nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-019-0422-3

Guided nuclear exploration increases CTCF target search efficiency

Anders S. Hansen[1,2,3,4,5,7], Assaf Amitai[6,7], Claudia Cattoglio[1,2,3,4], Robert Tjian[1,2,3,4]* and Xavier Darzacq[1,2,3]*

The enormous size of mammalian genomes means that for a DNA-binding protein the number of nonspecific, off-target sites vastly exceeds the number of specific, cognate sites. How mammalian DNA-binding proteins overcome this challenge to efficiently locate their target sites is not known. Here, through live-cell single-molecule tracking, we show that CCCTC-binding factor, CTCF, is repeatedly trapped in small zones that likely correspond to CTCF clusters, in a manner that is largely dependent on an internal RNA-binding region (RBR_i). We develop a new theoretical model called anisotropic diffusion through transient trapping in zones to explain CTCF dynamics. Functionally, transient RBR_i-mediated trapping increases the efficiency of CTCF target search by ~2.5-fold. Overall, our results suggest a 'guided' mechanism where CTCF clusters concentrate diffusing CTCF proteins near cognate binding sites, thus increasing the local ON-rate. We suggest that local guiding may allow DNA-binding proteins to more efficiently locate their target sites.

Mammalian nuclei are organized by a myriad of biophysical forces into subcompartments with distinct functions. At the micrometer-scale, nuclear compartments such as nucleoli, speckles and Cajal bodies, carry out specialized biochemical functions that are spatially segregated[1]. Below the micrometer-scale and diffraction limit of conventional optical microscopy, many proteins interact dynamically to form local high concentration clusters or hubs[2]. Thus, many proteins exhibit a nonrandom nuclear distribution. Similarly, many proteins display anomalous and non-Brownian diffusion inside the nucleus[3–5], which has been proposed to be due to molecular crowding and transient interactions[2,6,7]. It is still unclear what mechanisms allow proteins to form clusters and control their diffusion and target search mechanism in vivo.

Since the kinetics of a reaction or binding event depend on the diffusive properties and nuclear organization of the reactants[2,8], understanding the molecular interactions that control a nuclear protein's target search mechanism and its overall distribution is essential to understanding its function. Target search mechanisms have been extensively probed in prokaryotes[9], where studies have emphasized the importance of colocalization between genes encoding transcription factors and their binding sites[10–13]. In prokaryotes, proteins are thought to locate their binding sites through facilitated diffusion[14,15], which involves two types of motion: (1) sliding in one dimension along DNA and (2) disassociating from DNA, diffusing in three dimensions and rebinding at a proximal DNA site. It is thought that a mixture of these two modes of motion serves as the optimal search strategy[16]. In contrast to bacteria, however, mammalian genomes are enormous and chromatinized. Specifically, nucleosomes would seem to rule out one-dimensional sliding on DNA in mammals, which raises the question of how DNA-binding proteins find their targets in mammalian nuclei.

Here we investigate how mammalian DNA-binding proteins find their nuclear targets and focus on the nuclear target search mechanism of CTCF. CTCF, together with the cohesin complex, folds mammalian genomes into spatial domains known as topologically associating domains (TADs), which regulate enhancer-promoter contacts and gene expression[17]. CTCF is an 11 zinc finger (ZF) DNA-binding protein with unstructured N- and C-terminal domains. The function of these domains remains poorly understood[17]. Cohesin is thought to be a ring-shaped multi-subunit complex, which holds together CTCF-demarcated TADs as chromatin loops[18]. Although CTCF and cohesin have emerged as key regulators of genome organization and function, their nuclear target search mechanisms have not been studied.

Using single-particle tracking (SPT) and theoretical modeling, we show here that both CTCF and cohesin exhibit unusual nuclear target search mechanisms, where anisotropic diffusion arises from being transiently trapped in specific zones with a characteristic size of ~200 nm. Our results indicate that these zones correspond to CTCF clusters and we find that trapping inside them is largely due to an RBR_i in CTCF. Functionally, transient RBR_i-mediated trapping in zones increases the efficiency of CTCF's DNA-target search mechanism by about 2.5-fold. More generally, we suggest that 'guiding' could be an effective way to control and regulate the local concentration of proteins in the nucleus around specific sites.

Results

CTCF exhibits anisotropic nuclear diffusion. We noticed that CTCF exhibits anomalous diffusion (Supplementary Fig. 1 and Supplementary Videos 1–4). This motivated us to systematically investigate how CTCF, cohesin and other nuclear proteins explore the mammalian nucleus using SPT. Using our established mouse embryonic stem cell (mESC) and human osteosarcoma (U2OS) cell lines, where CTCF and the cohesin subunit Rad21 have been endogenously Halo-tagged[19], we applied stroboscopic photo-activation SPT (spaSPT, see Fig. 1a and refs. [19,20]) to track single protein molecules over time in live cells. spaSPT overcomes common biases in SPT by using stroboscopic excitation to minimize 'motion-blur'

[1]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA. [2]Li Ka Shing Center for Biomedical and Health Sciences, Berkeley, CA, USA. [3]CIRM Center of Excellence, University of California, Berkeley, Berkeley, CA, USA. [4]Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA, USA. [5]Department of Biological Engineering, MIT, Cambridge, MA, USA. [6]Department of Chemical Engineering, MIT, Cambridge, MA, USA. [7]These author contributed equally: Anders S. Hansen, Assaf Amitai. *e-mail: jmlim@berkeley.edu; darzacq@berkeley.edu

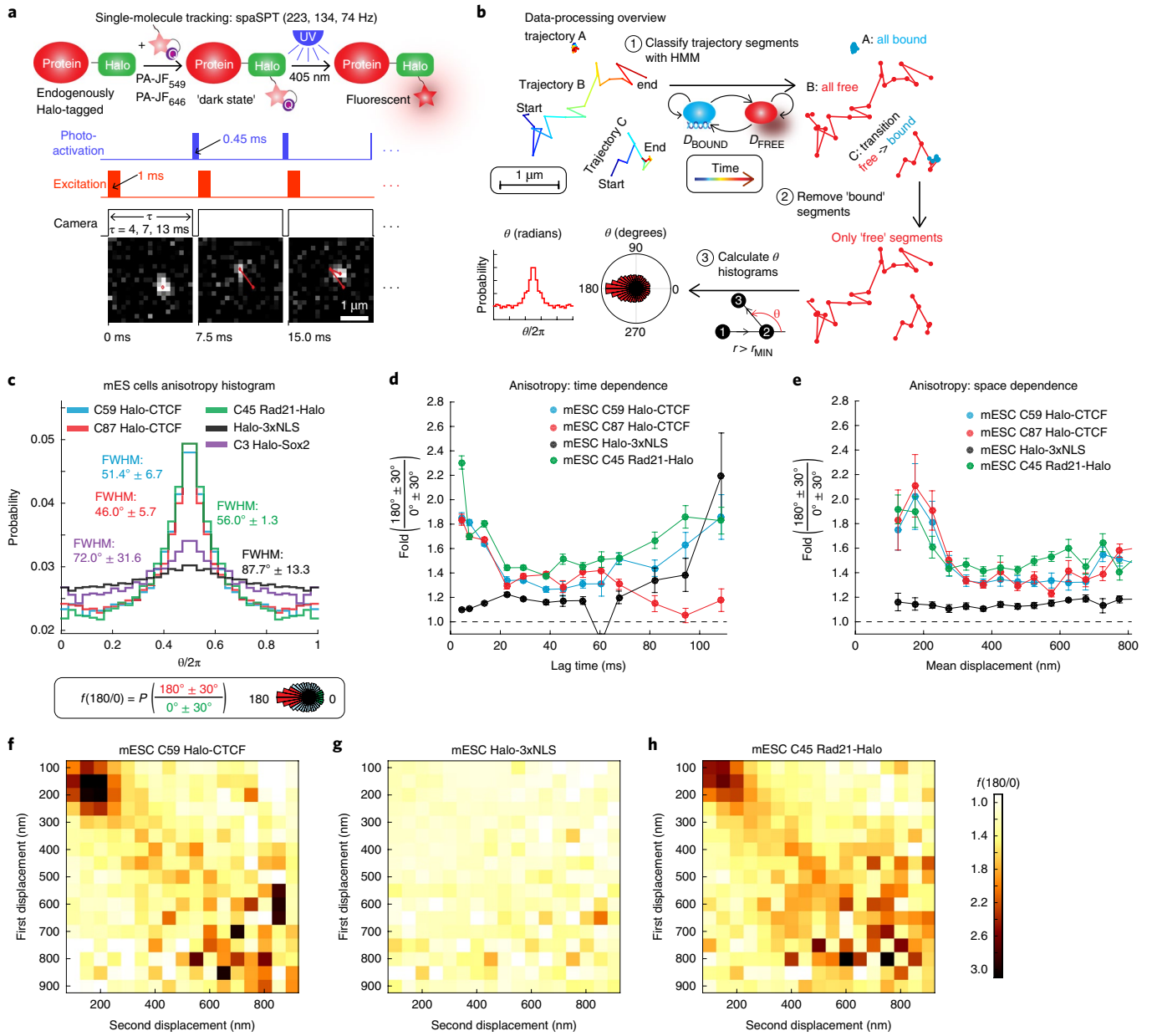

Fig. 1 | spaSPT reveals anisotropic CTCF diffusion in the nucleus. a, Overview of spaSPT. Sketch of HaloTag-labeling with PA-JF$_{549}$/PA-JF$_{646}$, photo-activation and overview of laser excitation pattern. Below, representative raw images with tracking overlaid. Scale bar, 1 µm. **b**, Overview of data-processing. spaSPT trajectories are classified into bound and free segments using a HMM[25]. The bound trajectory segments were then removed and angles are only calculated from free segments. Scale bar, 1 µm. **c**–**h**, As an additional criterion, the angle was only considered when both displacements making up the angle were >200 nm (**c**) or >125 nm (**d**–**h**). **c**, Bulk histograms of the distribution of angles for CTCF (C59, C87), Sox2 (C3) and Halo-3xNLS in mESCs. Definition of fold anisotropy, $f_{180/0}$. FWHM, full-width half maximum. **d**, Plot of $f_{180/0}$ versus lag time averaging over all displacement lengths. **e**, Plot of $f_{180/0}$ versus mean displacement length averaging over all lag times. Error bars in **d**,**e** show standard deviation from 50 subsamplings with replacement using 50% of the data and center values show value using 100% of the data. **f**–**h**, Anisotropy heatmaps showing $f_{180/0}$ versus length of the first and second displacement for C59 Halo-CTCF (**f**), Halo-3xNLS (**g**) and C45 mRad21-Halo (**h**, in the S/G2 phase of the cell cycle), averaging over all lag times.

bias[19–22] and by using photo-activation to track at low densities of ~0.5–1.0 molecules per nucleus per frame, which minimizes tracking errors[19,20,23,24]. Most nuclear proteins are in either a 'bound' chromatin-associated state (for example, trajectory A in Fig. 1b) or a seemingly 'free' diffusing state (for example, trajectory B in Fig. 1b). To explore the nuclear diffusion mechanism, we need to analyze exclusively the free trajectory segments. We applied a Hidden Markov model (HMM)[25] to classify trajectories into bound and free segments, removed the bound segments and calculated the angle[26–28] between three consecutive localizations provided that both displacements making up the angle were much larger than our localization error of ~35 nm (Fig. 1b and Supplementary Note 1). Finally, to comprehensively analyze the data at multiple spatio-temporal scales, we generated a large data set by imaging almost 2,000 single cells at three different frame rates (223, 134 and 74 Hz, see Data availability for raw data).

We first analyzed two independent mESC Halo-CTCF clones, C59 and C87. The angle distribution from diffusing CTCF trajectories showed a large peak at ~180° (Fig. 1c). This indicates that CTCF displays a directional bias: once CTCF has moved in one

direction, it is substantially more likely to move backward in the opposite direction than to continue forward. Brownian motion, which is isotropic, cannot explain this behavior. To quantify this effect, we define a 'fold anisotropy' metric, $f_{180/0}$: how many-fold more likely a step backward is relative to a step forward, which for CTCF was 1.77 in mouse embryonic stem cells (mESCs). While confinement inside the nucleus is expected to cause some anisotropy due to collisions with the nuclear envelope, the free, nuclear Halo-3xNLS protein was almost isotropic, $f_{180/0}=1.12$, ruling out nuclear confinement as the explanation (Fig. 1c). Could this be a general effect of all DNA-binding proteins? To test this, we analyzed Halo-Sox2 knock-in mESCs[29], but Sox2 was also much less anisotropic than CTCF ($f_{180/0}=1.27$, Fig. 1c). Thus, CTCF exhibits anisotropic diffusion distinct from other nuclear DNA-binding proteins.

