Anti-phase clustering of regulatory factors shapes gene bursting

Bitong Li[1,2], Yew Yan Wong[1,3], Neftali Flores-Rodriguez[4], Tara Davidson[1], Matthew S Graus[1,3], Valeriia Smialkovska[5], Hiroaki Ohishi[6], Angelika Feldmann[5], Hiroshi Ochiai [6]† and Mathias Francois[1,3,7]†

[1] Centenary Institute, The University of Sydney, Royal Prince Alfred Hospital, Missenden Rd, Sydney, New South Wales, 2006, Australia.

[2] Department of Molecular Medicine, School of Biomedical Sciences, The University of New South Wales, Sydney, New South Wales, Kensington, 2052, Australia.

[3] The Genome Imaging Centre, Centenary Institute, The University of Sydney, Royal Prince Alfred Hospital, Missenden Rd, Sydney, New South Wales, 2006, Australia.

[4] Sydney Microscopy and Microanalysis, The University of Sydney, Sydney, New South Wales, 2006, Australia.

[5] German Cancer Research Center (DKFZ), Im Neuenheimer Feld 581, 69120 Heidelberg, Germany.

[6] Division of Gene Expression Dynamics, Medical Institute of Bioregulation, Kyushu University, Fukuoka, Japan.

[7] School of Medical Sciences, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, 2006, Australia.

†Correspondence should be address to

Hiroshi Ochiai (Tel: +81-92-642-6882)
ochiai.hiroshi.403@m.kyushu-u.ac.jp

Mathias Francois (Tel: +61 28627 9497)
m.francois@centenary.org.au

One-sentence summary: Different types of SOX2 clustering activities modulate *Nanog* transcription burst in pluripotent stem cells.

Keywords: transcription burst, SOX2 transcription factor, live imaging, single molecule, stem cell, enhancer RNA, gene locus, clustering dynamics

Abstract

The ability of stem cells to divide and self-renew depends on a complex choreography of molecular events that maintain the transcriptional oscillation of pluripotency genes. Only a handful of transcription factors (TFs) are necessary to preserve pluripotency and reprogram differentiated cells into stem cells. Paradoxically, while the protein players are known, the challenge remains to decipher the series of steps that TFs undertake to modulate "on" and "off" fluctuations of gene transcription. Here, we use single-molecule tracking combined with the STREAMING-tag transcriptional reporter systems to reveal temporal clustering patterns of endogenous SOX2 occupancy at the *Nanog* locus in relation to its nascent mRNA synthesis in live embryonic stem cells. These patterns distinctively outline multifaceted regulatory behaviours of SOX2 associated with various stages of the *Nanog* transcription cycle. This study exposes that SOX2 clustering activity is out-of-phase with regulatory factors that engage with transcription burst at the *Nanog* gene locus.

Main text

The exquisite regulation of gene expression in space and time dictates cell fate decisions and lineage specification at the molecular level. Two components are essential to this process: 1) the communication between genome regulatory regions, including promoters and enhancers; 2) a molecular machinery able to coordinate the activity of the genome control panel. The latter include transcription factors (TFs) that regulate messenger RNA (mRNA) synthesis in *trans* via binding to DNA *cis*-regulatory elements.

Recent progress in imaging and single-cell analysis technologies has revealed that transcription, production of RNA molecules from a corresponding gene, occurs in a discontinuous and dynamic manner. Nascent mRNA synthesis oscillates between sustained production in an "on" state and minimal synthesis in an "off" state (*1*). The timescale of this on-and-off cycle varies between genes but typically ranges from several minutes to hours (*2*). While transcriptional activity dictates the levels of gene mRNA synthesis, the relationship between TFs dynamics and transcription burst kinetics is not yet fully understood.

TFs activity is highly dependent on chromatin region accessibility, DNA affinity and the recruitment of cell-type specific co-factors (*3–5*). In addition, multiple biophysical constraints shape TFs mobility and fine-tune their behaviour. TFs move through a complex chromatin landscape, which displays a sophisticated higher-order organisation. Importantly, TFs navigate a crowded environment while eliciting interactions with different classes of molecules, such as proteins, RNA, and DNA. Recent advances in high-resolution imaging, such as single-molecule tracking (SMT) (*6–9*), have revealed key biophysical features, such as a diffusion rate or confined mobility, that drive TF-specific activity. These quantifiable behaviours have determined TF search pattern mechanism, offering new insights into TF biology, on how a TF select its target gene. For instance, a landmark study has demonstrated, that SOX2, which is essential for stem cell pluripotency, searches for a specific target through the

combination of diffusion and sliding (*10*). Different diffusion rates of SOX2 were shown to modulate asymmetric gene expression programs during early embryo formation, limiting stem cell fates to either extra-embryonic annexes or the embryo proper (*11*). Despite a clear role for TF biophysical parameters in controlling cell fate decision, there is still a gap in our understanding of how these behavioural features directly contribute to the modulation of gene expression levels.

To unravel the mechanisms that underpin the dynamics of transcriptional activation, it is critical to be able to determine the TF/DNA binding kinetics requirement and establish a functional relationship of this characteristic with the nascent mRNA synthesis rate. Despite important findings brought with SMT-based methods, they have several limitations. Firstly, it is difficult to biologically interpret the activity kinetics of different TF molecule sub-populations (e.g. different types of confined mobility). Furthermore, the resolution of SMT-based techniques is not sufficient to determine the precise chromosomal coordinates of TF interaction.

To overcome these limitations, complementary methods that capture genome locations have been used to assess the activity of a protein of interest in a specific region. Some studies based on locus-centric approaches have taken advantage of anchor-based genetic reporter systems to label specific genome locations (*12*). While this approach enables the visualisation of specific DNA regions and the analysis of protein behaviour through SMT, the information about these regions, such as whether gene transcription is on or off, remains obscure. As a result, it is difficult to discriminate between how a specific molecular behaviour is functionally related to a specific step in the transcription cycle.

Investigating the relationship between TF binding kinetics and the rate of transcriptional activity is complicated by significant differences in the timescales of these two biological processes. We addressed this challenge by using a multimodal imaging approach that enables us to explore the

coordinated activity of regulatory factors and nascent mRNA synthesis at a specific gene locus in live cells, or by measuring them correlatively in fixed cells.

In this study, we use the *Nanog* gene locus, well known to exhibit robust transcriptional bursts (*13*), as a proof-of-principle that it is possible to quantify its coordinated transcriptional activity along with its direct regulator, the SOX2 TF, in real-time in pluripotent stem cells. Importantly, SOX2 has been shown to directly regulate *Nanog* transcription to maintain pluripotency with both loss- and gain-of-function leading to impaired Nanog expression (*14–18*).

We have achieved this by combining STREAMING-tag and SMT-based imaging techniques. While STREAMING-tag reporter system allows for profiling of locus mobility and transcription bursts at allelic resolution via labelling with different fluorescence markers (*19*), SMT provides a readout of TF behaviour at the locus of interest. This work provides a time-based picture of multifaceted SOX2 behaviours linked to distinct phases of the *Nanog* gene locus transcription cycle.

Multi-modal microscopy to quantify SOX2 occupancy at the *Nanog* gene locus relative to its transcriptional on-state

To address the main limitation of converting biophysical observations into a relevant biological interpretation in the context of TF mobility and gene transcription, we opted for a paired measurement that combines two live imaging methods: SMT and the STREAMING-tag reporter system. To measure TF long dwell times on the chromatin (e.g. over 4 sec or more) also known as confined mobility in parallel with processive transcriptional stages, we used a multimodal imaging platform. The assessment of STREAMING-tag reporter activity via Spinning Disk Confocal (SDC) module allows us to measure nascent mRNA synthesis and locus mobility, while the slow tracking SMT using the HILO module displays TF chromatin occupancy dynamics at a single locus resolution (see Methods). This readout gives access to TF behaviours at a gene-specific locus relative to mRNA nascent transcription and allows quantifying TF binding dynamics and dwell time relative to the "on" and "off" states. This approach was established in mouse embryonic stem cells (mESCs) by introduction of the STREAMING-tag reporter cassette (MS2 and TetO repeats) at the *Nanog* locus with expression of reporter proteins (MCP-RFP and mTetR-GFP), and the SNAP-tag knocked-in into either the *Sox2* or *Brd4* locus (*19*) (fig. 1a).

