conditions. Bars are mean ± standard error of ≥ 3 independent experiments. *Sox2*$^{MS2/+}$ indicates cell line harbors the *Sox2*-MS2 reporter allele. SCR$^{del/+}$ indicates presence of an SCR deletion *in cis* with the *Sox2*-MS2 reporter. DRB indicates treatment with the transcriptional inhibitor 5,6-Dichloro-1-β-D-ribofuranosylbenzimidazole (DRB).

Figure 6. *Sox2* Transcription Is Not Associated with SCR Proximity. A) Schematic illustrating the expected relation between *Sox2*/SCR distance and MS2 transcription for a looping enhancer model. B) Maximum-intensity projection images centered on the *Sox2* promoter (cuO) show transcriptional activity without correlation to *Sox2*/SCR distance changes. The measured distance and MS2 signal are shown at bottom. The mean separation distance across the cell population is shown as a dotted red line. Scale bar is 1 um. C) Percent time with *Sox2* transcriptional burst as a function of *Sox2*/SCR distance. Weighted mean + SE for 7 experiments are shown. Weights were determined based on the proportion of frames in each bin contributed by individual experiments. D) Mean separation distance per cell, separated into bursting and non-bursting frames. (Mann-Whitney, p = 0.72). E) Mean separation distance across a 25 minute window for all transcriptional bursts (black) or randomly select time points (red), aligned according the burst initiation frame. Values plotted are mean ± 95% CI. F) Single cell trajectories of *Sox2* transcriptional bursts ranked by number of bursting frames per cell. At right, matched mean separation distances for each cell shown at left. Spearman's correlation coefficient for each is shown. G) Mean separation distance per cell for transcribing and non-transcribing cells. (Mann-Whitney, p = 0.12). H) Potential models of SCR regulation of *Sox2* that would uncouple *Sox2*/SCR proximity from transcriptional activity. Above, SCR leads to long-lived activation of the *Sox2* promoter that can persist long after *Sox2*/SCR contact is

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

disassembled. Below, SCR nucleates a large hub of activator proteins that can modify the *Sox2* promoter environment despite large distances between *Sox2* and SCR.

SUPPLEMENTAL FIGURE LEGENDS

Figure 1—figure supplement 1. Characterization of Modified Embryonic Stem Cell Lines. A) Schematic of modified cell lines used in this study. Primer sets used to amplify recombination arms for tetO- and cuO- integration are shown. **B,C)** PCR genotyping of ESC lines shown in A.

Figure 1—figure supplement 2. *Sox2* Expression Characterization for Modified Embryonic Stem Cell Lines. A) Ratio of *Sox2* expression from the 129 allele and the CastEiJ allele measured by qPCR for modified ESC lines. B) *Sox2* expression relative to control gene (Tbp) for various cell lines. E14 mESCs, which lack the CastEiJ allele, are included to demonstrate specificity of allele-specific qPCR assay. Deletion of SCR region leads to loss of expression from the *Sox2* allele *in cis*. N.D. is not detected.

Figure 2—figure supplement 1. Estimate Localization Precision for cuO and tetO. A-B) Histograms of X, Y, and Z position uncertainty for fluorescent beads with signal-to-noise ratios comparable to cuO/CymR-GFP (A) or tetO/TetR-tdTom (B). Data plotted are the standard deviation values measured using 10 frame sliding windows collected from 9-10 beads. Error bars show mean and standard deviation of the computed position uncertainties, which are reported in the upper right of each panel.

Figure 2—figure supplement 2. Tracking Lengths for tetO and cuO Spots Across Cell Lines. A-B) Histograms of the cuO-array (A) or tetO-array (B) Track lengths for cell lines used in the study as ESCs, NPCs, and MES. Tracking lengths were often shorter in NPCs or MES due to increased nuclear movement in these cell types compared to ESCs.

Figure 3—figure supplement 1. Characterization of ESC-derived Neural Progenitor Cell Lines. A) Immunofluorescence of fixed neural progenitor cells (NPCs) for the NPC markers SOX2 and PAX6. B-C) Immunofluorescence for the neuron marker β3-tubulin (B) or the astrocyte marker GFAP (C) on fixed cultures after 12 days of differentiation towards neurons or astrocytes, respectively. Scale bar is 100um.

Figure 3—figure supplement 2. SCR Inactivation Does Not Drive Locus Compaction Upon Differentiation. A) Potential models for *Sox2* locus compaction observed upon differentiation to NPCs or MES. At left, cellular differentiation may lead to global changes in chromatin structure that are not dependent of *Sox2*/SCR activation status. Alternatively, *Sox2* and SCR inactivation could lead to changes to chromatin structure within the Sox2 locus, driving locus-specific compaction. **B)** Strategy for CRISPR/Cas9-mediated SCR deletion. Two gRNAs were designed to flank the SCR region and generate a large deletion of SCR. Below, the SCR deletion allele shows a novel junction near the locations of expected Cas9 cutting, indicating a loss of the intervening SCR sequence. **C)** Scatterplot of mean and standard deviation of 3D distance measurements for each cell line visualizes similarity between *Sox2* label pairs across cell types. **D)** Dendrogram visualizing hierarchical clustering of Earth Mover's distances between 3D separation distance histograms of distinct Sox2 label pairs across cell types. SCR-deleted ESCs show greatest similarity to other ESCs as opposed to differentiated cells with inactivation of the SCR element.

Figure 4—figure supplement 1. Dynamics Statistics for Each Sox2 Locus Pair in ESCs. A-B) Normalized histograms of relative step size and change in 3D separation distance for adjacent frames. Mean value is highlighted by a red line. C) Measurement of the XY plane angle made between two successive displacement vectors demonstrates pronounced bias towards a locus "bouncing back" to its position preceding a given step. Statistical test for bias was performed using Kuiper's Test for Uniformity. ** is $p < 0.01$.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 5—figure supplement 1. Generation and Characterization of Sox2-MS2 Transcriptional Reporter ESCs. A) Targeting strategy for *Sox2* transcriptional reporter. A targeting plasmid was used with *Sox2* homology arms and a P2A peptide puromycin resistance gene cassette (2Apuro) inserted in frame with *Sox2*. Downstream of 2A puro is a 24x MS2 stem loop array, which is inserted into the 3' UTR. **B)** PCR genotyping assay to identify a targeted *Sox2* allele. A primer set was used that recognized the MS2 stem loop array and a genomic region downstream of the 3' homology arm. **C)** Western blotting for SOX2 protein in parental 129/CastEiJ ESCs or ESCs heterozygous for the Sox2-MS2 allele. Actin was used as a loading control. **D)** Normalized histogram of the percentage of time individual cells have a detectable *Sox2* transcriptional burst.

SUPPLEMENTAL FILE LEGENDS

Supplementary File 1. Primer Sequences Used in Cell Line Characterization. List of PCR primer sequences and expected amplicon size used in the study. Brief description of the purpose of each primer pair is included.

Supplementary File 2. 20bp Guide RNA Sequences Used in CRISPR/Cas9 Genome Engineering. List of 20bp sequences homologous to the mouse 129 genome designed into CRISPR/Cas9 sgRNAs. Targeted genomic location (mm9 coordinates), genome strand, and brief description of purpose for sgRNA is included.

