







Mesoscale chromatin confinement facilitates target search of pioneer transcription factors in live cells

Received: 8 June 2023

Accepted: 7 August 2024

Published online: 04 October 2024

 Check for updates

Zuhui Wang ^{1,2,3,12}, Bo Wang ^{1,3,4,12}, Di Niu^{1,3,4,12}, Chao Yin^{1,3,4}, Ying Bi^{1,2}, Claudia Cattoglio^{5,6}, Kyle M. Loh ^{7,8}, Luke D. Lavis ⁹, Hao Ge ^{1,10} & Wulan Deng ^{1,2,3,4,11} ✉

Pioneer transcription factors (PTFs) possess the unique capability to access closed chromatin regions and initiate cell fate changes, yet the underlying mechanisms remain elusive. Here, we characterized the single-molecule dynamics of PTFs targeting chromatin in living cells, revealing a notable ‘confined target search’ mechanism. PTFs such as FOXA1, FOXA2, SOX2, OCT4 and KLF4 sampled chromatin more frequently than non-PTF MYC, alternating between fast free diffusion in the nucleus and slower confined diffusion within mesoscale zones. Super-resolved microscopy showed closed chromatin organized as mesoscale nucleosome-dense domains, confining FOXA2 diffusion locally and enriching its binding. We pinpointed specific histone-interacting disordered regions, distinct from DNA-binding domains, crucial for confined target search kinetics and pioneer activity within closed chromatin. Fusion to other factors enhanced pioneer activity. Kinetic simulations suggested that transient confinement could increase target association rate by shortening search time and binding repeatedly. Our findings illuminate how PTFs recognize and exploit closed chromatin organization to access targets, revealing a pivotal aspect of gene regulation.

Cell fate transitions rely on specific transcription factors (TFs) to activate new gene expression programs. However, the presence of nucleosomes and compaction of closed chromatin hinder most TFs and the transcription machinery from accessing DNA templates^{1,2}. Pioneer TFs (PTFs), such as FOXA factors, exhibit remarkable abilities to access DNA targets even when bound to nucleosomes, facilitating the opening of closed chromatin and triggering cell fate changes^{3–6}. Despite extensive studies focusing on the structural and functional

aspects of PTFs binding to nucleosomes^{6–12}, the precise mechanisms underlying their initial search for and binding to closed chromatin, setting PTFs apart from other TFs, remain unclear.

Single-molecule localization microscopy enables the direct visualization of individual molecules in cells with spatial resolution surpassing the diffraction limit¹³. Live-cell single-molecule tracking (SMT) has revealed dynamic interactions between TFs^{14–16} or transcription machinery^{17,18} and chromatin, challenging the notion of stable

¹Biomedical Pioneering Innovation Center (BIOPIIC), Peking University, Beijing, China. ²Peking University-Tsinghua University-National Institute of Biological Sciences Joint Graduate Program, Peking University, Beijing, China. ³Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China. ⁴Peking-Tsinghua Center for Life Sciences (CLS), Peking University, Beijing, China. ⁵Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. ⁶Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA, USA. ⁷Department of Developmental Biology, Stanford University, Stanford, CA, USA. ⁸Institute for Stem Cell Biology & Regenerative Medicine, Stanford University, Stanford, CA, USA. ⁹Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA. ¹⁰Beijing International Center for Mathematical Research, Peking University, Beijing, China. ¹¹Beijing Advanced Innovation Center for Genomics (ICG), Peking University, Beijing, China. ¹²These authors contributed equally: Zuhui Wang, Bo Wang, Di Niu. ✉e-mail: wdeng@pku.edu.cn

protein complexes on chromatin^{19,20}. Understanding TF kinetics in the genome-wide target search is crucial for deciphering transcription regulation^{21,22}. Eukaryotic TFs primarily use three-dimensional (3D) diffusion in the nucleus but it is unclear if they use additional mechanisms for efficient chromatin target search, especially in closed chromatin where PTFs operate. Prokaryotic TFs use local sampling of DNA, known as facilitated diffusion, alongside 3D diffusion to enhance target search efficiency^{14,23}. While facilitated diffusion may occur in the open chromatin of eukaryotic cells^{16,24}, it falls short in explaining target search in closed chromatin, densely packed with nucleosomes. Despite studies on PTF kinetics^{25–28}, questions remain about the mechanism of efficient target binding in closed chromatin.

In this study, we combined live-cell fast SMT, fixed-cell multicolor photoactivated localization microscopy (PALM), computational simulations, kinetic modeling and epigenomic profiling to elucidate the dynamic process of PTFs searching for targets on chromatin in living cells. We revealed a notable ‘confined target search’ mechanism that enables PTFs to access targets in closed chromatin. In this mechanism, the PTF alternates between fast free diffusion in the nucleus and slower confined diffusion within mesoscale compacted chromatin domains in closed chromatin. This behavior, facilitated by the interaction of disordered non-DNA-binding regions with nucleosomes, enables shorter search time and repetitive binding trials to nucleosomal targets. Our findings suggest that the physical organization of chromatin influences TF kinetics in living cells and that TFs exploit local chromatin organization to enhance their activities.

Results

Fast stroboscopic photoactivation (spaSMT) reveals frequent chromatin binding by PTFs

To capture both bound molecules with low mobility and those in fast diffusion in living cells, we set up fast spaSMT microscopy at millisecond temporal resolution (Fig. 1a,b). This technique uses highly inclined and laminated optical (HILO) sheet illumination²⁹ to enhance the signal-to-noise ratio of imaging individual molecules in the nucleus. The protein of interest is tagged with HaloTag (Halo for short) and labeled with photoactivatable Janelia Fluor dyes³⁰ (Supplementary Fig. 1a). In living cells, single fluorophores are then sparsely and stochastically activated by a 405-nm laser and recorded with stroboscopic illumination to obtain sharp images. A large population of single-molecule trajectories are generated and analyzed with the kinetic modeling algorithm Spot-On³¹ to infer chromatin-bound fractions and diffusion constants.

We first used fast spaSMT to investigate a number of PTFs in comparison to non-PTFs. The examined PTFs included FOXA1 (ref. 3), FOXA2 (ref. 32) and reprogramming factors OCT4, SOX2 and KLF4, while MYC was recognized as a non-PTF⁴. Murine OCT4, SOX2, KLF4 and MYC were expressed in mouse embryonic fibroblasts (MEFs) and human FOXA1, FOXA2 and MYC were expressed in human bone osteosarcoma U2OS cells. These cells lacked endogenous expression and the Halo-tagged fusion proteins were exogenously expressed to a moderate level with the Tet-on inducible system. Additionally, the freely diffusing nuclear localization signal (NLS) and core histone H2B were assessed in U2OS cells as controls. A substantial number of fast spaSMT trajectories (10^5 – 10^6) were collected per factor and subsequently analyzed by Spot-On with two-state kinetic modeling (Fig. 1c,d, Supplementary Fig. 1b, Supplementary Video 1 and Supplementary Table 1) or vbSPT³³ (Supplementary Table 2). As anticipated, the majority of H2B (81.4%) and a minor fraction of NLS (7.9%) were found to be bound to chromatin. While the non-PTFs mMYC and hMYC showed a low chromatin-bound fraction at 12.5% and 18.7%, respectively, the PTFs exhibited higher values, with FOXA factors exhibiting the highest (FOXA1, 59.1%; FOXA2, 55.1%) even with varying protein expression levels (Supplementary Fig. 1c). Deletion of the DNA-binding domain (DBD) reduced the bound fraction to background levels, indicating that these PTFs frequently bound chromatin to sample for cognate targets.

FOXA2 has pioneering roles in formation of endoderm lineages^{32,34,35}. To further elucidate the dynamics of endogenous PTF–chromatin interactions, we generated homozygous FOXA2-Halo knock-in (KI) human embryonic stem cell (hES cell) clones and differentiated them to anterior primitive streak (APS) and subsequently definitive endoderm (DE)³⁶ (Fig. 1e). Throughout this process, FOXA2 was rapidly induced. Importantly, the introduction of the Halo did not interfere with FOXA2 expression or DE differentiation (Supplementary Fig. 1d–h and Supplementary Video 2). We next induced the KI cells to APS and DE stages, capturing 1.8×10^5 and 3.9×10^5 fast spaSMT trajectories of FOXA2-Halo, respectively. The results consistently revealed a substantial fraction of FOXA2 molecules (55–61%) in the chromatin-bound state in both APS and DE cells, as observed in two independent FOXA2-Halo KI clones (Fig. 1f and Supplementary Fig. 1i), aligning with the findings from U2OS cells. Therefore, these results indicate that the PTF molecules frequently interact with chromatin.

PTFs exhibit locally confined diffusion on chromatin

To further investigate the chromatin sampling mechanism of PTFs, we analyzed the directional preference of molecule diffusion^{37,38} (Fig. 2a). Specifically, we used a hidden Markov model (HMM) to classify segments (displacement between two sequential localizations) of molecule trajectories into bound and free states and measured the angles between two consecutive free segments. We found that all examined PTFs exhibited markedly anisotropic diffusion, with a higher probability to move backward than other directions (peaking around 180°) (Supplementary Fig. 2a,b). We quantified the degree of anisotropic diffusion by the ‘fold anisotropy’, $f_{180/0}$, and found that diffusion of PTFs showed generally high anisotropy (Fig. 2b). In contrast, nuclear proteins including non-PTF MYC showed typical isotropic diffusion. Deletion of the FOXA2 DBD reduced the degree of anisotropic diffusion to background levels, indicating that the observed anisotropic diffusion was associated with DNA-binding activities.

We next examined the degree of anisotropic diffusion as a function of the length of displacements flanking the angle (Fig. 2c,d and Supplementary Fig. 2c,d). All PTFs showed clear displacement-dependent anisotropy, peaking at displacements around 200 nm. This indicated that PTFs diffused anisotropically only at short displacements but diffused isotropically at longer displacements, suggestive of transiently confined diffusion within zones around the 200-nm mesoscale. Deletion of the DBD significantly reduced diffusion anisotropy at this scale (Fig. 2c). In contrast, diffusion of NLS and non-PTF MYC showed little anisotropy at any length of displacement. These results suggest that the PTF molecules uniquely adopt two alternating modes of target search in cells: free diffusion in the nucleus (isotropic) and confined diffusion within mesoscale zones (anisotropic).

We found that the level of confinement positively correlated with the fraction of chromatin-bound molecules (Fig. 2e). The correlation of the two measurements suggested that confined diffusion actively facilitates frequent chromatin binding. To test this, we selected state-switching trajectories (containing both free and bound states in one trajectory) and calculated the anisotropy of consecutive free segments before or after a bound state in trajectories, defined as ‘local anisotropy’, as opposed to the ‘global anisotropy’ measured from all free segments (Fig. 2a). Interestingly, PTFs showed a high level of local anisotropy (that is, confined diffusion) surrounding their chromatin-binding events (Fig. 2f and Supplementary Fig. 2e). By visualizing randomly selected state-switching trajectories, we observed that PTF molecules tended to diffuse around their previous bound sites (‘confined’ diffusion), while non-PTF MYC molecules tended to travel away from their previous bound sites randomly (‘unconfined’ diffusion) (Fig. 2g, Supplementary Fig. 2f,g and Supplementary Video 3). We also noticed that PTFs became unconfined after a few frames (7.5 ms per frame), indicating that these PTF molecules were only transiently confined in the local chromatin zone rather than permanently trapped

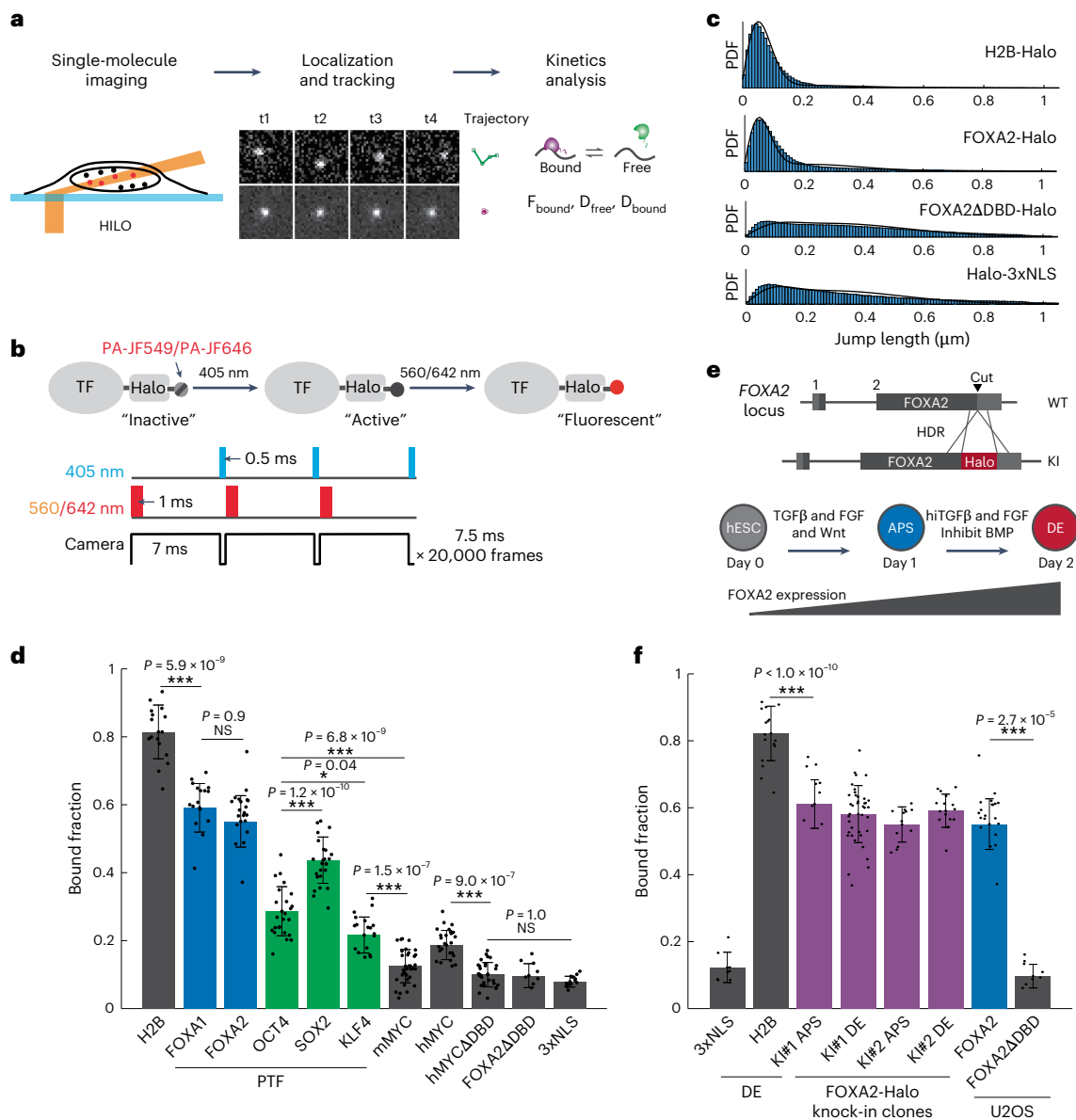


Fig. 1 PTFs sample chromatin frequently for targets by fast spaSMT.