Previous studies on fixed cells using super resolution-based approaches have shown that around 10 to 20 molecules of regulatory factors cluster on enhancers scaffolds that come into proximity of their target gene locus (*20*, *21*). In this study, we also opted for the term "clustering" to acknowledge the potential influence of protein oligomerisation and/or protein/RNA aggregates on the observed dynamics of TF occupancy, in addition to binding at specific sites.

To further validate the dynamics for this type of SOX2 behaviour in our reporter system, we performed a simultaneous two-colour SMT experiment with sparse labelling (JF549) in one channel and saturated labelling (JF646) in another channel in the same cells (supp fig. 1a-c). Fluctuations in

fluorescent signal intensity under saturated labelling conditions were used to construct masks to define regions of different SOX2 concentration, with high as the top 8%, following 30% as medium and the remaining signal as low (see Methods). The sparse SOX2 labelling signal was overlaid onto the mask to assess the dark fraction contribution to individual SMT-detected molecules (supp fig. 1b and c). This approach further confirm work on fixed cells and suggests that on average over 50% of the SMT-detected SOX2 molecules may accumulate in clusters of high or medium intensity. This indicates that imaged single SOX2 molecules are involved with foci made of at least two or more SOX2 molecules (supp fig. 1c and video supp 1).

Because TF chromatin occupancy dynamics occur on a shorter time scale (seconds) compared to transcription bursts (minutes), we chose a two-step measurement method that continuously monitors SMT interspersed with a regular assessment of transcriptional activity at the *Nanog* locus. To accomplish this, we utilised a dual image acquisition technique that discriminates between SOX2 chromatin binding dynamics and *Nanog* mRNA elongation kinetics (3 minutes) (fig. 1b, image acquisition mode 1). This imaging approach relies on continuously measuring by SMT for 30 frames (exposure time 500 ms, no interval time), alternating with 1 frame that records both GFP and RFP signals over a 3-minute period. Fluorescent beads were used to evaluate the changes in light path occurring between shifts from SDC to HILO modalities. Post-image acquisition, we applied a custom script to correct for image misalignment (see Methods) (supp fig. 2a-f). The STREAMING-tag system has been designed to denote transcription at a gene locus. This is done through tagging a specific gene locus with a GFP signal and actively transcribed mRNA with RFP. With this reporter system, the presence of a GFP-only signal denotes that the transcription of the *Nanog* locus is "off". Co-localisation of the GFP and RFP spots within a 440 nm region of interest (ROI, 4-pixel radius) determines that nascent *Nanog* transcription is occurring, defining an “on” state. Of importance, the STREAMING-tag cassette is knocked at around 300 bp downstream from the transcription start site

(TSS), while transcriptional pausing has been shown to occur about 50-60 bp downstream of the TSS (*22–24*). Therefore, with the STREAMING-tag reporter system, only the productive transcription after pause-release is observed (supp fig. 2g). This means that the "off" state, as we define it in this study, also includes the early steps of transcriptional initiation up to the release stage. The "on" state enables the detection of nascent mRNA produced during transcriptional elongation. When the RFP signal is lost, this means that splicing has occurred, or termination has been completed.

To measure the SOX2 occupancy dynamics in a locus-specific manner while *Nanog* is in an "on" state (fig. 1c), we used an ROI of 4-pixel radius, (440 nm) that centres on the GFP spot when both GFP and RFP signals are detected (fig. 1d and e). Such a distance was defined based on the previous observation that elevated concentration of regulatory factors around the *Nanog* locus ranges between 200 and 400 nm (*19*, *21*). Unsurprisingly, we find multiple locations with SOX2 enrichments all throughout the nucleus (fig. 1d). Interestingly, within 440 nm of the GFP detection, we systematically observe an enrichment of SOX2 events (fig. 1e), suggesting that clustering activity is temporally correlated with the *Nanog* locus activity.

Distinct types of SOX2 occupancy dynamics at the *Nanog* gene locus during on and off-state

Here, we measured the SOX2 visiting frequency at the *Nanog* locus to determine its clustering dynamics profile while the transcription is "on". It revealed distinct modes of occupancy that occur at different rates throughout transcription elongation (fig. 2a and b and video supp 2). The first mode indicates no to low clustering (fig. 2a), while the second shows persistent locus occupancy of moderate intensity (fig. 2b).

It is known that the SOX2 protein binds to enhancers at the *Nanog* locus to induce transcriptional transactivation (*25*, *26*); however, the timing of this event in relation to transcription cycle remains elusive. Hence, to understand how SOX2 clustering dynamics correlates with the timing of

transcription burst, we next set out to evaluate the occupancy rate of this TF during the "off" state period.

The assessment of SOX2 visiting frequency while the locus is not actively transcribing RNA uncovers a third type of occupancy behaviour characterised by a high rate of visiting events over a short number of frames (fig. 2c-e and video supp 3, video supp 4 and video supp 5). Quantitation of the different types of temporal clustering kinetics (fig. 2f) revealed that low to medium visiting frequency are the most prevalent at both on and off state, and that a discrete population of SOX2 molecules display high clustering activity only at off state. This may indirectly reflect a specific oligomeric status and recruitment of protein partners as well as RNA products likely to influence this TF high-frequency clustering activity (*20*, *27–30*).

To show that the different SOX2 clustering activities are not restricted to the *Nanog* locus only, we took advantage of the dual SMT dataset described above to analyse the presence of different types of visiting frequency relative to the fluctuation of SOX2 local concentration across the nucleus. This analysis has shown that all three types of clustering activity (low, medium and high) are broadly detected across the nucleus in regions that show elevated SOX2 concentrations (supp fig. 3a). Quantitation of these three types of visiting frequency revealed that, the level of SOX2 high temporal clustering activity is a broad mechanism used on a genome-wide scale most likely across many gene loci (supp fig. 3b).

As we have shown above, SOX2 occupies the *Nanog* gene locus more frequently when transcription is in "off" state. To further explore the relationship between SOX2 and RNA Polymerase II, we quantified the rate of co-localisation between the transcriptionally active form of RNA Pol II (phospho-Ser5 RNA Pol II) and SOX2 clusters using dSTORM super-resolution microscopy. This approach indicated that on average SOX2 cluster population co-localises (cluster centre mass to centre mass, d < 50 nm) with active RNA Pol II foci at a low rate (5%, supp fig. 4a-c). Additional analysis determined that active RNA Pol II is mostly in presence of a vast majority of SOX2 clusters

that are the smallest one detected (< 1 μm) (supp fig. 4d). The findings from both types of imaging approaches revealed the spatiotemporal activity of SOX2 during transcription elongation. Mainly, SOX2 visiting frequency increases while elongation is off, which results in small SOX2 cluster sizes when active RNA Pol II is near.

Chromatin organisation and epigenetic modifications both play important roles in the activation of transcriptional initiation and are modulated in a cyclical, stepwise fashion. These periodic series of ordered and coordinated molecular events have been included in the theoretical modelling of transcription (*31–34*). Further, a mathematical model proposed a molecular framework for the transition between "off" and "on" states that considers transcriptional activator residence time relative to the burst frequency (*34*). Such a conceptual framework prompted us to assess SOX2 chromatin occupancy dynamics in relation to the progression through a full cycle of transcription.

High SOX2 clustering frequency is anti-phased with *Nanog* mRNA elongation

Based on the population average of SOX2 long dwell time in mESC (4-5 sec, supp fig. 5) (*11*, *35*), and the average time of transcription burst at the *Nanog* locus (3 min) (*13*), we designed a second mode of acquisition to image continuously through a full transcription cycle (fig. 3a, image acquisition mode 2). Briefly, this approach alternates confocal imaging of the *Nanog* locus for 3 seconds and SMT measurement for 500 ms with a repeat of this cycle over 9 minutes (150 cycles total, see Methods, video supp 6). Such a design allowed us to infer that a detected SOX2 molecule (long-dwell time ≥ 4 sec) prior to and after the *Nanog* locus imaging is the same molecule. To be conservative, in addition to the detection of long-lived fraction (2-frame detection, supp fig. 6a, schematic), we classified SOX2 short dwell time events that corresponded to the capture solely over one frame period (supp fig. 6b, schematic). This approach enables us to measure the presence of SOX2 clusters that persist near the *Nanog* locus over time.

To fully cover a complete event of transcription burst, we analysed SOX2 survival rates during a period that encompasses a step of transcriptional elongation and its transition to a refractory period, until the nascent RNA signal disappeared (fig. 3a). Analysis of SOX2 visiting frequencies with acquisition mode 2 over a 9-minute period while the *Nanog* locus is in off-state further confirmed the presence of the three types of occupancy rates (low, medium and high) reported above. The analysis of the number of SOX2 detected molecules and their relative survival rates at the *Nanog* locus revealed that, in addition to different types of visiting frequencies, SOX2 clusters display both short-lived and long-lived fractions (fig. 3b). From there it is possible to classify SOX2 clustering activity into three distinct types: type I - low to no clustering and short lived; type II - medium to high clustering and short lived; and type III - high clustering and long lived (fig. 3b and supp fig. 7).