Supplementary File 3. Data Table from 3D Tracking of cuO/CymR and tetO/TetR labels. All data used in the study for cuO/CymR and tetO/TetR localization. C1 refers to Channel 1 (cuO/CymR). C2 refers to Channel2 (tetO/TetR). Columns are as follows:

- **Cell_Line** – label used to identify cell line
- **Batch** – unique microscopy session identifier
- **C1_T_Step-sec** – step size between frames
- **Locus_ID** – unique identifier for each Sox2 locus
- **C1_TrackID** – track identifier from TrackMate
- **C1_Track_Length** – track length from TrackMate
- **C1_SpotID** – spot identifier from TrackMate
- **C1_X_Value_pixel** – X position in pixels for C1 spot
- **C1_Y_Value_pixel** – Y position in pixels for C1 spot
- **C1_Z_Value_slice** – Z position in slices for C1 spot
- **C1_T_Value_frame** – frame of measurement
- **C1_X_Value_um** – X position in microns for C1 spot
- **C1_Y_Value_um** – Y position in microns for C1 spot
- **C1_Z_Value_um** – Z position in microns for C1 spot
- **C1_T_Value_sec** – time point in seconds for measurement
- **C2_TrackID** – track identifier from TrackMate
- **C2_Track_Length** – track length from TrackMate
- **C2_SpotID** – spot identifier from TrackMate
- **C2_X_Value_pixel** – X position in pixels for C2 spot
- **C2_Y_Value_pixel** – Y position in pixels for C2 spot
- **C2_Z_Value_slice** – Z position in slices for C2 spot
- **C2_T_Value_frame** – imaging frame
- **C2_X_Value_um** – X position in microns for C2 spot
- **C2_Y_Value_um** – Y position in microns for C2 spot
- **C2_Z_Value_um** – Z position in microns for C2 spot
- **C2_T_Value_sec** – time point in seconds
- **X_Distance_um** – X distance between C1 and C2 labels
- **Y_Distance_um** – Y distance between C1 and C2 labels
- **Z_Distance_um** – Z distance between C1 and C2 labels
- **XY_Distance_um** – XY distance between C1 and C2 labels
- **XYZ_Distance_um** – XYZ distance between C1 and C2 labels,

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

C1_Corrected_X_Value_um – X position in microns for C1 spot after correcting for chromatic aberration,
C1_Corrected Y_Value_um – Y position in microns for C1 spot after correcting for chromatic aberration
C1_Corrected Z_Value_um – Z position in microns for C1 spot after correcting for chromatic aberration
Corrected_X_Distance_um – X distance after correcting for chromatic aberration
Corrected_Y_Distance_um – Y distance after correcting for chromatic aberration
Corrected_Z_Distance_um – Z distance after correcting for chromatic aberration
Corrected_XY_Distance_um – XY distance after correcting for chromatic aberration
Corrected_XYZ_Distance_um – XYZ distance after correcting for chromatic aberration
Relative_C1_Corrected_X_Value_um – X position of C1 label relative to the position of C2 in microns
Relative_C1_Corrected_Y_Value_um – Y position of C1 label relative to the position of C2 in microns
Relative_C1_Corrected_Z_Value_um – Z position of C1 label relative to the position of C2 in microns
Relative_XY_Displacement_um – Relative XY distance traveled by the C1 label between adjacent frames
Relative_XYZ_Displacement_um – Relative XYZ distance traveled by the C1 label between adjacent frames
Relative_XY_Angle_radians – Relative angle between two successive displacements for the C1 label in the XY plane

Supplementary File 4. Data Table for MS2 Transcription Analysis for All Loci. All data used in transcriptional analysis of *Sox2* locus. Columns are as follows:

Cell_Line – label used to identify cell line
Locus_ID – unique identifier for each Sox2 locus
Gauss_Filter – whether the MS2 Gaussian fit passed the knn filter step
Noise_Filter – whether the MS2 Gaussian fit passed a high frequency noise filter step
Pass_Filter – whether the MS2 signal for the given frame was classified as transcriptional signal. Required both Gauss_Filter = TRUE and Noise_Filter = TRUE
Gaussian_Height_Threshold – minimum relative height above background allowed for Gaussian fit
Gaussian_Width_Threshold – maximum Gaussian variance allowed for Gaussian fit
Background – Offset calculated from Gaussian fit. If no Gaussian fit was found, set to median pixel intensity of ROI
Gaussian Height – Amplitude calculated from Gaussian fit. If no Gaussian fit was found, set to 0
Gaussian_Volume – Volume under fitted Gaussian. If no Gaussian fit was found, set to 0
Local_Median – Median pixel intensity of ROI
Norm_MS2_Signal – Relative height of MS2 gaussian normalized to background. For frames that did not pass filter, local median value was used in pace of gaussian height. See MATERIALS and METHODS for more details.
R_Squared – Coefficient of determination between 2D gaussian fit and experimental data
T_Value_frame – imaging frame
X_Value_pixel – X position in pixels for C2 spot (cuO/CymR)
X_Location – X position of peak of fit Gaussian
X_Sigma – X dimension variance of fit Gaussian
Y_Value_pixel – Y position in pixels for C2 spot (cuO/CymR)
Y_Location – Y position of peak of fit Gaussian
Y_Sigma – Y dimension variance of fit Gaussian

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Z_Value_slice – Z position in slices for C2 spot (cuO/CymR)
Batch – unique microscopy session identifier

Supplementary File 5. Data Table for MS2 Transcription Analysis and 3D Localization for Sox2-SCR Singlets. Data used to compare transcriptional activity of *Sox2* locus to 3D distances between *Sox2* and SCR. C1 refers to Channel 1 (tetO/TetR). C2 refers to Channel2 (cuO/CymR). Columns are as in Supplementary Files 3 and 4 with one additional column:

Active_Transcribing – Whether the locus demonstrated any MS2 signal that passed filter during imaging session.

Supplementary File 6. Data Table of Statistical Comparison of Distances Centered on Transcriptional Bursts. Summary statistics and associated Mann-Whitney p-values for pairwise comparisons between burst centered and random centered distances.

SUPPLEMENTAL VIDEO LEGENDS

Video 1. Visualization of Sister Chromatids at *Sox2* Locus. Maximum-intensity Z projection of 3D confocal Z-stacks of cuO/CymR-GFP (left) and tetO/TetR-tdTom (middle) labeling the *Sox2* promoter region and SCR, respectively demonstrate two clear spots for the SCR label, suggesting cells in S/G2. These cells were excluded from analysis. Scale bar is 1 um.

Video 2. Variability in *Sox2* Locus Organization Across Cells. Maximum-intensity Z projection of 3D confocal Z-stacks of cuO/CymR (green) and tetO/TetR (magenta) labeling the *Sox2* promoter region and SCR, respectively for three individual cells highlighted in Figure 3. The distance range explored by Cell1 and Cell2 is limited, while Cell3 shows large, abrupt changes in distance. Scale bar is 1 um.

Video 3. Identification of *Sox2* Transcriptional Bursts in mESCs. Maximum-intensity Z projection of 3D confocal Z-stacks of a tandem dimer of MS2 coat protein fused with two copies of tagRFP-T. The dashed yellow box highlights the ROI used for burst detection in our automated analysis pipeline, centered on the location of the Sox2 promoter (cuO/CymR location, not shown). Detected bursts are highlighted by red circles centered on the burst location, with color intensity indicating burst intensity. Scale bar is 1 um.

Video 4. High Transcriptional Output from *Sox2* Locus. Maximum-intensity Z projection of 3D confocal Z-stacks of a tandem dimer of MS2 coat protein fused with two copies of tagRFP-T demonstrate a period of high transcriptional activity for the highlighted *Sox2* gene. The dashed yellow box highlights the ROI used for burst detection in our automated analysis pipeline, centered on the location of the *Sox2* promoter (cuO/CymR location, not shown). Detected bursts are highlighted by red circles centered on the burst location, with color intensity indicating burst intensity. Scale bar is 1 um.

Video 5. *Sox2* Transcriptional Bursts in the Absence of SCR Proximity. Maximum-intensity Z projection of 3D confocal Z-stacks of cuO/CymR (green) and tetO/TetR (magenta) labeling the *Sox2* promoter region and SCR, respectively (left), and MS2 coat protein highlighting *Sox2* transcriptional activity (right). We detect clear *Sox2* transcriptional bursts despite no colocalization of the Sox2/SCR labels. Scale bar is 1 um.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

ACKNOWLEDGMENTS

We thank Elphege Nora, Geoffrey Fudenberg, Brian Black, and members of the Lomvardas and Weiner labs for helpful discussion; Elphege Nora, Benoit Bruneau, and Kirstin Meyer for a critical reading of the manuscript; Lena Bengtsson for experimental help and technical assistance; and Edith Heard, Patrick Devine, Feng Zhang, Elphege Nora, Michele Calos, Robert Singer, Philippe Soriano, Barbara Panning, Ali Brivanlou and Luke Lavis for helpful reagents. This work was supported by an American Heart Association Postdoctoral Fellowship (#16POST309100006, JMA); NIH Grants R21EB022787 (ODW), R35GM118167 (ODW), R01DC013560 (SL), and R21EB021453 (BH); and the WM Keck Foundation Medical Research Grant (BH). BH is a Chan Zuckerberg Biohub Investigator.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

Afgan, E., Baker, D., Batut, B., van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., et al. (2018). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research *46*, W537–W544.