a, Schematic diagram of SMT study. **b**, Schematic diagram of fast spaSMT with stroboscopic illumination and photoactivable fluorophore PA-JF549 or PA-JF646. **c**, The probability distribution function of jump length from all pooled trajectories of indicated factors at an imaging frequency of 133 Hz. **d**, Bound fractions (mean \pm s.d.; dots represent data points) derived from two-state Spot-On analyses on fast spaSMT trajectories of indicated Halo-tagged factors exogenously expressed in MEF cells (murine OCT4, SOX2, KLF4 and MYC) or U2OS cells (human FOXA1, FOXA2, MYC, DBD deletion mutants, H2B and 3xNLS) with Tet-on system. Colored molecules are PTFs. **e**, Top, illustration of clustered regularly interspaced short palindromic repeats (CRISPR)-Cas9-mediated KI of HaloTag to the C terminus of the endogenous FOXA2 locus in hES cells. Bottom,

differentiation of hES cell to APS and DE. TGF, transforming growth factor; FGF, fibroblast growth factor; BMP, bone morphogenetic protein. FOXA2 is not expressed in hES cells and is induced to express upon differentiation. **f**, Bound fractions (mean \pm s.d.; dots represent data points) of indicated factors in the following cell conditions: stable hES cell lines expressing Halo-tagged 3xNLS or H2B were induced to the DE state; two FOXA2-Halo homozygous KI clones were induced to the APS or DE state; U2OS cells expressed FOXA2 variants. Significance was calculated using a one-way analysis of variance (ANOVA) followed by Tukey's multiple comparison test. * $P < 0.05$ and *** $P < 0.001$. NS, not significant. For **c**, **d** and **f**, $n = 16$ –58 biologically independent cells for each sample (Supplementary Table 1).

(Fig. 2f). These results suggest that confined diffusion of PTFs is closely associated with frequent chromatin binding.

FOXA2's confined diffusion mapped to histone-interacting C-terminal domain (CTD)

We next investigated the cause and consequence of chromatin confinement by focusing on FOXA2. FOXA2 is composed of a well-structured Forkhead DBD and two intrinsically disordered regions (IDRs) known as the N-terminal domain (NTD) and the CTD (Fig. 3a). Previous studies

demonstrated that the CTDs of murine FOXA1 and FOXA2 directly interact with purified core histones and are required for nucleosome binding³. Impairment of this interaction significantly affects mouse embryonic development, altering gene activation and chromatin opening³². However, it is unclear how the CTD-histone interaction mechanistically promotes pioneer activity. We confirmed CTD-histone interactions by coimmunoprecipitation (co-IP) experiments using nuclear lysates. FOXA2 interacted with core histone H3 and H4 and deletion of its CTD abolished this interaction (Fig. 3b). Additionally,

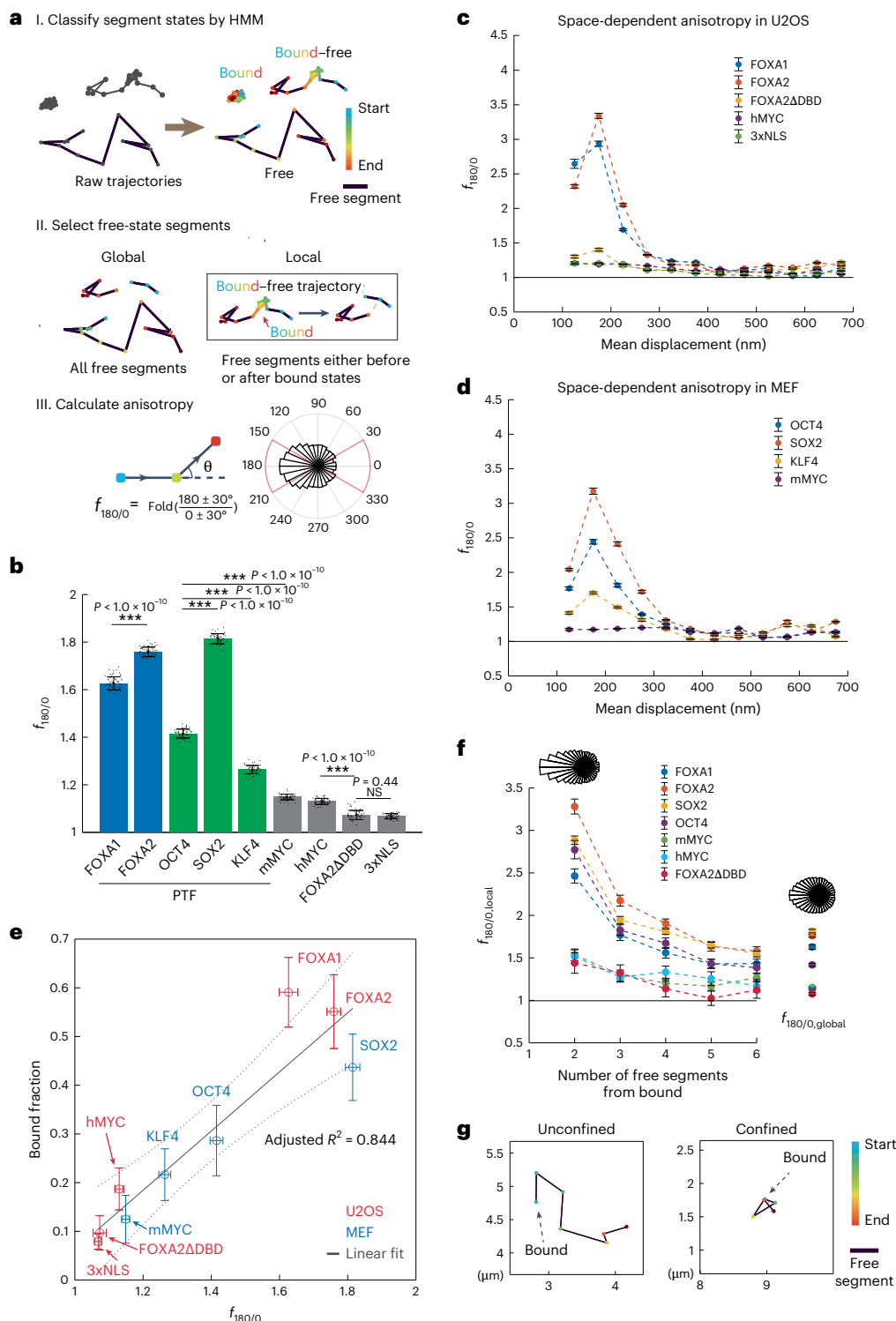


Fig. 2 | PTFs exhibit anisotropic diffusion and confined target search on chromatin. **a**, Illustration of the analysis pipeline of global and local diffusion anisotropy. The molecule localizations of exemplary trajectories are color-coded illustrating start to end and the free segments are particularly colored black. **b**, The $f_{180/0}$ values (mean \pm s.d.; dots represent data points) of indicated molecules calculated from all free segments (global anisotropy) in U2OS cells and MEF cells. Colored molecules are PTFs. **c, d** Space-dependent anisotropy in U2OS (**c**) and MEF (**d**) cells examined by $f_{180/0}$ against mean displacement length of selected free segments. Error bars show the s.e.m. from 50 subsamplings of segment angles with 50% replacement. **e**, Plots of global anisotropy $f_{180/0}$ versus bound fraction of indicated molecules. Horizontal and vertical error bars are the s.d. of $f_{180/0}$ and bound fraction, with the center denoting the mean. The gray

solid line represents the linear regression and the gray dotted line represents the 95% confidence interval of the regression. The adjusted R^2 was calculated by adjusting the number of data points. **f**, The $f_{180/0, \text{local}}$ values (mean \pm s.d.) were calculated from a variable number of free segments before and after a binding event (local anisotropy, left) and compared to the $f_{180/0}$ values of all free segments (global anisotropy, right). Top, the representative angle distribution histograms of FOXA2. **g**, Representative single-molecule trajectories showing diffusion in either a confined or an unconfined manner after a binding event. Localizations and segments are colored as in **a**. Significance was calculated using a one-way ANOVA followed by Tukey's multiple comparison test. * $P < 0.05$ and *** $P < 0.001$. For **b–f**, $n = 13–58$ biologically independent cells for each sample (Supplementary Table 1).

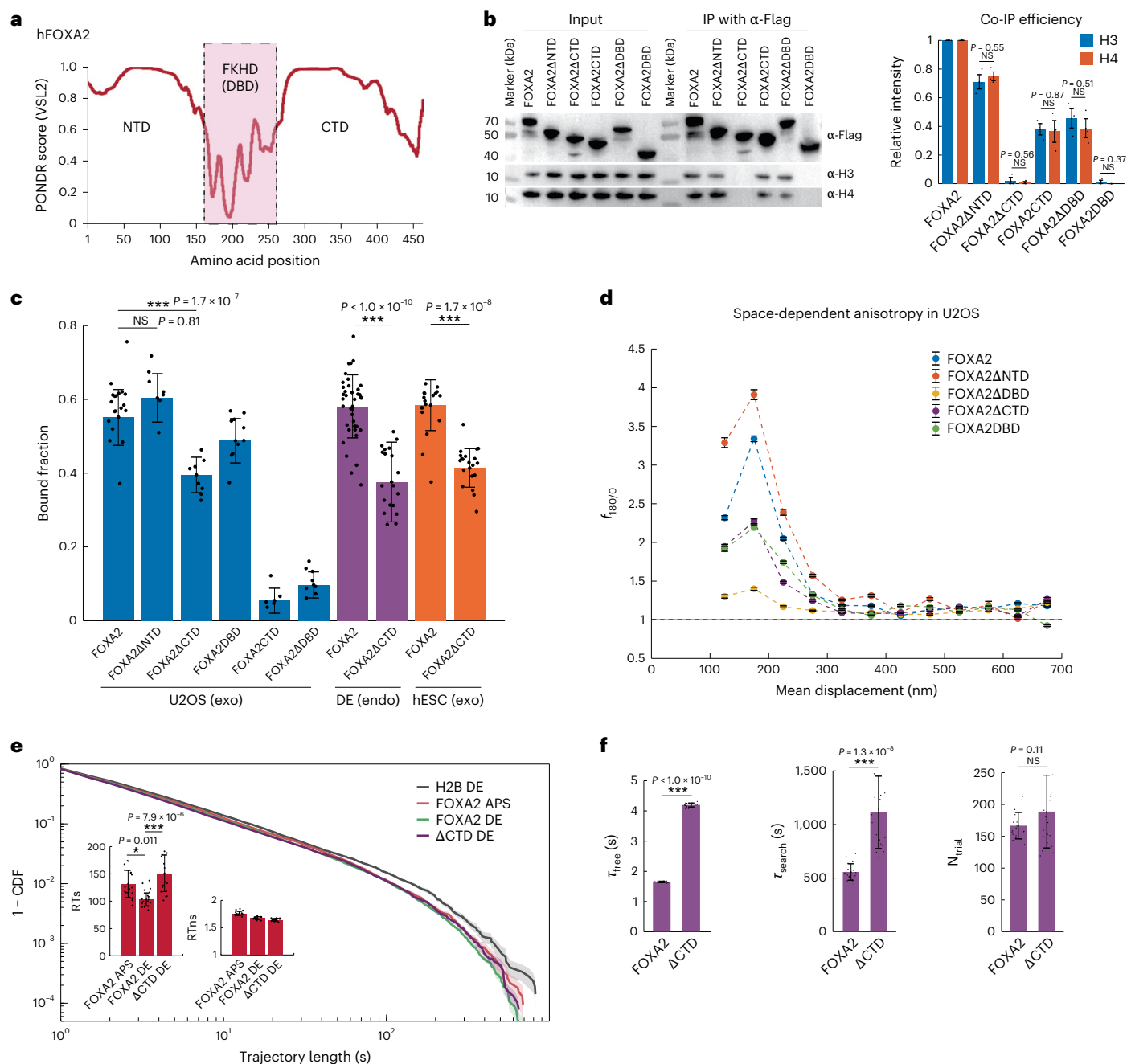


Fig. 3 | CTD_{FOXA2} is essential for FOXA2 anisotropic diffusion and target search. **a**, Predictor of naturally disordered regions (POND) scores of the human FOXA2 protein. **b**, Co-IP using nuclear lysates of U2OS cells expressing Flag-tagged FOXA2 variants. Representative western blots (left) and quantification (mean \pm s.e.m.; dots represent data points) of three replicates (right) are shown; $n = 3$ biologically independent samples. **c**, Bound fractions (mean \pm s.d.; dots represent data points) derived from two-state Spot-On analyses of fast spaSMT trajectories for indicated molecules in indicated cells. The terms ‘exo’ and ‘endo’ indicate exogenous and endogenous expression with KI clones, respectively. **d**, Space-dependent anisotropy examined by $f_{180/0}$

against mean displacement length of selected free segments. Error bars show the s.e.m. from 50 subsamplings of segment angles with 50% replacement. **e**, Left, log–log survival curves of slowSMT trajectories of the indicated molecules. Inset, corrected RTs and RTns (mean \pm s.e.m.; dots represent data points) of endogenously tagged FOXA2 or Δ CTD in DE cells. **f**, Comparison of FOXA2 and Δ CTD target search kinetics (mean \pm s.d.; dots represent data points). Significance was calculated using a one-way ANOVA followed by Tukey’s multiple comparison test (**c,e**) or a two-sided t -test (**b,f**). * $P < 0.05$ and *** $P < 0.001$. For **c–f**, $n = 16–58$ biologically independent cells for each sample (Supplementary Table 1).