To further understand the relationship between the temporality of SOX2 occupancy and mRNA synthesis, we next imaged the activity of a well-known elongation factor, the BRD4 protein. Here, we used a reporter cell line that harbours the STREAMING-tag reporter system combined with a SNAP-tag that has been knocked into the *Brd4* locus (*19*) and employed the imaging acquisition mode 2 described for SOX2. This approach revealed that BRD4 occupancy is mostly concurrent with nascent mRNA production (fig. 3c and supp fig. 8) (*19*). Importantly, this approach further confirmed that the long-lived and highly active SOX2 clustering dynamics (type III) is out-of-phase with the elongation step that involves BRD4 and RNA polymerase II activity (fig. 3d and supp fig. 9).

Since earlier studies showed that open chromatin regions drift relatively slowly compared to closed ones that diffuse faster (*19*, *36*), this prompted us to investigate further how SOX2 occupancy behaviours affect the *Nanog* locus mobility.

To refine the temporal hierarchy between the distinct types of SOX2 behaviours and transcriptional elongation, we measured changes in *Nanog* locus mobility in relation to the presence or absence of SOX2 occupancy. This quantitative approach revealed that the *Nanog* locus displays opposite diffusion kinetics that are correlated to different types of SOX2 residency profiles. During SOX2

occupancy type II, the *Nanog* locus harbours a slower diffusion rate while SOX2 is present compared to when there is no SOX2 detected (fig. 3e). Conversely, the locus diffusion kinetics increases in the presence of high SOX2 occupancy (type III) compared to when there is no SOX2 detected (fig. 3f). Therefore, by combining information gained on change in locus mobility with occupancy rate measurement, we posit that SOX2 type II behaviour is associated with locus remodelling and occurs prior to active nascent mRNA synthesis as opposed to type III likely to follow the elongation step and being associated with locus repackaging.

In addition to interacting with DNA, SOX2 has been shown to recruit RNA (*28*) and more broadly, TFs have been suggested to harbour an RNA-binding arm, which also influences their mobility (*37*). Since this study shows that high SOX2 occupancy activity with longer dwell time at the *Nanog* locus occurs in the absence of *Nanog* nascent transcripts, this prompted us to investigate further the potential role of alternative RNA species such as enhancer RNAs (eRNAs) from well-characterised SOX2-dependent regulatory elements (*26*, *28*).

Nanog eRNAs are mostly transcribed at "off" state

eRNAs function involves stabilising chromatin loops and recruiting co-factor such as p300 and acting as a sponge for the negative elongation factor (NELF) (*38*). Since their role in modulating transcription is a possible mechanism, this prompted us to specifically investigate the role of *Nanog* eRNAs involvement in the transcription cycle. The *Nanog* locus is one of the most well-characterised genomic regions in the context of embryonic stem cell pluripotency, with the presence of three super-enhancers that frame the gene body within a 100 kb region (*39*, *40*). These regulatory regions are located at -5 kb, -45 kb, and +60 kb from the transcription start site (TSS), respectively (*39*) and are actively transcribed (*41*). The functional hierarchy of these three enhancer regions in mESCs has been validated using genome editing with -5 kb being the most critical one, -45 kb of modest influence and

+60 kb with no impact on the expression (*40*). Using enhancer transcription as a proxy for TF activity at these regulatory regions, we interrogated the presence of eRNAs molecules in relation to *Nanog* nascent transcripts. For this, we employed an RNA-FISH method to map eRNAs and transcribed *Nanog* introns in parallel with the STREAMING-tag reporter as a readout of elongated *Nanog* expression. We devised an approach for quantifying eRNA/nascent mRNA expression in a high-throughput manner over thousands of cells (supp fig. 10).

Taking advantage of GRO-seq datasets (*39*), we designed probe sets that detect both sense and/or anti-sense eRNA strands for each of the three *Nanog* enhancers (fig. 4a). This analysis revealed that the expression of at least one or more sense or anti-sense eRNAs occurs in more than 80% of the stem cell population (fig. 4b). Co-expression of all three eRNAs was observed in approximately 15-20% of the cells (fig. 4b), with the +60 kb region being by far the most highly transcribed enhancer on the anti-sense strand (70%) as opposed to 30-35% for the -45 kb and -5 kb (fig. 4b). Pairwise analysis of each combination of eRNAs (sense and anti-sense) revealed that at least 30% of the stem cell population co-express an eRNA pair, with a set of sense eRNAs (+60 kb/-5 kb or +60 kb/-45 kb) being co-expressed in around 15% of the cells. Due to eRNAs rapid degradation rate, these co-expression events suggest that enhancers were probably transcribed almost simultaneously.

The analysis of chromatin conformation by Capture-C has demonstrated the contacts between *Nanog* promoter and all three enhancers (supp fig. 11), allowing us to posit that eRNA co-localisation may be correlated with a co-activation of their transcription in the proximity of the *Nanog* gene locus. To further assess whether transcription at these enhancers occurs synergistically, we quantified the co-localisation of these eRNAs either by pair or for all three transcripts (fig. 4c). This analysis reveals that all pairs tested display (5-15%) of sub-nuclear co-localisation. In some instances, co-localisation was also found for all three eRNAs for both the sense and anti-sense strands (fig. 4c-d). Altogether,

these data suggest that transcriptional activity of *Nanog* enhancers is synchronised within a scaffold region for at least 30-45% of the mESCs population.

To explore the temporal coordination between *Nanog* eRNA transcription and nascent mRNA production, we next investigated the expression of eRNAs relative to the *Nanog* active transcription. All three eRNA transcripts exhibit a positive correlation between their expression and the accumulation of endogenous *Nanog* introns products (fig. 4e). By contrast, regardless of the transcriptional direction of eRNAs, the quantitation of eRNAs' co-localisation with *Nanog* mRNA showed that only a small number of enhancer transcripts are present during the synthesis of *Nanog* nascent RNA (fig. 4f and supp fig. 12, - 45kb eRNA), suggesting that transcription of enhancer and gene body does not occur at the same time. Because the co-labelling of *Nanog* eRNAs and *Nanog* mRNA introns has limitations in defining the order of steps between these two transcriptional events, due to a possible intron retention mechanism (*42*), we further refine the temporal relationship of their synthesis using eRNA RNA-FISH combined with RFP detection from the STREAMING tag cassette. This enables us to quantify more accurately the abundance of *Nanog* enhancer transcripts in relation to the progression of *Nanog* elongation (fig. 4g and h). This approach shows that in the “on” state, while nascent *Nanog* mRNA is actively produced, eRNA presence is rarely co-detected for any of these enhancers. By contrast, in the “off” states, there is a wide range of eRNA molecules found in the nucleus (fig. 4g and h), suggesting that the timing of enhancer transcription is not concurrent to the one of *Nanog* mRNA elongation. This finding on the asynchronous activity of between enhancer and gene body RNA synthesis is consistent with a previous report suggesting that proximal enhancer transcription is mutually exclusive with gene body transcription (*43*).

SOX2 has been shown to directly recruit *Nanog* enhancers (*25*, *40*, *44*), however, the timing of this recruitment relative to transcriptional activation was unclear. Therefore, combining live observations of its clustering dynamics with eRNA quantitation relative to *Nanog* mRNA elongation, we propose

that the enhancer transcription surge coincides with SOX2's densest clustering activity while the gene locus is in a "off" mode.

In conclusion, our findings strengthen the link between the *Nanog* gene transcription cycle and the activity of its regulatory factors by integrating spatial and temporal SOX2 and BRD4 clustering dynamics, locus mobility, and enhancer transcription. This study establishes that SOX2 temporal clustering is out-of-phase with the active basal-transcriptional machinery to determine either the basis of permissive transcriptional "off" state that precedes transcription burst or a step that contributes to elongation termination (supp fig. 13, model).

Discussion

This study demonstrates that SOX2 medium and high intensity-clustering activities at the *Nanog* gene locus frame transcription burst events and do not appear to directly cooperate with the BRD4-associated elongation step. This finding fits with the observation that, in stem cells, SOX2 gain-of-function leads to a reduction in *Nanog* mRNA levels and a loss of pluripotency (*14*, *17*). In that instance, overcrowding of the *Nanog* locus may interfere with the elongation kinetics that does not require frequent SOX2 occupancy, disturb the pre-mRNA splicing (*45*) and derail the cycling pace of *Nanog* transcription required to drive pluripotency.