Alberts, B., Johnson, A.D., Lewis, J., Morgan, D., Raff, M., Roberts, K., and Walter, P. (2014). Molecular Biology of the Cell (W. W. Norton & Company).

Banerji, J., Olson, L., and Schaffner, W. (1983). A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. Cell *33*, 729–740.

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.

Beagan, J.A., Duong, M.T., Titus, K.R., Zhou, L., Cao, Z., Ma, J., Lachanski, C.V., Gillis, D.R., and Phillips-Cremins, J.E. (2017). YY1 and CTCF orchestrate a 3D chromatin looping switch during early neural lineage commitment. Genome Research *27*, 1139–1152.

Belmont, A.S., and Straight, A.F. (1998). In vivo visualization of chromosomes using lac operator-repressor binding. Trends Cell Biol *8*, 121–124.

Benabdallah, N.S., Williamson, I., Illingworth, R.S., Boyle, S., Grimes, G.R., Therizols, P., and Bickmore, W.A. (2017). PARP mediated chromatin unfolding is coupled to long-range enhancer activation. Biorxiv.org

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S.M., Singer, R.H., and Long, R.M. (1998). Localization of ASH1 mRNA particles in living yeast. Mol Cell *2*, 437–445.

Bickmore, W.A. (2013). The Spatial Organization of the Human Genome. Annu. Rev. Genom. Hum. Genet. *14*, 67–84.

Blum, R., Vethantham, V., Bowman, C., Rudnicki, M., and Dynlacht, B.D. (2012). Genome-wide identification of enhancers in skeletal muscle: the role of MyoD1. Genes Dev *26*, 2763–2779.

Bonev, B., Mendelson Cohen, N., Szabo, Q., Fritsch, L., Papadopoulos, G.L., Lubling, Y., Xu, X., Lv, X., Hugnot, J.-P., Tanay, A., et al. (2017). Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell *171*, 557–572.e24.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Bothma, J.P., Garcia, H.G., Esposito, E., Schlissel, G., Gregor, T., and Levine, M. (2014). Dynamic regulation of eve stripe 2 expression reveals transcriptional bursts in living Drosophila embryos. Proc Natl Acad Sci U S A *111*, 10598–10603.

Boyle, A., Davis, S., Shulha, H., Meltzer, P., Margulies, E., Weng, Z., Furey, T., and Crawford, G.E. (2008). High-Resolution Mapping and Characterization of Open Chromatin across the Genome. Cell *132*, 311–322.

Buecker, C., Srinivasan, R., Wu, Z., Calo, E., Acampora, D., Faial, T., Simeone, A., Tan, M., Swigut, T., and Wysocka, J. (2014). Reorganization of Enhancer Patterns in Transition from Naive to Primed Pluripotency. Stem Cell *14*, 838–853.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nature Methods *10*, 1213–1218.

Bulger, M., and Groudine, M. (2010). Enhancers: the abundance and function of regulatory sequences beyond promoters. Dev Biol *339*, 250–257.

Carlton, P.M., Boulanger, J., Kervrann, C., Sibarita, J.-B., Salamero, J., Gordon-Messer, S., Bressan, D., Haber, J.E., Haase, S., Shao, L., et al. (2010). Fast live simultaneous multiwavelength four-dimensional optical microscopy. Proc Natl Acad Sci U S A *107*, 16016–16022.

Chen, B., Gilbert, L.A., Cimini, B.A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., et al. (2013). Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. Cell *155*, 1479–1491.

Chen, H., Levo, M., Barinov, L., Fujioka, M., Jaynes, J.B., and Gregor, T. (2018). Dynamic interplay between enhancer–promoter topology and gene activity. Nature Genetics *290*, 1–13.

Chen, X., Vega, V.B., and Ng, H.H. (2008a). Transcriptional regulatory networks in embryonic stem cells. Cold Spring Harb Symp Quant Biol *73*, 203–209.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008b). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

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.

Chubb, J.R., Trcek, T., Shenoy, S.M., and Singer, R.H. (2006). Transcriptional pulsing of a developmental gene. Curr Biol *16*, 1018–1025.

Cremer, T., Cremer, M., Dietzel, S., Müller, S., Solovei, I., and Fakan, S. (2006). Chromosome territories – a functional nuclear landscape. Current Opinion in Cell Biology *18*, 307–316.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci USA *107*, 21931–21936.

de Wit, E., Vos, E.S.M., Holwerda, S.J.B., Valdes-Quezada, C., Verstegen, M.J.A.M., Teunissen, H., Splinter, E., Wijchers, P.J., Krijger, P.H.L., and de Laat, W. (2015). CTCF Binding Polarity Determines Chromatin Looping. Mol Cell *60*, 676–684.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Dekker, J. (2016). Mapping the 3D genome: Aiming for consilience. Nat Rev Mol Cell Biol *17*, 741–742.

Dekker, J., and Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. Cell *164*, 1110–1121.

Deng, W., Lee, J., Wang, H., Miller, J., Reik, A., Gregory, P.D., Dean, A., and Blobel, G.A. (2012). Controlling Long-Range Genomic Interactions at a Native Locus by Targeted Tethering of a Looping Factor. Cell *149*, 1233–1244.

Deng, W., Rupon, J.W., Krivega, I., Breda, L., Motta, I., Jahn, K.S., Reik, A., Gregory, P.D., Rivella, S., Dean, A., et al. (2014). Reactivation of developmentally silenced globin genes by forced chromatin looping. Cell *158*, 849–860.

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.

Dowen, 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., et al. (2014). Control of Cell Identity Genes Occurs in Insulated Neighborhoods in Mammalian Chromosomes. Cell *159*, 374–387.

Durand, N.C., Robinson, J.T., Shamim, M.S., Machol, I., Mesirov, J.P., Lander, E.S., and Aiden, E.L. (2016). Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Systems *3*, 99–101.

Edelstein, A., Amodaj, N., Hoover, K., Vale, R., and Stuurman, N. (2010). Computer Control of Microscopes Using µManager. Curr Protoc Mol Biol *92*, 14.20.1–14.20.17.

Fudenberg, G., and Imakaev, M. (2017). FISH-ing for captured contacts: towards reconciling FISH and 3C. Nature Methods *14*, 1–7.

Fukaya, T., Lim, B., and Levine, M. (2016). Enhancer Control of Transcriptional Bursting. Cell *166*, 358–368.

Germier, T., Kocanova, S., Walther, N., Bancaud, A., Shaban, H.A., Sellou, H., Politi, A.Z., Ellenberg, J., Gallardo, F., and Bystricky, K. (2017). Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. Biophys J *113*, 1383–1394.

Giorgetti, L., and Heard, E. (2016). Closing the loop: 3C versus DNA FISH. Genome Biol *17*, 1–9.

Grimm, J.B., English, B.P., Chen, J., Slaughter, J.P., Zhang, Z., Revyakin, A., Patel, R., Macklin, J.J., Normanno, D., Singer, R.H., et al. (2015). A general method to improve fluorophores for live-cell and single-molecule microscopy. Nature Methods *12*, 244–250.

Gröschel, S., Sanders, M.A., Hoogenboezem, R., de Wit, E., Bouwman, B.A.M., Erpelinck, C., van der Velden, V.H.J., Havermans, M., Avellino, R., van Lom, K., et al. (2014). A Single Oncogenic Enhancer Rearrangement Causes Concomitant EVI1 and GATA2 Deregulation in Leukemia. Cell *157*, 369–381.

Gu, B., Swigut, T., Spencley, A., Bauer, M.R., Chung, M., Meyer, T., and Wysocka, J. (2018). Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. Science *359*, 1050–1055.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910.

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., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

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., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet *39*, 311–318.

Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. Cell *169*, 13–23.

Horta, A., Monahan, K., Bashkirova, L., and Lomvardas, S. (2018). Cell type-specific interchromosomal interactions as a mechanism for transcriptional diversity. bioRxiv 287532.