CTD alone or deletion of DBD exhibited weak interaction with core histones compared to full-length FOXA2, indicating that DNA binding enhances CTD–histone interactions.

We investigated the contribution of FOXA2 domains to its chromatin association dynamics using fast spaSMT. Deletion of the DBD eliminated specific chromatin association, as expected for a TF (Fig. 3c). Remarkably, deletion of the CTD significantly decreased the bound

fraction of FOXA2 on chromatin (from 55.1% to 39.5%) (Fig. 3c) and also reduced confined diffusion ($f_{180/0,200\text{nm}}$: from 3.34 to 2.27) in U2OS cells (Fig. 3d). The global and local anisotropy of its diffusion were similarly reduced (Supplementary Fig. 3a–c). In comparison, deletion of the NTD did not decrease but even slightly increased the chromatin-bound fraction and diffusion anisotropy, suggesting inhibitory roles of NTD in chromatin association. Post-translation modification in the NTD

such as acetylation has been reported to attenuate FOXA1 nucleosome binding and remodeling binding³⁹. Our focus remained on the CTD. Therefore, we generated homozygous hES cell clones with CTD deletion (Δ CTD-Halo) based on the FOXA2-Halo KI hES cell clone (Supplementary Fig. 3d). The expression level of Δ CTD-Halo was the same as that of full-length FOXA2-Halo in stage-matched DE cells. Consistent with the results from U2OS cells, CTD deletion significantly reduced the chromatin-bound fraction (from 58.1% to 37.6%) (Fig. 3c) and confined diffusion in DE cells (Supplementary Fig. 3e). Experiments with exogenously expressed FOXA2 and Δ CTD in hES cells also produced consistent results (Fig. 3c and Supplementary Fig. 3f). These results indicate that the CTD has essential roles in frequent chromatin association and confined diffusion of FOXA2.

We further explored the contribution of CTD to stable target binding using slowSMT. The slowSMT imaging module, with a long exposure time (500 ms) and low laser power, selectively captures stably bound molecules while blurring fast-diffusing ones (Supplementary Videos 4 and 5). Using two-exponential fitting and photobleaching correction with H2B (Supplementary Fig. 3g), we derived the residence time of FOXA2-specific binding (RTs) around 2 min and that of nonspecific binding (RTns) around 2 s (Fig. 3e). Notably, our results indicated that FOXA2 has a more extended RTs than other reported TFs²⁰, similar to that of the GAGA factor²⁸. Importantly, deletion of the CTD did not reduce RTs and RTns, showing that, while the CTD has essential roles in target searching, it has little effect on the stability of target binding. We observed a slight increase in RTs upon CTD deletion, which could be attributed to the Δ CTD favoring open chromatin and exhibiting a stronger binding affinity for naked DNA compared to nucleosomal DNA. This finding is distinct from previous observations showing that IDRs increase the TF binding time on targets^{28,40}. By calculating the target search kinetics with the 3D diffusion-dominant model¹⁸, we found that deletion of the CTD resulted in a twofold increase in both free diffusion time (τ_{free}) and target search time (τ_{search}) between two specific binding events (Fig. 3f, Supplementary Fig. 3h and Supplementary Table 3), suggesting a less efficient target search. These findings indicate that, while the DBD of FOXA2 determines the DNA-binding affinity, its non-DNA-binding and histone-interacting CTD promotes its confined diffusion, frequent chromatin sampling and efficient target search.

Disordered CTD boosts pioneer activity

The unique kinetics of the PTF–chromatin interaction prompts the question of its contribution to pioneer activities, including chromatin binding and chromatin opening. To avoid indirect effects, we expressed Flag-tagged FOXA2 and Δ CTD with the Tet-on system for 4 days in U2OS cells and performed chromatin IP with sequencing (ChIP-seq) and assay for transposase-accessible chromatin with sequencing (ATAC-seq) (Fig. 4a,b, Supplementary Fig. 4a–f and Supplementary Table 4). U2OS cells were chosen as they did not express any of the FOXA family proteins. FOXA2 functioned as a typical PTF in U2OS cells, as 86% of its binding sites were initially in closed chromatin and 43% of them were induced to an open state (Fig. 4a). In comparison, its Δ CTD mutant showed clear defects in pioneer activity, losing stable binding at 48.7% of FOXA2 sites (Fig. 4b and Supplementary Fig. 4a) and displaying a deficiency in inducing chromatin opening, even at Δ CTD-bound sites (Fig. 4a and Supplementary Fig. 4f,g). Notably, the impaired binding upon CTD deletion was preferentially in closed chromatin rather than open chromatin (Fig. 4a). These results show that the CTD of FOXA2 greatly enhances its pioneer activity of binding and opening closed chromatin target sites.

The mammalian FOXA factors, FOXA1, FOXA2 and FOXA3, share well-conserved DBDs but have less conserved CTDs, with CTD_{FOXA3} being the shortest and sharing the least similarity (Fig. 4c). Using fast spaSMT and ChIP-seq, we investigated the chromatin binding of FOXA factors and corresponding Δ CTD variants in U2OS cells

(Supplementary Fig. 4h,i). Our results showed that FOXA3 had a lower chromatin-bound fraction (FOXA1, 58.0%; FOXA2, 58.9%; FOXA3, 30.9%) and weaker space-dependent anisotropic diffusion compared to FOXA1 and FOXA2 (Fig. 4d,e). Deletion of CTD_{FOXA1} and CTD_{FOXA2} led to reduced chromatin-bound fractions, while deletion of CTD_{FOXA3} did not. The ChIP-seq profiles of the three FOXA factors were also distinct (Supplementary Fig. 4j,k). FOXA1 and FOXA2 bound extensively to closed chromatin sites, whereas FOXA3 had fewer binding sites and tended to bind open chromatin or relatively accessible regions (Fig. 4f and Supplementary Fig. 4l). FOXA1 and FOXA2 bound to closed chromatin lacking distinguishing histone marks (Supplementary Fig. 4m), consistent with previous observations¹². Hence, FOXA1 and FOXA2 appeared to have stronger pioneer activity than FOXA3 at least in this cellular context. Like FOXA2, deletion of CTD_{FOXA1} led to substantial loss of specific FOXA1 binding preferentially in closed chromatin (Fig. 4f). Thus, the disordered CTDs differentiated FOXA factors in their single-molecule confinement dynamics and the pioneer capability of binding targets in closed chromatin. These results again emphasize the enhancing roles of the disordered CTDs in chromatin confinement and targeting closed chromatin.

Lastly, FOXA2 has a pioneering role in the formation and differentiation of DE and their derivatives, both in hES cell in vitro differentiation models³⁵ and in mouse models³². Deletion of the histone-interacting region of CTD in mouse FOXA2 impaired embryonic development and altered gene activation and the accessible chromatin landscape³². Thus, we aimed to examine the roles of FOXA2 and its CTD in hES cell differentiation to DE. For this purpose, we generated FOXA2 knockout (KO) hES cell clones in addition to Δ CTD clones and conducted RNA sequencing (RNA-seq) analyses on these cells at the hES cell, APS or DE stages (Supplementary Fig. 4n,o). In examining the gene network transition from hES cells to APS and DE, we found that KO cells displayed minor defects in activating APS-specific genes and more apparent defects in activating DE-specific genes (Fig. 4g) and that CTD was required for the normal expression of a subset of them, including *SOX17*, *FOXA1* and *CXCR4* (groups I and III; Fig. 4h and Supplementary Table 6). These results highlight the importance of CTD for normal DE-specific gene expression.

Interaction of FOXA2 with mesoscale chromatin domains

We next examined the spatial relationship of FOXA2 with chromatin by super-resolution microscopy. Specifically, we used two-color PALM to simultaneously detect FOXA2 (tagged with Halo and stained by photoactivatable PA-JF646) and the core histone protein H2B (tagged with photoconvertible fluorescent protein mEosEM⁴¹) in paraformaldehyde-fixed U2OS cells (Fig. 5a and Supplementary Fig. 5a–c). Cells with a comparable expression level were sorted and selected for imaging experiments. H2B PALM exhibited a higher localization density at the nuclear periphery (Fig. 5a and Supplementary Fig. 5d), where heterochromatin was frequently located, indicating a high degree of chromatin compaction. To quantitatively analyze the cluster sizes of different FOXA2 variants, we used SR-Tesseler⁴², a robust and unbiased segmentation algorithm, and found that full-length FOXA2 formed larger clusters than FOXA2 Δ CTD and FOXA2 Δ DBD (Supplementary Fig. 5e). Notably, measured by Manders correlation, FOXA2 PALM in the same cells showed significantly colocalized distribution with H2B PALM, especially at the nuclear periphery (Fig. 5b and Supplementary Fig. 5f). In contrast, Δ CTD showed a lower degree of periphery colocalization with H2B and Δ DBD led to even nuclear distribution. Therefore, the two-color PALM results agree with the genomic results of FOXA2 targeting closed chromatin regions in a CTD-dependent manner.

We further investigated the super-resolved characteristics of FOXA2-bound chromatin with PALM. While nucleosomes displayed discrete nanodomains with a width of a few tens of nanometers (Fig. 5a)⁴³, we hypothesized that the observed chromatin confinement of FOXA2

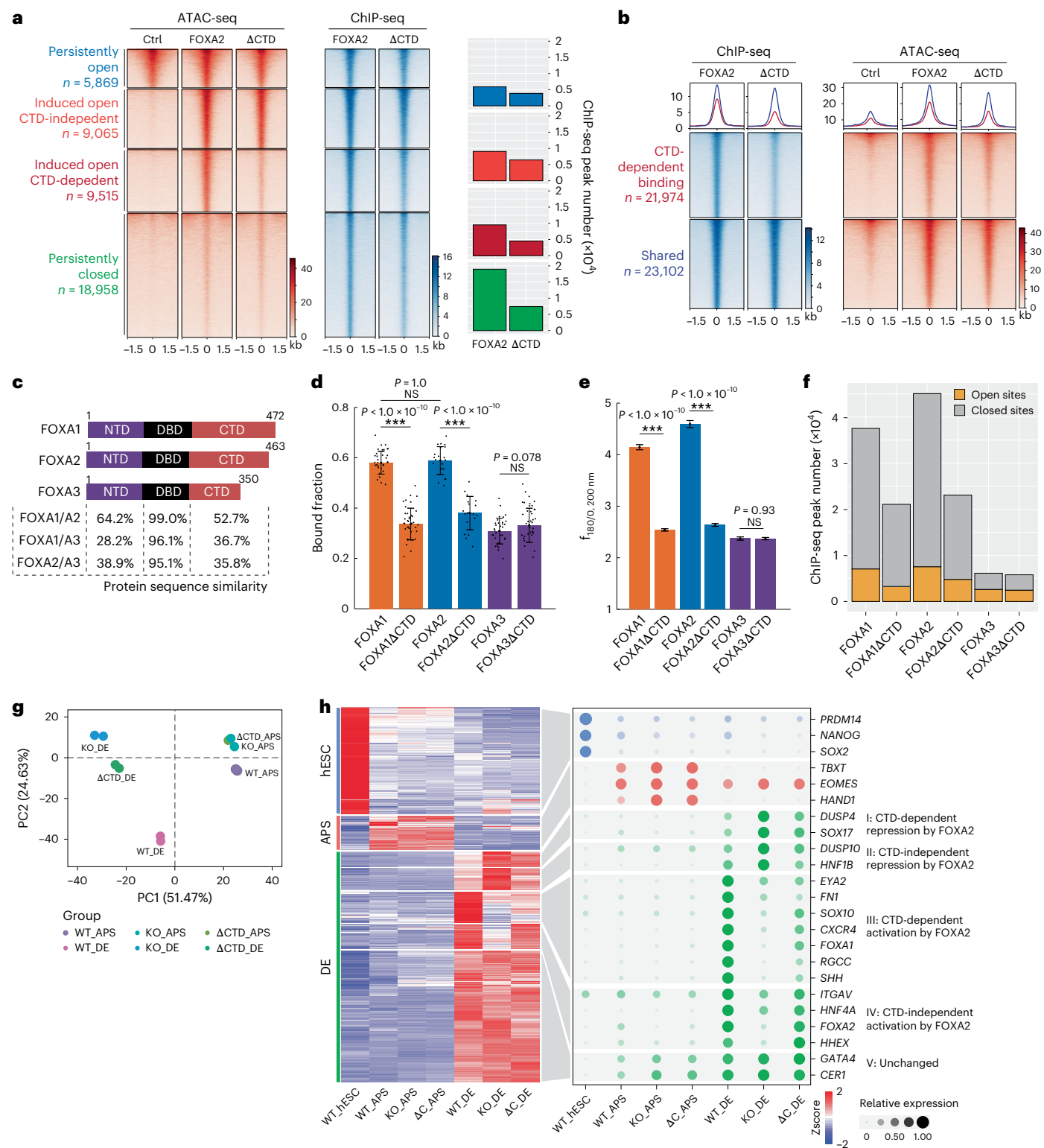


Fig. 4 | FOXA CTD enhances pioneering closed chromatin and is required for normal DE-specific gene expression. a, Heat maps of ATAC-seq signal and ChIP-seq signal of indicated samples centered by the FOXA2-binding sites, classified by the change in ATAC-seq signal after 4-day expression of FOXA2. The number of ChIP-seq peaks in each group is plotted on the right ($n = 2$ biological replicates). **b**, Heat maps of ChIP-seq and ATAC-seq signal of indicated samples centered by the FOXA2-binding sites, classified by the change in ChIP-seq signal. **c**, Protein sequence similarity of human FOXA factors. **d**, Bound fractions (mean \pm s.d.; dots represent data points) derived from two-state Spot-On analyses of fast spaSMT trajectories for indicated molecules in U2OS cells. **e**, The $f_{180/0,200nm}$ values (peak values of space-dependent anisotropy at 200 nm;

mean \pm s.e.m.) of indicated molecules in U2OS cells. Error bars show the s.e.m. from 50 subsamplings of segment angles with 50% replacement. **f**, Numbers of ChIP-seq peaks in either open or closed chromatin based on ATAC-seq of WT U2OS cells. Shared ChIP-seq peaks of FOXA and FOXA Δ CTD are plotted for the Δ CTD samples. **g**, PCA of indicated RNA-seq samples with differentially expressed genes. **h**, Left, RNA-seq heat map of FOXA2 putative targets classified by stage-specific expression in hES cells (259 genes), APS (83 genes) and DE (546 genes). Right, bubble plot of representative genes in five subgroups of DE genes ($n = 2$ biological replicates). Significance was calculated using a one-way ANOVA followed by Tukey's multiple comparison test. *** $P < 0.001$. For **d** and **e**, $n = 33$ –58 biologically independent cells for each sample (Supplementary Table 1).

was related to chromatin structure at a larger physical scale, as the confined diffusion zones had an averaged radius of ~200 nm (Fig. 2c). Thus, we used the hidden Markov random field model (HMRF) to classify chromatin in the nucleus into seven compaction levels and subsequently into three functional compartments representing different chromatin activities on the basis of previous characterizations⁴⁴ (Fig. 5c,d and Supplementary Fig. 5g). The resulting chromatin compaction landscape featured chains of mesoscale chromatin domain (CD) regions having high nucleosome density (classes 4–7), broader perichromatin (PC) regions at the outer rim of the CD having lower nucleosome density (classes 2 and 3) and interchromatin (IC) regions depleted of nucleosomes (class 1). By measuring FOXA2 molecule localizations or cluster centroids in chromatin regions of different compaction levels, we found that FOXA2 could bind and even preferentially bound the compacted closed chromatin regions (Fig. 5e and Supplementary Fig. 5h,i). In comparison, the Δ CTD showed less density enrichment in CDs, consistent with its impaired pioneer activity.