It has been shown that variation in gene locus mobility is informative about the transcription status even more so than the promoter to enhancer 3D distance (*46*). Our findings provide additional evidence for the involvement of SOX2 into different phases of the *Nanog* gene transcription cycle, as shown by the different locus mobility responses observed during the "off" state in the absence or presence of SOX2 clusters. These findings are consistent with the observation that binding of the NuRD repressor complex increases *Nanog* locus mobility through nucleosome repackaging (*47*). Since it has been shown that SOX2 and NuRD interact (*48*, *49*), it is tempting to speculate that SOX2 type III clustering behaviour that occurs post-elongation might be associated with NuRD and might contribute to chromatin compaction of that region. More broadly, these distinct SOX2 clustering activities are likely to indirectly represent changes in protein partners, variations in RNA concentration and chromatin epigenetic state at the locus that modulate SOX2 occupancy behaviour (*10*, *28*, *50–52*).

This study further informs earlier work that found a role for the interaction between SOX2 and RNA (*28*) and the finding that most TFs have an RNA-binding arm that helps generate an RNA-containing condensate that affects the diffusion and dwell time of genome effectors (*37*). Interestingly, the transcription of *Nanog* enhancer RNA adopts a temporal pattern akin to SOX2 high-clustering activities, with low to no expression during the elongation stage but higher concentration while the

locus is in an "off" state. This observation suggests that SOX2 TF has to keep operating at the *Nanog* locus in an environment continuously crowded with RNA, either in the presence of eRNA or mRNA. Yet, SOX2 occupancy dynamics at the *Nanog* locus display multiple behaviours (type I-III), depending on whether elongation is "on" or "off" and whether it is preceding or occurring after the elongation step. This means that different types of RNA at the gene site, which change in concentration over time, are critical in determining how SOX2 behaves (*28*, *37*). Here, we provide multiple lines of evidence that different SOX2-dependent molecular reactions represent varying "off" states at the *Nanog* locus region. These molecular reactions involve eRNAs, which have been shown to shape chromatin loops and recruit co-activators (*38*, *53*, *54*), likely influencing transcriptional output *Nanog* in a cooperative manner with SOX2 clustering dynamics.

In conclusion, we posit that cell-to-cell *Nanog* transcriptional fluctuation may be driven by asynchronous variations of SOX2 occupancy dynamics, eRNA levels, and mRNA bursts frequencies, producing a continuum of *Nanog* expression levels across the stem cell population. These varying synchronicities between genome effectors may give rise to a range of transcription rates, leading either towards self-renewal or differentiation.

Figures

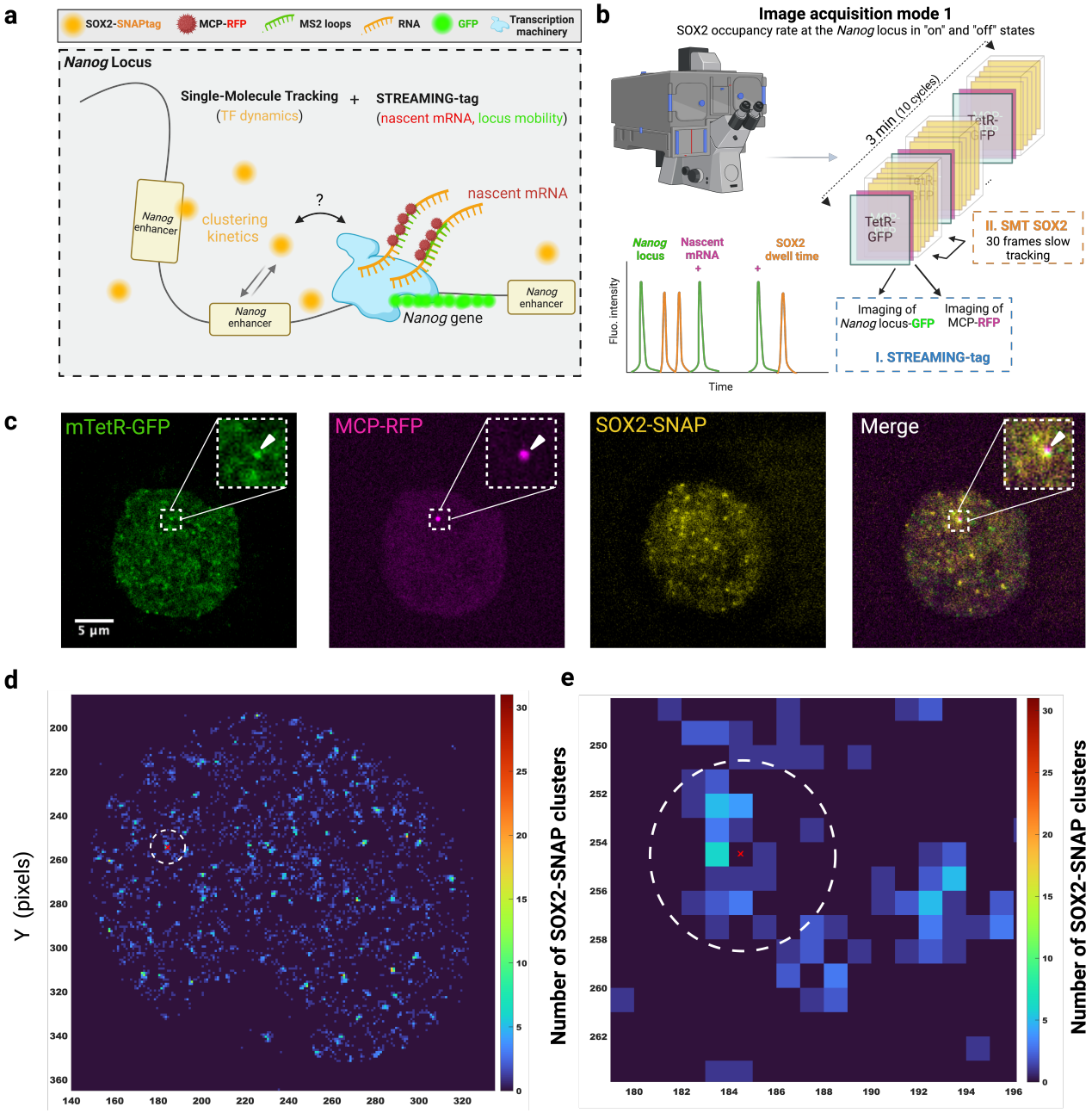

Figure 1. Multimodal imaging method to visualise transcription factor activity at a single locus resolution relative to its transcriptional state. (a) Schematic of the principle for dual imaging to capture SOX2 clustering dynamics at the *Nanog* locus relative to the event of nascent mRNA synthesis. The STREAMING-tag cassette at the *Nanog* locus enables the detection of the *Nanog* mRNA synthesis (MS2/MCP-RFP) along with the *Nanog* locus mobility (TetO/mTetR-GFP) while SNAP-tag knocked-in at the *Sox2* locus reveals chromatin occupancy rate of this TF via single

molecule tracking (SMT). (**b**) Principle of the imaging acquisition mode 1 set up to capture the fluorescent signals from STREAMING-tag and SMT detection methods in a live cell. Multimodal image acquisition uses a combination of SDC (STREAMING-tag) and HILO (SMT) microscopy (**c**) Fluorescent signals from the *Nanog* locus (mTetR-GFP), *Nanog* mRNA (MCP-RFP) and SOX2-SNAPtag. (**d**) Representative heat map of most visited locations accumulated over a 3 min period for the detected SOX2 molecules across the cell nucleus. The heat map reveals an overlap between SOX2 long dwell time and the region of interest for the *Nanog* locus (red cross). (**e**) Higher magnification of the dashed circle shown in panel **d** demonstrates SOX2 clustering activity within a 4-pixel radius ROI (circle dashed line) centred on the *Nanog* locus when both GFP (red cross) and RFP signals are detected.