A previous study reported that PTEFb exhibits scale-free anisotropic diffusion, with a magnitude of anisotropy that remains constant in space and time[26]. To see whether a similar mechanism holds for CTCF, we analyzed $f_{180/0}$ as a function of the lag time between frames (Fig. 1d) and the mean displacement length (Fig. 1e). After an initial decline, $f_{180/0}$ was relatively constant in time (Fig. 1d, measurements become noisy at long lag times (>60 ms) due to few long trajectories). CTCF showed a clear anisotropy peak at ~200 nm displacements (Fig. 1e). The spatial dependence was clearer when we plotted $f_{180/0}$ as a function of the length of the first and second displacements: CTCF showed a prominent peak (Fig. 1f), which was not seen for Halo-3xNLS (Fig. 1g). Using simulations (Supplementary Figs. 2 and 3), we verified that the observed anisotropy was not due to our localization uncertainty of ~35 nm (Supplementary Fig. 2a–c. However, note that at >60 nm localization uncertainty artifactual anisotropy appears, see Supplementary Fig. 3). We verified that our pipeline (Fig. 1b) removed the bound population almost completely (Supplementary Fig. 2d–f and Supplementary Table 1). Naturally, a protein bound to the chromatin polymer will fluctuate back and forth, creating apparent anisotropic motion[30,31]. Indeed, the velocity autocorrelation function of a chromatin-bound protein has a negative dip[32] because chromatin diffusion is subdiffusive. However, because of the approximately two to three orders of magnitude slower movement of chromatin compared to CTCF, such motion exhibits short displacements. We verified using simulations of chromatin polymer diffusion (Supplementary Fig. 2g) that our HMM-approach (Fig. 1b) fully filtered out this polymer-bound population. Thus, chromatin fluctuations cannot explain CTCF anisotropy. To further confirm that this result was also not due to tracking errors, we performed two-color spaSPT control experiments: we labeled Halo-CTCF 1:1 with two distinguishable dyes, PA-JF_{549} and PA-JF_{646}, which enabled us to identify tracking errors (for example, red to green switches in the same trajectory). As expected, tracking errors were almost nonexistent at the low densities of our imaging (Supplementary Fig. 4). However, we found that tracking errors increase exponentially with displacement length reaching ~5% around ~800 nm displacements (Supplementary Fig. 4a,f). We thus limited our analysis to displacements below this length (full discussion in Supplementary Note 2). Finally, we asked whether this mechanism is conserved between species and analyzed CTCF diffusion in human U2OS cells. Human CTCF similarly showed highly anisotropic diffusion peaking at ~200 nm displacements (Supplementary Fig. 5a–d). Notably, when we analyzed cohesin anisotropy, we found a similar anisotropic diffusion mechanism peaking around ~200 nm displacements (Fig. 1d,e,h and Supplementary Fig. 5e–j). We conclude that CTCF and cohesin exhibit a previously unreported mode of anisotropic nuclear diffusion that is conserved between mouse stem cells and human cancer cells.

Transient trapping in zones can explain CTCF dynamics. Whereas anisotropic PTEFb diffusion was scale-free and thus explainable with a fractal model[26], CTCF anisotropy is not uniform with displacement length (that is, it peaks at ~200 nm). Hence, a different mechanism is needed to explain CTCF dynamics. We hypothesized that weak and transient interactions could govern CTCF motion in the nucleus. If the probability of binding transiently at a given position was uniform throughout the nucleus, the resulting protein diffusion should be isotropic and Brownian, but with an effective (reduced) diffusion coefficient. However, if the source of CTCF anisotropic diffusion is transient trapping and retention in domains/zones of a characteristic size (Fig. 2a,b), this might explain the ~200 nm scale (Fig. 1e).

To test this hypothesis, we simulated chromatin as a coarse-grained self-avoiding polymer confined inside the nucleus[33] (Supplementary Fig. 6a). CTCF undergoes Brownian motion and interacts with chromatin in three distinct ways. First, CTCF can bind cognate binding sites (CBSs) (Fig. 2a) with a Poisson-distributed residence time for specific binding ($\tau_{CBS}=1$ min)[19], which is effectively infinitely stable on the time-scale of our SPT experiments (milliseconds) and will thus be filtered out by our analysis pipeline (Fig. 1b). Second, on encountering a transiently trapping zone (TTZ) (monomer, arriving at a distance ε_{TTZ} from its center) CTCF has a certain probability, P_{trap}, of becoming absorbed and trapped. We model TTZs as spherical domains with radius $\varepsilon_{TTZ}=200$ nm (Fig. 2a). CTCF then diffuses inside the TTZ and can exit it with probability P_{exit} every time it hits its boundary (Supplementary Fig. 6b,c). Thereby, TTZs trap the protein for a much shorter time than the CBSs. The model parameters are chosen such that the protein has a probability to escape a zone on release rather than rebinding immediately. While diffusing inside the TTZ, the protein can transiently bind at any position. We will discuss a third class of zones, power-law distributed trapping zones (PTZs), later.

We then tested our model by simulating Brownian diffusion using Euler's scheme subject to cognate and transient interactions (see Methods for simulation parameters). Although protein diffusion is Brownian in our model, transient trapping of CTCF in the TTZs faithfully reproduced our experimental observations (Fig. 2c–f), including high anisotropy (Fig. 2c,d) and a clear peak of anisotropy at ~200 nm mean displacements (Fig. 2f). $f_{180/0}$ shows a relatively constant trend in time (Figs. 1d and 2e and Supplementary Fig. 5b), suggesting that the interaction time of a protein within a zone has a broad distribution. We note that since CTCF escape from the TTZ is diffusion limited, the interaction time does not have a power-law distribution. The location of the peak in Fig. 2g provides information about the size of the TTZ. Our simulations reproduced this behavior only at high trapping probability—if we lowered the probability P_{trap} of trapping CTCF in these zones, the diffusion remained anisotropic, but the anisotropy peak largely disappeared (Fig. 2f,h). We refer to this model as anisotropic diffusion through transient trapping in zones (ADTZ).

What is the underlying mechanism of ADTZ? When a protein transiently interacts with chromatin, its dynamics will be governed by rapid reattachment to the release site[34]. If a protein is bound to a site, on dissociation it is more likely to re-attach to the same site rather than bind to another site. Protein diffusion therefore appears as if biased toward its starting position. When the time to return is on the order of the frame rate of our SPT experiments (milliseconds) the protein will appear to take a step back following a forward step. We call this mechanism that will contribute to the angle anisotropy re-attachment (Fig. 2b). Additionally, while the protein is trapped in the TTZ it is reflected from the domain boundary with some probability. We call this mechanism retention (Fig. 2b). Hence, we suggest that anisotropy may originate from a combination of reattachment and retention.

Finally, diffusion is also often analyzed using the mean squared displacement (MSD), which grows as a power law with time $\mathrm{MSD}_i(\tau) = (r_i(t+\tau) - r_i(t)) \sim \tau^{\alpha}$, with $\alpha<1$ for subdiffusive

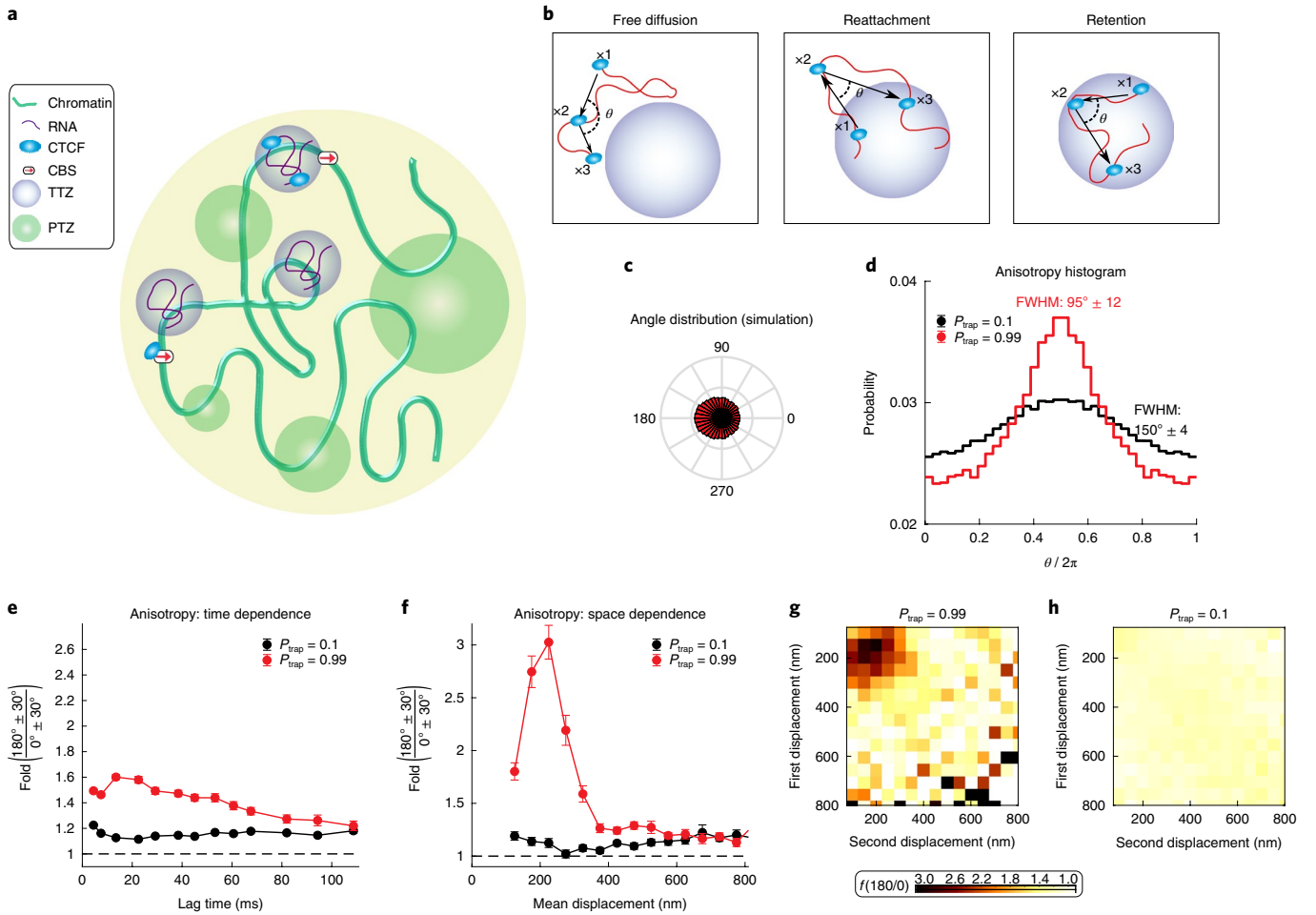

Fig. 2 | A model wherein CTCF diffusion in the nucleus is governed by its interaction with trapping zones can explain the experimental data. a, CTCF (light blue ellipse) diffuses inside the nucleus (large yellow circle). It interacts with one of three zone types: (1) small, cognate DNA binding sites (CBSs) (red arrow), (2) transiently trapping zones (TTZs) of size 200 nm (purple) and (3) trapping zones of different dimensions, whose size is derived from a power-law distribution (power-law distributed trapping zones, PTZs) (green, for details about PTZs, please see Fig. 3 and below). **b**, The protein (light blue) is observed at three time points (X_1, X_2, X_3) along its stochastic trajectory (red curve). Left, outside the zones, CTCF performs free diffusion. The angle of two consecutive steps is uniformly distributed for this mode of motion. Middle, if CTCF leaves a TTZ (X_1), diffuses freely (X_2) and then reattaches back to the TTZ (X_3), the angle distribution is anisotropic. Right, while inside a zone, CTCF can be reflected from its boundary. The angle distribution, in this case, will be anisotropic. **c**, The angle distribution for a protein performing diffusion in a nucleus containing only CBSs and TTZs. FWHM, full-width half maximum. See Supplementary Table 2 for the simulation parameters. **d**, Angle distribution of the trajectory computed from the simulation when the nucleus contains only zones of type 1 and 2 (CBSs and TTZs). CTCF is trapped with high probability ($P_{trap} = 0.99$) when touching the zones (red curve). The protein is trapped with low probability ($P_{trap} = 0.1$) when touching the zone (black curve). **e**, Plot of $f_{180/0}$ versus lag time averaging over all displacement lengths of the simulation data shown in **d**. **f**, Plot of $f_{180/0}$ versus mean displacement length averaging over all lag times. The data shown in **e**,**f** is the result of $n = 30$ independent simulations. In each simulation, the trajectory of CTCF was recorded until its 10,000s capture event. Error bars in **e**,**f** show standard deviation from 50 subsamplings with replacement using 50% of the data and center values show value using 100% of the data. **g**,**h**, Anisotropy heatmaps showing $f_{180/0}$ versus length of the first and second displacement for high binding probability ($P_{trap} = 0.99$) (**g**) and for smaller binding probability ($P_{trap} = 0.1$) (**h**). See Supplementary Table 2 for the full parameters used.

proteins[3,35]. For CTCF, the exponent was in the range $0.83 < \alpha < 0.92$ (Supplementary Fig. 1, the inferred α-value is sensitive to mean squared displacement-fit conditions), which is comparable to what we get with the ADTZ model ($\alpha = \sim 0.92$, Supplementary Fig. 7). Subdiffusion is often modeled phenomenologically using continuous time random walk or Fractional Langevin motion (fLm) models[30,36], but through analysis of the normalized velocity autocorrelation function ($C \equiv C_v^{\Delta t}(\tau)/C_v^{\Delta t}(0)$, Supplementary Fig. 8) we find that neither model can explain CTCF dynamics (see Supplementary Note 3). We conclude that the ADTZ model can better explain CTCF dynamics than the other tested models.