We further subdivided CD regions into single CDs and identified their centroids with the strongest nucleosome intensity. Each cell at the focal plane had ~3,000 CDs (Supplementary Fig. 5j) with an average area of 4.4×10^4 nm² and diameter of 240 nm (Fig. 5f), which was similar to the physical size of confinement zones (Fig. 2c). The coincidence suggested that organization of individual CDs might underlie observed chromatin confinement of FOXA2, in consideration of the tight correlation among chromatin confinement, chromatin-bound fraction and successful binding at closed chromatin sites. By quantifying the distances between the CD centroids and FOXA2 centroids, we showed that FOXA2 spanned the entire CD entity with a slight preference to the periphery (Fig. 5g,h). Notably, we observed the confinement trajectories of single FOXA2 molecules within and surrounding the CDs by combining live-cell PALM and fast spaSMT techniques (Fig. 5i and Supplementary Fig. 5k). In addition to using model-based chromatin segmentation, we took a parallel analysis approach by dividing the nucleus into nine equal-area regions with increasing H2B density and obtained consistent conclusions (Supplementary Fig. 5l,m). Together, these results provide super-resolution imaging evidence that FOXA2 targets mesoscale compacted chromatin domains and that the CTD facilitates this process. Importantly, our findings suggest that the interaction of FOXA2 with mesoscale chromatin domains might underlie the diffusion confinement within a characteristic space and result in efficient pioneer activity.

Kinetic advantage of confined target search

We next investigated how PTFs might gain kinetic advantages from confined diffusion in binding energetically unfavorable and nucleosome-embedded targets. To simulate the confined target search process, we conducted coarse-grained computer simulations of chromatin as a chain of monomers, partially containing a cognate binding site (CBS) of the TF^{38,45}. The simulations assigned a monomer radius of 200 nm to model the spatial volume of transient chromatin confinement observed in experiments (Fig. 6a, Supplementary Fig. 6a, Supplementary Table 7 and Supplementary Video 6). The TF underwent Brownian motion in the nuclear space between chromatin monomers (free nuclear diffusion); once it entered a monomer, it diffused inside at a lower diffusion coefficient and was confined with a certain probability to exit ($P_{\text{exit}} < 1$) (confined diffusion). While diffusing inside the chromatin monomers, the TF could pause briefly at a given rate to model nonspecific binding and have chances to contact the CBS within.

Synthetic TF trajectories were generated by assigning variable P_{exit} levels to model different levels of confinement, with the lowest P_{exit} representing the strongest confinement. The space-dependent diffusion anisotropy from synthetic TF trajectories showed striking resemblance to that from the experimental data of TFs, peaking around 200 nm (Fig. 6b,c). The pattern of space-dependent anisotropy remained with a variable virtual frame rate (Supplementary Fig. 6b)

and showed an inverse correlation with P_{exit} (Fig. 6d), confirming that the space-dependent trajectory anisotropy measured diffusion confinement. With an increasing level of confinement, the target search mechanism of TFs gradually switched from being dominated by free diffusion to being dominated by confined diffusion (Fig. 6e), resulting in a higher target search efficiency (Fig. 6f) and repetitive trials of target binding (Fig. 6g), which might be essential for PTF to disturb local nucleosome compaction and increase the chance of binding a partially exposed motif on nucleosomes. Using $f_{180/0,200\text{nm}}$ values measured from experiments, we extrapolated the corresponding P_{exit} of TFs and calculated their target search kinetics (Fig. 6d and Supplementary Table 8). The PTFs FOXA1, FOXA2 and SOX2 showed the strongest diffusion confinement, with a predicted P_{exit} of FOXA2 at 0.17 and a fraction of confined diffusion at 23% (Fig. 6e). In contrast, the predicted P_{exit} values of MYC, NLS and FOXA2 Δ DBD were close to 1, indicating little confinement. Compared to FOXA2, its Δ CTD mutant showed a higher predicted P_{exit} at 0.3, a lower fraction of confined diffusion at 15%, a lower on rate to CBS and less repetitive binding to the CBS. The changes in target search kinetics may explain its impaired pioneer activity in binding targets in closed chromatin.

We speculated that extremely strong confinement (a very small P_{exit}) could trap TFs unproductively within monomers without a CBS for extended periods, negatively impacting the overall efficiency of target search throughout the nucleus. Therefore, we calculated the search time between binding two different CBSs in separate monomers and found that extremely small or large P_{exit} values indeed resulted in a long search time (Fig. 6h). Importantly, we identified an optimal P_{exit} that minimized the search time, thereby maximizing efficiency. Notably, the predicted P_{exit} values of FOXA1, FOXA2 and SOX2 fell around the optimal P_{exit} . These results suggest that PTFs may have evolved their protein–nucleosome interaction to achieve an optimal degree of confinement for efficient target searching without becoming overly trapped.

Additionally, we explored an alternative model wherein the observed confinement could be attributed to repetitive binding to specific target sites that are clustered. To test this hypothesis, we modified the chromatin simulation by placing multiple binding sites (5, 50 or 100 CBSs) clustered within a single 200-nm-radius chromatin monomer (Supplementary Fig. 6a). Subsequently, we performed the diffusion simulation without confinement by setting P_{exit} to 1 and examined the anisotropy of the diffusion fraction in simulated molecule trajectories as described earlier. The results revealed an absence of spatial confinement of the diffusing molecules (Supplementary Fig. 6c), despite the default clustering of bound molecules in this case. Therefore, it became evident that clustered target sites could not explain the observed anisotropy of diffusing PTFs.

Modulation of target search by non-DNA-binding regions

Our results establish the essential roles of non-DNA-binding disordered regions in target search dynamics and pioneer activity. Thus, we speculated that such activity could be used to alter the pioneer activity of other TFs through protein fusion. Indeed, fusion of CTD_{FOXA1} or CTD_{FOXA2} to FOXA3 Δ CTD greatly increased the chromatin-bound fraction (from 33.2% to 45.8% and 51.7%) and space-dependent anisotropy ($f_{180/0,200\text{nm}}$ from 2.37 to 2.94 and 3.23) as measured by fast spaSMT (Fig. 6i). This finding indicated that CTD_{FOXA1/2} could increase the chromatin interaction and confined diffusion of a homologous DNA-binding protein. To further test whether this strategy can be applied to TFs from a different family, we used the Tet-on system to express SOX2 and a fusion protein of SOX2 and CTD_{FOXA2} (SOX2-FC) in U2OS cells (Supplementary Fig. 6d). According to fast spaSMT in living cells, SOX2-FC showed a higher chromatin-bound fraction and higher diffusion anisotropy than SOX2 (Fig. 6j), indicating that addition of CTD further enhanced chromatin interaction and chromatin confinement. We further examined the TFs using ChIP-seq and ATAC-seq. Notably, while SOX2 functioned as a typical PTF, binding targets that

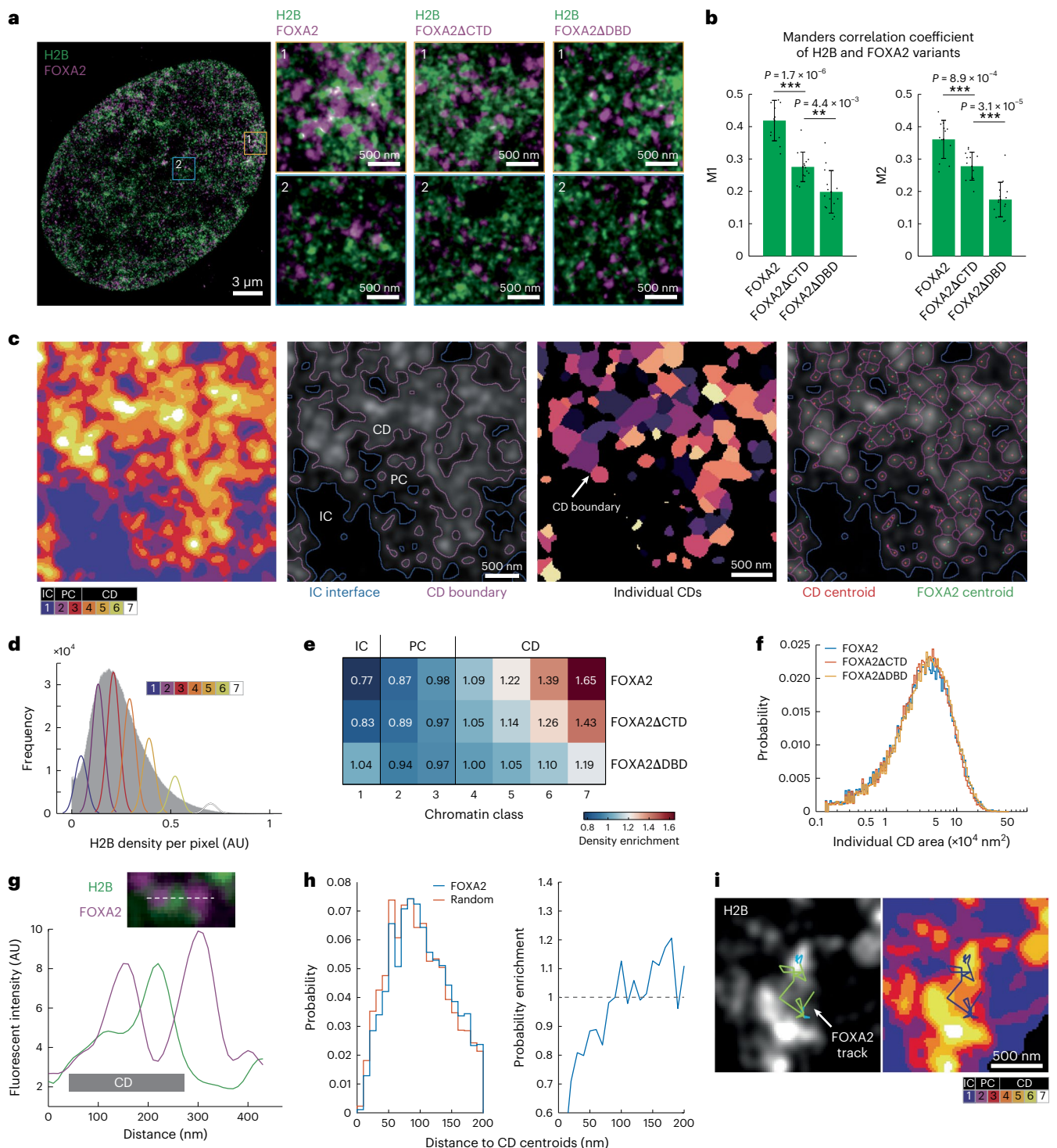


Fig. 5 | Two-color PALM reveals that FOXA2 targets compacted chromatin domains enriched for nucleosomes. a, Two-color PALM images of fixed U2OS cells that express mEosEM-tagged H2B and Halo-tagged FOXA2 variants labeled with PA-JF646. Blue and yellow boxes indicate representative peripheral and central regions showing the spatial relationship between chromatin and FOXA2 variants. **b**, Manders correlation coefficients (mean \pm s.d.; dots represent data points) for the whole cell show the colocalization of FOXA2 variants and H2B in both directions with auto thresholds; $n = 12$ (FOXA2) or $n = 14$ (FOXA2ΔCTD and FOXA2ΔDBD) biologically independent cells. **c**, Process of model-based chromatin classification and analysis. Column 1, segmentation of chromatin into seven intensity classes based on HMRF model. Column 2, blue and magenta contours partition chromatin into three compartments: IC (class 1), PC (classes 2 and 3) and CD (classes 4–7). Column 3, separation of individual chromatin domains. Column 4, CD centroids (red) and FOXA2 centroids (green)

relative to chromatin domains. **d**, Histogram of the H2B intensity of the example nucleus image in gray, overlaid with probability density functions of the seven Gaussians generated by the HMRF model. AU, arbitrary units. **e**, Enrichment or depletion of the density of FOXA2 variants relative to a random distribution for each chromatin class. **f**, Area distribution of individual CDs. **g**, Representative image and intensity line plots of a chromatin domain in relation to FOXA2. **h**, Distribution of distance to the CD centroids from FOXA2 centroids compared to (left) or divided by (right) random localizations. **i**, Overlaid image of live-cell PALM of H2B and fast spaSMT of FOXA2. The trajectory of FOXA2 is color-coded in green and blue representing the ‘free’ and ‘bound’ states, respectively. Significance was calculated using a one-way ANOVA followed by Tukey’s multiple comparison test. $**P < 0.01$ and $***P < 0.001$. For **d–h**, $n = 12$ (FOXA2) or $n = 14$ (FOXA2ΔCTD and FOXA2ΔDBD) biologically independent cells.