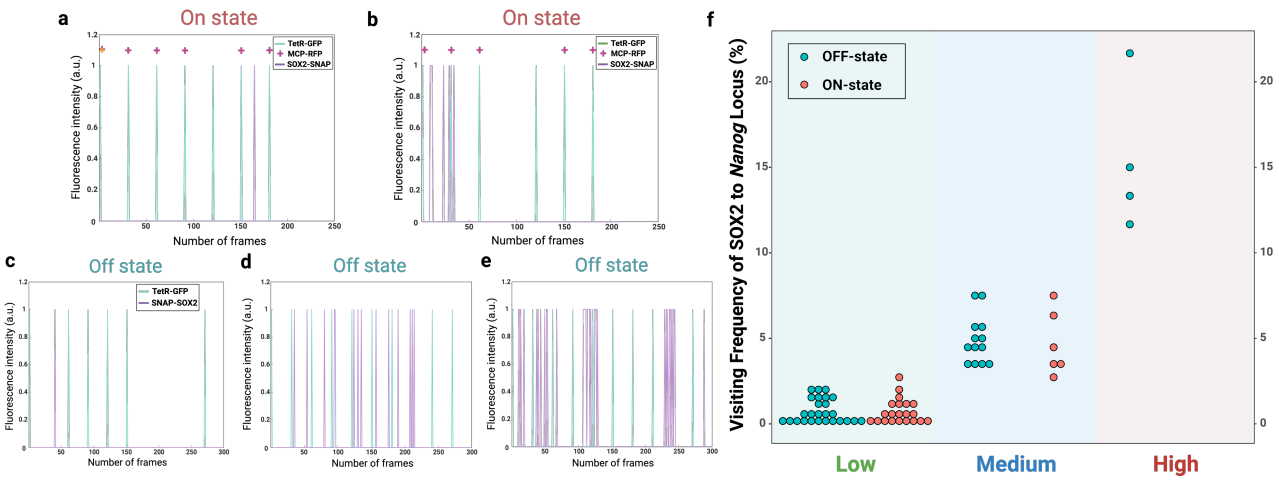

Figure 2. Multiple SOX2 clustering dynamics behaviours at the *Nanog* locus during "off" state. (**a-b**) On-state. Signal intensity for GFP (green) at the *Nanog* locus, SOX2 SMT signal (purple) and RFP (magenta +) for *Nanog* mRNA detection (on-state). Examples of no to low (a) and medium (b) visiting frequency. **(c-e)** Off-state. Signal intensity for GFP (green) at the *Nanog* locus, SOX2 SMT signal (purple) showing the visiting frequency of SOX2 when the locus is available, but production of nascent mRNA is off. Examples of no to low (c), medium (d) and high (e) visiting frequency. **(f)** Quantitation of the different SOX2 behaviours during the on-state (n = 27) and off-state (n = 43)

classified based on the visiting frequency (the total number of SOX2 molecules that have visited this GFP positive spot normalised to the total number of frames). The visiting frequency accounts for the number of SOX2 molecules that have visited the GFP spot.

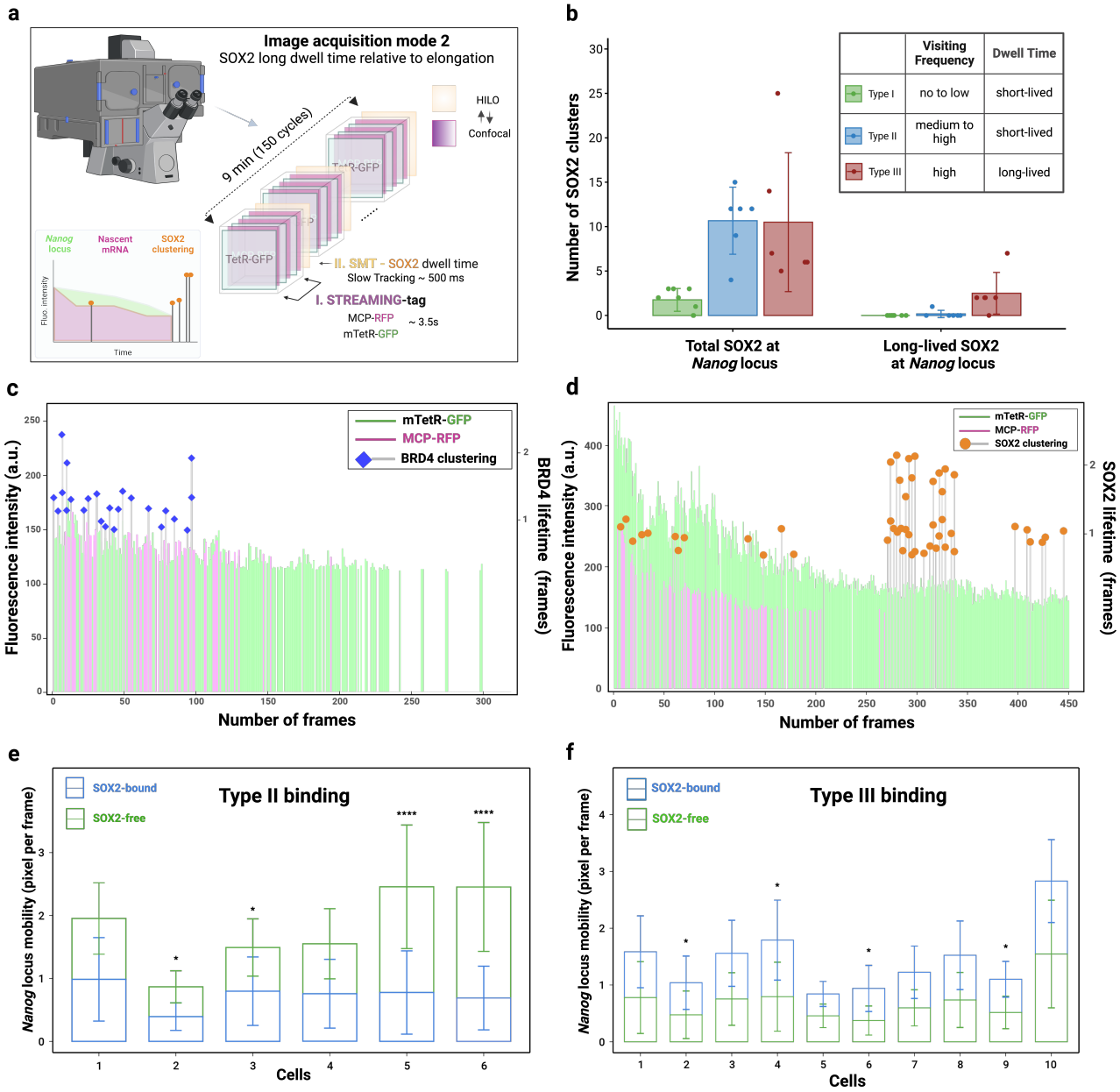

Figure 3. Distinct types of SOX2 burst dynamics associate with different locus mobility and frame the elongation process at the *Nanog* locus. (a) Image acquisition mode 2 set up to capture a full transcription cycle at the *Nanog* locus (STREAMING-tag) in parallel to SOX2 or BRD4

chromatin binding dynamics (SMT). **(b)** Quantification of detected SOX2 clusters and their relative survival rate at the *Nanog* locus. (**c** and **d**) BRD4 **(c)** or SOX2 **(d)** dwell time relative to the elongation process. Fluorescent intensity at the *Nanog* locus is shown relative to time. *Nanog* locus signal is labelled with GFP (green), and *Nanog* mRNA synthesis (magenta). SOX2 (orange) and BRD4 (blue) long dwell times (over 2 frames) and short dwell times (over 1 frame) reveal diverse binding kinetics relative to the elongation process. (**e** and **f**) 2D *Nanog* locus mobility in the presence (blue) or absence (green) of SOX2 binding.