Huang, J., Liu, X., Li, D., Shao, Z., Cao, H., Zhang, Y., Trompouki, E., Bowman, T.V., Zon, L.I., Yuan, G.-C., et al. (2016). Dynamic Control of Enhancer Repertoires Drives Lineage and Stage-Specific Transcription during Hematopoiesis. Developmental Cell *36*, 9–23.

Jin, F., Li, Y., Dixon, J.R., Selvaraj, S., Ye, Z., Lee, A.Y., Yen, C.-A., Schmitt, A.D., Espinoza, C.A., and Ren, B. (2013). A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature *503*, 290–294.

Kamachi, Y., and Kondoh, H. (2013). Sox proteins: regulators of cell fate specification and differentiation. Development *140*, 4129–4144.

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., and Haussler, D. (2002). The human genome browser at UCSC. Genome Research *12*, 996–1006.

Kervrann, C., and Boulanger, J. (2006). Optimal spatial adaptation for patch-based image denoising. IEEE Trans Image Process *15*, 2866–2878.

Kieffer-Kwon, K.-R., Tang, Z., Mathe, E., Qian, J., Sung, M.-H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., et al. (2013). Interactome Maps of Mouse Gene Regulatory Domains Reveal Basic Principles of Transcriptional Regulation. Cell *155*, 1507–1520.

Kim, T.-K., Hemberg, M., and Gray, J.M. (2015). Enhancer RNAs: A Class of Long Noncoding RNAs Synthesized at Enhancers: Figure 1. Cold Spring Harb Perspect Biol *7*, a018622–a018623.

Krijger, P.H.L., and de Laat, W. (2016). Regulation of disease-associated gene expression in the 3D genome. Nat Rev Mol Cell Biol *17*, 771–782.

Lacoste, A., Berenshteyn, F., and Brivanlou, A.H. (2009). An efficient and reversible transposable system for gene delivery and lineage-specific differentiation in human embryonic stem cells. Cell Stem Cell *5*, 332–342.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.

Lefebvre, V., Dumitriu, B., Penzo-Méndez, A., Han, Y., and Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int J Biochem Cell Biol *39*, 2195–2214.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive Promoter-Centered Chromatin Interactions Provide a Topological Basis for Transcription Regulation. Cell *148*, 84–98.

Li, Y., Rivera, C.M., Ishii, H., Jin, F., Selvaraj, S., Lee, A.Y., Dixon, J.R., and Ren, B. (2014). CRISPR Reveals a Distal Super-Enhancer Required for Sox2 Expression in Mouse Embryonic Stem Cells. PLoS ONE *9*, e114485.

Lieberman-Aiden, E., Van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science *326*, 289–293.

Lim, B., Heist, T., Levine, M., and Fukaya, T. (2018). Visualization of Transvection in Living Drosophila Embryos. Mol Cell *70*, 287–296.e6.

Lionnet, T., Czaplinski, K., Darzacq, X., Shav-Tal, Y., Wells, A.L., Chao, J.A., Park, H.Y., de Turris, V., Lopez-Jones, M., and Singer, R.H. (2011). A transgenic mouse for in vivo detection of endogenous labeled mRNA. Nature Methods *8*, 165–170.

Long, H.K., Prescott, S.L., and Wysocka, J. (2016). Ever-Changing Landscapes: Transcriptional Enhancers in Development and Evolution. Cell *167*, 1170–1187.

Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective Inhibition of Tumor Oncogenes by Disruption of Super-Enhancers. Cell *153*, 320–334.

Lucas, J.S., Zhang, Y., Dudko, O.K., and Murre, C. (2014). 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. Cell *158*, 339–352.

Lupiáñez, D.G., Kraft, K., Heinrich, V., Krawitz, P., Brancati, F., Klopocki, E., Horn, D., Kayserili, H., Opitz, J.M., Laxova, R., et al. (2015). Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell *161*, 1012–1025.

Markenscoff-Papadimitriou, E., Allen, W.E., Colquitt, B.M., Goh, T., Murphy, K.K., Monahan, K., Mosley, C.P., Ahituv, N., and Lomvardas, S. (2014). Enhancer interaction networks as a means for singular olfactory receptor expression. Cell *159*, 543–557.

Marshall, W.F., Straight, A., Marko, J.F., Swedlow, J., Dernburg, A., Belmont, A., Murray, A.W., Agard, D.A., and Sedat, J.W. (1997). Interphase chromosomes undergo constrained diffusional motion in living cells. Curr Biol *7*, 930–939.

Martin, R.M., Rino, J., Carvalho, C., Kirchhausen, T., and Carmo-Fonseca, M. (2013). Live-Cell Visualization of Pre-mRNA Splicing with Single-Molecule Sensitivity. CellReports *4*, 1144–1155.

Masui, O., Bonnet, I., Le Baccon, P., Brito, I., Pollex, T., Murphy, N., Hupé, P., Barillot, E., Belmont, A.S., and Heard, E. (2011). Live-Cell Chromosome Dynamics and Outcome of X Chromosome Pairing Events during ES Cell Differentiation. Cell *145*, 447–458.

Michaelis, C., Ciosk, R., and Nasmyth, K. (1997). Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. Cell *91*, 35–45.

Morgan, S.L., Mariano, N.C., Bermudez, A., Arruda, N.L., Wu, F., Luo, Y., Shankar, G., Jia, L., Chen, H., Hu, J.-F., et al. (2017). Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. Nat Commun *8*, 1–9.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Mullick, A., Xu, Y., Warren, R., Koutroumanis, M., Guilbault, C., Broussau, S., Malenfant, F., Bourget, L., Lamoureux, L., Lo, R., et al. (2006). The cumate gene-switch: a system for regulated expression in mammalian cells. BMC Biotechnol *6*, 43.

Mumbach, M.R., Rubin, A.J., Flynn, R.A., Dai, C., Khavari, P.A., Greenleaf, W.J., and Chang, H.Y. (2016). HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nature Methods *13*, 919–922.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzoni, E.O., and Reinberg, D. (2015). CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science *347*, 1017–1021.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930–933.e22.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., Van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Ochiai, H., Sugawara, T., Sakuma, T., and Yamamoto, T. (2014). Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells. Sci. Rep. *4*, 7125.

Okamoto, R., Uchikawa, M., and Kondoh, H. (2014). Sixteen additional enhancers associated with the chicken Sox2locus outside the central 50-kb region. Dev Growth Differ *57*, 24–39.

Pevny, L.H., and Nicolis, S.K. (2010). Sox2 roles in neural stem cells. Int J Biochem Cell Biol *42*, 421–424.

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.-T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural Protein Subclasses Shape 3D Organization of Genomesduring Lineage Commitment. Cell *153*, 1281–1295.

Quinodoz, S.A., Ollikainen, N., Tabak, B., Palla, A., Schmidt, J.M., Detmar, E., Lai, M.M., Shishkin, A.A., Bhat, P., Takei, Y., et al. (2018). Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell *174*, 744–757.e24.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2010). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, –283.

Rao, S.S.P., Huang, S.-C., Hilaire, B.G.S., Engreitz, J.M., Perez, E.M., Kieffer-Kwon, K.-R., Sanborn, A.L., Johnstone, S.E., Bascom, G.D., Bochkov, I.D., et al. (2017). Cohesin Loss Eliminates All Loop Domains. Cell *171*, 305–309.e324.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., et al. (2014). A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. Cell *159*, 1665–1680.

Raymond, C.S., and Soriano, P. (2007). High-efficiency FLP and PhiC31 site-specific recombination in mammalian cells. PLoS ONE *2*, e162.

Robinett, C.C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A., and Belmont, A.S. (1996). In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. J Cell Biol *135*, 1685–1700.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Roukos, V., Voss, T.C., Schmidt, C.K., Lee, S., Wangsa, D., and Misteli, T. (2013). Spatial Dynamics of Chromosome Translocations in Living Cells. Science *341*, 660–664.

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., et al. (2018). Coactivator condensation at super-enhancers links phase separation and gene control. Science *361*, eaar3958.

Sanyal, A., Lajoie, B.R., Jain, G., and Dekker, J. (2012). The long-range interaction landscape of gene promoters. Nature *489*, 109–113.

Sarkar, A., and Hochedlinger, K. (2013). The Sox Family of Transcription Factors: Versatile Regulators of Stem and Progenitor Cell Fate. Stem Cell *12*, 15–30.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., et al. (2012). Fiji: an open-source platform for biological-image analysis. Nature Methods *9*, 676–682.

Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nature Methods *9*, 671–675.

Schwarzer, W., and Spitz, F. (2014). The architecture of gene expression: integrating dispersed cis-regulatory modules into coherent regulatory domains. Current Opinion in Genetics & Development *27*, 74–82.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. Cell *148*, 458–472.

Sigal, A., Milo, R., Cohen, A., Geva-Zatorsky, N., Klein, Y., Liron, Y., Rosenfeld, N., Danon, T., Perzov, N., and Alon, U. (2006). Variability and memory of protein levels in human cells. Nature *444*, 643–646.

The ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *488*, 57–74.

Thurman, R.E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M.T., Haugen, E., Sheffield, N.C., Stergachis, A.B., Wang, H., Vernot, B., et al. (2012). The accessible chromatin landscape of the human genome. Nature *489*, 75–82.

Thyagarajan, B., Olivares, E.C., Hollis, R.P., Ginsburg, D.S., and Calos, M.P. (2001). Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. Molecular and Cellular Biology *21*, 3926–3934.

Tinevez, J.-Y., Perry, N., Schindelin, J., Hoopes, G.M., Reynolds, G.D., Laplantine, E., Bednarek, S.Y., Shorte, S.L., and Eliceiri, K.W. (2017). TrackMate: An open and extensible platform for single-particle tracking. Methods *115*, 80–90.

Tomioka, M., Nishimoto, M., Miyagi, S., Katayanagi, T., Fukui, N., Niwa, H., Muramatsu, M., and Okuda, A. (2002). Identification of Sox-2 regulatory region which is under the control of Oct-3/4-Sox-2 complex. Nucleic Acids Research *30*, 3202–3213.

Uchikawa, M., Ishida, Y., Takemoto, T., Kamachi, Y., and Kondoh, H. (2003). Functional Analysis of Chicken Sox2 Enhancers Highlights an Array of Diverse Regulatory Elements that Are Conserved in Mammals. Dev Cell *4*, 509–519.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Vierstra, J., Rynes, E., Sandstrom, R., Zhang, M., Canfield, T., Hansen, R.S., Stehling-Sun, S., Sabo, P.J., Byron, R., Humbert, R., et al. (2014). Mouse regulatory DNA landscapes reveal global principles of cis-regulatory evolution. Science *346*, 1007–1012.

Wamstad, J.A., Alexander, J.M., Truty, R.M., Shrikumar, A., Li, F., Eilertson, K.E., Ding, H., Wylie, J.N., Pico, A.R., Capra, J.A., et al. (2012). Dynamic and Coordinated Epigenetic Regulation of Developmental Transitions in the Cardiac Lineage. Cell *151*, 1–29.

Weina, K., and Utikal, J. (2014). SOX2 and cancer: current research and its implications in the clinic. Clin Transl Med *3*, 1–10.

Weintraub, A.S., Li, C.H., Zamudio, A.V., Sigova, A.A., Hannett, N.M., Day, D.S., Abraham, B.J., Cohen, M.A., Nabet, B., Buckley, D.L., et al. (2017). YY1 Is a Structural Regulator of Enhancer-Promoter Loops. Cell *171*, 1573–1588.e28.

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.

Wuebben, E.L., and Rizzino, A. (2017). The dark side of SOX2: cancer - a comprehensive overview. Oncotarget *8*, 44917–44943.

Xu, Z., Thomas, L., Davies, B., Chalmers, R., Smith, M., and Brown, W. (2013). Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. BMC Biotechnol *13*, 87.

Young, R.A. (2011). Control of the Embryonic Stem Cell State. Cell *144*, 940–954.

Zappone, M.V., Galli, R., Catena, R., Meani, N., De Biasi, S., Mattei, E., Tiveron, C., Vescovi, A.L., Lovell-Badge, R., Ottolenghi, S., et al. (2000). Sox2 regulatory sequences direct expression of a (beta)-geo transgene to telencephalic neural stem cells and precursors of the mouse embryo, revealing regionalization of gene expression in CNS stem cells. Development *127*, 2367–2382.

Zhang, Y., Wong, C.-H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., et al. (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature *504*, 306–310.

Zhou, H.Y., Katsman, Y., Dhaliwal, N.K., Davidson, S., Macpherson, N.N., Sakthidevi, M., Collura, F., and Mitchell, J.A. (2014). A Sox2 distal enhancer cluster regulates embryonic stem cell differentiation potential. Genes Dev *28*, 2699–2711.

Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. Cell *152*, 642–654.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 1--figure supplement 1

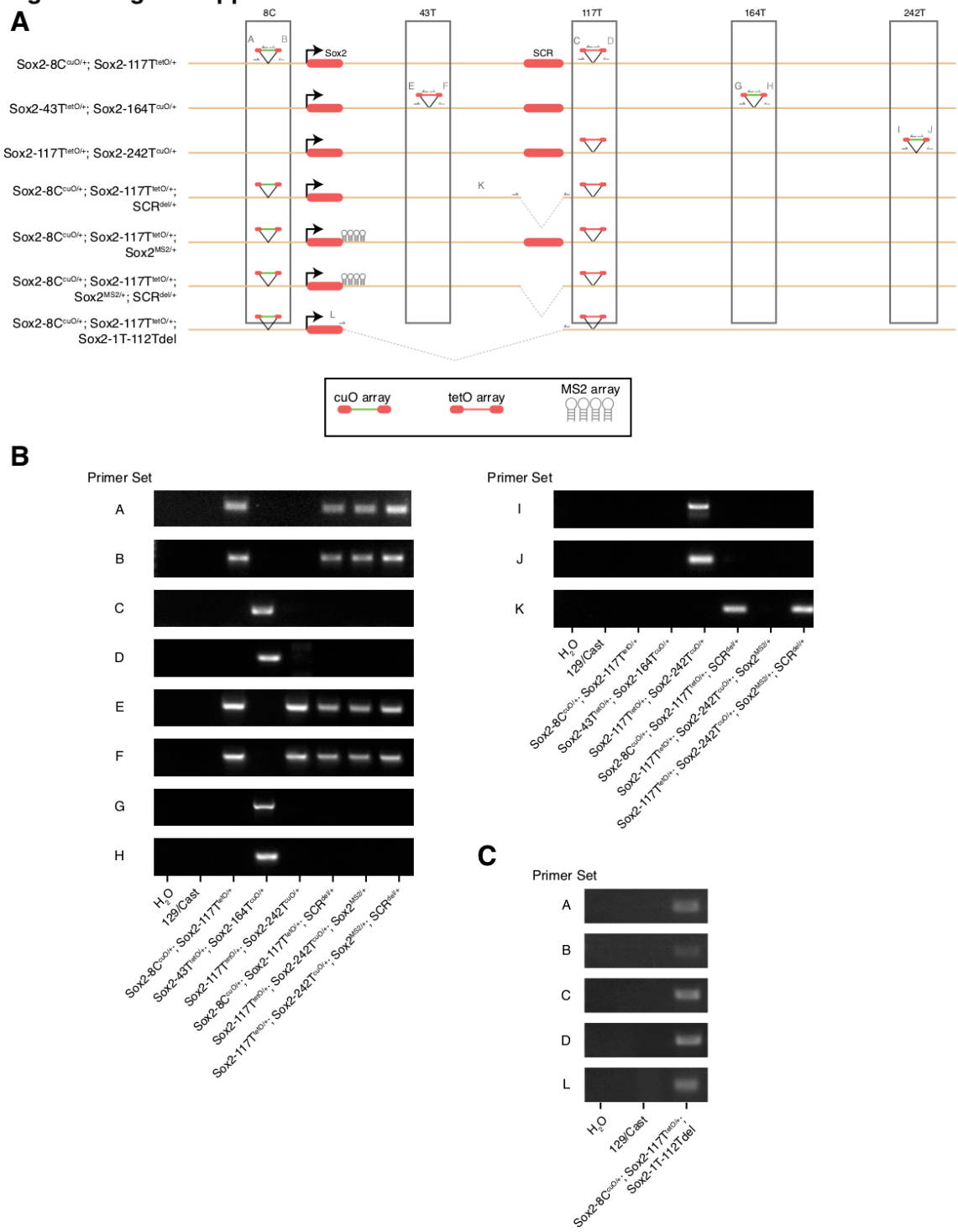

Figure 1—figure supplement 1. Characterization of Modified Embryonic Stem Cell Lines. A) Schematic of modified cell lines used in this study. Primer sets used to amplify recombination arms for tetO- and cuO- integration are shown. **B,C)** PCR genotyping of ESC lines shown in A.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 1--figure supplement 2