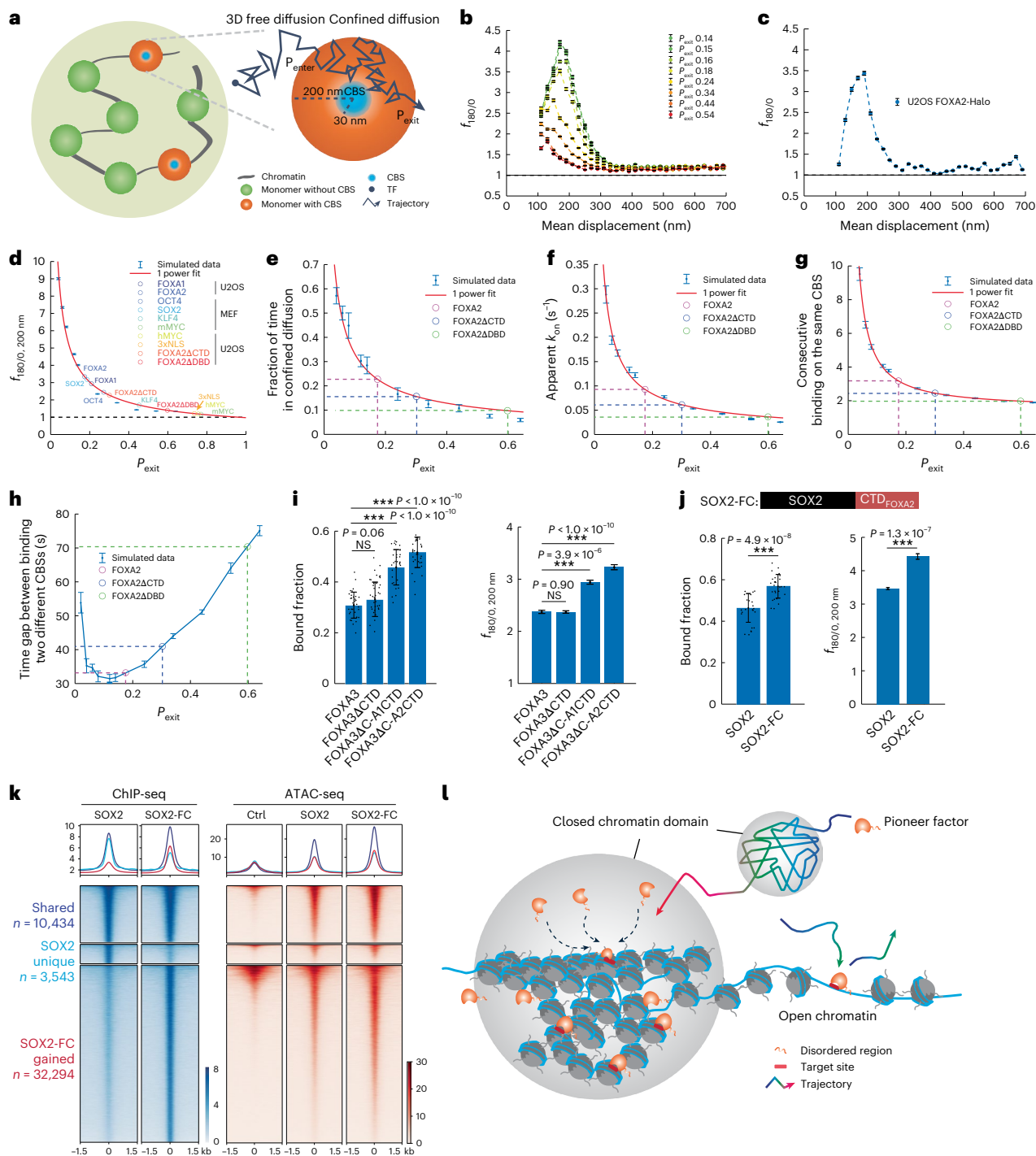


Fig. 6 | Computer simulations reveal the kinetic advantage of confined target search. **a**, Chromatin is modeled as a chain of monomers with or without CBS. TF undergoes 3D free diffusion outside of the monomers and confined diffusion within the monomers with certain probabilities to enter (P_{enter}) and exit (P_{exit}). **b,c**, Space-dependent anisotropy examined by $f_{180/0}$ against mean displacement length of selected free segments, using simulation-generated trajectories with variable P_{exit} (**b**) or experimental data of FOXA2 (**c**). **d**, The plot of space-dependent anisotropy, measured by $f_{180/0,200\text{nm}}$ (mean \pm s.e.m.), as a function of P_{exit} . The plot was fitted by a one-power model and the P_{exit} of indicated molecules was extrapolated using experimental $f_{180/0,200\text{nm}}$ values. **e-g**, Indicated kinetics calculated using simulated trajectories at various P_{exit} levels. The simulated data were well fitted by the one-power model and the corresponding y-axis values of indicated molecules were calculated with extrapolated P_{exit} from **d**; $n = 12$ independent simulations. **h**, Time gap between binding two different targets (two CBSs in different monomers) calculated using the simulated trajectories

at various P_{exit} levels, overlaid with estimated values of indicated molecules using the extrapolated P_{exit} ; $n = 12$ independent simulations in **b** and **d-h**. For each round of simulations, there were 5,000 diffusion steps per round (**b,d,e**) or 500 iterations of CBS binding per round (**f-h**). Data are presented as the mean \pm s.d. in **e-h,i,j**. Bound fractions (mean \pm s.d., dots represent data points) and $f_{180/0,200\text{nm}}$ (mean \pm s.e.m.) derived from fast spaSMT trajectories for indicated molecules in U2OS cells. FOXA3 Δ C-A1CTD and FOXA3 Δ C-A2CTD represent the fusion proteins of FOXA3 Δ C with the CTDs of FOXA1 and FOXA2, respectively. SOX2-FC represents the fusion protein of SOX2 and the CTD of FOXA2. **k**, The heat maps and averaged profiles of ChIP-seq (left) and ATAC-seq (right) of U2OS cells expressing indicated proteins for 4 days. Ctrl, WT U2OS cells ($n = 2$ biological replicates). **l**, The working model of PTF target search. Significance was calculated using a one-way ANOVA followed by Tukey's multiple comparison test (**i**) or a two-sided t -test (**j**). *** $P < 0.001$. For **i,j**, $n = 23-36$ biologically independent cells for each sample (Supplementary Table 1).

were originally in a closed state and increasing their chromatin accessibility, SOX2-FC exhibited remarkably enhanced pioneer activity. This was evident in its ability to bind significantly more SOX2 target sites in closed chromatin and result in a higher level of chromatin opening (Fig. 6k and Supplementary Fig. 6e,f). Importantly, the gained binding sites of SOX2-FC were bona fide SOX2 targets that contained a SOX2 consensus motif and shared little overlap with FOXA2-binding sites (Supplementary Fig. 6g). These results highlighted a significant enhancement of SOX2's pioneer activity through fusion with FC. Thus, a non-DNA-binding disordered region could promote TF targeting of closed chromatin, offering a method to improve pioneer activity by altering the target searching process.

Discussion

Here, we revealed a previously unrecognized confined target search mechanism used by PTFs to target compacted closed chromatin (Fig. 6l). Unlike the previously identified 3D diffusion-dominant model, PTFs alternate between nonexhaustive 3D free diffusion in the nucleus and confined diffusion in mesoscale compacted chromatin domains. This reduces the target search time and enables PTFs to repetitively probe nucleosomal targets, thereby enhancing pioneer activity in closed chromatin. Importantly, this process depends on the non-DNA-binding disordered regions of PTFs, facilitating confined diffusion through flexible and frequent interactions with core histones enriched in closed chromatin. Our findings also provide insights for identifying new natural PTFs and engineering synthetic ones.

Our results underscore the crucial role of chromatin spatial organization in the target search of TFs⁴⁶, particularly PTFs. While the epigenetic determinants guiding PTF binding to nucleosomal targets remain unclear¹², our super-resolution imaging revealed the presence of spatially organized mesoscale compacted chromatin domains, acting as landing pads that attract PTFs through interactions with their disordered regions. Subsequent transient diffusion confinement within these chromatin domains further reduces the search space required to find targets. Additionally, it increases the frequency of binding attempts by PTFs, potentially alleviating local nucleosome compaction and transiently exposing buried motifs for stable PTF binding^{47,48}. This mechanism aligns with prior observations including slow nuclear mobility of FOXA factors^{25,26}, nonspecific chromatin sampling of FOXA2 (ref. 12), Zelda hubs²⁷ and OCT4 interactions with histones⁴⁸, suggesting a shared mechanism among PTFs in leveraging chromatin organization for target search in closed chromatin. Hence, our study unveils a remarkable mechanism through which PTFs exploit chromatin spatial architecture to access closed chromatin.

Our findings highlight the essential role of disordered regions in pioneer function. The nucleosomal binding capability of PTFs is generally thought to be encoded within the well-structured DBD and the fine structural features of PTF–nucleosome interactions have been characterized^{2,49}. As the ternary complex of TF-bound nucleosome is relatively unstable, a TF with strong DNA affinity could likely stabilize its binding to a partially dissociated nucleosome. Consistently, our results revealed that FOXA2 remained bound to targets in living cells for a relatively long period of time. This observation is consistent with the extended RT of *Drosophila* GAGA factor, which is also a PTF²⁸. However, the DBD alone may not guarantee pioneer activity, as TFs from the same family with highly similar DBDs do not always function as PTFs; in line with this notion, FOXA1, FOXA2 and FOXA3 showed distinct levels of capacity to bind closed chromatin in the same cellular context. TFs commonly contain IDRs that mediate multivalent interactions and facilitate a wide variety of biological activities, including their binding dynamics^{40,50–53}. The potential contribution of diverse disordered regions to pioneer activity has only recently begun to be unraveled^{3,28,48,54,55}. In this study, we found that the IDR of FOXA2 had a surprisingly important role in the dynamic process of target search, particularly for closed targets. While it promoted

target association rate, its deletion did not seem to affect dissociation rate. This is different from the IDR of GAGA factor²⁸, suggesting that it is unlikely all PTFs use exactly the same mechanism. Moreover, the structured DBD and the disordered non-DBD regions could function as modules, contributing to stable target binding and target search in closed chromatin, respectively, although one indeed influences the other. Given that PTFs are often mutated in both DBD and non-DBD regions across various cancer types, comprehending the fundamental properties of PTFs holds notable value in human cancer research. Our study encourages further exploration of the diverse disordered regions within other PTFs.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41594-024-01385-5>.

References

- Klemm, S. L., Shipony, Z. & Greenleaf, W. J. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207–220 (2019).
- Zhu, F. et al. The interaction landscape between transcription factors and the nucleosome. *Nature* **562**, 76–81 (2018).
- Cirillo, L. A. et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
- Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
- Zaret, K. S. Pioneer transcription factors initiating gene network changes. *Annu. Rev. Genet.* **54**, 367–385 (2020).
- Balsalobre, A. & Drouin, J. Pioneer factors as master regulators of the epigenome and cell fate. *Nat. Rev. Mol. Cell Biol.* **23**, 449–464 (2022).
- Luzete-Monteiro, E. & Zaret, K. S. Structures and consequences of pioneer factor binding to nucleosomes. *Curr. Opin. Struct. Biol.* **75**, 102425 (2022).
- Michael, A. K. et al. Mechanisms of OCT4–SOX2 motif readout on nucleosomes. *Science* **368**, 1460–1465 (2020).
- Echigoya, K. et al. Nucleosome binding by the pioneer transcription factor OCT4. *Sci. Rep.* **10**, 11832 (2020).
- Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J. & Cramer, P. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **580**, 669–672 (2020).
- Tanaka, H. et al. Interaction of the pioneer transcription factor GATA3 with nucleosomes. *Nat. Commun.* **11**, 4136 (2020).
- Donaghey, J. et al. Genetic determinants and epigenetic effects of pioneer-factor occupancy. *Nat. Genet.* **50**, 250–258 (2018).
- Lelek, M. et al. Single-molecule localization microscopy. *Nat. Rev. Methods Prim.* **1**, 39 (2021).
- Elf, J., Li, G.-W. & Xie, X. S. Probing transcription factor dynamics at the single-molecule level in a living cell. *Science* **316**, 1191–1194 (2007).
- Gebhardt, J. C. M. et al. Single-molecule imaging of transcription factor binding to DNA in live mammalian cells. *Nat. Methods* **10**, 421–426 (2013).
- Chen, J. et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014).
- Cisse, I. I. et al. Real-time dynamics of RNA polymerase II clustering in live human cells. *Science* **341**, 664–667 (2013).
- Nguyen, V. Q. et al. Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. *Mol. Cell* **81**, 3560–3575 (2021).