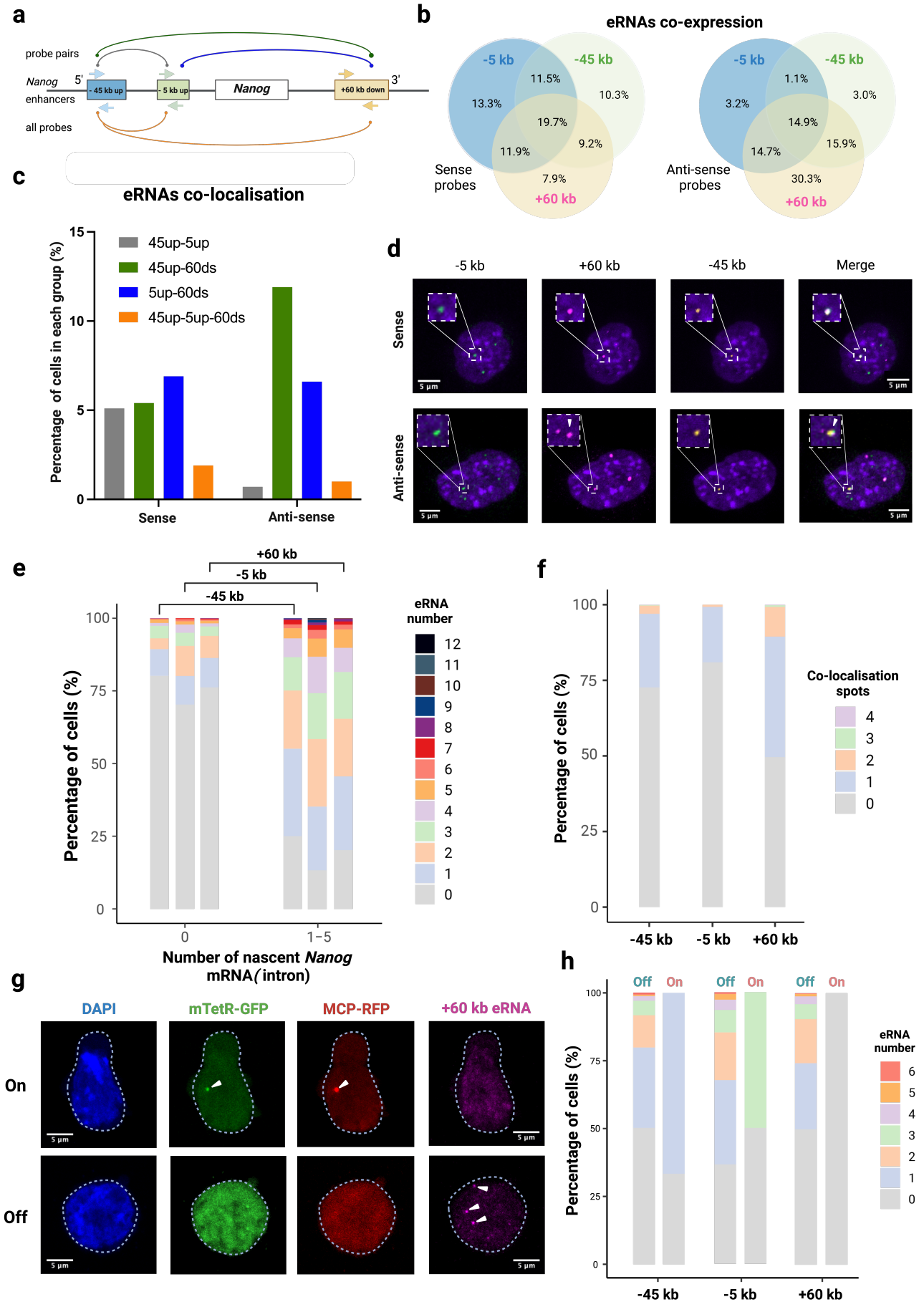
a
probe pairs
Nanog
5'
3'
enhancers
all probes
- 45 kb up
- 5 kb up
Nanog
+60 kb down
b
eRNAs co-expression
-5 kb
-45 kb
+60 kb
Sense probes
Anti-sense probes
c
eRNAs co-localisation
45up-5up
45up-60ds
5up-60ds
45up-5up-60ds
Percentage of cells in each group (%)
Sense
Anti-sense
d
-5 kb
+60 kb
-45 kb
Merge
Sense
Anti-sense
e
+60 kb
-5 kb
-45 kb
Percentage of cells (%)
eRNA number
Number of nascent Nanog mRNA(intron)
f
Percentage of cells (%)
Co-localisation spots
-45 kb
-5 kb
+60 kb
g
DAPI
mTetR-GFP
MCP-RFP
+60 kb eRNA
On
Off
h
Off
On
Percentage of cells (%)
eRNA number
-45 kb
-5 kb
+60 kb

Figure 4. *Nanog* enhancers are individually or synergistically transcribed while *Nanog* nascent transcription is off. (**a**) Map of eRNAs probes (coloured arrows indicate sense and anti-sense orientations) at three regulatory regions involved with *Nanog* gene transcription. Distances from transcription start sites are -45 kb, -5 kb and +60 kb. Coloured lines connecting the different enhancers represent the different probe combinations used to assess co-localisation: grey [-45 kb/-5 kb]; green [-45 kb/+60 kb]; blue [-5 kb/+60 kb]; orange, all three probes. (**b**) Venn diagram showing the co-expression levels of different eRNAs in embryonic stem cells. **(c)** Quantitation of three eRNAs co-localisation either assessed by pairs or in sets of three for both sense and anti-sense transcripts (≥ 4000 cells imaged per condition). **(d)** Single-molecule FISH showing sub-nuclear co-localisation of *Nanog* eRNAs for both sense and anti-sense products (green -5 kb, magenta +60 kb, -45 kb orange), scale bar 5μm. **(e)** Abundance of three different *Nanog* eRNAs correlated to the *Nanog* intron expression. **(f)** Quantitation of co-localisation events between *Nanog* eRNAs and *Nanog* intron mRNA. **(g)** Single-molecule FISH for the +60 kb eRNA combined with STREAMING tag imaging shows that the +60 kb eRNA is expressed while *Nanog* mRNA nascent transcription is off (green *Nanog* locus, red *Nanog* mRNA, magenta +60 kb eRNA, blue DAPI). **(h)** Temporal relationship between individual eRNAs from three different *Nanog* enhancers relative to the co-expression of the *Nanog* STREAMING tag reporter system.

References

1. J. Rodriguez, D. R. Larson, Transcription in Living Cells: Molecular Mechanisms of Bursting. *Annu. Rev. Biochem.* **89**, 189–212 (2020).

2. W. F. Beckman, M. Á. L. Jiménez, P. J. Verschure, Transcription bursting and epigenetic plasticity: an updated view. *Epigenetics Commun.* **1**, 6 (2021).

3. J. H. Gibcus, J. Dekker, The hierarchy of the 3D genome. *Mol. Cell* **49**, 773–782 (2013).

4. M. Jordan Rowley, V. G. Corces, Organizational Principles of 3D Genome Architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).

5. E. Miron, R. Oldenkamp, J. M. Brown, D. M. S. Pinto, C. S. Xu, A. R. Faria, H. A. Shaban, J. D. P. Rhodes, C. Innocent, S. de Ornellas, H. F. Hess, V. Buckle, L. Schermelleh, Chromatin arranges in chains of mesoscale domains with nanoscale functional topography independent of cohesin. *Sci. Adv.* **6**, eaba8811 (2020).

6. M. Tokunaga, N. Imamoto, K. Sakata-Sogawa, Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 159–161 (2008).

7. K. I. Mortensen, L. S. Churchman, J. A. Spudich, H. Flyvbjerg, Optimized localization analysis for single-molecule tracking and super-resolution microscopy. *Nat. Methods* **7**, 377–381 (2010).

8. I. Izeddin, V. Récamier, L. Bosanac, I. I. Cissé, L. Boudarene, C. Dugast-Darzacq, F. Proux, O. Bénichou, R. Voituriez, O. Bensaude, M. Dahan, X. Darzacq, Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife* **3**, e02230 (2014).

9. Z. Liu, L. D. Lavis, E. Betzig, Imaging Live-Cell Dynamics and Structure at the Single-Molecule Level. *Mol. Cell* **58**, 644–659 (2015).

10. J. Chen, Z. Zhang, L. Li, B.-C. Chen, A. Revyakin, B. Hajj, W. Legant, M. Dahan, T. Lionnet, E. Betzig, R. Tjian, Z. Liu, Single-Molecule Dynamics of Enhanceosome Assembly in Embryonic Stem Cells. *Cell* **156**, 1274–1285 (2014).

11. M. D. White, J. F. Angiolini, Y. D. Alvarez, G. Kaur, Z. W. Zhao, E. Mocskos, L. Bruno, S. Bissiere, V. Levi, N. Plachta, Long-Lived Binding of Sox2 to DNA Predicts Cell Fate in the Four-Cell Mouse Embryo. *Cell* **165**, 75–87 (2016).

12. L. Dahal, N. Walther, R. Tjian, X. Darzacq, T. G. W. Graham, Single-molecule tracking (SMT): a window into live-cell transcription biochemistry. *Biochem. Soc. Trans.* **51**, 557–569 (2023).

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

14. B. Boer, J. Kopp, S. Mallanna, M. Desler, H. Chakravarthy, P. J. Wilder, C. Bernadt, A. Rizzino, Elevating the levels of Sox2 in embryonal carcinoma cells and embryonic stem cells inhibits the expression of Sox2:Oct-3/4 target genes†. *Nucleic Acids Res.* **35**, 1773–1786 (2007).