RBR_i and ZF domains control CTCF diffusion. Our simulations suggest an intriguing mechanistic model (ADTZ) where zones of an unknown nature transiently trap diffusing CTCF with high probability. What could be the nature of these TTZs? Having previously shown using super-resolution photo-activated localization microscopy (PALM) imaging that CTCF and cohesin form small colocalizing clusters in the nucleus[19], here we hypothesized that the TTZs could correspond to CTCF clusters. This is because clustering is due to self-interaction and self-interaction might also explain transient CTCF trapping. In a companion paper (ref. [37]), we show that CTCF self-association is largely RNA-mediated and DNA-independent consistent with a previous report[38] and that CTCF clustering is substantially reduced after endogenous deletion of an internal RNA-binding region in CTCF (RBR_i; mESC C59D2 ΔRBR_i-Halo-CTCF; amino acids 576–611 have been replaced with 3xHA). Thus, if the ADTZ model hypothesis is correct and TTZs correspond to CTCF

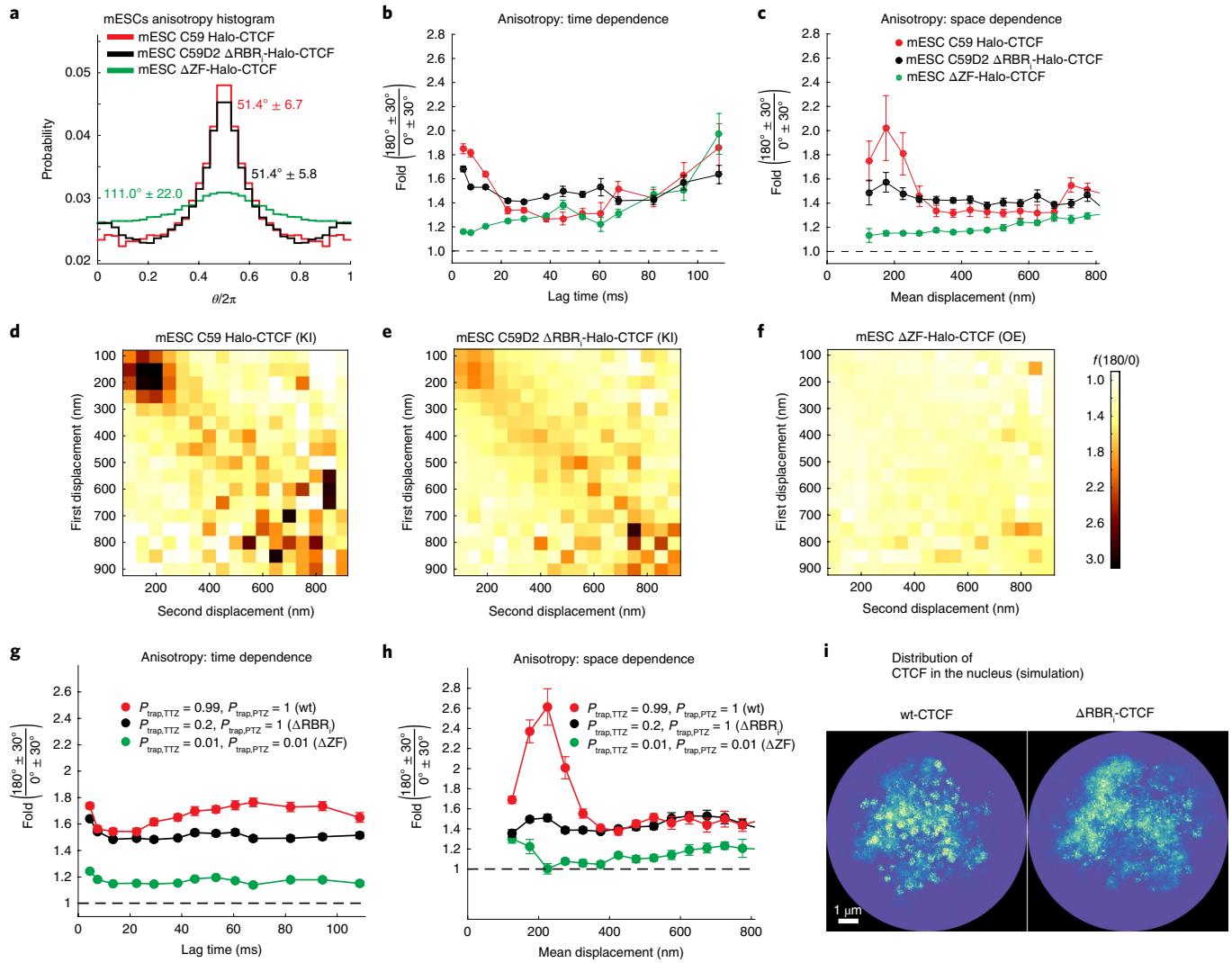

Fig. 3 | Anisotropy and nuclear distribution of ΔRBR$_i$-CTCF. a, Bulk angle distribution histograms for CTCF (C59) (red), an endogenous CTCF mutant lacking the RBR$_i$ domain (C59D2) (black), CTCF with a deletion of CTCF's 11 ZF domain (ΔZF) and that likely binds neither DNA nor RNA (ΔZF-Halo-CTCF) (green). **b**, Plot of $f_{180/0}$ versus lag time averaging over all displacement lengths. **c**, Plot of $f_{180/0}$ versus mean displacement length averaging over all lag times. **a**–**c**, show standard deviation from 50 subsamplings with replacement using 50% of the data and center values show value using 100% of the data. **d**–**f**, Anisotropy heatmaps showing $f_{180/0}$ versus length of the first and second displacement for the three cell lines, averaging over all lag times. In **a**–**f**, KI refers to endogenous knock-in cell lines (C59, C59D2) and OE refers to mutants studied under exogenous over-expression conditions. Over-expression conditions were optimized (see Supplementary Fig. 10) to minimize the well-known artifacts associated with CTCF over-expression[19,39]. Simulations. **g**,**h**, Plot of $f_{180/0}$ versus lag time averaging over all displacement lengths (**g**) or versus mean displacement length averaging over all lag times (**h**). The data shown in **g**,**h** is the result of $n = 30$ independent simulations. In each simulation, the trajectory of CTCF was recorded until its 10,000s capture event. Error bars and center values are computed in the same way as in **b**,**c**. Red, full model representing wt-CTCF, which interacts with high trapping probability with all three zones (CBS, TTZ, PTZ: see Fig. 2a). Black, model of ΔRBR$_i$-CTCF, which interacts with high trapping probability with CBS and PTZ but has a very weak affinity for the TTZ. Green, model representing ΔZF-CTCF, which has a very low affinity to the CBSs, TTZs and PTZs. **i**, The distribution of CTCF in the nucleus estimated from the simulation (full model corresponding to wt-CTCF) (left), and the model where CTCF weakly interacts with TTZ (corresponding to ΔRBR$_i$-CTCF) (right). Scale bar, 1 μm. See Supplementary Table 2 for the full details of the parameters used. Each image (left and right) is the two-dimensional projection of the probability distribution function of simulated CTCF, estimated from one representative simulation out of the 30 simulations whose results are shown in **g**,**h**. In each simulation, the trajectory of CTCF was recorded until its 10,000s capture event.

clusters, it should be possible to abolish TTZ-mediated trapping of CTCF in two ways: (1) by reducing the trapping probability of CTCF (Fig. 2d–f) or (2) by reducing the number of CTCF clusters.

To test these hypotheses, we performed spaSPT for a series of CTCF mutants in mESCs (Fig. 3a–f) and U2OS cells (Supplementary Fig. 9). The ΔRBR$_i$-Halo-CTCF mutant is a knock-in, but since the other mutants (for example, ΔZF-CTCF) would likely be lethal, we developed a low transient over-expression protocol to minimize the artifacts that are otherwise observed with strong over-expression of CTCF[19,39] (Supplementary Fig. 10). Whereas a full deletion of CTCF's 11 ZF domain (ΔZF), which is required for DNA-binding, caused CTCF diffusion to become nearly isotropic, ΔRBR$_i$-CTCF remained almost as anisotropic as wild-type- (wt-)CTCF at the bulk level (Fig. 3a). However, analysis of the fold anisotropy ($f_{180/0}$) as a function of time (Fig. 3b) and space (Fig. 3c) revealed two surprising results. First, the anisotropy peak largely, although not entirely, disappeared in ΔRBR$_i$-CTCF (Fig. 3c–e). Our ADTZ theory (Fig. 2) predicts that this can occur in two ways: (1) by reducing

the number/fraction of TTZs (see Supplementary Fig. 11 for full parameter scan of the model) and (2) by reducing the trapping probability (Fig. 2a,b), P_{trap}, to the TTZs. Clustering (TTZs) is clearly reduced in ΔRBR_i-CTCF mESCs[37] (see below) and it is likely that P_{trap} is too. These results paint a mechanistic picture for the ADTZ model: the RBR_i domain mediates CTCF clustering, which serve as TTZs that transiently trap diffusing CTCF in zones of a defined size (~200 nm), resulting in highly anisotropic CTCF diffusion. However, this points to the second surprising result: ΔRBR_i-CTCF diffusion is still highly anisotropic (Fig. 3a), but with a magnitude ($f_{180/0}$) that is approximately constant in both time and space (Fig. 3b,c). Thus, ΔRBR_i-CTCF exhibits approximately scale-free anisotropy similar to PTEFb[26].

Since ΔZF-CTCF is essentially isotropic, these results suggest a more complicated model wherein CTCF anisotropy arises through a combination of two mechanisms. First, RBR_i-mediated interactions transiently trap CTCF in small TTZs, resulting in a peak of anisotropy at ~200 nm. Second, ZF-mediated interactions, perhaps due to transient DNA-interactions, generally trap CTCF without a clearly defined scale. Lending further support to this interpretation, we observed similar ΔZF-, ΔRBR_i- and wt-CTCF behavior in human U2OS cells (Supplementary Fig. 9).

We next studied whether a new class of zones might explain the scale-free anisotropy of ΔRBR_i-CTCF (Fig. 3a–e). Because the data suggests a 'scale-free' anisotropy, we added to the nucleus zones of different sizes that are sampled from a power-law distribution (Fig. 2a) ($P_{PTZ}(\varepsilon) \sim \varepsilon^{-\gamma}$) and term them PTZs. To reproduce our experiments, we choose the PTZs such that large zones are rarer (see Methods). We then performed Brownian simulations of an interacting protein, with chromatin represented as a polymer where the majority of its monomers are TTZs (of size 200 nm), a smaller fraction is PTZs, and a third fraction consists of CBSs. The results recapitulate both the peak of anisotropy at ~200 nm and the large ($f_{180/0}$ = ~1.5) and scale-free anisotropy at longer displacements (Fig. 3g,h). Thus, the ADTZ model can explain the behavior of wt-CTCF and when we 'computationally mutate' CTCF such that it has weak interaction with TTZs, the model can also explain the behavior of ΔRBR_i-CTCF. Moreover, we computationally estimated the distribution and clustering of wt-CTCF and ΔRBR_i-CTCF in the nucleus as it results from our simulations (Fig. 3i). Finally, we analyzed the velocity autocorrelation function for our data and model (Supplementary Fig. 12) as well as probability distribution function[40] of displacements (Supplementary Fig. 13), and found the probability distribution function to be non-Gaussian both for the experimental data and in our model (full analysis in Supplementary Note 3).

Direct evidence that TTZs correspond to CTCF clusters. Next, we attempted to directly test our model that TTZs correspond to CTCF clusters and are the mechanistic origin of the higher anisotropy of wt-CTCF compared to ΔRBR_i-CTCF (Fig. 3a–f), by simultaneously visualizing CTCF diffusion (spaSPT) and clusters (PALM). We labeled Halo-CTCF in mESCs with two distinguishable photoactivatable dyes: ~10% of CTCF with PA-JF_{549} for spaSPT to follow anisotropic diffusion and ~90% of CTCF with PA-JF_{646} for PALM to visualize clustering in live cells (Fig. 4a). We then performed two-color simultaneous spaSPT/PALM, tracked and localized particles at ~134 Hz, and corrected for drift[41] (Fig. 4b). We analyzed spaSPT data as above (Fig. 1b), removing the bound trajectory segments and kept only the localizations corresponding to anisotropic trajectory segments (Fig. 4b–d). For PALM, we merged single molecules appearing in multiple frames and/or blinking and then assigned clusters using DBSCAN[42]. Our PALM localization precision was ~23 nm (standard deviation, Supplementary Fig. 14a). Anisotropic displacements will occur both by chance and, perhaps, due to TTZ-mediated transient trapping, in which case anisotropic localizations should be enriched near CTCF clusters for wt-CTCF, but not for ΔRBR_i-CTCF. We tested this by calculating the pair cross-correlation function[43] between anisotropic localizations and cluster localizations. Indeed, we found that anisotropic and cluster localizations were significantly more likely to occur close to each other ($\lessapprox$250 nm) for wt-CTCF (Fig. 4e). In contrast, anisotropic and cluster localizations were neither more nor less likely to occur close to each other for ΔRBR_i-CTCF. We speculate that the pair cross-correlation does not fully decay to 1 (completely spatial randomness) for either wt-CTCF or ΔRBR_i-CTCF due to slight nucleolar exclusion. Thus, these observations provide direct evidence that CTCF clusters likely correspond to TTZs and are a cause of RBR_i-mediated anisotropy.

However, simultaneous two-color spaSPT and PALM is challenging and subject to limitations including: significant cell movement during the experiment, difficulties assigning clusters in live cells due to movement and incomplete labeling, two-color registration errors and a limited number of anistropic localizations in a single-cell experiment. Nevertheless, most biases would either degrade our ability to detect colocalization and/or apply equally to wt-CTCF and ΔRBR_i-CTCF. Moreover, we observe the same colocalization even without assigning clusters (Supplementary Fig. 14b). We therefore cautiously conclude that these observations are consistent with our model that TTZs correspond to clusters and are the source of RBR_i-mediated anisotropic diffusion.