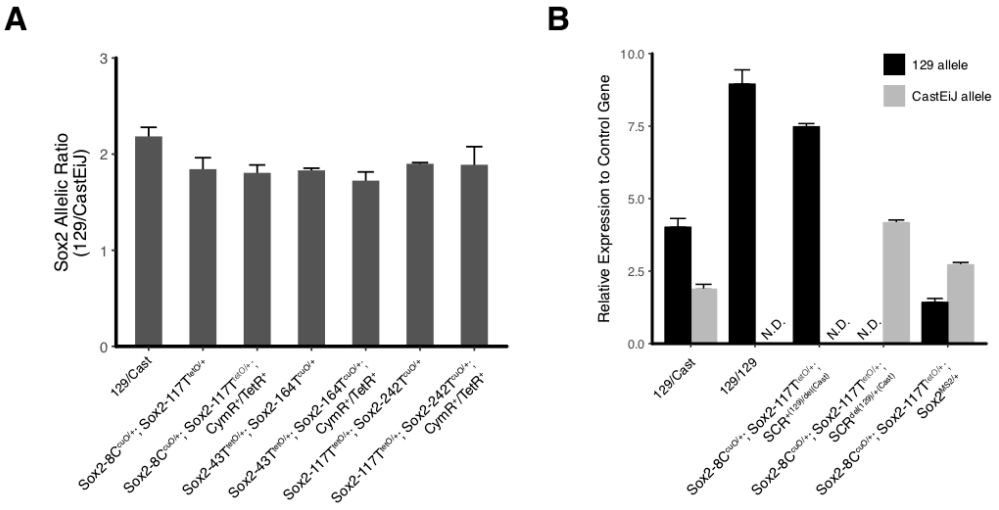

Figure 1—figure supplement 2. *Sox2* Expression Characterization for Modified Embryonic Stem Cell Lines. A) Ratio of *Sox2* expression from the 129 allele and the CastEiJ allele measured by qPCR for modified ESC lines. B) *Sox2* expression relative to control gene (Tbp) for various cell lines. E14 mESCs, which lack the CastEiJ allele, are included to demonstrate specificity of allele-specific qPCR assay. Deletion of SCR region leads to loss of expression from the *Sox2* allele *in cis*. N.D. is not detected.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 2--figure supplement 1

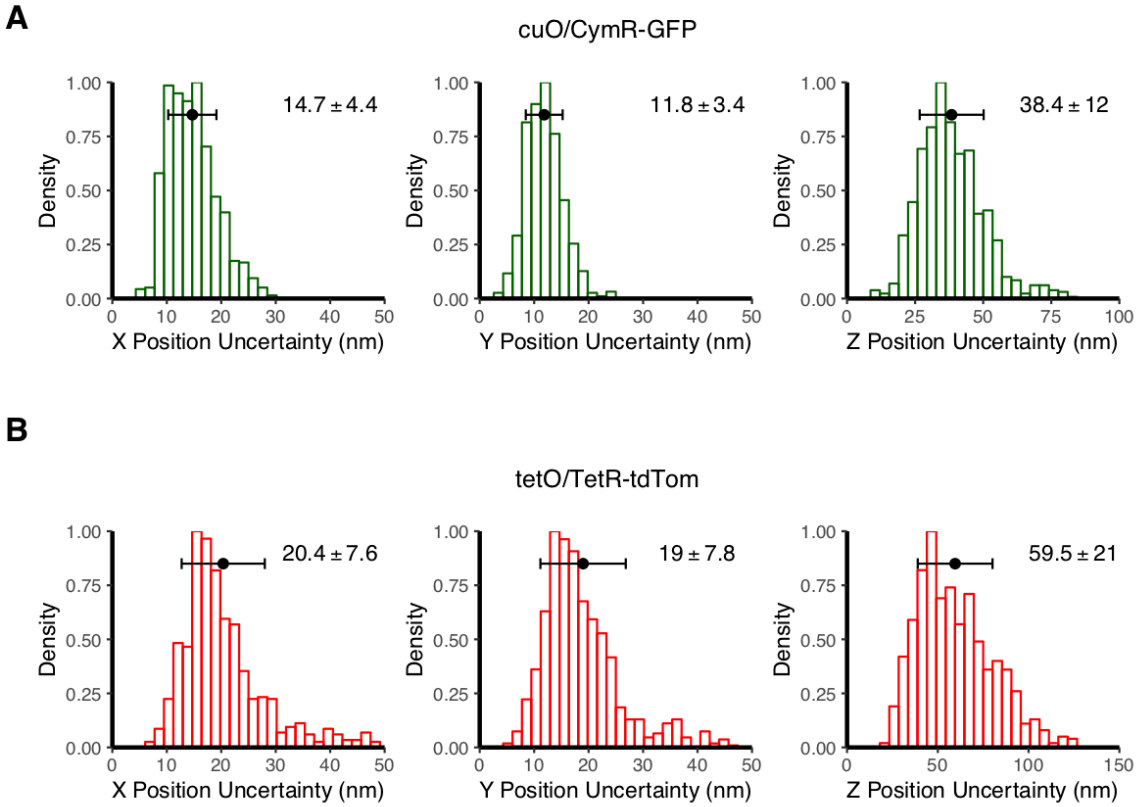

Figure 2—figure supplement 1. Estimate Localization Precision for cuO and tetO. A-B) Histograms of X, Y, and Z position uncertainty for fluorescent beads with signal-to-noise ratios comparable to cuO/CymR-GFP (A) or tetO/TetR-tdTom (B). Data plotted are the standard deviation values measured using 10 frame sliding windows collected from 9-10 beads. Error bars show mean and standard deviation of the computed position uncertainties, which are reported in the upper right of each panel.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 2--figure supplement 2