19. Liu, Z. & Tjian, R. Visualizing transcription factor dynamics in living cells. *J. Cell Biol.* **217**, 1181–1191 (2018).
20. Wang, Z. & Deng, W. Dynamic transcription regulation at the single-molecule level. *Dev. Biol.* **482**, 67–81 (2022).
21. Halford, S. E. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* **32**, 3040–3052 (2004).
22. Mazzocca, M., Fillot, T., Loffreda, A., Gnani, D. & Mazza, D. The needle and the haystack: single molecule tracking to probe the transcription factor search in eukaryotes. *Biochem. Soc. Trans.* **49**, 1121–1132 (2021).
23. von Hippel, P. H. & Berg, O. G. Facilitated target location in biological systems. *J. Biol. Chem.* **264**, 675–678 (1989).
24. Leith, J. S. et al. Sequence-dependent sliding kinetics of p53. *Proc. Natl Acad. Sci. USA* **109**, 16552–16557 (2012).
25. Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V. & Zaret, K. S. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev.* **23**, 804–809 (2009).
26. Lerner, J., Katznelson, A., Zhang, J. & Zaret, K. S. Different chromatin-scanning modes lead to targeting of compacted chromatin by pioneer factors FOXA1 and SOX2. *Cell Rep.* **42**, 112748 (2023).
27. Mir, M. et al. Dynamic multifactor hubs interact transiently with sites of active transcription in *Drosophila* embryos. *eLife* **7**, e40497 (2018).
28. Tang, X. et al. Kinetic principles underlying pioneer function of GAGA transcription factor in live cells. *Nat. Struct. Mol. Biol.* **29**, 665–676 (2022).
29. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 159–161 (2008).
30. Grimm, J. B. et al. Bright photoactivatable fluorophores for single-molecule imaging. *Nat. Methods* **13**, 985–988 (2016).
31. Hansen, A. S. et al. Robust model-based analysis of single-particle tracking experiments with Spot-On. *eLife* **7**, e33125 (2018).
32. Iwafuchi, M. et al. Gene network transitions in embryos depend upon interactions between a pioneer transcription factor and core histones. *Nat. Genet.* **52**, 418–427 (2020).
33. Persson, F., Lindén, M., Unoson, C. & Elf, J. Extracting intracellular diffusive states and transition rates from single-molecule tracking data. *Nat. Methods* **10**, 265–269 (2013).
34. Burtscher, I., Barkey, W. & Lickert, H. Foxa2–Venus fusion reporter mouse line allows live-cell analysis of endoderm-derived organ formation. *Genesis* **51**, 596–604 (2013).
35. Lee, K. et al. FOXA2 is required for enhancer priming during pancreatic differentiation. *Cell Rep.* **28**, 382–393 (2019).
36. Loh, K. M. et al. Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations. *Cell Stem Cell* **14**, 237–252 (2014).
37. Izeddin, I. et al. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife* **3**, e02230 (2014).
38. Hansen, A. S., Amitai, A., Cattoglio, C., Tjian, R. & Darzacq, X. Guided nuclear exploration increases CTCF target search efficiency. *Nat. Chem. Biol.* **16**, 257–266 (2020).
39. Kohler, S. & Cirillo, L. A. Stable chromatin binding prevents FoxA acetylation, preserving FoxA chromatin remodeling. *J. Biol. Chem.* **285**, 464–472 (2010).
40. Chong, S. et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* **361**, eaar2555 (2018).
41. Fu, Z. et al. mEosEM withstands osmium staining and Epon embedding for super-resolution CLEM. *Nat. Methods* **17**, 55–58 (2020).
42. Levet, F. et al. SR-Tesseler: a method to segment and quantify localization-based super-resolution microscopy data. *Nat. Methods* **12**, 1065–1071 (2015).
43. Ricci, M. A., Manzo, C., García-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**, 1145–1158 (2015).
44. Miron, E. et al. Chromatin arranges in chains of mesoscale domains with nanoscale functional topography independent of cohesin. *Sci. Adv.* **6**, eaba8811 (2020).
45. Amitai, A. Chromatin configuration affects the dynamics and distribution of a transiently interacting protein. *Biophys. J.* **114**, 766–771 (2018).
46. Mazzocca, M. et al. Chromatin organization drives the search mechanism of nuclear factors. *Nat. Commun.* **14**, 6433 (2023).
47. Mondal, A., Mishra, S. K. & Bhattacharjee, A. Kinetic origin of nucleosome invasion by pioneer transcription factors. *Biophys. J.* **120**, 5219–5230 (2021).
48. Sinha, K. K., Bilokapic, S., Du, Y., Malik, D. & Halic, M. Histone modifications regulate pioneer transcription factor cooperativity. *Nature* **619**, 378–384 (2023).
49. Fernandez Garcia, M. et al. Structural features of transcription factors associating with nucleosome binding. *Mol. Cell* **75**, 921–932 (2019).
50. Sabari, B. R. et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**, eaar3958 (2018).
51. Garcia, D. A. et al. An intrinsically disordered region-mediated confinement state contributes to the dynamics and function of transcription factors. *Mol. Cell* **81**, 1484–1498 (2021).
52. Darzacq, X. & Tjian, R. Weak multivalent biomolecular interactions: a strength versus numbers tug of war with implications for phase partitioning. *RNA* **28**, 48–51 (2022).
53. Chen, Y. et al. Mechanisms governing target search and binding dynamics of hypoxia-inducible factors. *eLife* **11**, e75064 (2022).
54. Boller, S. et al. Pioneering activity of the C-terminal domain of EBF1 shapes the chromatin landscape for B cell programming. *Immunity* **44**, 527–541 (2016).
55. Frederick, M. A. et al. A pioneer factor locally opens compacted chromatin to enable targeted ATP-dependent nucleosome remodeling. *Nat. Struct. Mol. Biol.* **30**, 31–37 (2023).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2024

Methods

Cell culture and stable cell line

U2OS cells, HEK293T cells and MEF cells were cultured in the complete DMEM (Gibco, C11995500BT) supplemented with 10% FBS (VisTech, SE100-011) and 1% penicillin–streptomycin (Hyclone, SV30010). The stable cell lines of U2OS cells with the Tet-on inducible system were constructed with the PiggyBac Transposon system (System Biosciences) containing a puromycin selection gene. Briefly, the Tet-on plasmid and supertransposase plasmid were cotransfected into U2OS cells using lipofectamine 3000 (Thermo Fisher, L3000075) according to the manufacturer's instructions. Then, 48 h after transfection, $1\ \mu\text{g ml}^{-1}$ puromycin (Selleck, S7417) were added to the medium to select for positively integrated cells.

The lentivirus Tet-on inducible lentiviral system was packaged in HEK293T cells with standard protocol of calcium phosphate transfection, collected at 36 h after transfection and filtered through 0.45- μm pore size filters (Pall, 4614). To image infected MEF cells, 2×10^4 early-passage MEF cells were first seeded on a four-well Cellvis chamber (Cellvis, C4-1.5H-N) for 24 h before viral infection and then infected with viral supernatant of OCT4-Halo, SOX2-Halo, KLF4-Halo or MYC-Halo, rtTA viral supernatants and $8\ \mu\text{g ml}^{-1}$ polybrene (MacGene, MC032). The viral-containing medium was replaced with fresh medium 12 h after infection and doxycycline was added at 24 h after infection. MEF cells were stained and imaged 48 h after infection. All plasmids are available upon request.

hES cell culture, differentiation and genome editing

H7 hES cells (WiCell, WAe007-A) were cultured on plates precoated with 1% (v/v) Geltrex (Thermo Fisher, A1413302) with complete Essential 8 medium (Thermo Fisher, A1517001). Accutase (Thermo Fisher, A1110501) was used for cell dispersion and ROCK inhibitor thiazovivin (Tocris, 3845) at a final concentration of $1\ \mu\text{M}$ was added to the culture medium for the first day after replating cells. hES cells were differentiated to APS and DE with a 2-day protocol using a pluripotent stem cell DE induction kit (Thermo Fisher, A3062601) or homemade medium³⁶. Cells were stained with surface marker APC-CXCR4 (BD 555976) at a 1:5 dilution for 30 min on ice and examined by flow cytometry to assess differentiation rate.

For genome editing, single guide RNAs (sgRNAs) were designed using CRISPOR (<http://crispor.tefor.net/crispor.py>) and cloned into a plasmid expressing Cas9 and Venus fluorescent protein. Sequences for genome-editing sgRNAs are listed in Supplementary Table 9. hES cells were transfected using Eugene 6 (Promega, E2691) following the manufacturer instructions. For KI, a Cas9–sgRNA plasmid and a donor plasmid were cotransfected; for KO or domain deletion, two Cas9–sgRNA plasmids were cotransfected. Then, 48 h after transfection, Venus-positive cells were sorted by fluorescence-activated cell sorting (FACS), plated at low density and cultured for about 10 days to generate discrete cell clones. Individual cell clones were picked out and genotyped by PCR to identify desired homozygous clones. The desired clones were further verified by Sanger sequencing and western blots.

Microscopy setup

All SMT and PALM experiments were performed on a custom-built Nikon Ti2 microscope as before³⁶. It is equipped with two electron-multiplying charge-coupled device (EM-CCD) cameras (Andor, iXon Ultra 897), a $\times 100$ (1.49 numerical aperture) oil-immersion total internal reflection fluorescence (TIRF) objective (Nikon apochromat CFI Apo TIRF $\times 100$ oil), a perfect focusing system (Applied Scientific Instrumentation) and a stable-top chamber for maintaining $37\ ^\circ\text{C}$ and 5% CO_2 (Tokai Hit). The HILO illumination was achieved by a Nikon TIRF module, multiple lasers (405 nm, maximum 140 mW, OBIS, Coherent; 560 nm, maximum 1 W, MPB; 642 nm, maximum 1.5 W, MPB) controlled by the acousto-optic tunable filter (AOTF) system (AA Opto-Electronic, AOTFnC-VIS-TN), a multiband dichroic mirror (405, 488, 561 and 633 nm quad-band,

Semrock) and emission filters (TMR/PA-JF549, Semrock 593/40-nm bandpass filter; JF646/PA-JF646; Semrock 676/37-nm bandpass filter).

Cell labeling for SMT, PALM and confocal imaging

U2OS and MEF cells were plated on Mat-Tek dishes (Mat-Tek, P35G-1.5-14-C) or eight-well Cellvis chambers (Cellvis, C8-1.5H-N) with no. 1.5 high-performance cover glass and protein expression was induced with $1\ \mu\text{g ml}^{-1}$ doxycycline (Selleck, S4163) for 1 day. For hES cells and their differentiation, a glass-bottom culture dish was first coated with Geltrex (Thermo Fisher, no. A1413301) overnight. To label Halo fusion protein in living cells, Halo ligand-conjugated fluorescent dyes at the indicated final concentration were added to culture medium (fast spaSMT, 25 nM PA-JF646 or 2.5 nM PA-JF549; slowSMT, 25 pM JF646; PALM, 25 nM PA-JF646; confocal imaging, 200 nM TMR) before incubating for 30 min. The medium was then removed and two rounds of incubation with fresh medium for 30 min were performed to remove free dyes. Right before live-cell imaging of SMT, the medium was changed to phenol red-free medium. For two-color PALM and confocal imaging of fixed cells, after Halo staining, cells were crosslinked with prewarmed 4% PFA for 10 min at $37\ ^\circ\text{C}$, permeabilized with 0.1% v/v Triton X-100 in PBS for 10 min at room temperature and then washed with PBS. The medium used in each step was prewarmed and caution was taken to avoid cells detaching from the glass bottom.

Acquisition of SMT data

The 133-Hz fast spaSMT experimental settings were as follows: in one frame of camera exposure time (7 ms), 1 ms of 642-nm excitation (100% AOTF) was delivered at the beginning of the frame; 405-nm photoactivation pulses were delivered during the camera integration time ($\sim 447\ \mu\text{s}$) to minimize background and their intensity was optimized to achieve a mean density of 1–2 molecules per frame per nucleus. In addition to 7 ms (~ 133 Hz), the camera exposure times were set at 13 ms (~ 74 Hz) to obtain additional data for analyzing the space dependency of anisotropy. A total of 20,000 frames were recorded per cell per experiment. For each sample, we performed two or three independent replicates and recorded data of multiple cells in each replicate.

The slowSMT experiments used a long exposure time (500 ms) and continuous illumination with low laser intensity to capture the position of slow-moving molecules. Under this mode, fast-moving molecules will be blurred into the background noise. Caution was taken to ensure that cells did not have gross movement during the process of imaging to minimize tracking error. Generally, each cell was captured for 1,500 frames with 500-ms exposure time. For each sample, we performed two or three independent replicates and recorded data of multiple cells in each replicate.

Acquisition of simultaneous two-color PALM data

A 405-nm laser was used during the camera integration time ($\sim 447\ \mu\text{s}$) to photoconvert mEosEM and photoactivate PA-JF646 at sparse density. The 560-nm and 642-nm lasers were used to illuminate cells for continuous 30 ms and two colors were detected simultaneously with two EM-CCD cameras at an exposure time of 30 ms. A total of 30,000 frames of images were taken for each cell and at least 12 cells were recorded for each sample. In the combined live-cell PALM and fast spaSMT experiments, PALM data were collected with an exposure time of 30 ms; fast spaSMT data were collected immediately with an exposure time of 7.5 ms.

Single-molecule localization and trajectory linking

Both fast spaSMT movies and slowSMT movies were analyzed with custom-made MATLAB scripts (https://gitlab.com/tjian-darzaccq-lab/SPT_LocAndTrack) implementing a multiple-target tracking (MTT) algorithm⁵⁷ to generate single-molecule trajectories. The settings used in the MATLAB scripts for fast spaSMT were as follows: localization error, $10^{-6.25}$; deflation loops, 0; blinking (frames), 0; maximum

competitors, 3; maximum D ($\mu\text{m}^2 \text{s}^{-1}$), 20. The settings used in the MATLAB scripts for slowSMT were as follows: localization error, $10^{-6.25}$; deflation loops, 0; blinking (frames), 1; maximum competitors, 3; maximum D ($\mu\text{m}^2 \text{s}^{-1}$), 0.7.

Estimation of bound fraction and diffusion coefficient from fast spaSMT data

Fraction of bound molecules and diffusion coefficient of unbound molecules were determined by fitting the two-state Spot-On kinetic model³¹ (version 1.05; GitLab tag 6a3e885c) with the spaSMT trajectories obtained with a 7.5-ms exposure time. The parameters used in the fitting were as follows: $dZ = 0.7 \mu\text{m}$; GapsAllowed = 1; Time Points = 8; JumpsToConsider = 4; Model Fit = 2; NumberOfStates = 2; FitLocError = 0; LocError = $0.040 \mu\text{m}$; $D_{\text{Free_2State}} = [0.5, 25]$; $D_{\text{Bound_2State}} = [0.0001, 0.04]$.

RT inference and survival analysis of slowSMT data

The trajectory lengths of stably bound molecules from slowSMT experiments were plotted as a survival curve (1 – cumulative distribution function). We adopted the method from a previous study⁵⁶ to calculate the RT of TFs on targets.

Calculation of target search kinetics

To investigate how FOXA2 explores the nucleus, we adopted the method described by Nguyen et al.¹⁸ to derive various target searching parameters.