CTCF target search mechanism is RBR_i-guided. Having elucidated a complicated mode of CTCF diffusion, we next asked if the function could be to regulate the efficiency of the search for CBSs. To study CTCF target search efficiency, we performed spaSPT and fluorescence recovery after photobleaching (FRAP) experiments. First, spaSPT experiments coupled with two-state model-based analysis[19,20] revealed that the apparent free diffusion coefficient (D_{FREE}) was largely unaffected, whereas the total bound fraction decreased significantly from ~62.4 to 42.0% (Fig. 5a–c). The total bound fraction captures both specific binding to CBSs (~1–4 min residence time[19]) and 'nonspecifically bound' CTCF. Both FRAP and spaSPT independently revealed that the specifically bound fraction was substantially reduced, whereas the nonspecifically bound fraction barely changed (Fig. 5d and Supplementary Fig. 15a–c). However, model-fitting of the FRAP data revealed that the apparant residence time for binding CBSs was unchanged for ΔRBR_i-CTCF. Consistently, we confirmed that RBR_i-deletion does not affect CTCF affinity for its canonical DNA-binding site in vitro, using a fluorescence polarization assay with FAM-labeled DNA duplexes and recombinant proteins (Fig. 5e and Supplementary Fig. 15d–g).

Thus, these results demonstrate a surprising function of CTCF RBR_i: the strength of CTCF specific binding to cognate DNA sites is unaffected by the RBR_i, but the amount of CTCF engaged in specific binding is strongly impacted. In a simplified two-state model, the fraction of CTCF specifically bound to chromatin is controlled by the balance between its binding (ON) and dissociation (OFF) rates ($F_{BOUND} = \frac{k^*_{ON}}{k^*_{ON}+k_{OFF}}$). Within this framework and using our measurements, we find that the RBR_i increases k^*_{ON} by ~2.5-fold (see Methods for calculation). The ON-rate is the inverse of the time it takes the protein to find its CBS (search time). Thus, the RBR_i increases the frequency or rate of CTCF finding a specific DNA-binding site by ~2.5-fold, without affecting CTCF affinity for specific binding sites. This suggests that the ADTZ-diffusion mechanism serves to increase the efficiency of CTCF's target search mechanism. Additionally, this suggests that TTZs serve to 'load' or 'guide' CTCF toward CBSs. Hence, we speculate that TTZs contain CTCF CBSs (Fig. 5f, left), which would allow the TTZs to increase the local concentration of CTCF around them, thus increasing the local ON-rate.

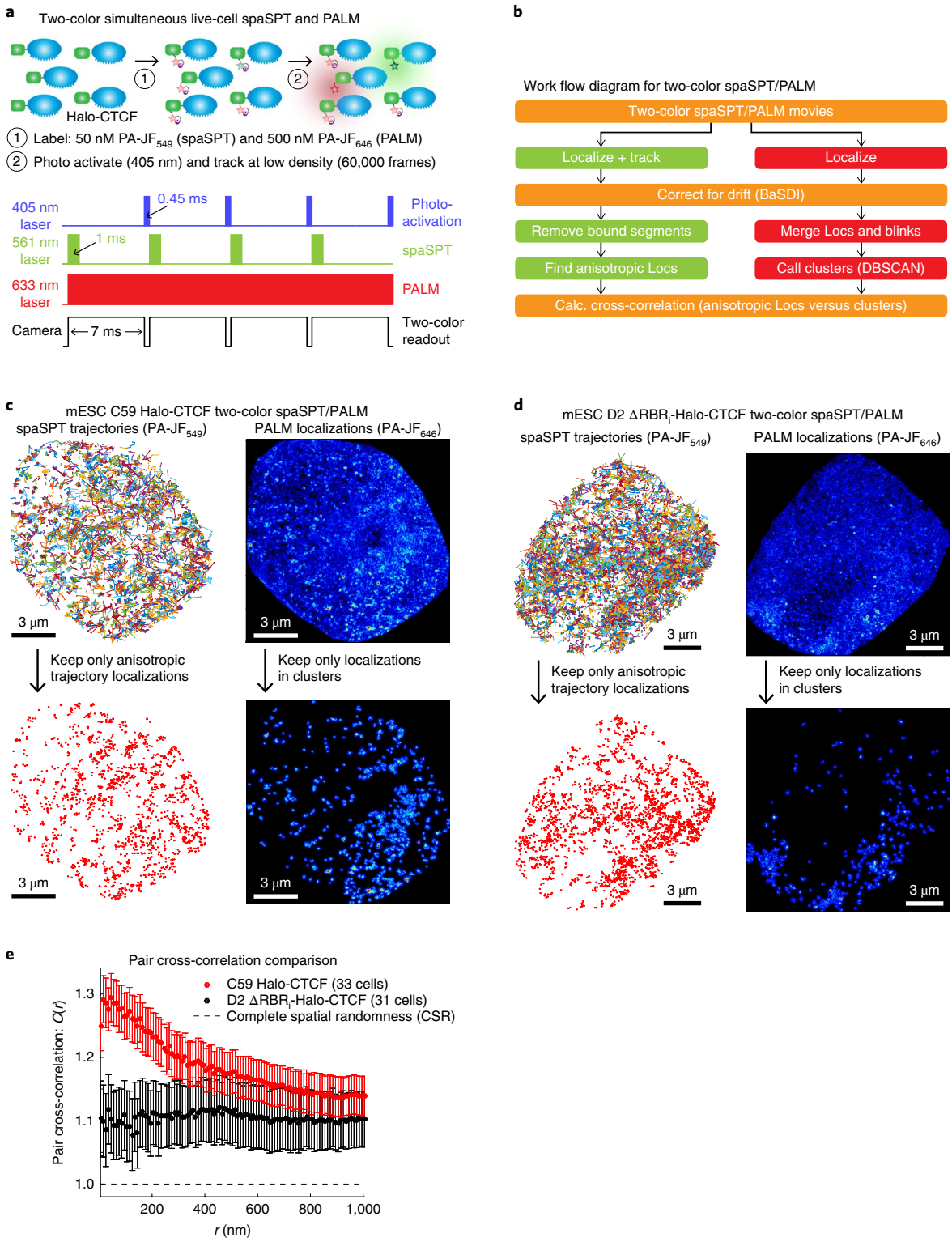

Fig. 4 | Direct evidence that TTZs correspond to CTCF clusters. a, Overview of two-color live-cell simultaneous spaSPT (using Halo-PA-JF_{549}) and PALM (using Halo-PA-JF_{646}) as well as laser and camera settings. Cells were imaged until near completion (60,000 frames; almost all particles photo-activated and bleached). Labeling, ~90% Halo-PA-JF_{646} for PALM to see clusters; ~10% Halo-PA-JF_{549} for spaSPT to see anisotropic trajectories. **b**, Outline of the computational analysis workflow. **c**,**d**, Representative raw spaSPT (left) and raw PALM reconstruction (right) data for C59 Halo-CTCF (**c**) and D2 ΔRBR_i-Halo-CTCF (**d**). Below, the three localizations making up each anisotropic trajectory segment (angle ∈ (150; 210°)) are shown (left). Right, clusters assigned based on PALM data using DBSCAN (radius, $\varepsilon = 100$ nm). Out of a total of 33 cells for C59 and 31 cells for D2 from $n = 3$ biologically independent replicates. Scale bars, 3 μm. **e**, Pair cross-correlation analysis. We calculated the pair cross-correlation[43] between localizations making up the anisotropic segments of SPT trajectories and the localizations making up clusters determined from PALM. A value of 1 is expected for complete spatial randomness (CSR), whereas a value above 1 indicates 'enrichment'. Error bars show standard error of the mean from $n = 3$ biologically independent replicates for the indicated number of cells, where 8–12 cells were collected per replicate.

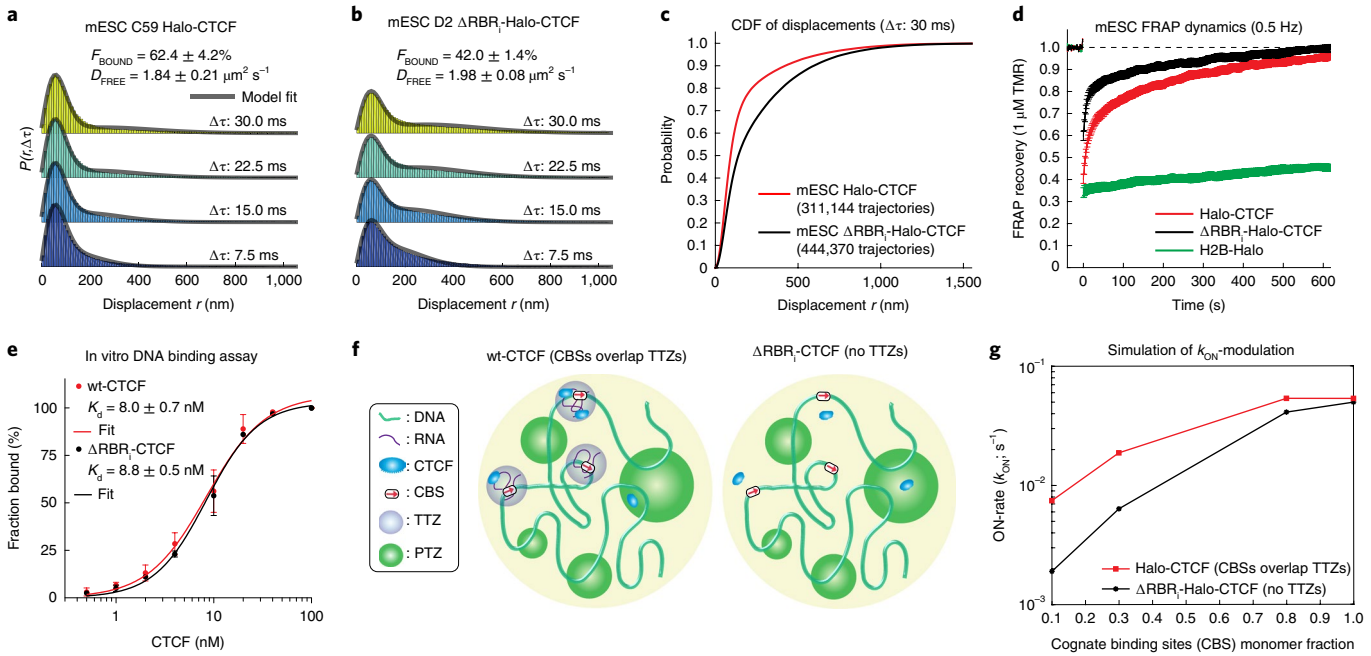

Fig. 5 | RBR_i-guided CTCF target search mechanism. a,**b**, spaSPT displacement histograms for C59 wt-CTCF (**a**) and D2 ΔRBR_i-CTCF (**b**). Raw displacement data for four different lag times are shown with a two-state Spot-On model fit (bound versus free) overlaid[19,20]. The inferred bound fraction and free diffusion coefficients are shown (mean across at least $n = 4$ biological replicates) as is the standard error (±). **c**, Cumulative probability function of displacement lengths ($\Delta\tau = 30$ ms) for the same data as in **a**,**b**. **d**, FRAP data for wt-CTCF (18 cells), ΔRBR_i-CTCF (18 cells) and histone H2B (control). Model-fits and inferred residence times are shown in Supplementary Fig. 15. The data show mean and s.e.m. for 18 single cells per condition from $n = 3$ biologically independent replicates. **e**, DNA affinity of recombinant wt-CTCF and ΔRBR_i-CTCF measured by fluorescence polarization binding assays. A labeled DNA duplex containing the core CTCF binding site (5 nM) was incubated with increasing amounts of protein (0.5, 1, 2, 4, 10, 20, 40, 100 nM). Changes in fluorescence polarization were used to compute binding curves, fitted here to a Hill equation to estimate dissociation constant (K_d) and Hill coefficient (h) for wt and mutant CTCF (see Methods and Supplementary Fig. 15d–g). Plotted are mean values and standard deviations ($n = 3$ biologically independent experiments). **f**, Overview of simulation to test effect of TTZs on k_{ON}. Illustration of a model where each CBS is surrounded by a TTZ (left) corresponding to wt-CTCF. Right, illustration of a model where the CBSs are not surrounded by TTZs. The nucleus contains only PTZs and CBSs. This model corresponds to ΔRBR_i-CTCF. **g**, The ON-rate to find a CBS (inverse of the mean first encounter time). The ON-rate was found for the two models as a function of the fraction of monomers that are CBSs out of all monomers. The other monomers are PTZs (size drawn from a power-law distribution—see Supplementary Table 2). The ON-rate was estimated for the wt-CTCF model (red curve) where each CBS monomer (of radius 30 nm) is surrounded by a TTZ (radius 200 nm). The ON-rate was also estimated for the ΔRBR_i-CTCF model (blue curve) where the CBS monomers are not surrounded by the TTZ. The size distribution of the PTZs and the other parameters used were the same as in Fig. 3g,h, except for the total number of monomers, which was 20.