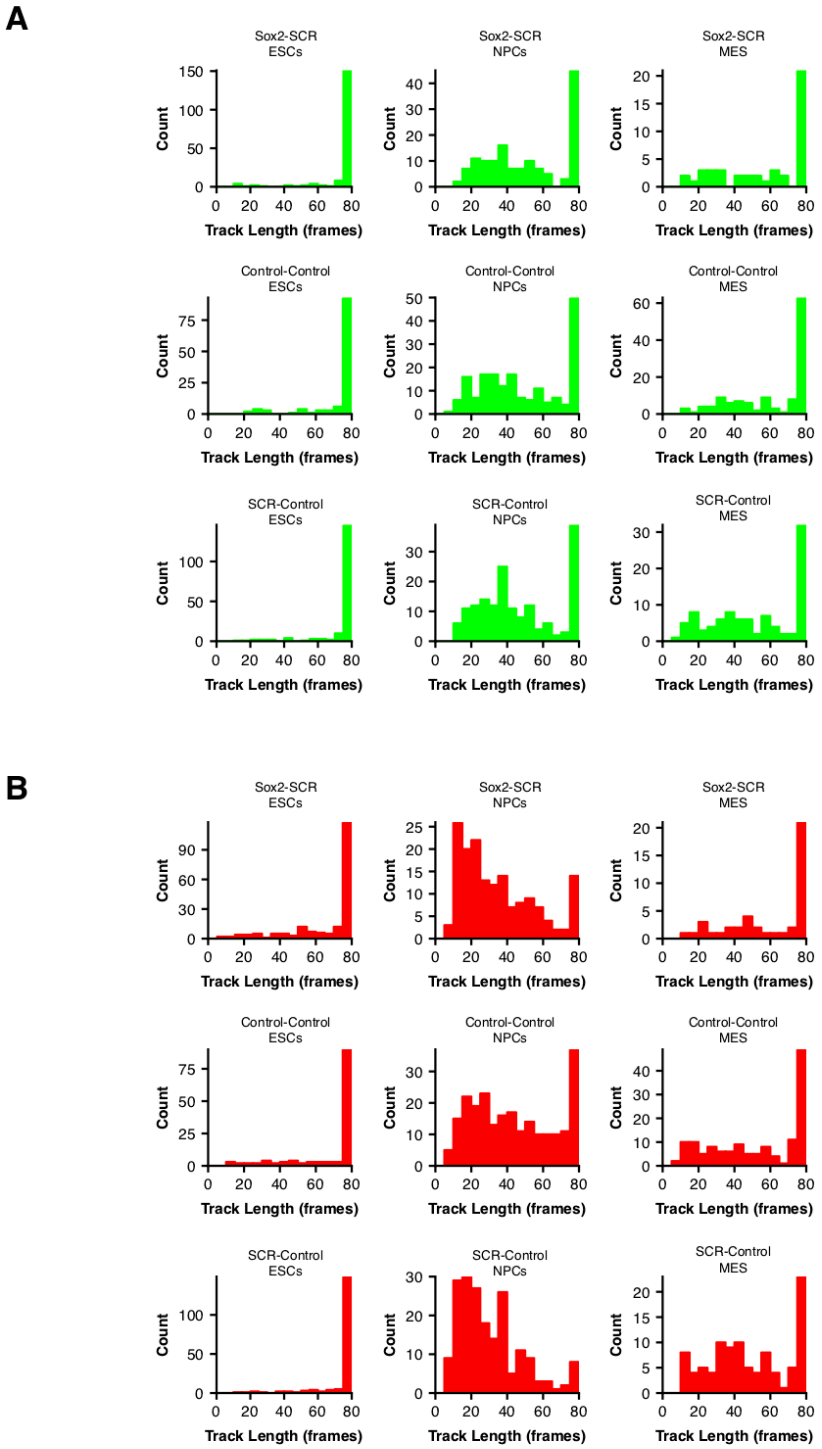

Figure 2—figure supplement 2. Tracking Lengths for tetO and cuO Spots Across Cell Lines. A-B) Histograms of the cuO-array (A) or tetO-array (B) Track lengths for cell lines used in the study as ESCs, NPCs, and MES. Tracking lengths were often shorter in NPCs or MES due to increased nuclear movement in these cell types compared to ESCs.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 3--figure supplement 1

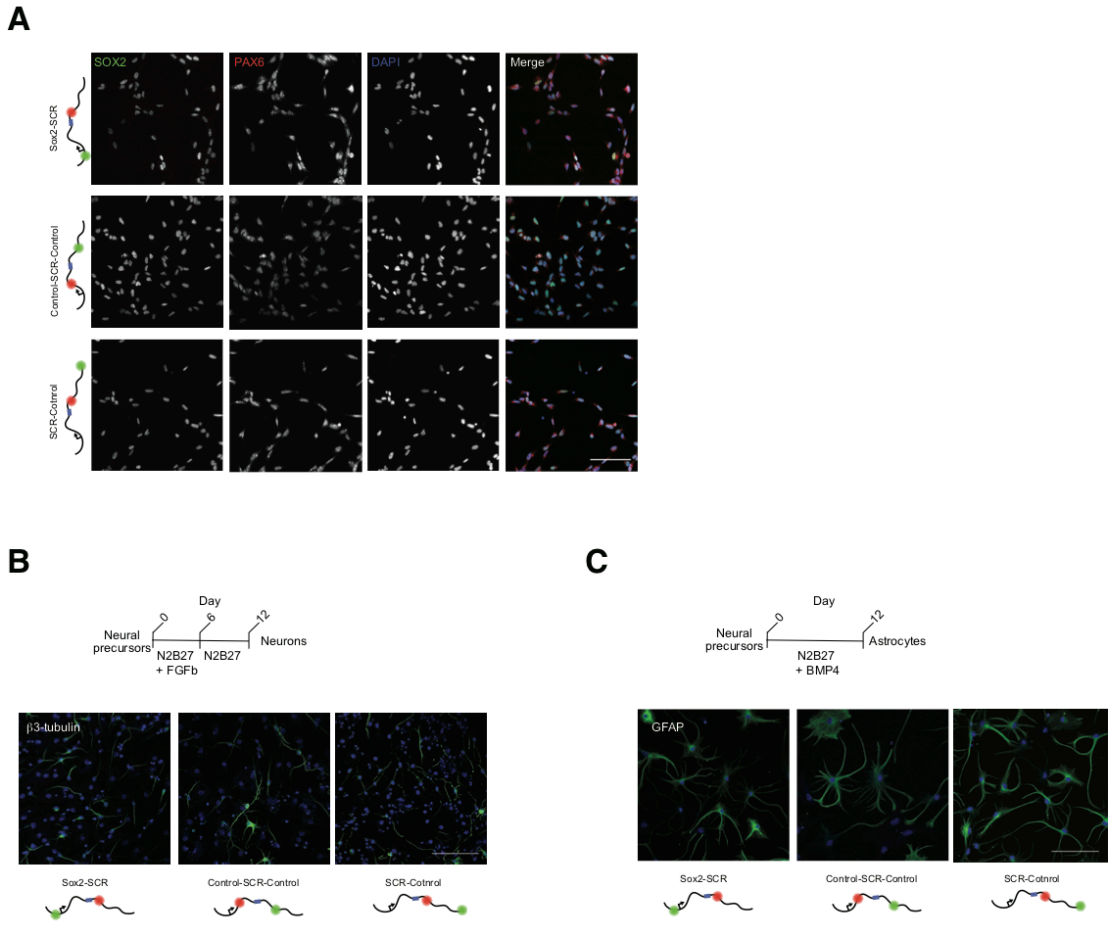

Figure 3—figure supplement 1. Characterization of ESC-derived Neural Progenitor Cell Lines. A) Immunofluorescence of fixed neural progenitor cells (NPCs) for the NPC markers SOX2 and PAX6. B-C) Immunofluorescence for the neuron marker β3-tubulin (B) or the astrocyte marker GFAP (C) on fixed cultures after 12 days of differentiation towards neurons or astrocytes, respectively. Scale bar is 100um.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 3--figure supplement 2