Angular analysis of fast spaSMT data

We used the Matlab scripts from a previous study (<https://gitlab.com/anders.sejr.hansen/anisotropy>)³⁸ with some changes to analyze the distribution of the molecule trajectories obtained from our fast spaSMT experiments. Briefly, we merged trajectories from different cells obtained from the fast spaSMT datasets (133 Hz and 74 Hz). In the merging step, we also performed a quality control step to further ensure that the downstream analysis was not subject to the wrong connection of two different molecules observed in consecutive frames during reconnection steps in the MTT algorithm. ‘ClosestDist’ was set to 2 so that, if any two detected localizations in the same frame were within $2 \mu\text{m}$, their trajectories would be discarded from downstream analysis. The merged trajectories were HMM-classified to label each segment of a trajectory as either ‘bound’ or ‘free’ and the bound segments of each trajectory were removed. The angles formed by two consecutive ‘free’ segments with a minimum length of 100 nm (‘GlobalMinJumpThres’ set to 0.1, equivalent to 100 nm) were calculated. The angles from all free segments were pooled and the fold of the angles ranging from $180 \pm 30^\circ$ versus $0 \pm 30^\circ$, $f_{180/0}$ (that is, global anisotropy) were calculated to measure the anisotropic behavior of the molecules.

In addition to analyzing angles from all free segments, we analyzed the angles composed of free segments right before or after one bound segment (that is, local anisotropy). Specifically, after classifying the segments of each trajectory into bound or free states by the HMM model as above, we selected a defined number of the consecutive free segments either before or after a bound segment. Then, we discarded any free segments less than 100 nm to calculate angles between the two adjacent free segments and pooled all angles to calculate the $f_{180/0, \text{local}}$.

Processing and rendering PALM data

Raw imaging data with single-molecule signals were analyzed using the ImageJ plug-in ThunderStorm³⁸ to detect single-molecule localizations. Low-quality single-molecule signals (point spread function too small or too large or with high uncertainty) were filtered out. Single-molecule localization data consisting of x and y localization lists were used to render PALM images by assigning pixel size as 10 nm and pixel value equal to the number of localizations within the pixel area.

Two-color PALM data processing and model-based chromatin segmentation

Two-color PALM raw data were analyzed by ThunderStorm to generate molecule localizations and merge reappearing localizations in subsequent frames. The resulted localization data were further analyzed by a self-written MATLAB script. Sample drift was determined by calculating the spatial cross-correlation function between sequential subsets of localizations and corrected. Data with major sample drift were discarded to ensure high data quality. We generated rendered super-resolution images with pixel size at 10 nm and pixel value as the number of molecular localizations within the pixel. Manders correlation coefficients were calculated by the ImageJ Colocalization Threshold plug-in using H2B PALM and FOXA2 PALM images as input and a default threshold.

A Voronoi diagram was created by the SR-Tesseler analysis⁴², after which clusters were identified by setting the density factor to 2.5. Clusters were filtered by minimum area and minimum number of locs (min area, 20; min no. locs, 50). The following information was directly exported: area, number of detections, circularity and diameter.

Chromatin compaction levels were determined by the intensity of rendered H2B super-resolution images with 10-nm pixel size. The nuclear border was determined by Gaussian smoothing and threshold binarization. The resulting images were further analyzed by a hidden Markov random field model to segment images to seven classes⁴⁴ on the basis of H2B intensity.

After the chromatin compaction segmentation into seven classes, localization densities of different FOXA2 variants in each chromatin class were calculated and divided by densities of localizations following a random distribution; the relative density enrichments were presented as heat maps. The seven chromatin intensity classes were further divided into three biologically relevant regions: IC (class 1), PC (classes 2 and 3) and CD (classes 4–7), as described in the literature⁴⁴. The CD region was composed of chains of individual CDs. Their centroids were found by the regional maximum algorithm and the borders of individual CDs were determined by the intensity valley calculated from intensity gradients. The centroids of FOXA2 variant clusters were also determined by the regional maximum algorithm; only centroids located within an individual CD were selected out and their distance to the corresponding CD centroid was measured.

As a parallel method, the cell nucleus was also divided into nine equal-area parts from low to high density on the basis of the intensity of H2B and FOXA2 variants, respectively, and the localization number and percentage of H2B and FOXA2 variants within each H2B region, as well as the overlapped areas between H2B and FOXA2 at different densities, were calculated. Heat maps were plotted with x and y axes representing the index of nine density regions of the FOXA2 variants and H2B from low to high, respectively.

In combined live-cell PALM and fast spaSMT experiments, 1,000 frames of images with a 30-ms exposure time were used to render a PALM image and fast spaSMT data were collected immediately after PALM and overlaid.

Simulation of chromatin and the TF target search

We used a previously developed chromatin model^{38,59} that treats chromatin as a coarse-grained, self-avoiding polymer confined within the nucleus. Briefly, we modeled the chromatin as a series of n monomers with a radius of 200 nm, with positions (R_1, \dots, R_n). A small portion of randomly chosen monomers had CBSs placed in their middle (for confined target search model) or filled with the defined number of CBSs randomly (for clustering target sites model). The chromatin chain first reached equilibrium with the defined spring potential and Lennard–Jones forces; then, each monomer position was fixed (Supplementary Fig. 6a). All monomers were confined within a sphere with radius R to simulate the cell nucleus and each rigid monomer was reflected in the normal direction of the tangent plane of the sphere boundary.

For the simulation of confined target search model, the simulated TF interacted with the modeled chromatin in three ways. First, the TF underwent free Brownian motion when not interacting with monomers. Second, when the TF touched the boundary of monomer (for a distance between the TF and the center of monomer smaller than ϵ_{mono}), it had a certain probability (P_{enter}) of entering the monomer. Upon entering the monomer, the TF diffused inside the monomer and exited it with certain probability (P_{exit}) once hitting the boundary of the monomer. When diffusing inside the monomer, the TF could transiently bind with a Poissonian on rate k_{on} and then be released with a Poissonian off rate k_{off} to mimic nonspecific binding. If the monomer contained a CBS in its middle, the TF would bind to the CBS when its distance to the center of the CBS (same as the center of monomer) was smaller than ϵ_{CBS} . The RT of CBS binding followed a Poisson distribution with a mean time $\tau_{\text{CBS}} = 12.5$ s. The value of RT did not affect the simulation results because we only focused on the target search (that is, diffusion) rather than target binding section of the final trajectory. After finishing binding to the CBS, the TF continued diffusing inside the monomer. Of note, we used Euler's scheme to generate the Brownian motion of the protein.

In the simulation of the clustering target sites model, the TF underwent free Brownian motion when not interacting with CBSs. The TF bound to the CBS when its distance to the center of the CBS was smaller than ϵ_{CBS} . The RT of CBS binding followed a Poisson distribution with a mean time of τ_{CBS} (12.5 s). After completing the binding on CBS, the TF continued to diffuse inside the nucleus.

Each round of simulation generated a long Brownian trajectory with a time gap of 0.1 ms. The protein was placed at the release radius (α_{mono}) of a randomly chosen monomer and started diffusing. Each round of simulation stopped when the total number of steps reached 5,000 or after 500 iterations of CBS binding, depending on the downstream analysis. To simulate the localization error of real imaging conditions, white Gaussian noise with an s.d. of 30 nm and zero mean was added to each trajectory point. The generated trajectories were then projected into two dimensions (discarding the z position) and split into short trajectories with the desired time gap (for example, 7.5 ms) to mimic the exposure time used in fast spaSMT. We performed 12 simulations for each condition. The generated trajectories were then pooled and analyzed using the same pipeline as previously described (see Methods, 'Angular analysis of fast spaSMT data'). More detailed parameters used in the simulation are listed in Supplementary Table 7.

Co-IP

Stable U2OS cell lines were induced with $1 \mu\text{g ml}^{-1}$ doxycycline for 48 h, collected and gently lysed with ice-cold NP-40 lysis buffer (10 mM HEPES pH 7.5, 10 mM KCl, 0.2 mM EDTA, 1 mM DTT and 0.5% NP-40) on ice for 20 min. The pelleted nuclei were further extracted with high-salt buffer (500 mM NaCl) by incubation on ice for 60 min. The nuclear lysate was immunoprecipitated with preactivated Flag-Magnetic Dynabeads (Thermo Fisher, 88847) at 4 °C overnight, followed by one wash with balance lysis buffer (25 mM HEPES pH 7.5, 0.2 mM EDTA and 0.5% NP-40), three washes with high-salt wash buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 0.2 mM EDTA and 0.5% NP-40), one wash with low-salt wash buffer (25 mM HEPES pH 7.5, 150 mM NaCl, 0.2 mM EDTA and 0.5% NP-40) and boiling in SDS-PAGE loading buffer at 100 °C for 10 min. The antibodies used are listed in Supplementary Table 10.

ChIP-seq and ATAC-seq

ChIP-seq followed the standard protocols⁶⁰. ChIP-seq experiments were performed in two replicates. Primary antibodies for ChIP-seq are listed in Supplementary Table 10.

OmniATAC-seq was performed according to a previously described protocol⁶¹. ATAC-seq experiments were performed in two replicates.

For ChIP-seq and ATAC-seq, raw reads were quality-checked with FastQC (version 0.10.1) and trimmed by the trim_galore function (version 0.6.10) with paired mode and default parameters.

Reads were aligned to human genome (hg38) using Bowtie 2 (version 2.4.4). The resulting SAM files were converted to BAM files with SAMtools (version 0.1.18). The BAM files of the two replicates of one treatment were peak-called individually using MACS2 (version 2.2.7.1) with default parameters and the resulting narrowPeak files were used to call peaks with the IDR (version 2.0.4.2) method. The bamCoverage function in deepTools (version 2.4.1) was used to convert BAM files to bigWig files for visualization and normalized by reads per kilobase per million mapped reads (RPKM). A high correlation above 0.8 was observed between replicates for all samples. The R package intervene (version 0.6.5) was used to determine overlaps among two or more peak sets. The function computeMatrix in the deepTools suite (version 2.4.1) was used with reference-point mode to calculate scores for each genomic region. Motif analysis of differential peaks from each group was performed using HOMER (version 4.11). Genome-wide chromatin state predictions for each biological condition were performed using ChromHMM (version 1.24) software with default parameters. To compare the ChIP-seq intensity from FOXA2 and FOXA2 Δ CTD, ChIP-seq libraries were normalized to input-calibrated ChIP-qPCR of 12 sites that comprised varying signal intensities, using a linear regression model (Supplementary Fig. 4b and Supplementary Table 5) as described previously⁶².

RNA-seq

Poly(A) RNA-seq was performed for undifferentiated hES cells or for hES cells differentiated to APS and DE. The purified total RNA was added to ERCC RNA spike-in mix 2 (Thermo Fisher, 4456740). mRNA purification, RNA fragmentation, first-strand and second-strand complementary DNA (cDNA) synthesis were performed according to the TruSeq RNA sample preparation v2 kit (Illumina, RS-122-2001) but using Superscript III for reverse transcription instead of Superscript II (50 °C for 50-min incubation time). cDNA was purified with AMPure XP beads (Beckman, A63882) and subjected to a Solexa rapid library protocol. PCR reactions were cleaned up once with AMPure XP beads. Library quality and fragment size were assessed by qPCR and Fragment Analyzer and sequenced on the Illumina HiSeq4000 sequencing platform (single-end reads, 50 bp long). RNA-seq experiments were performed in two replicates.

Analysis of RNA-seq data

Reads were quality-checked with FastQC and aligned to the transcriptome of the human genome (hg38) by STAR version 2.7.1a with the following parameters: `--outSJfilterReads Unique --outFilterMultimapNmax 1 --outFilterIntronMotifs RemoveNoncanonical --outSAMstrandField intronMotif`. The gene expression level was calculated using HTSeq version 2.0.2 based on the gencode hg38 annotation (GENCODE version 32) with the following parameters: `htseq-count --stranded=no -f bam --additional-attr=gene_name -m union`. All subsequent analyses were performed in R.

Differential gene expression analysis was performed on three cell stages (hES cell, APS and DE) with two replicates using DESeq2 version 1.30.1. Stage-specific genes were selected with $\log_2(\text{fold change}) > 1$ and false discovery rate (FDR) < 0.05 . FDR was calculated by the Wald test implemented by DESeq2 with Benjamini-Hochberg adjustment. Principal component analysis (PCA) of differentially expressed genes was performed using DESeq2. Putative FOXA2 target genes were identified as genes having ± 5 -kb FOXA2-binding sites flanking the gene body at the DE stage using clusterProfiler version 3.18.1. DE stage-specific genes were further sorted on the basis of gene expression changes after FOXA2 KO and their dependency on CTD deletion using DESeq2-normalized RNA-seq read counts. Genes with the fold change > 1.5 after FOXA2 deletion were classified as 'repressed' by FOXA2, while genes with fold change < 0.66 were classified as 'activated' by FOXA2. Within these FOXA2-dependent genes (repressed or activated), their dependency on CTD was further assessed to

identify the 'CTD-dependent' genes. Therefore, these genes were classified into five groups: I (CTD-dependent gene repression with fold change(KO/wild type (WT)) > 1.5 and fold change(Δ CTD/WT) > 1.5), II (CTD-independent gene repression with fold change(KO/WT) > 1.5 and $2/3 < \text{fold change}(\Delta\text{CTD/WT}) < 1.5$), III (CTD-dependent gene activation with fold change(KO/WT) < 2/3 and fold change(Δ CTD/WT) < 2/3), IV (CTD-independent gene activation with fold change(KO/WT) < 2/3 and fold change(Δ CTD/WT) > 2/3) and V (unchanged with $2/3 < \text{fold change}(\Delta\text{CTD/WT}) < 1.5$). Gene Ontology (GO) term enrichment analysis was performed using clusterProfiler and GO terms with a *P* value < 0.05 were considered significantly enriched. All volcano plots and bubble charts were plotted using ggplot2 and the heat map was generated using pheatmap.

Reporting summary

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

Data availability

All data needed to reproduce the results are available in the main text or Supplementary Information. Imaging data are available from Zenodo (<https://doi.org/10.5281/zenodo.10427790>). Genomic data were deposited to the National Center for Biotechnology Information's Gene Expression Omnibus (accession number [GSE203650](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE203650)). Source data are provided with this paper.

Code availability

The analysis code was deposited to GitHub (<https://github.com/denglabpku/>).