To test computationally whether having CBSs inside TTZs could accelerate the search, we performed simulations where a fraction of the monomers were CBSs (with a capture radius of 30 nm, Supplementary Table 2) and the rest of the monomers were PTZs that have a wide size distribution (off-target trapping zones). In the wt-CTCF model (Fig. 5f, left), each CBS is surrounded by a TTZ of size 200 nm. By contrast, in the ΔRBR_i-CTCF model (Fig. 5f, right), there are no TTZs, only CBSs and PTZs. Using simulations, we estimated the ON-rate for CTCF protein encountering a CBS for the first time (inverse of the mean first passage time). In the limit of no off-target sites (no PTZs, only CBSs), TTZ-mediated trapping did not significantly accelerate the ON-rate (Fig. 5g). In contrast, when the fraction of CBS was low (many off-target PTZs), the presence of TTZs accelerated the search fourfold. We conclude that RBR_i-mediated trapping in TTZs accelerates the CBS-target search by effectively increasing the cross-section of target sites (TTZs (200 nm) are much larger than CBSs (30 nm)) and reducing trapping in off-target PTZs.

Discussion

It has long been known that diffusion in cells can be anomalous or non-Brownian[3] and anomalous subdiffusive behavior has previously been reported for DNA[44], RNA[45] and proteins[26], but our mechanistic and functional understanding has been limited. Here, we discover a different mode of nuclear exploration: ADTZ (Fig. 2). Unlike phenomenological models of subdiffusion such as fractional Langevin motion or continuous time random walk[3], our model offers a mechanistic explanation for CTCF dynamics.

We suggest that CTCF diffusion is largely regulated by two mechanisms. First, the RBR_i domain in CTCF contributes to CTCF forming clusters in the nucleus, and at the same time, results in transient trapping of CTCF in TTZs of a small size (~200 nm, Fig. 6a). Our model that TTZs correspond to CTCF clusters is corroborated by our direct observation of anisotropic diffusion near CTCF clusters for wt-CTCF, but not for ΔRBR_i-CTCF (Fig. 4). Speculatively, since the RBR_i domain mediates both CTCF clustering and RNA-binding[37], we suggest that if RNA(s) directly or indirectly holds together CTCF clusters, weak CTCF-RNA interactions mediated by the RBR_i domain may repeatedly trap diffusing CTCF. Nevertheless, we emphasize that the molecular architecture of CTCF clusters remains unknown, that ZF1 and Z10 of CTCF may also bind RNA in addition to the RBR_i domain[37,46] and that putative CTCF-RNA interactions would need to be relatively weak and transient (milliseconds to tens of milliseconds) to result in anisotropic diffusion.

The second mechanism that regulates CTCF dynamics is through its 11 ZF domains, which is also responsible for cognate DNA-binding. ΔRBR_i-CTCF still displays anisotropic diffusion with a scale-free magnitude, but ΔZF-CTCF exhibits isotropic

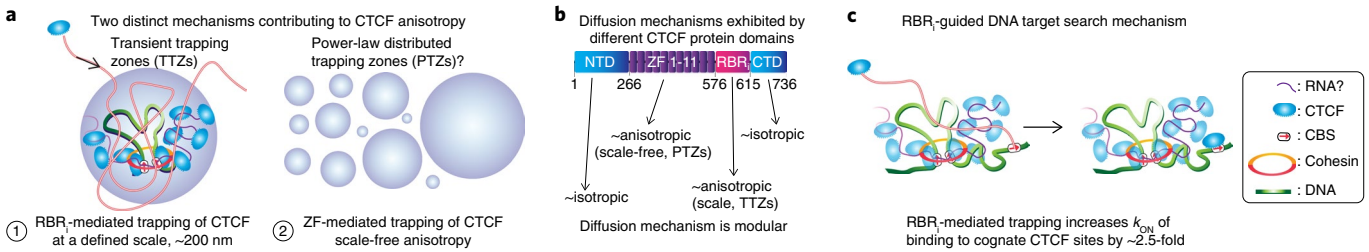

Fig. 6 | Model. a, Two distinct mechanisms contribute to CTCF anisotropy. First, RBR_i-mediated trapping of CTCF in TTZs of a characteristic size (~200 nm). TTZs are likely to correspond to clusters of CTCF that form in a largely RBR_i-dependent manner[37] and therefore, perhaps, are held together by RNA(s) of a currently unknown identity. CBSs reside within the TTZ, as does a piece of DNA strand (green). Second, ΔRBR_i-CTCF still displays scale-free anisotropy, which may be due to nonspecific interactions with DNA. We speculatively model this as arising from trapping in PTZs. **b**, Diffusion mechanism is modular. By mutating individual protein domains, the effect of each individual domain can be determined. CTCF contains four major protein domains: an N-terminal domain (NTD) of largely unknown function, 11 ZFs essential for DNA-binding, a short RBR_i and a C-terminal domain (CTD) of largely unknown function. The ZF domain and the RBR_i domain appear to mediate anisotropic diffusion via PTZ and TTZ type mechanisms, respectively. **c**, Target search mechanism. The RBR_i increases the apparent rate at which CTCF locates a CBS. Given that the RBR_i also mediates CTCF clustering, we speculate that RBR_i-mediated CTCF clusters, perhaps near loop boundaries, help to guide diffusing CTCF toward its CBSs (DNA is shown in green).

diffusion (Fig. 3) This suggests that the ZF domain mediates 'scale-free anisotropic diffusion'. Conceptually, this suggests that protein diffusion is modular, tunable and programmable (Fig. 6b)[2]. Specifically, some protein domains exhibit essentially isotropic diffusion (for example, the N- and C-terminal domains of CTCF, Halo-3xNLS and so on). CTCF's RBR_i domain mediates anisotropic diffusion through an ADTZ-type mechanism and CTCF's 11 ZF domain mediates anisotropic diffusion with a scale-free magnitude perhaps through a PTZ-type mechanism (Fig. 3). Since the RBR_i and ZF domains interact primarily with RNA and DNA, respectively, we suggest that TTZ-mediated trapping in ~200 nm zones may be mediated by CTCF-RNA interactions and that scale-free anisotropic diffusion, observed for ΔRBR_i-CTCF, is most likely due to DNA-mediated interactions (Fig. 6b). We model scale-free DNA-mediated anisotropy as originating from transient interaction of CTCF with zones of different sizes (Fig. 6a, PTZs), which we speculate have a power-law distribution. We speculate that PTZs may correspond to TAD or A/B-compartment structures, which vary both in range and size[17], but we stress that we have no clear data to suggest a physical or biological origin of PTZs and that several other models—including diffusion on a fractal[26], or diffusion in a disordered medium[47]—can all equally well explain the scale-free anisotropy observed for ΔRBR_i-CTCF.

We conclude that both ZF-domain- and RBR_i-domain-mediated interactions govern CTCF diffusion in a modular manner (Fig. 6b) and these interactions result in anisotropy that manifests itself at different scales. Thus, by mixing and matching protein domains with defined diffusion mechanisms, it should in principle be possible to design a protein with a desired diffusion mechanism. This could be exploited both in synthetic biology approaches to engineer proteins with desired target search mechanisms or by the cell during evolution to fine-tune function.

Our results suggest a strong link between CTCF clustering, diffusion and target search mechanism (Fig. 6c). Since the RBR_i increases the rate at which CTCF locates CBSs by ~2.5-fold, this suggests a model where the CTCF DNA-target search is guided by its RBR_i domain and where CTCF clusters guide diffusing CTCF proteins toward nearby CBSs (Fig. 6c). A guided target search may be advantageous in mammalian cells, which have many more off-target sites on their genomes. This model significantly changes the view on the mechanism of target search and protein localization in mammalian cells from the facilitated diffusion picture from bacteria. Guided search has also been observed in bacterial protein-DNA-binding sites, which tend to be flanked by AT-rich DNA, serving as an 'energetic funnel' guiding proteins to their target sites[48]. Similarly, on infection of human cells with Herpes Simplex Virus, RNA Pol II becomes enriched in viral replication compartments, where its diffusion becomes anisotropic[6]. These observations are further consistent with the ADTZ model, where viral replication compartments may serve as TTZs for Pol II.

CTCF and cohesin regulate genome organization together[17]. It is therefore striking that both CTCF (Fig. 1) and cohesin (Fig. 1h and Supplementary Fig. 5e–j) exhibit similar ADTZ-type anomalous diffusion. It will be interesting to explore in the future if TTZs also contribute to topological loading of cohesin on chromatin. Consistent with an RBR_i-connection between CTCF, cohesin and genome organization, we show in a companion paper that a significant fraction of CTCF loops are lost in ΔRBR_i-CTCF mESCs[37]. Thus, the same protein domain simultaneously regulates CTCF diffusion, clustering, target search mechanism and function. Our results highlight the power of SPT, theory and analysis, when coupled with genome-editing and mutations, as an approach to discover protein domains engaging in important interactions. It may be informative to apply a similar approach to other nuclear proteins in the future—especially for proteins that also form clusters or hubs[49,50].

Online content

Any methods, additional references, Nature Research 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/s41589-019-0422-3.

Received: 20 April 2019; Accepted: 30 October 2019;
Published online: 02 December 2019

References

1. Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends Genet.* **27**, 295–306 (2011).
2. Woringer, M. & Darzacq, X. Protein motion in the nucleus: from anomalous diffusion to weak interactions. *Biochem. Soc. Trans.* **46**, 945–956 (2018).
3. Metzler, R., Jeon, J.-H., Cherstvy, A. G. & Barkai, E. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Phys. Chem. Chem. Phys.* **16**, 24128–24164 (2014).
4. Höfling, F. & Franosch, T. Anomalous transport in the crowded world of biological cells. *Reports Prog. Phys.* **76**, 046602 (2013).
5. Rhodes, J., Mazza, D., Nasmyth, K. & Uphoff, S. Scc2/Nipbl hops between chromosomal cohesin rings after loading. *eLife* **6**, e30000 (2017).

6. McSwiggen, D. T. et al. Evidence for DNA-mediated nuclear compartmentalization distinct from phase separation. *eLife* **8**, e47098 (2019).
7. Bancaud, A., Lavelle, C., Huet, S. & Ellenberg, J. A fractal model for nuclear organization: current evidence and biological implications. *Nucleic Acids Res.* **40**, 8783–8792 (2012).
8. Rice, S. A. Diffusion-limited reactions. *Compr. Chem. Kinet.* **25**, 3–46 (1985).
9. Kapanidis, A. N., Uphoff, S. & Stracy, M. Understanding protein mobility in bacteria by tracking single molecules. *J. Mol. Biol.* **430**, 4443–4455 (2018).
10. Pulkkinen, O. & Metzler, R. Distance matters: the impact of gene proximity in bacterial gene regulation. *Phys. Rev. Lett.* **110**, 198101 (2017).
11. Kolesov, G., Wunderlich, Z., Laikova, O. N., Gelfand, M. S. & Mirny, L. A. How gene order is influenced by the biophysics of transcription regulation. *Proc. Natl Acad. Sci. USA* **104**, 13948–13953 (2007).
12. van den Broek, B. et al. Coiling enhances target localization by proteins. *Proc. Natl Acad. Sci. USA* **105**, 15738–15742 (2008).
13. Di Stefano, M., Rosa, A., Belcastro, V., di Bernardo, D. & Micheletti, C. Colocalization of coregulated genes: a steered molecular dynamics study of human chromosome 19. *PLoS Comput. Biol.* https://doi.org/10.1371/journal.pcbi.1003019 (2013).
14. Bauer, M. & Metzler, R. Generalized facilitated diffusion model for DNA-binding proteins with search and recognition states. *Biophys. J.* **102**, 2321–2330 (2012).
15. Slutsky, M. & Mirny, L. A. Kinetics of protein-DNA interaction: facilitated target location in sequence-dependent potential. *Biophys. J.* **87**, 4021–4035 (2004).
16. Lomholt, M., Ambjörnsson, T. & Metzler, R. Optimal target search on a fast-folding polymer chain with volume exchange. *Phys. Rev. Lett.* **95**, 260603 (2005).
17. Rada-Iglesias, A., Grosveld, F. G. & Papantonis, A. Forces driving the three-dimensional folding of eukaryotic genomes. *Mol. Syst. Biol.* **14**, e8214 (2018).
18. Hassler, M., Shaltiel, I. A. & Haering, C. H. Towards a unified model of SMC complex function. *Curr. Biol.* **28**, R1266–R1281 (2018).
19. 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).
20. Hansen, A. S. et al. Robust model-based analysis of single-particle tracking experiments with spot-on. *eLife* **7**, e33125 (2018).
21. 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).
22. Di Rienzo, C., Piazza, V., Gratton, E., Beltram, F. & Cardarelli, F. Probing short-range protein Brownian motion in the cytoplasm of living cells. *Nat. Commun.* **5**, 5891 (2014).
23. Manley, S. et al. High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nat. Methods* **5**, 155–157 (2008).
24. GrimmJ. B. et al. Bright photoactivatable fluorophores for single-molecule imaging. *Nat. Methods* **13**, 985–988 (2016).
25. 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).
26. Izeddin, I. et al. Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife* **2014**, 1–27 (2014).
27. Liao, Y., Yang, S. K., Koh, K., Matzger, A. J. & Biteen, J. S. Heterogeneous single-molecule diffusion in one-, two-, and three-dimensional microporous coordination polymers: directional, trapped, and immobile guests. *Nano Lett.* **12**, 3080–3085 (2012).
28. Burov, S. et al. Distribution of directional change as a signature of complex dynamics. *Proc. Natl Acad. Sci. USA* **110**, 19689–19694 (2013).
29. Teves, S. S. et al. A dynamic mode of mitotic bookmarking by transcription factors.*eLife* **6**, 22280 (2016).
30. Weber, S. C., Spakowitz, A. J. & Theriot, J. A. Bacterial chromosomal loci move subdiffusively through a viscoelastic cytoplasm.*Phys. Rev. Lett.* **104**, 238102 (2010).
31. Weber, S. C., Thompson M. A., Moerner, W. E., Spakowitz, A. J. & Theriot, J. A. Analytical tools to distinguish the effects of localization error, confinement, and medium elasticity on the velocity autocorrelation function. *Biophys. J.* **102**, 2443–2450 (2012).
32. Weber, S. C., Theriot, J. A. & Spakowitz, A. J. Subdiffusive motion of a polymer composed of subdiffusive monomers. *Phys. Rev. E* **82**, 11913 (2010).
33. 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).
34. Amitai, A. Chromatin configuration affects the dynamics and distribution of a transiently interacting protein. *Biophys. J.* **114**, 766–771 (2018).
35. Saxton, M. J. A biological interpretation of transient anomalous subdiffusion. I. Qualitative model. *Biophysj* **92**, 1178–1191 (2007).
36. Metzler, R. & Klafter, J. The random walk's guide to anomalous diffusion: a fractional dynamics approach. *Phys. Rep.* **339**, 1–77 (2000).
37. Hansen, A. S. et al. Distinct classes of chromatin loops revealed by deletion of an RNA-binding region in CTCF. *Mol. Cell* **76**, 396–411 (2019).
38. Saldaña-Meyer, R. et al. CTCF regulates the human p53 gene through direct interaction with its natural antisense transcript, Wrap53. *Genes Dev.* **28**, 723–734 (2014).
39. Rasko, J. E. J. et al. Cell growth inhibition by the multifunctional multivalent zinc-finger factor CTCF. *Cancer Res.* **61**, 6002–6007 (2001).
40. Lampo, T. J., Stylianidou, S., Backlund, M. P., Wiggins, P. A. & Spakowitz, A. J. Cytoplasmic RNA-protein particles exhibit non-Gaussian subdiffusive behavior. *Biophys. J.* **112**, 532–542 (2017).
41. Elmokadem, A. & Yu, J. Optimal drift correction for superresolution localization microscopy with Bayesian inference. *Biophys. J.* **109**, 1772–1780 (2015).
42. Ester, M., Kriegel, H.-P., Sander, J. & Xu, X. In *Proc. 2nd International Conference on Knowledge Discovery and Data Mining* (ed Fayyad, Usama M.) 226–231 (1996).
43. Stone, M. B. & Veatch, S. L. Steady-state cross-correlations for live two-colour super-resolution localization data sets. *Nat. Commun.* **6**, 7347 (2015).
44. Bronstein, I. et al. Transient anomalous diffusion of telomeres in the nucleus of mammalian cells. *Phys. Rev. Lett.* **103**, 18102 (2009).
45. Golding, I. & Cox, E. C. Physical nature of bacterial cytoplasm. *Phys. Rev. Lett.* **96**, 98102 (2006).
46. Saldaña-Meyer, R. et al. RNA interactions are essential for CTCF-mediated genome organization. *Mol. Cell* **76**, 412–422 (2019).
47. Havlin, S. & Ben-Avraham, D. Diffusion in disordered media. *Adv. Phys.* **51**, 187–292 (2002).
48. Cencini, M. & Pigolotti, S. Energetic funnel facilitates facilitated diffusion. *Nucleic Acids Res.* **46**, 558–567 (2017).
49. Mir, M. et al. Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos. *eLife* **7**, e40497 (2018).
50. Tsai, A. et al. Nuclear microenvironments modulate transcription from low-affinity enhancers. *elife* **6**, e28975 (2017).