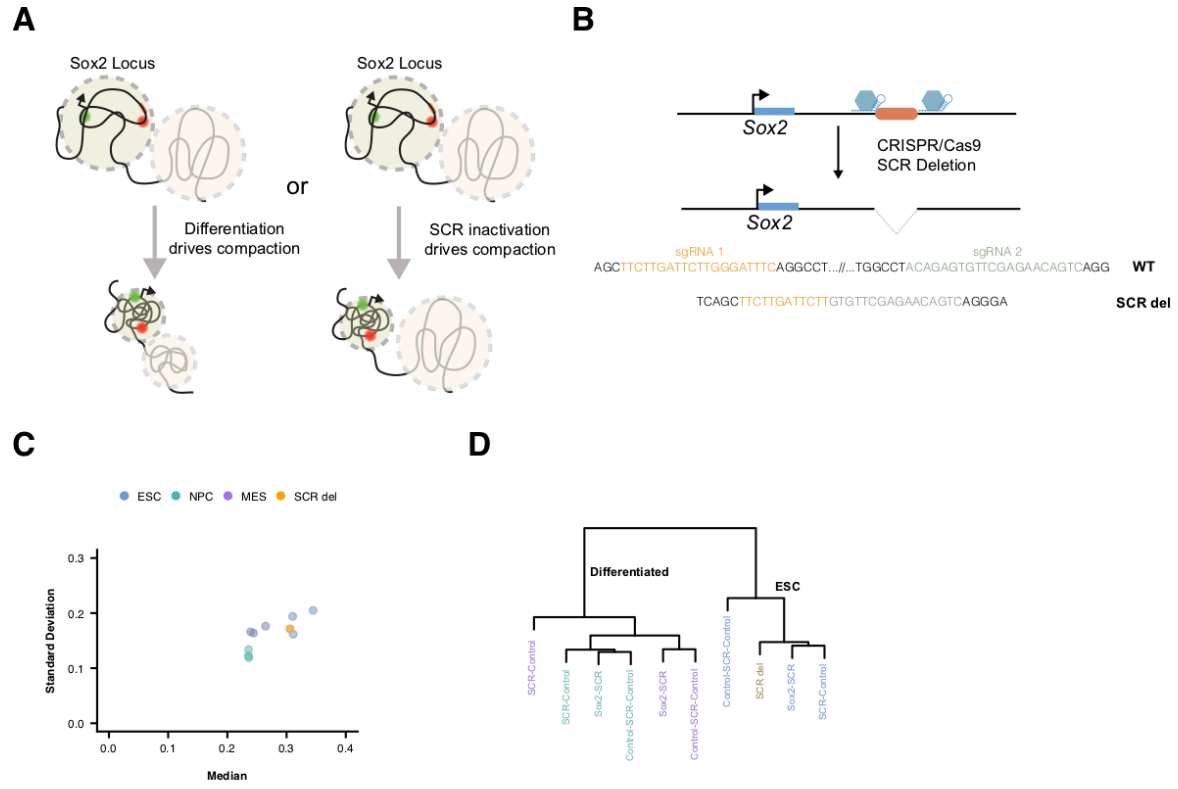

Figure 3—figure supplement 2. SCR Inactivation Does Not Drive Locus Compaction Upon Differentiation. A) Potential models for *Sox2* locus compaction observed upon differentiation to NPCs or MES. At left, cellular differentiation may lead to global changes in chromatin structure that are not dependent of *Sox2*/SCR activation status. Alternatively, *Sox2* and SCR inactivation could lead to changes to chromatin structure within the Sox2 locus, driving locus-specific compaction. **B)** Strategy for CRISPR/Cas9-mediated SCR deletion. Two gRNAs were designed to flank the SCR region and generate a large deletion of SCR. Below, the SCR deletion allele shows a novel junction near the locations of expected Cas9 cutting, indicating a loss of the intervening SCR sequence. **C)** Scatterplot of mean and standard deviation of 3D distance measurements for each cell line visualizes similarity between *Sox2* label pairs across cell types. **D)** Dendrogram visualizing hierarchical clustering of Earth Mover's distances between 3D separation distance histograms of distinct Sox2 label pairs across cell types. SCR-deleted ESCs show greatest similarity to other ESCs as opposed to differentiated cells with inactivation of the SCR element.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 4--figure supplement 1

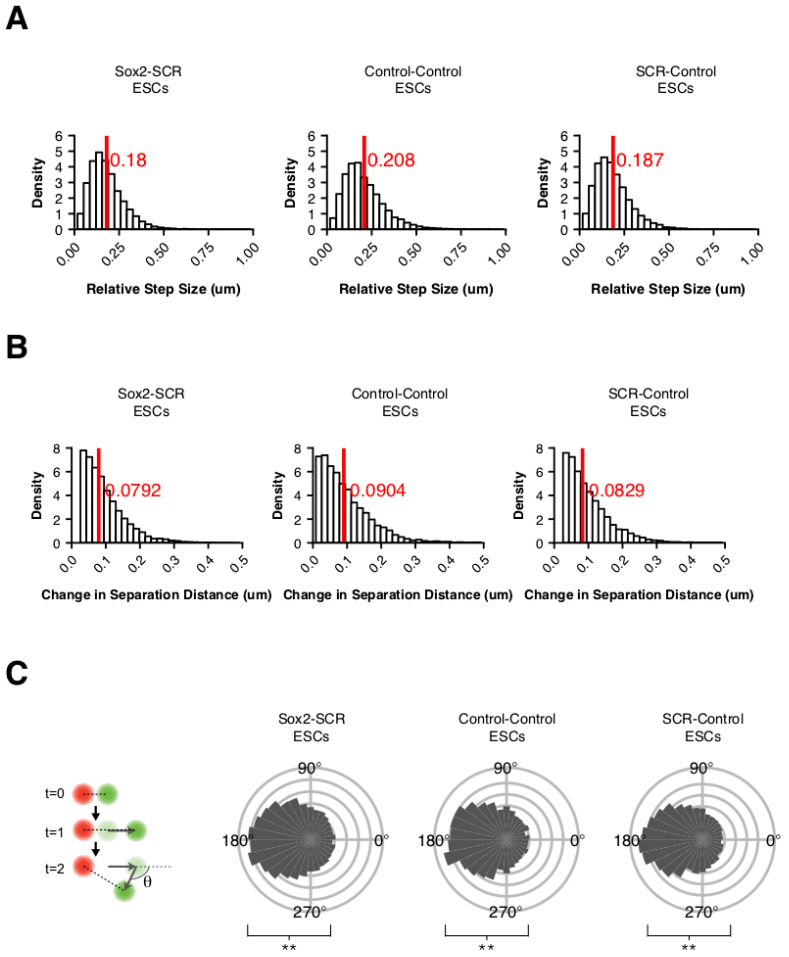

Figure 4—figure supplement 1. Dynamics Statistics for Each Sox2 Locus Pair in ESCs. A-B) Normalized histograms of relative step size and change in 3D separation distance for adjacent frames. Mean value is highlighted by a red line. C) Measurement of the XY plane angle made between two successive displacement vectors demonstrates pronounced bias towards a locus "bouncing back" to its position preceding a given step. Statistical test for bias was performed using Kuiper's Test for Uniformity. ** is $p < 0.01$.

bioRxiv preprint doi: https://doi.org/10.1101/409672; this version posted September 5, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 5--figure supplement 1

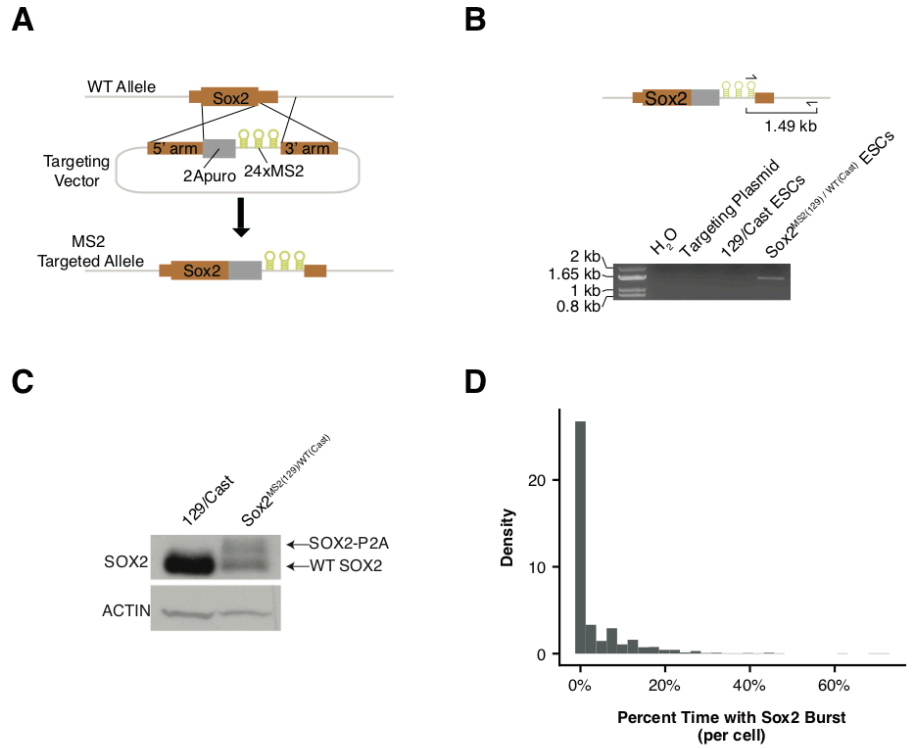

Figure 5—figure supplement 1. Generation and Characterization of Sox2-MS2 Transcriptional Reporter ESCs. A) Targeting strategy for *Sox2* transcriptional reporter. A targeting plasmid was used with *Sox2* homology arms and a P2A peptide puromycin resistance gene cassette (2Apuro) inserted in frame with *Sox2*. Downstream of 2A puro is a 24x MS2 stem loop array, which is inserted into the 3' UTR. **B)** PCR genotyping assay to identify a targeted *Sox2* allele. A primer set was used that recognized the MS2 stem loop array and a genomic region downstream of the 3' homology arm. **C)** Western blotting for SOX2 protein in parental 129/CastEiJ ESCs or ESCs heterozygous for the Sox2-MS2 allele. Actin was used as a loading control. **D)** Normalized histogram of the percentage of time individual cells have a detectable *Sox2* transcriptional burst.