References

- Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R. & Darzacq, X. CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *eLife* **6**, e25776 (2017).
- Sergé, A., Bertaux, N., Rigneault, H. & Marguet, D. Dynamic multiple-target tracing to probe spatiotemporal cartography of cell membranes. *Nat. Methods* **5**, 687–694 (2008).
- Ovesný, M., Křížek, P., Borkovec, J., Švindrych, Z. & Hagen, G. M. ThunderSTORM: a comprehensive ImageJ plug-in for PALM and STORM data analysis and super-resolution imaging. *Bioinformatics* **30**, 2389–2390 (2014).
- Amitai, A., Seeber, A., Gasser, S. M. & Holcman, D. Visualization of chromatin decompaction and break site extrusion as predicted by statistical polymer modeling of single-locus trajectories. *Cell Rep.* **18**, 1200–1214 (2017).
- Ji, X. et al. Chromatin proteomic profiling reveals novel proteins associated with histone-marked genomic regions. *Proc. Natl Acad. Sci. USA* **112**, 3841–3846 (2015).
- Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
- Luan, J. et al. Distinct properties and functions of CTCF revealed by a rapidly inducible degron system. *Cell Rep.* **34**, 108783 (2021).

Acknowledgements

We are grateful to R. Tjian, X. Darzacq and their laboratory members for assistance and discussions in initiating this study. We thank R. Salomon and G. Dailey for assistance in cell culture and cloning. We thank A. Amitai for providing computer code and discussion on kinetic simulation of target search. We thank P. Xu for sharing the mEosEM plasmid. We thank the National Center for Protein Sciences at Peking University for providing the FACS service and confocal microscopes. We thank the following for funding support: the National Key Research and Development Program of China (2020YFA0509502 to W.D.), the National Natural Science Foundation of China (32170566 to W.D.; T2225001 to H.G.), the Beijing Advanced Innovation Center for Genomics at Peking University (to W.D.) and the Peking-Tsinghua Center for Life Sciences (to W.D.).

Author contributions

Design of experiments, W.D., Z.W., B.W. and D.N. Live-cell SMT and analysis, Z.W. and W.D. ChIP-seq and ATAC-seq, D.N. RNA-seq, W.D. and C.C. Genomic analysis, B.W., D.N. and Y.B. Two-color PALM and analysis, B.W. Computer simulation of target search, H.G., Z.W. and B.W. Lentivirus production and MEF SMT, Y.C. hES cell culture and differentiation, W.D., Y.C. and K.M.L. Cell line construction, W.D., Z.W., D.N., B.W. and Y.B. Halo dye synthesis, L.D.L. Writing—original draft, W.D., Z.W., B.W. and D.N. Writing—review and editing, C.C., K.M.L. and H.G.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41594-024-01385-5>.

Correspondence and requests for materials should be addressed to Wulan Deng.

Peer review information *Nature Structural & Molecular Biology* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Sara Osman, in collaboration with the *Nature Structural & Molecular Biology* team.

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

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- | n/a | Confirmed |
|-------------------------------------|--|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> A description of all covariates tested |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals) |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
<i>Give P values as exact values whenever suitable.</i> |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated |

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Data analysis https://github.com/denglabpku/).

The genomic data were analyzed using FastQC(v 0.10.1) for quality control of raw reads of ChIP-seq, ATAC-seq and RNA-seq; TrimGalore(v0.6.10) remove reads of low quality and sequencing adaptors; bowtie2 (v2.4.4) for reads alignment to reference genome (human: hg38); samtools (v.0.1.18) for converting SAM files to BAM files; MACS2(2.2.7.1) for peak calling of individual samples; IDR(2.0.4.2) for peak calling of two replicates; intervene (v0.6.5) for venn plot; deepTools suite (v.2.4.1) for converting BAM files to bigwig files, and plotting heatmap; HOMER (v4.11) for motif analysis; ChromHMM (v1.24) for genome-wide chromatin state predictions; STAR (v2.7.1a) for reads alignment to reference genome (human: hg38); HTSeq (v2.0.2) for gene expression matrix calculation; DESeq2(v1.30.1) for differentially expressed gene analysis; ggplot2(v3.3.5) for charts plotting including scatter plot, bubble plot and bar plot; clusterProfiler(v3.18.1): for identifying ChIP-seq peaks's putative target genes, and GO enrichment analysis.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All data are available in the main text or the supplementary materials. Imaging data are available at Zenodo (accession number 10427790). Genomic data have been deposited in NCBI's Gene Expression Omnibus (accession number: GSE 203650). ChIP-seq/ATAC-seq/RNA-seq reads were aligned to human genome hg38.

Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	NA
Reporting on race, ethnicity, or other socially relevant groupings	NA
Population characteristics	NA
Recruitment	NA
Ethics oversight	NA

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

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

Sample size	The sample size was chosen large enough to get robust analysis results. We imaged 841 cells in total for fast spaSMT (around 10 million trajectories in total), 124 cells for slowSMT (around a half million trajectories in total), and 40 cells for PALM. We do not use any statistical method to determine sample size.
Data exclusions	In the live-cell imaging data, only cells with large movements have been exclude. No other data exclusions.
Replication	In fast spaSMT experiments, we conducted two or three independent replicates for each sample and recorded data from multiple cells in each replicate. The final results were generated from hundreds of thousands to millions of molecule trajectories. In PALM experiments, we recorded data from at least twelve cells for each sample. Genomic experiments were performed in two replicates. We confirmed that our claims and figure results are reproducible.
Randomization	To calculate the relative density of molecules in PALM experiments, we employed a randomization approach by randomizing the spatial distribution of molecules. We then divided the actual density by the randomized density to obtain the relative density. For confocal imaging, we randomly select cells when recording. For fast spaSMT and slowSMT experiments, cells were selected randomly, and the trajectories were pooled together to fit the model. For western-blot, Co-IP, and genomic experiments, since millions of cells were collected, no randomization is required.
Blinding	Blinding was not applicable to this study. We did not work with human subjects.

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> Antibodies
<input type="checkbox"/>	<input checked="" type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

Methods

n/a	Involvement in the study
<input type="checkbox"/>	<input checked="" type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

Antibodies

Antibodies used

anti-FOXA2 antibody (1:500, R&D, AF2400-SP), anti-SOX17(1:1000, R&D, AF1924-SP), anti-Halo(1:1000, Promega, G9211), anti-TBP(1:1000, Abcam, ab51841), anti-H3(1:2000, Beyotime, AH433), anti-H4(1:500, Beyotime, AH458), anti-Flag(1:1000, Sigma, A01868-40/ Sigma, F1804), Goat Anti-Rabbit IgG (H&L)-HRP Conjugated secondary antibody(1:10000, EASYBIO, BE0101), Goat Anti-Mouse IgG (H&L)-HRP Conjugated secondary antibody (1:10000, EASYBIO, BE0102), Rabbit Anti-Goat IgG (H&L)-HRP Conjugated secondary antibody(1:10000, EASYBIO, BE0103). This list is also included in Supplementary Table 10.

Validation

anti-FOXA2 antibody (R&D, AF2400-SP):
Validation for FOXA2 band in HepG2 human hepatocellular carcinoma cell line, Jurkat human acute T cell leukemia cell line, and K562 human chronic myelogenous leukemia cell line.
https://www.rndsystems.com/cn/products/human-hnf-3beta-foxa2-antibody_af2400

anti-SOX17(R&D, AF1924-SP):
Validation for SOX17 band in lysates of BG01V human embryonic stem cells untreated (-) or endoderm differentiated (+).
https://www.rndsystems.com/cn/products/human-sox17-antibody_af1924

anti-Halo(Promega, G9211)
Validation of HaloTag® fusion protein
<https://www.promega.com.cn/resources/msds/msdss/g9000/g9211/>

anti-TBP(Abcam, ab51841)
Validation for TBP band in HeLa Whole Cell Lysate and HepG2 (Human hepatocellular liver carcinoma cell line) Whole Cell Lysate
<https://www.abcam.com/products/primary-antibodies/tata-binding-protein-tbp-antibody-nuclear-loading-control-and-chip-grade-ab63766.html#lb>

anti-H3(Beyotime, AH433)
<https://www.beyotime.com/product/AH433.html>

anti-H4(Beyotime, AH458)
<https://beyotime.com/msds/AH458.pdf>

anti-Flag(Sigma, A01868-40)
Validation of Flag-tag in HeLa cell lysates containing DYKDDDDK fusion protein
<https://www.thermofisher.cn/cn/zh/antibody/product/DYKDDDDK-Tag-Antibody-clone-114F12C8-Monoclonal/A01868-40>

anti-Flag(Sigma,F1804)
Validation of Flag-tag in HeLa cell lysates containing DYKDDDDK fusion protein
<https://www.sigmaaldrich.cn/CN/zh/product/sigma/f1804>

HRP-conjugated secondary antibodies(EASYBIO, BE0101)
http://bioeasytech.com/product/2901.html?goods_id=5786

Eukaryotic cell lines

Policy information about [cell lines and Sex and Gender in Research](#)

Cell line source(s)

H7 hESCs (WiCell, WAe007-A) , HEK293T (CRL-1573™) from ATCC, HeLa(CCL-2™) from ATCC, U-2OS (HTB-96™)from ATCC

Authentication

Authenticated by STR profiling at Peking University

Mycoplasma contamination

Tested negative for Mycoplasma contamination.

Commonly misidentified lines
(See [ICLAC](#) register)

none

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA

ChIP-seq

Data deposition

- Confirm that both raw and final processed data have been deposited in a public database such as [GEO](#).
- Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links
May remain private before publication.

We have deposited the raw as well as processed ATAC, RNA and ChIP sequencing files to the GEO repository with accession #: GSE203650

Files in database submission

GSM6181133 WT_ES
 GSM6181134 WT_APS
 GSM6181135 WT_DE
 GSM6181136 KO1_APS
 GSM6181137 KO1_DE
 GSM6181138 KO21_APS
 GSM6181139 KO21_DE
 GSM6181140 dC11_APS
 GSM6181141 dC11_DE
 GSM6181142 dCdN10_APS
 GSM6181143 dCdN10_DE
 GSM7996338 U2OS_Input
 GSM7996341 U2OS_FOXA2FL_day4_biol rep1
 GSM7996342 U2OS_FOXA2FL_day4_biol rep2
 GSM7996345 U2OS_FOXA2dCTD_day4_biol rep1
 GSM7996346 U2OS_FOXA2dCTD_day4_biol rep2
 GSM7996347 U2OS_FOXA1FL_day4_biol rep1
 GSM7996348 U2OS_FOXA1FL_day4_biol rep2
 GSM7996349 U2OS_FOXA3FL_day4_biol rep1
 GSM7996350 U2OS_FOXA3FL_day4_biol rep2
 GSM7996351 U2OS_FOXA1dCTD_day4_biol rep1
 GSM7996352 U2OS_FOXA1dCTD_day4_biol rep2
 GSM7996353 U2OS_FOXA3dCTD_day4_biol rep1
 GSM7996354 U2OS_FOXA3dCTD_day4_biol rep2
 GSM7996355 U2OS_SOX2FL_day4_biol rep1
 GSM7996356 U2OS_SOX2FL_day4_biol rep2
 GSM7996357 U2OS_SOX2FLFOXA2CTD_day4_biol rep1
 GSM7996358 U2OS_SOX2FLFOXA2CTD_day4_biol rep2
 GSM7996359 ATAC_U2OS_WT_biol rep1
 GSM7996360 ATAC_U2OS_WT_biol rep2
 GSM7996363 ATAC_U2OS_FOXA2FL_day4_biol rep1
 GSM7996364 ATAC_U2OS_FOXA2FL_day4_biol rep2
 GSM7996367 ATAC_U2OS_FOXA2dCTD_day4_biol rep1
 GSM7996368 ATAC_U2OS_FOXA2dCTD_day4_biol rep2
 GSM7996369 ATAC_U2OS_SOX2FL_day4_biol rep1
 GSM7996370 ATAC_U2OS_SOX2FL_day4_biol rep2
 GSM7996371 ATAC_U2OS_SOX2FLFOXA2CTD_day4_biol rep1
 GSM7996372 ATAC_U2OS_SOX2FLFOXA2CTD_day4_biol rep2
 GSM7996373 ATAC_U2OS_WT_SOX2_CellLine_biol rep1
 GSM7996374 ATAC_U2OS_WT_SOX2_CellLine_biol rep2

Genome browser session
 (e.g. [UCSC](#))

No longer applicable.

Methodology

Replicates	Two biological duplicates for each sample.
Sequencing depth	ATAC-seq: Sequenced to 45-60M total reads, paired-end mode, 150bp read lengths. Over 90% of uniquely mapped read. ChIP-seq: sequenced to 30-40 total reads, paired end mode, 150bp read lengths, Over 90% of uniquely mapped reads. RNA-seq: Sequenced to 25-30M total reads, single-end mode, 150bp read lengths. Over 90% of uniquely mapped reads.

Antibodies	See Supplementary Table 10.
Peak calling parameters	IDR(version 2.0.4.2) call peak was used for performing peak calling with the following option: <code>idr --input-file-type narrowPeak --rank p.value -soft-idr-threshold 0.05</code> .
Data quality	FastQC was used to quality check the raw sequencing data using standard metrics and default thresholds
Software	For ChIP-seq, raw reads were quality-checked with FastQC (version 0.10.1) and trimmed by trim_galore function (v0.6.10), with --paired mode and the default parameters. Reads were aligned to human genome (hg38) using bowtie2 (v2.4.4). The resulted sam files were converted to bam files with samtools (v.0.1.18). The bam files of the two replicates of one treatment were called peaks individually using MACS2(v2.2.7.1) with the default parameters, and the resulted narrowPeak files were used to call peaks by the IDR(2.0.4.2) method. The bamCoverage function in deeptools(v2.4.1) was used to convert bam files to bigwig files for visualization, and normalized by RPKM. A high correlation above 0.8 was observed between replicates for all samples. The R package intervene (v0.6.5) was used to determine overlaps among two or more peak sets. The function computeMatrix in the deepTools suite (v.2.4.1) was used with reference-point mode to calculate scores for each genomic region. Motif analysis of differential peaks from each group were performed by software HOMER (v4.11). Genome-wide chromatin state predictions for each biological condition were performed using ChromHMM (v1.24) software with default parameters. To compare the ChIP-seq intensity from FOXA2 and FOXA2 Δ CTD, ChIP-seq libraries were normalized to input-calibrated ChIP-qPCR of 12 sites that comprised varying signal intensities, using a linear regression model.