15. S. Masui, Y. Nakatake, Y. Toyooka, D. Shimosato, R. Yagi, K. Takahashi, H. Okochi, A. Okuda, R. Matoba, A. A. Sharov, M. S. H. Ko, H. Niwa, Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nat. Cell Biol.* **9**, 625–635 (2007).

16. H. Fong, K. A. Hohenstein, P. J. Donovan, Regulation of Self-Renewal and Pluripotency by Sox2 in Human Embryonic Stem Cells. *Stem Cells* **26**, 1931–1938 (2008).

17. J. L. Kopp, B. D. Ormsbee, M. Desler, A. Rizzino, Small increases in the level of Sox2 trigger the differentiation of mouse embryonic stem cells. *Stem Cells Dayt. Ohio* **26**, 903–911 (2008).

18. S. Zhang, W. Cui, Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World J. Stem Cells* **6**, 305–311 (2014).

19. H. Ohishi, S. Shimada, S. Uchino, J. Li, Y. Sato, M. Shintani, H. Owada, Y. Ohkawa, A. Pertsinidis, T. Yamamoto, H. Kimura, H. Ochiai, STREAMING-tag system reveals spatiotemporal relationships between transcriptional regulatory factors and transcriptional activity. *Nat. Commun.* **13**, 7672 (2022).

20. J. Li, A. Dong, K. Saydaminova, H. Chang, G. Wang, H. Ochiai, T. Yamamoto, A. Pertsinidis, Single-Molecule Nanoscopy Elucidates RNA Polymerase II Transcription at Single Genes in Live Cells. *Cell* **178**, 491-506.e28 (2019).

21. J. Li, A. Hsu, Y. Hua, G. Wang, L. Cheng, H. Ochiai, T. Yamamoto, A. Pertsinidis, Single-gene imaging links genome topology, promoter–enhancer communication and transcription control. *Nat. Struct. Mol. Biol.* **27**, 1032–1040 (2020).

22. J. M. Tome, N. D. Tippens, J. T. Lis, Single-molecule nascent RNA sequencing identifies regulatory domain architecture at promoters and enhancers. *Nat. Genet.* **50**, 1533–1541 (2018).

23. F. X. Chen, E. R. Smith, A. Shilatifard, Born to run: control of transcription elongation by RNA polymerase II. *Nat. Rev. Mol. Cell Biol.* **19**, 464–478 (2018).

24. L. Core, K. Adelman, Promoter-proximal pausing of RNA polymerase II: a nexus of gene regulation. *Genes Dev.* **33**, 960–982 (2019).

25. D. J. Rodda, J.-L. Chew, L.-H. Lim, Y.-H. Loh, B. Wang, H.-H. Ng, P. Robson, Transcriptional Regulation of *Nanog* by OCT4 and SOX2*. *J. Biol. Chem.* **280**, 24731–24737 (2005).

26. P. Agrawal, S. Blinka, K. Pulakanti, M. H. Reimer, C. Stelloh, A. E. Meyer, S. Rao, Genome editing demonstrates that the −5 kb Nanog enhancer regulates Nanog expression by modulating RNAPII initiation and/or recruitment. *J. Biol. Chem.* **296**, 100189 (2021).

27. W.-K. Cho, J.-H. Spille, M. Hecht, C. Lee, C. Li, V. Grube, I. I. Cisse, Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* **361**, 412–415 (2018).

28. Z. E. Holmes, D. J. Hamilton, T. Hwang, N. V. Parsonnet, J. L. Rinn, D. S. Wuttke, R. T. Batey, The Sox2 transcription factor binds RNA. *Nat. Commun.* **11**, 1805 (2020).

29. L. Hou, Y. Wei, Y. Lin, X. Wang, Y. Lai, M. Yin, Y. Chen, X. Guo, S. Wu, Y. Zhu, J. Yuan, M. Tariq, N. Li, H. Sun, H. Wang, X. Zhang, J. Chen, X. Bao, R. Jauch, Concurrent binding to

DNA and RNA facilitates the pluripotency reprogramming activity of Sox2. *Nucleic Acids Res.* **48**, 3869–3887 (2020).

30. W. J. de Jonge, H. P. Patel, J. V. W. Meeussen, T. L. Lenstra, Following the tracks: How transcription factor binding dynamics control transcription. *Biophys. J.* **121**, 1583–1592 (2022).

31. R. Métivier, G. Reid, F. Gannon, Transcription in four dimensions: nuclear receptor-directed initiation of gene expression. *EMBO Rep.* **7**, 161–167 (2006).

32. J. M. Pedraza, J. Paulsson, Effects of Molecular Memory and Bursting on Fluctuations in Gene Expression. *Science* **319**, 339–343 (2008).

33. Y. Wang, F. Liu, J. Li, W. Wang, Reconciling the concurrent fast and slow cycling of proteins on gene promoters. *J. R. Soc. Interface* **11**, 20140253 (2014).

34. Y. Wang, T. Ni, W. Wang, F. Liu, Gene transcription in bursting: a unified mode for realizing accuracy and stochasticity. *Biol. Rev.* **94**, 248–258 (2019).

35. Z. Liu, W. R. Legant, B.-C. Chen, L. Li, J. B. Grimm, L. D. Lavis, E. Betzig, R. Tjian, 3D imaging of Sox2 enhancer clusters in embryonic stem cells. *eLife* **3**, e04236 (2014).

36. H. Ochiai, T. Sugawara, T. Yamamoto, Simultaneous live imaging of the transcription and nuclear position of specific genes. *Nucleic Acids Res.* **43**, e127 (2015).

37. O. Oksuz, J. E. Henninger, R. Warneford-Thomson, M. M. Zheng, H. Erb, A. Vancura, K. J. Overholt, S. W. Hawken, S. F. Banani, R. Lauman, L. N. Reich, A. L. Robertson, N. M. Hannett, T. I. Lee, L. I. Zon, R. Bonasio, R. A. Young, Transcription factors interact with RNA to regulate genes. *Mol. Cell* **83**, 2449-2463.e13 (2023).

38. P. R. Arnold, A. D. Wells, X. C. Li, Diversity and Emerging Roles of Enhancer RNA in Regulation of Gene Expression and Cell Fate. *Front. Cell Dev. Biol.* **7** (2020).

39. S. Blinka, M. H. Reimer, K. Pulakanti, S. Rao, Super-Enhancers at the *Nanog* Locus Differentially Regulate Neighboring Pluripotency-Associated Genes. *Cell Rep.* **17**, 19–28 (2016).

40. S. Blinka, S. Rao, Nanog Expression in Embryonic Stem Cells - An Ideal Model System to Dissect Enhancer Function. *BioEssays News Rev. Mol. Cell. Dev. Biol.* **39** (2017).

41. D. B. Mahat, N. D. Tippens, J. D. Martin-Rufino, S. K. Waterton, J. Fu, S. E. Blatt, P. A. Sharp, Single-cell nascent RNA sequencing unveils coordinated global transcription. *Nature* **631**, 216–223 (2024).

42. S. Das, S. Jena, D. N. Levasseur, Alternative Splicing Produces Nanog Protein Variants with Different Capacities for Self-renewal and Pluripotency in Embryonic Stem Cells*. *J. Biol. Chem.* **286**, 42690–42703 (2011).

43. K. Hamamoto, Y. Umemura, S. Makino, T. Fukaya, Dynamic interplay between non-coding enhancer transcription and gene activity in development. *Nat. Commun.* **14**, 826 (2023).

44. M. A. Lodato, C. W. Ng, J. A. Wamstad, A. W. Cheng, K. K. Thai, E. Fraenkel, R. Jaenisch, L. A. Boyer, SOX2 Co-Occupies Distal Enhancer Elements with Distinct POU Factors in ESCs and NPCs to Specify Cell State. *PLOS Genet.* **9**, e1003288 (2013).

45. X. Rambout, F. Dequiedt, L. E. Maquat, Beyond Transcription: Roles of Transcription Factors in Pre-mRNA Splicing. *Chem. Rev.* **118**, 4339–4364 (2018).

46. A. Platania, C. Erb, M. Barbieri, B. Molcrette, E. Grandgirard, M. A. de Kort, K. Meaburn, T. Taylor, V. M. Shchuka, S. Kocanova, G. M. Oliveira, J. A. Mitchell, E. Soutoglou, T. L. Lenstra, N. Molina, A. Papantonis, K. Bystricky, T. Sexton, Competition between transcription and loop extrusion modulates promoter and enhancer dynamics. bioRxiv [Preprint] (2023). https://doi.org/10.1101/2023.04.25.538222.