Methods

Cell culture. JM8.N4 mESCs[51] (male, research resource identifier (RRID) CVCL_J962; from UC Davis Knockout Mouse Project) and human U2OS osteosarcoma cells (RRID CVCL_0042) were cultured as described previously[19]. Briefly, mESCs were grown on plates precoated with a 0.1% autoclaved gelatin solution (Sigma-Aldrich, G9391) under feeder free conditions in knockout DMEM with 15% FBS and leukemia inhibitory factor (LIF) (full recipe: 500 ml knockout DMEM (ThermoFisher, no. 10829018), 90 ml fetal bovine serum (HyClone, FBS SH30910.03 lot no. AXJ47554)), 6 ml MEM NEAA (ThermoFisher no. 11140050), 6 ml GlutaMax (ThermoFisher no. 35050061), 5 ml penicillin-streptomycin (ThermoFisher no. 15140122), 4.6 µl 2-mercapoethanol (Sigma-Aldrich M3148) and LIF) and half the medium removed and replenished daily. mESCs were passaged every 2 d by trypsinization. The homozygous endogenously Halo-tagged cell lines (for example, C45 mRad21-Halo) and the cell lines stably over-expressing a transgene (for example, H2B-Halo or Halo-3xNLS) have been described previously as well[19,37]. Likewise, female U2OS cells (RRID CVCL_0042) were grown in low glucose DMEM with 10% FBS (full recipe: 500 ml DMEM (ThermoFisher no. 10567014), 50 ml fetal bovine serum (HyClone FBS SH30910.03 lot no. AXJ47554) and 5 ml penicillin-streptomycin (ThermoFisher no. 15140122)) and were passaged every 2–4 d before reaching confluency. The homozygous endogenously Halo-tagged cell lines (for example, C32 Halo-hCTCF) and the cell lines stably over-expressing a transgene (for example, H2B-Halo or Halo-3xNLS) have been described previously as well[19]. Cells were grown in a Sanyo copper alloy IncuSafe humidified incubator (MCO-18AIC(UV)) at 37 °C/5.5% CO_2. An otherwise identical medium, but without phenol red (ThermoFisher no. 31053028), was used for live-cell imaging.

Transient transfections of cells plated on plasma-cleaned 25 mm circular no. 1.5H cover glasses (High-Precision 0117650) either directly (U2OS cells) or MatriGel coated (mESCs, Fisher Scientific, no. 08–774–552) were also performed as previously described[19]. After overnight growth and ~14–32 h before the imaging experiment, cells were then transfected with a plasmid encoding the protein of interest using 1,000 ng of plasmid per well in a six-well plate using Lipofectamine 3000 (ThermoFisher no. L3000008). Given the known issues with over-expression of CTCF[39], which are known to alter the dynamics of CTCF[19], we developed a low-over-expression system for transient transfections, which is outlined in Supplementary Fig. 10. All plasmids used were identical except for the protein of interest and are as follows: the protein of interest (for example, wt-Halo-mCTCF or wt-mRad21-Halo) was expressed from a L30 promoter (we previously found L30 to only slightly change the behavior compared to endogenously tagged proteins, unlike CMV or EF1a promoters that strongly changed the behavior) and followed by an SV40 poly(A) signal. Downstream of the protein of interest, EGFP-3xNLS was expressed from a PGK promoter and followed by a bGH poly(A) signal. GFP-NLS facilitated finding transfected cells and outlining the nucleus. Crucially, the green fluorescent protein (GFP) signal was generally proportional to the expression level of the protein of interest and therefore made it possible to identify cells with a robust, but not too high, expression level. Since we previously showed that strong over-expression of CTCF alters its dynamics and greatly decreases its chromatin-bound fraction[19], we had performed a small screen for promoters that had the smallest effect on the CTCF bound fraction when over-expressed and identified the L30 promoter as the optimal one in this screen. In summary, for transient transfection experiments, cells were plated 2 d before the imaging experiment and transfected 1 d before the imaging experiment.

As described previously[19], key cell lines were pathogen tested (IMPACT II test for mESC C59, PCR-based mycoplasma assay for U2OS C32) and authenticated through short tandem repeat profiling (U2OS).

spaSPT. For spaSPT experiments, cells were grown overnight on plasma-cleaned 25 mm circular no. 1.5H cover glasses (High-Precision 0117650) either directly (U2OS) or MatriGel coated (mESCs, Fisher Scientific, no. 08–774–552). The next day, cells were labeled with 5–50 nM PA-JF_{549} or PA-JF_{646} (ref. [24]) for ~15–30 min, washed twice (one wash: remove medium; PBS wash; add fresh medium), and the the medium was changed to phenol red-free medium at the end of the final wash. Cells were then loaded on the microscope using an incubation chamber that maintains a humidified 37 °C atmosphere with 5% CO_2 and the objective was also heated to 37 °C (Okolabs stage top chamber). Single-molecule imaging was performed on a custom-built Nikon TI microscope controlled through NIS-Elements software (Nikon) and equipped with two EM-CCD cameras (Andor, iXon Ultra 897), a ×100/numerical aperture 1.49 oil-immersion total internal reflection fluorescence (TIRF) objective (Nikon apochromat CFI Apo TIRF ×100 oil), a perfect focusing system to correct for axial drift and motorized laser illumination (Ti-TIRF, Nikon), to achieve HILO illumination[52]. Excitation lasers were: 561 nm (1 W, Genesis Coherent) for PA-JF_{549} with emission filter, Semrock 593/40 nm; 633 nm (1 W, Genesis Coherent) for PA-JF_{646} with emission filter, Semrock 676/37 nm; 405 nm (140 mW, OBIS, Coherent) for all photo-activation experiments. Lasers were AOTF-controlled (AA Opto-Electronic, AOTFnC-VIS-TN) and the laser light was optical fiber coupled, reflected using a multi-band dichroic (405/488/561/633 nm quad-band, Semrock) and focused in the back focal plane of the objective.

For each cell line or condition, we recorded data for around 20–40 cells over at least four replicates performed on different days. The spaSPT experimental settings were as follows: 1 ms 561 nm or 633 nm excitation (100% AOTF) of PA-JF_{549} or PA-JF_{646} was delivered at the beginning of the frame; 405 nm photo-activation pulses were delivered during the camera integration time (~447 µs) to minimize background and their intensity optimized to achieve a mean density of ≤1 molecule per frame per nucleus. 20,000 frames (4 ms) or 30,000 frames (7 or 13 ms) were recorded per cell per experiment. The camera exposure times were: 4 ms (~223 Hz), 7 ms (~134 Hz) or 13 ms (~74 Hz).

spaSPT images were converted into trajectories using a custom-written MATLAB implementation of the MTT-algorithm (ref. [53], code: https://gitlab.com/tjian-darzacq-lab/SPT_LocAndTrack). Settings were: localization error, $10^{-6.25}$; deflation loops, 0; Blinking (frames), 1; maximum competitors, 3; maximum D ($\mu m^2 s^{-1}$), 20.

Two-color simultaneous live-cell spaSPT and PALM and analysis. The protocol and microscope were essentially identical to our one-color spaSPT experiments, except with the following modifications. After overnight growth, cells were labeled simultaneously with 50 nM PA-JF_{549} and 500 nM PA-JF_{646} (ref. [24]) for ~30 min. The two cameras (both Andor iXon Ultra 897 EM-CCD), synchronized using a DAQ board (NI-DAQ PCI-6723), were aligned using 100 nm fluorescent beads (TetraSpeck, ThermoFisher Scientific, T7279) to a less than 60 nm offset. Subsequently, the remaining offset was measured using beads and corrected for after single-molecule localization microscopy. A single-edge dichroic beamsplitter (Di02-R635–25x36, Semrock) was used to separate two emission ranges of wavelengths and a distinct emission filter placed in front of the two cameras (Semrock FF01–676/37–25 and FF01–593/40–25, respectively). Then, 60,000 frames were recorded with 7 ms exposure times (~134 Hz including ~447 µs camera integration time). Photo-activation of both PA-JF_{549} and PA-JF_{646} was achieved using a 405 nm laser (OBIS) and reached near completion after 60,000 frames (very few molecules remained photo-activatable). For spaSPT, we use 1 ms excitation pulses of 561 nm as above. For PALM, we used continuous 633 nm illumination.

Localization and tracking were performed using the MTT-algorithm (https://gitlab.com/tjian-darzacq-lab/SPT_LocAndTrack) and the following settings: localization error, $10^{-6.25}$; deflation loops, 0; Blinking (frames), 1; maximum competitors, 3; maximum D ($\mu m^2 s^{-1}$), 20 (JF_{549} channel) or 1 (JF_{646} channel). The trajectories were then analyzed as outlined in Fig. 4b. Localizations from both colors were used for BaSDI-mediated[41] drift correction (we modified the core BaSDI code to iteratively ensure convergence). spaSPT data was then analyzed as above (https://gitlab.com/anders.sejr.hansen/anisotropy): trajectories were HHM-classified, bound segments removed, anisotropic trajectory segments identified where both displacements were at least 125 nm and the three localizations making up the angle saved. PALM data was analyzed as previously described (https://gitlab.com/anders.sejr.hansen/palm_pipeline/): after drift correction, we merged localizations appearing in consecutive frames using nearest-neighbor tracking as well as blinking molecules (maximum linking distance, 75 nm; maximum blink, two frames; localization uncertainty, ~23 nm) into a single localization (taking the mean x, y coordinates), segmented the nuclei using polygon-segmentation, called clusters using DBSCAN[42] ($\varepsilon = 100$ nm, minimum number of points = 40 (re-scaled according to total number of localizations)) and kept the localizations making up the clusters. Finally, we calculated the pair cross-correlation function[43] between the localizations making up the anisotropic trajectory segments (spaSPT) and the localizations making up the clusters (PALM).