47. S. Basu, O. Shukron, D. Hall, P. Parutto, A. Ponjavic, D. Shah, W. Boucher, D. Lando, W. Zhang, N. Reynolds, L. H. Sober, A. Jartseva, R. Ragheb, X. Ma, J. Cramard, R. Floyd, J. Balmer, T. A. Drury, A. R. Carr, L.-M. Needham, A. Aubert, G. Communie, K. Gor, M. Steindel, L. Morey, E. Blanco, T. Bartke, L. Di Croce, I. Berger, C. Schaffitzel, S. F. Lee, T. J. Stevens, D. Klenerman, B. D. Hendrich, D. Holcman, E. D. Laue, Live-cell three-dimensional single-molecule tracking reveals modulation of enhancer dynamics by NuRD. *Nat. Struct. Mol. Biol.* **30**, 1628–1639 (2023).

48. S. K. Mallanna, B. D. Ormsbee, M. Iacovino, J. M. Gilmore, J. L. Cox, M. Kyba, M. P. Washburn, A. Rizzino, Proteomic analysis of Sox2-associated proteins during early stages of mouse embryonic stem cell differentiation identifies Sox21 as a novel regulator of stem cell fate. *Stem Cells Dayt. Ohio* **28**, 1715–1727 (2010).

49. S. Wang, P. Xia, B. Ye, G. Huang, J. Liu, Z. Fan, Transient activation of autophagy via Sox2-mediated suppression of mTOR is an important early step in reprogramming to pluripotency. *Cell Stem Cell* **13**, 617–625 (2013).

50. M. Mazzocca, E. Colombo, A. Callegari, D. Mazza, Transcription factor binding kinetics and transcriptional bursting: What do we really know? *Curr. Opin. Struct. Biol.* **71**, 239–248 (2021).

51. G. Kaur, M. W. Costa, C. M. Nefzger, J. Silva, J. C. Fierro-González, J. M. Polo, T. D. M. Bell, N. Plachta, Probing transcription factor diffusion dynamics in the living mammalian embryo with photoactivatable fluorescence correlation spectroscopy. *Nat. Commun.* **4**, 1637 (2013).

52. J. Hörberg, A. Reymer, Decoding the dual recognition mechanism of the glucocorticoid receptor for DNA and RNA: sequence versus shape. *Sci. Rep.* **13**, 16125 (2023).

53. M. U. Kaikkonen, N. J. Spann, S. Heinz, C. E. Romanoski, K. A. Allison, J. D. Stender, H. B. Chun, D. F. Tough, R. K. Prinjha, C. Benner, C. K. Glass, Remodeling of the Enhancer Landscape during Macrophage Activation Is Coupled to Enhancer Transcription. *Mol. Cell* **51**, 310–325 (2013).

54. L. Pnueli, S. Rudnizky, Y. Yosefzon, P. Melamed, RNA transcribed from a distal enhancer is required for activating the chromatin at the promoter of the gonadotropin α-subunit gene. *Proc. Natl. Acad. Sci.* **112**, 4369–4374 (2015).

55. J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P.

Tomancak, A. Cardona, Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

56. D. R. Stirling, M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, A. Goodman, CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics* **22**, 433 (2021).

57. The MathWorks Inc., Optimization Toolbox, version 9.4 (R2022b) (2022); https://www.mathworks.com.

58. C. T. Rueden, M. C. Hiner, E. L. Evans, M. A. Pinkert, A. M. Lucas, A. E. Carpenter, B. A. Cimini, K. W. Eliceiri, PyImageJ: A library for integrating ImageJ and Python. *Nat. Methods* **19**, 1326–1327 (2022).

59. C. Stringer, T. Wang, M. Michaelos, M. Pachitariu, Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).

60. A. Imbert, W. Ouyang, A. Safieddine, E. Coleno, C. Zimmer, E. Bertrand, T. Walter, F. Mueller, FISH-quant v2: a scalable and modular tool for smFISH image analysis. *RNA* **28**, 786–795 (2022).

61. Scientific Volume Imaging, Huygens Localiser, version 24.04 (2024); http://svi.nl.

62. A. J. McCann, J. Lou, M. Moustaqil, M. S. Graus, A. Blum, F. Fontaine, H. Liu, W. Luu, P. Rudolffi-Soto, P. Koopman, E. Sierecki, Y. Gambin, F. A. Meunier, Z. Liu, E. Hinde, M. Francois, A dominant-negative SOX18 mutant disrupts multiple regulatory layers essential to transcription factor activity. *Nucleic Acids Res.* **49**, 10931–10955 (2021).

63. J. R. Hughes, N. Roberts, S. McGowan, D. Hay, E. Giannoulatou, M. Lynch, M. De Gobbi, S. Taylor, R. Gibbons, D. R. Higgs, Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat. Genet.* **46**, 205–212 (2014).

64. D. J. Downes, A. L. Smith, M. A. Karpinska, T. Velychko, K. Rue-Albrecht, D. Sims, T. A. Milne, J. O. J. Davies, A. M. Oudelaar, J. R. Hughes, Capture-C: a modular and flexible approach for high-resolution chromosome conformation capture. *Nat. Protoc.* **17**, 445–475 (2022).

65. J. O. J. Davies, J. M. Telenius, S. J. McGowan, N. A. Roberts, S. Taylor, D. R. Higgs, J. R. Hughes, Multiplexed analysis of chromosome conformation at vastly improved sensitivity. *Nat. Methods* **13**, 74–80 (2016).

66. S. W. Wingett, P. Ewels, M. Furlan-Magaril, T. Nagano, S. Schoenfelder, P. Fraser, S. Andrews, HiCUP: pipeline for mapping and processing Hi-C data. F1000Research 4:1310 [Preprint] (2015). https://doi.org/10.12688/f1000research.7334.1.

67. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

68. J. Cairns, P. Freire-Pritchett, S. W. Wingett, C. Várnai, A. Dimond, V. Plagnol, D. Zerbino, S. Schoenfelder, B.-M. Javierre, C. Osborne, P. Fraser, M. Spivakov, CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biol.* **17**, 127 (2016).

69. A. Feldmann, E. Dimitrova, A. Kenney, A. Lastuvkova, R. J. Klose, CDK-Mediator and FBXL19 prime developmental genes for activation by promoting atypical regulatory interactions. *Nucleic Acids Res.* **48**, 2942–2955 (2020).

70. W. A. Whyte, D. A. Orlando, D. Hnisz, B. J. Abraham, C. Y. Lin, M. H. Kagey, P. B. Rahl, T. I. Lee, R. A. Young, Master Transcription Factors and Mediator Establish Super-Enhancers at Key Cell Identity Genes. *Cell* **153**, 307–319 (2013).

Acknowledgments: We thank Sydney Microscopy and Micro-Analysis core facilities for ongoing technical support with imaging. We thank Dr Alexandros Pertsinidis and Dr Jieru Li (MSK, NY) for their valuable input and feedback on reviewing the manuscript. We thank Dr Luke Lavis Janelia Farm (Virginia Tech) for providing reagents necessary to establish the SOX2 HaloTag knock-in cell lines.

Funding:

NHMRC idea Grant (APP 2029719) (MF)

NHMRC idea Grant (APP2019904) (MF)

NSW Cardio-vascular Research Capacity Program – Senior Researcher grant (MF)

Grants-in Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology JP24H02326 (HOc)

Grants-in Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science, and Technology JP22H02609 (HOc)

CREST from the Japan Science and Technology Agency to HOc (JPMJCR23N3)

NIG-JOINT 69A2024 (HOc)

NIG-JOINT 14R2024 (HOc)

Medical Research Centre Initiative for High Depth Omics, Kyushu University, Japan (HOc)

Author contributions:

Conceptualisation: HOc, MF

Methodology: HOc, BL, YW, NR, TD, HOh

Investigation: BL, MG, NR, YW, TD

Visualisation: YW, BL, MF

Funding acquisition: HOc, MF

Project administration: HOc, MF

Supervision: MF, TD

Writing – original draft: MF

Writing – review & editing: MF, BL, HOc

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data are available in the main text or the supplementary materials and have been deposited to Datadryad for access to raw files.

Accession number and link: https://doi.org/10.5061/dryad.wpzgmsbw4

Supplementary Materials

Figs. S1 to S13

Movies S1 to S6