FRAP. FRAP was performed exactly as previously described[19] for C59 mESCs (Halo-CTCF, Rad21-$SNAP_f$) and C59D2 mESCs (ΔRBR$_i$-Halo-CTCF, Rad21-$SNAP_f$) on an inverted Zeiss LSM 710 AxioObserver confocal microscope (330 frames, 2 s between frames, 100 nm pixels, circular bleach spot with a 10 pixel radius). We performed three biological replicates recording a total of 18 cells for C59 Halo-CTCF and 18 cells for C59D2 ΔRBR$_i$-Halo-CTCF and analyzed the data as previously described[19]. As demonstrated previously[19,54], Halo-CTCF and ΔRBR$_i$-Halo-CTCF fall in the 'reaction dominant' regime, where the recovery depends only on the k_{OFF} and we therefore fit the FRAP recoveries to the reaction dominant model below:

$$\mathrm{FRAP}(t) = 1 - Ae^{-k_a t} - Be^{-k_b t}$$

where FRAP is the fluorescence recovery, A and B are the fast and slow subpopulation fractions with their associated first order rate constants, k_a and k_b. We interpret the slower rate as specific binding to cognate sites. The fits are shown in Supplementary Fig. 15.

Fluorescence-based DNA-binding assays. The DNA-binding affinities of wt-CTCF and ΔRBRi-CTCF were compared in vitro using fluorescence polarization assays with fluorescein (FAM)-labeled double-stranded DNA oligo probes and recombinant proteins purified from Sf9 insect cells (Supplementary Fig. 15d).

We used pFastBAC (Invitrogen) to generate recombinant Bacmid DNAs for the fusion mouse proteins 3xFLAG-Halo-wt-CTCF-6xHis[55] (1,086 amino acids, 123.5 kDa) and 3xFLAG-Halo-ΔRBR$_i$-CTCF-6xHis (1,086 amino acids, 123.7 kDa). We used the Bac-to-Bac Baculovirus Expression System (Invitrogen)

to generate recombinant baculovirus and infected Sf9 cells (~2×10^6 per ml) with amplified baculoviruses expressing recombinant wt- or ΔRBR_i-CTCF. Then, 48 h after infection, Sf9 suspension cultures were collected, washed extensively with cold PBS, lyzed in five packed cell volumes of high salt lysis buffer (HSLB; 1.0 M NaCl, 50 mM HEPES pH 7.9, 0.05% NP-40, 10% glycerol, 10 mM 2-mercaptoethanol and protease inhibitors) and sonicated. Lysates were cleared by ultracentrifugation, supplemented with 10 mM imidazole and incubated at 4 °C with Ni-NTA resin (Qiagen) for 90 min. Bound proteins were washed extensively with HSLB with 20 mM imidazole, equilibrated with 0.5 M NaCl HGN buffer (50 mM HEPES pH 7.9, 10% glycerol, 0.01% NP-40) with 20 mM imidazole and eluted with 0.5 M NaCl HGN supplemented with 0.25 M imidazole. We analyzed eluted fractions by SDS–PAGE followed by staining with PageBlue Protein staining solution. Peak fractions were pooled and incubated with anti-FLAG M2 Affinity Gel (Sigma) for 3 h at 4 °C. Finally, bound proteins were washed extensively with HSLB, equilibrated to 0.2 M NaCl HGN, and eluted with 3xFLAG peptide (Sigma) at 0.4 mg ml^{-1}, and protein concentrations determined by PageBlue staining compared to a Bovine Serum Albumin (BSA) standard.

Fluorescence polarization-based equilibrium saturation DNA-binding reactions were as described[56], but with the following modifications. Binding reactions were assembled on ice in triplicates in 384-well black flat bottom microplates (Corning 3820) in 50 mM Tris-HCl (pH 7.5), 200 mM NaCl, 10% (v/v) glycerol, 0.5 mM TCEP and 5 µM zinc acetate. Reactions contained 5 nM (FAM)-labeled oligo probe (wild type or mutant core CTCF binding site) and increasing amounts of wt-CTCF or ΔRBRi-CTCF recombinant protein (0.5–100 nM) in a total volume of 20 µl per well. Buffer only and oligo-only controls were included and used to correct polarization measurements as detailed below. Plates were spun down briefly and incubated at 25 °C for 10–15 min to allow reactions to reach equilibrium before measuring fluorescence polarization on a TECAN Spark microplate reader. Fluorescence intensities were not altered by the addition of protein (Supplementary Fig. 15e). Protein amounts were checked by Page Blue staining to confirm equal concentration of wt-CTCF and ΔRBRi-CTCF in the binding assays (Supplementary Fig. 15g).

Probes with a 5′ 6-FAM (NHS Ester) modification (56-FAMN from IDT) and were as follows. Core CTCF binding site duplex: 5′-AGGACCAGCAGGGGGCGCA-3′ (forward) and 5′-/56-FAMN/TGCGCCCCCTGCTGGTCCT-3′ (reverse). Mutated core CTCF binding site duplex: 5′-AGGATTCTAATTTCGATCA-3′ (forward) and 5′-/56-FAMN/TGATCGAAATTAGAATCCT-3′ (reverse).

The polarization (P) values shown in Supplementary Fig. 15f are averages of $n=3$ independent experiments (using the same oligos and purified protein batches), each containing three technical replicates, calculated as:

$$P = \frac{\left(I_{\mathrm{par}}(\mathrm{spl}) - I_{\mathrm{par}}(\mathrm{blk})\right) - G\left(I_{\mathrm{per}}(\mathrm{spl}) - I_{\mathrm{per}}(\mathrm{blk})\right)}{\left(I_{\mathrm{par}}(\mathrm{spl}) - I_{\mathrm{par}}(\mathrm{blk})\right) + G\left(I_{\mathrm{per}}(\mathrm{spl}) - I_{\mathrm{per}}(\mathrm{blk})\right)}$$

where I_{par}(spl) I_{per}(spl), I_{par}(blk) and I_{per}(blk) are the fluorescence intensities measured in the parallel plane, perpendicular plane, parallel plane for buffer only and perpendicular plane for buffer only, respectively. G is the G factor computed by the TECAN Spark instrument using buffer only and oligo-only control wells

Fluorescence polarization anisotropy (A) was then calculated from polarization (P) as:

$$A = \frac{2P}{3 - P}$$

We then obtained the fraction of CTCF bound to the core CTCF motif at any given protein concentration as:

$$F_{\mathrm{bound}} = \frac{A - A_{\mathrm{free\ oligo}}}{A_{\mathrm{max}} - A_{\mathrm{free\ oligo}}}$$

where $A_{free\ oligo}$ is anisotropy of oligo-only control and A_{max} is anisotropy of the 100 nM protein sample.

The fraction-bound values obtained in $n=3$ independent experiments were then fitted to a Hill equation (Fig. 5e):

$$y = \frac{y_{\mathrm{max}} x^h}{K_{\mathrm{d}}^h + x^h}$$

wt-CTCF and ΔRBR_i-CTCF show comparable affinity for DNA (P value ~0.4, Prism 5 extra sum-of-squares F test).

Brownian motion simulations with simSPT. Even particles obeying ideal Brownian motion can appear anisotropic due to localization uncertainty. For example, a chromatin-bound protein subject to 35 nm localization error (defined as Gaussian standard deviation, roughly our experimental localization error) will appear to move around the true position due to the localization uncertainty and this movement will appear highly anisotropic. To test how well our anisotropy pipeline filters out spurious apparent anisotropy stemming from localization error, we performed 'matched Brownian simulations'. Briefly, for each protein and condition, we used Spot-On[20] to estimate the bound fraction, the free diffusion coefficient and the localization error/uncertainty (~35 nm). We then simulated 500,000 trajectories under realistic HiLo experimental conditions (Supplementary Figs. 2 and3) using simSPT (ref. [20], available on GitLab: https://gitlab.com/tjian-darzacq-lab/simSPT) matching the bound and free diffusion coefficients, the bound fraction and the localization error (20–75 nm) and simulated data at 223, 134 and 74 Hz to match the experiments. We then processed the simulated SPT data identically to how we processed the experimental spaSPT data using the anisotropy pipeline (https://gitlab.com/anders.sejr.hansen/anisotropy).

Calculation of relative ON-rates for wt-CTCF and ΔRBR_i-CTCF. Here we calculate the effect of the RBR_i on CTCF's search time within a simplified two-state model framework, within which the specifically bound fraction for a protein is given by:

$$F_{\mathrm{BOUND}} = \frac{k_{\mathrm{ON}}^*}{k_{\mathrm{ON}}^* + k_{\mathrm{OFF}}}$$

Here, k_{ON}^* corresponds to the pseudo-first-order rate constant and is related to the concentration of available specific binding sites $k_{ON}^* = k_{ON}[\mathrm{DNA}]$. The total bound fraction is the sum of nonspecific (ns) and specifically (s) bound protein: $F_{BOUND,tot} = F_{BOUND,ns} + F_{BOUND,s}$. Using the ~134 Hz data, we estimate (Fig. 5a–c) that $F_{BOUND,tot,wt-CTCF} = 62.4\%$ and $F_{BOUND,tot,\Delta RBR-CTCF} = 42.0\%$. Estimating the nonspecifically bound fraction is challenging since it is both difficult to define and measure. However, we have previously[19] used the bound fraction that remains after mutating the ZF domain of CTCF required for specific binding, as an approximate measure of the nonspecifically bound fraction, $F_{BOUND,ns}$. We used a CTCF mutant (Halo-ZF^{mut}-CTCF) with 11 His-to-Arg point mutations[19]—one mutation in each of the 11 ZFs—to estimate this nonspecifically bound fraction for wt-CTCF. To estimate this for ΔRBR_i-CTCF, we made the ΔRBR_i-deletion in this plasmid to generate Halo-ZF^{mut}-ΔRBR_i-CTCF. Using again the 134 Hz spaSPT data, we thus obtain $F_{BOUND,ns,wt-CTCF} = 25.3 \pm 6.0\%$ and $F_{BOUND,ns,\Delta RBR-CTCF} = 25.1 \pm 7.1\%$ (mean ± standard deviation across $n=3$ biological replicates, Supplementary Fig. 15c). Subtracting, we arrive at $F_{BOUND,s,wt-CTCF} = 37.1\%$ (this number differs by some percentage points from ref. [19], which is likely due to using a different frame rate and due to experimental variability) and $F_{BOUND,s,\Delta RBR-CTCF} = 16.9\%$. Noting that k_{OFF} does not appear to differ significantly between wt-CTCF and ΔRBR_i-CTCF (Fig. 5d and Supplementary Fig. 15a), we can thus calculate the ratio of $k_{ON,wt-CTCF}^*$ and $k_{ON,\Delta RBR-CTCF}^*$:

$$\frac{k_{\mathrm{ON,wt-CTCF}}^*}{k_{\mathrm{ON,}\Delta\mathrm{RBR}i-\mathrm{CTCF}}^*} = \frac{\frac{F_{\mathrm{BOUND,s,wt-CTCF}} k_{\mathrm{OFF}}}{1-F_{\mathrm{BOUND,s,wt-CTCF}}}}{\frac{F_{\mathrm{BOUND,s,}\Delta\mathrm{RBR}i-\mathrm{CTCF}} k_{\mathrm{OFF}}}{1-F_{\mathrm{BOUND,s,}\Delta\mathrm{RBR}i-\mathrm{CTCF}}}} = \frac{\frac{F_{\mathrm{BOUND,s,wt-CTCF}}}{1-F_{\mathrm{BOUND,s,wt-CTCF}}}}{\frac{F_{\mathrm{BOUND,s,}\Delta\mathrm{RBR}i-\mathrm{CTCF}}}{1-F_{\mathrm{BOUND,s,}\Delta\mathrm{RBR}i-\mathrm{CTCF}}}} = \frac{\frac{0.371}{1-0.371}}{\frac{0.169}{1-0.169}} = 2.90$$

We stress here that this calculation is associated with some uncertainty, since it is difficult both to define and to measure the nonspecifically bound fraction. Moreover, the two-state framework is simplified.

As a second and orthogonal estimate of the effect of the RBR_i on the k_{ON}^*, we estimated the specifically bound fraction from the FRAP data. While a full and quantitative description of the reaction-diffusion system[57] is beyond on the scope of this work, we here as an ad hoc and operational metric define the specifically bound fraction as the fraction that remains in the FRAP curve after 4 s of recovery. The initial FRAP recovery is dominated by recovery due to diffusion (free CTCF moving out of the bleached spotted and being replaced by unbleached CTCF) and nonspecifically bound CTCF (nonspecifically bound CTCF, presumably with dwell times in the range of tens to hundreds of milliseconds, being replaced by unbleached CTCF). An ad hoc time of approximately 4 s is chosen, after which most of the nonspecifically bound and free CTCF should have been replaced and what remains is then assumed to be an estimate of specifically bound CTCF as shown in Supplementary Fig. 15b. From this estimate we obtain: $F_{BOUND,s,wt-CTCF,FRAP} = 49.7\%$ and $F_{BOUND,s,\Delta RBR-CTCF,FRAP} = 29.0\%$. If we plug in these numbers we obtain:

$$\frac{k_{\mathrm{ON,wt-CTCF}}^*}{k_{\mathrm{ON,}\Delta\mathrm{RBR}i-\mathrm{CTCF}}^*} = \frac{\frac{F_{\mathrm{BOUND,s,wt-CTCF}}}{1-F_{\mathrm{BOUND,s,wt-CTCF}}}}{\frac{F_{\mathrm{BOUND,s,}\Delta\mathrm{RBR}i-\mathrm{CTCF}}}{1-F_{\mathrm{BOUND,s,}\Delta\mathrm{RBR}i-\mathrm{CTCF}}}} = \frac{\frac{0.497}{1-0.497}}{\frac{0.290}{1-0.290}} = 2.42$$

As can been seen, the FRAP-based estimate differs from the spaSPT-based estimate and we emphasize and stress that these numbers and calculations are associated with uncertainty. Nevertheless, given than the specifically bound fraction is substantially higher for wt-CTCF than for ΔRBR_i-CTCF and that the nonspecifically bound fractions are similar, the k_{ON}^* also has to be substantially higher. And this is the case regardless of whether we use spaSPT or FRAP to estimate these. In other words, the point is not whether the k_{ON}^* is precisely two-, 2.5- or threefold higher for wt-CTCF than for ΔRBR_i-CTCF, but simply that it must be significantly higher.

Brownian simulation of chromatin and the protein. To model the dynamics of a protein transiently interacting with chromatin we used a previously developed chromatin model[33]. We do not include in our model nucleosomes and crowding by bound proteins. We concentrate only on simulating the interaction between the diffusing protein (CTCF) with spherical elements (zones/monomers).

The chromatin molecule is represented as a long polymer with N monomer, whose positions are $(\mathbf{R}_1, ..., \mathbf{R}_N)$, diffusing in a large domain of radius A representing the nucleus. The protein is represented by a diffusing point particle. The chromatin chain is a flexible polymer with spring potential, and Lennard-Jones forces, describing self-avoidance of each monomer pairs. The polymer has potential energy of the form

$$U(\mathbf{R}_1, ..., \mathbf{R}_N) = U_{\text{spring}}(\mathbf{R}_1, ..., \mathbf{R}_N) + U_{\text{LJ}}(\mathbf{R}_1, ..., \mathbf{R}_N)$$

where the spring potential is

$$U_{\text{spring}}(\mathbf{R}_1, ..., \mathbf{R}_N) = \frac{1}{2}k\sum_{i=1}^{N-1}\left(|\mathbf{r}_{i+1,i}| - l_0\right)^2$$

where $\mathbf{r}_{i,j} = \mathbf{R}_i - \mathbf{R}_j$, l_0 is the equilibrium length of a bond, $k = \frac{3}{S_{l_0}^2}$ is the spring coefficient and S_{l_0} is the standard deviation of the bond length. We chose the empirical relation $S_{l_0} = 0.2l_0$. The Lennard-Jones potential is

$$U_{\text{LJ}}^{i,j}(\mathbf{r}_{i,j}) = \begin{cases} 4\left[\left(\frac{\sigma}{\mathbf{r}_{i,j}}\right)^{12} - 2\left(\frac{\sigma}{\mathbf{r}_{i,j}}\right)^{6} + \frac{1}{4}\right] & \text{for } |\mathbf{r}_{i,j}| < 2^{1/6}\sigma \\ 0 & \text{for } |\mathbf{r}_{i,j}| \geq 2^{1/6}\sigma \end{cases}$$

where σ is the size of the monomer. With the choice $l_0 = 2\sigma$, the springs that materialize bonds, cannot cross each other in stochastic simulations using the potential U. We do not account here for bending elasticity. Finally, we used Euler's scheme to generate Brownian simulations. At an impenetrable boundary, each rigid monomer is reflected in the normal direction of the tangent plane.

We recall here the interaction model between the chromatin sites (monomers/zones) and the protein (particle) developed in ref. [34] (see Supplementary Fig. 6). The protein diffuses in the nuclear domain until encountering one of the monomer sites by entering the trapping radius ε. The protein is then absorbed at the monomer (zone) with probability P_{trap} or reflected with probability $1 - P_{\text{trap}}$. While trapped, the particle is free to diffuse inside the zone, which is of radius ε, and is partially reflected from its boundary when it hits it. Every time it hits the boundary it can exit with probability P_{exit}. On release, the particle is placed at a distance a from absorbing monomer position with the same radial direction (see Supplementary Fig. 6c) from which it escaped and starts diffusing again. Since we take $a > \varepsilon$ and $P_{\text{trap}} < 1$, the particle has a finite probability to escape from the absorbing zone rather than rebinding back to it immediately. While inside a zone, the protein can bind at a point inside the zone with Poissonian ON-rate k_{on} and then be released with Poissonian off-rate k_{off} (see Supplementary Table 2).

As we observed in the experiment, in some of the protein's trajectory it is bounded at one point (up to our localization error) for the full length of the trajectory. To account for this behavior, we assume that in this state the protein is confined in a small site of size ε_{CBS}, which we denote the cognate binding site (CBS). Of the total N monomer site, a subset fraction f_{CBS} monomers are chosen randomly as CBSs. We denote other monomer sites as TTZs. Hence, the CBS (respectively, TTZ) has trapping radius ε_{CBS} (respectively, ε_{TTZ}), release radius a_{CBS} (respectively, a_{TTZ}), capture probability $P_{\text{abs,CBS}}$ (respectively $P_{\text{abs,TTZ}}$). The protein has a characteristic binding time τ_{CBS} within the CBSs. The third kind of monomers are the PTZs. Each has a different size, which is drawn out of a power-law distribution $P_{\text{PTZ}}(\varepsilon)$ with a size cutoff at 800 nm. The probability $P_{\text{trap,PTZ}}$ of binding to a PTZ on encountering one could be different than of the TTZs. Inside the TTZs and the PTZs, the protein diffuses and is partially reflected from its boundary when it hits it. The exit probability P_{exit} is the same for TTZs and PTZs. The release radius of each PTZs is different and is larger by 10 nm from its trapping radius ε.

At the equilibration stage of the system, the polymer is placed inside the nuclear domain and equilibrates for a time T longer than its longest relaxation mode. The end monomer remains fixed at the origin. After this initial phase, the polymer configuration is fixed (except for the simulation presented in Supplementary Fig. 2g). The protein is then placed at a random position in the nuclear domain. The protein position evolves in time according to the Langevin equation

$$\mathrm{d}\mathbf{x} = \sqrt{2D_i}\mathrm{d}\boldsymbol{\omega}, \qquad i = F, Z, \text{CBS}$$

where D_Z is the diffusion coefficient of the protein within a zone, D_F is the diffusion coefficient of a free protein (outside the zones), D_{CBS} is the diffusion coefficient within the small CBS and $\mathrm{d}\omega$ is white Gaussian noise. The diffusion coefficient of the protein in bulk is taken to be $D_F = 5\mu m^2/s$, of the order of the experimentally estimated diffusion coefficient taken from experiments[19]. The protein diffuses until encountering a trapping zone or a CBS as detailed above. For each unique polymer configuration, we simulated the protein motion for 10,000 absorption events. In each condition, we randomized many different polymer configurations.

Each simulation produced a long Brownian trajectory of the protein in the nuclear domain. White Gaussian noise with a standard deviation of 30 nm and mean zero is added to each trajectory point. The trajectories from the simulations were aggregated and then split to short trajectories with the same length statistics as that of the experimental trajectories. The trajectories were then subjected to the same classification procedure by an HMM as were the experimental trajectories. We then computed different statistics of these trajectories.

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

Data availability

Raw and processed SPT data is freely available at Zenodo: https://zenodo.org/record/2208323. All cell lines will be provided upon request.

Code availability

Raw code as well as a detailed description of how the data was analyzed is available on GitLab: https://gitlab.com/anders.sejr.hansen/anisotropy. The code for localization and tracking is also available on GitLab: https://gitlab.com/tjian-darzacq-lab/SPT_LocAndTrack. The code for performing Brownian motion simulations (Supplementary Figs. 2a–c and 3) is likewise available on GitLab: https://gitlab.com/tjian-darzacq-lab/simSPT. Finally, the PALM-analysis code is also available on GitLab: https://gitlab.com/anders.sejr.hansen/palm_pipeline/.

References

51. Pettitt, S. J. et al. Agouti C57BL/6N embryonic stem cells for mouse genetic resources. *Nat. Methods* **6**, 493–495 (2009).
52. Tokunaga, M., Imamoto, N. & Sakata-Sogawa, K. Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nat. Methods* **5**, 159–161 (2008).
53. 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).
54. Sprague, B. L., Pego, R. L., Stavreva, D. A. & McNally, J. G. Analysis of binding reactions by fluorescence recovery after photobleaching. *Biophys. J.* **86**, 3473–3495 (2004).
55. Cattoglio, C. et al. Determining cellular CTCF and cohesin abundances to constrain 3D genome models. *eLife* **8**, e40164 (2019).
56. Rossi, A. M. & Taylor, C. W. Analysis of protein-ligand interactions by fluorescence polarization. *Nat. Protoc.* **6**, 365 (2011).
57. Mueller, F., Mazza, D., Stasevich, T. J. & McNally, J. G. FRAP and kinetic modeling in the analysis of nuclear protein dynamics: what do we really know? *Curr. Opin. Cell Biol.* **22**, 403–411 (2010).

Acknowledgements

We thank L. Lavis for generously providing JF dyes, M. Woringer for insightful discussions and help with simSPT, A. Tangara and A. Robles for microscope assembly and maintenance, G. Dailey for help and assistance with cloning, K. Heydari at the Li Ka Shing Facility for flow cytometry assistance and A. deHart, L. Witowsky, A. Manford, L. Dahal and A. Basil Heckert for discussions and help with fluorescence polarization experiments. We thank A. Seeber, K. Dao Duc, D. McSwiggen and other members of the Tjian and Darzacq laboratories for comments on the manuscript. This work was performed in part at the CRL Molecular Imaging Center, supported by the Gordon and Betty Moore Foundation. A.S.H. was a postdoctoral fellow of the Siebel Stem Cell Institute and is supported by a National Institutes of Health (NIH) NIGMS K99 Pathway to Independence Award (no. K99GM130896). This work was supported by NIH grant nos. UO1-EB021236 and U54-DK107980 (X.D.), the California Institute of Regenerative Medicine grant no. LA1–08013 (X.D.) and by the Howard Hughes Medical Institute (003061, R.T.).

Author contributions

A.S.H., A.A., R.T. and X.D. conceived of the project. A.S.H. and A.A. conceived of the ADTZ model. A.S.H. performed the experiments, developed anisotropy analysis pipeline, analyzed the experimental data and performed Brownian motion simulations. A.A. developed the theoretical framework, performed and analyzed model simulations. A.S.H. and C.C. generated the C59D2 ΔRBR$_i$-Halo-CTCF mESC line. C.C. performed in vitro CTCF binding assays. A.S.H. and A.A. drafted the manuscript and all authors edited the manuscript. R.T. and X.D. supervised the project. A.S.H. and A.A. contributed equally to this project.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41589-019-0422-3.

Correspondence and requests for materials should be addressed to R.T. or X.D.

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

nature research

Corresponding author(s): Robert Tjian
Xavier Darzacq

Last updated by author(s): Oct 16, 2019

Reporting Summary

Nature Research 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 Research policies, see Authors & Referees 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	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐	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)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	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	The raw data is available freely on Zenodo: https://zenodo.org/record/2208323
Data analysis	Raw code as well as a detailed description of how the data was analyzed is available on GitLab: https://gitlab.com/anders.sejr.hansen/anisotropy. The code for localization and tracking is also available on GitLab: https://gitlab.com/tjian-darzacq-lab/SPT_LocAndTrack. The code for performing Brownian motion simulations (Supplementary Fig. 1b-d) is likewise available on GitLab: https://gitlab.com/tjian-darzacq-lab/simSPT. Finally, PALM analysis code can be found at: https://gitlab.com/anders.sejr.hansen/palm_pipeline and the microscope was controlled through the Nikon Elements Software.

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/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

Raw and precessed data is freely available at Zenodo under Accession code 2208323: https://zenodo.org/record/2208323

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

Life sciences study design

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

Sample size	Sample sizes was chosen to be large enough to enable robust analysis. In total, we performed spaSPT on almost two thousand single cells and analyzed around 100 million unique displacements. No statistical method was used to determine sample size.
Data exclusions	For accurate SPT analysis, the localization density needs to be low (i.e. we aimed for 0.5-1 in focus molecule per frame, on average). Due to cell-to-cell variation in expression levels and dye labeling and experimental variation, we occasionally, but rarely, observed cells with higher localization densities. These movies were not analyzed in order to minimize tracking errors. The exclusion criteria were pre-determined and applied equally to all samples.
Replication	In general, we performed 3-4 biological replicates for each spaSPT experiment and recorded 6-10 single cells per replicate. The main figures shows the merged data for these 3-4 replicates. We generally compared means and standard errors across the 3-4 biological replicates between conditions when drawing conclusions. No claims in the manuscript failed to reproduce, but we generally did not repeat the analysis for an additional 3-4 replicates to test reproducibility.
Randomization	Randomization and blinding are not applicable here. We did not working with human subjects.
Blinding	Randomization and blinding are not applicable here. We did not working 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	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	JM8.N4 mouse embryonic stem cells were obtained from the KOMP Repository at UC Davis (male mESCs; Research Resource Identifier: RRID:CVCL_J962). Human U2OS osteosarcoma cells (Research Resource Identifier: RRID:CVCL_0042) were a gift from David Spector's lab at CSHL.
Authentication	The U2OS cell lines were authenticated using STR profiling as described in Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., & Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife, 6, e25776.
Mycoplasma contamination	mESC C59 and C87 cell line were pathogen tested using an IMPACT II assay and the U2OS C32 cell line were tested for mycoplasma contamination using a PCR-based mycoplasma assay as described in Hansen, A. S., Pustova, I., Cattoglio, C., Tjian, R., & Darzacq, X. (2017). CTCF and cohesin regulate chromatin loop stability with distinct dynamics. Elife, 6, e25776. The other cell lines were not directly tested, but C59D2 derives from the tested cell line.
Commonly misidentified lines (See ICLAC register)	None of the cell lines used are in the ICLAC register and the U2OS cell line has been authenticated