

In silico nanoscope to study the interplay of genome organization and transcription regulation

Soundhararajan Gopi^{1,2}, Giovanni B. Brandani¹, Cheng Tan³, Jaewoon Jung^{3,4},
 Chenyang Gu¹, Azuki Mizutani¹, Hiroshi Ochiai⁵, Yuji Sugita^{3,4,6}, Shoji Takada^{1,*}

¹Department of Biophysics, Graduate School of Science, Kyoto University, Kyoto 606-8502, Japan

²Present address: Department of Biochemistry, University of Zurich, Zurich 8057, Switzerland

³Computational Biophysics Research Team, RIKEN Center for Computational Science, Kobe 650-0047, Japan

⁴Theoretical Molecular Science Laboratory, RIKEN Cluster for Pioneering Research, Saitama 351-0198, Japan

⁵Division of Gene Expression Dynamics, Medical Institute of Bioregulation, Kyushu University, Fukuoka 812-0054, Japan

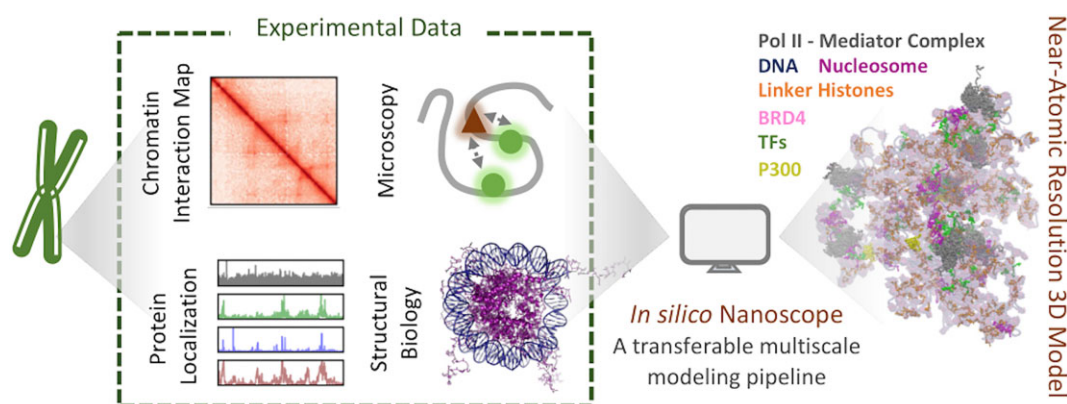
⁶Laboratory for Biomolecular Function Simulation, RIKEN Center for Biosystems Dynamics Research, Kobe 650-0047, Japan

*To whom correspondence should be addressed. Email: takada@biophys.kyoto-u.ac.jp

Abstract

In eukaryotic genomes, regulated access and communication between *cis*-regulatory elements (CREs) are necessary for enhancer-mediated transcription of genes. The molecular framework of the chromatin organization underlying such communication remains poorly understood. To better understand it, we develop a multiscale modeling pipeline to build near-atomistic models of the 200 kb *Nanog* gene locus in mouse embryonic stem cells comprising nucleosomes, transcription factors, co-activators, and RNA polymerase II–mediator complexes. By integrating diverse experimental data, including protein localization, genomic interaction frequencies, cryo-electron microscopy, and single-molecule fluorescence studies, our model offers novel insights into chromatin organization and its role in enhancer–promoter communication. The models equilibrated by high-performance molecular dynamics simulations span a scale of ~350 nm, revealing an experimentally consistent local and global organization of chromatin and transcriptional machinery. Our models elucidate that the sequence-regulated chromatin accessibility facilitates the recruitment of transcription regulatory proteins exclusively at CREs, guided by the contrasting nucleosome organization compared to other regions. By constructing an experimentally consistent near-atomic model of chromatin in the cellular environment, our approach provides a robust framework for future studies on nuclear compartmentalization, chromatin organization, and transcription regulation.

Graphical abstract



Introduction

The genome of eukaryotic organisms is strategically organized and compartmentalized inside the nucleus to provide regulated access to *cis*-regulatory elements (CREs) for controlled transcription necessary for cell survival, replication, differentiation, and maturation [1–4]. Chromatin organization inferred by Hi-C experiments and its variants report a modular genome organization comprised of coarse chromatin compartments (A/B compartments) with distinct levels of tran-

scriptional activity (high/low, respectively) and a finer organization called topologically associated domains (TADs) [5–8]. TADs are insulated from other genomic regions and are characterized by *cis*-interacting DNA segments often demonstrated to regulate the expression of encompassed genes [4, 9]. The CREs are enriched in transcription factor (TF) binding and active epigenetic signatures, suggesting a strong interplay between genome organization and transcription regulation underlying the cell fate decisions [10–12].

Received: November 27, 2024. Revised: February 10, 2025. Editorial Decision: February 22, 2025. Accepted: March 8, 2025

© The Author(s) 2025. Published by Oxford University Press on behalf of Nucleic Acids Research.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Despite extensive studies, the mechanistic details of the interplay between genome organization and transcription regulation remain elusive, with conflicting molecular picture in the literature [9, 10, 12, 13]. These discrepancies have been discussed in the context of a range of transcription regulation models, such as contact [8], network [14], or diffusion [15, 16]-driven transcription regulation differing in the extent of contribution from three-dimensional (3D) genome organization. To address the ambiguity, live-cell time-resolved nanoscopy experiments measure the transcriptional output together with the relative distance of distal genomic locations and molecular factors from the gene locus [10, 13, 17]. Such advanced *in-situ* imaging approaches reveal a hierarchical organization of the molecular factors at the transcription site, offering insights into the mechanistic details of transcription activation [18].

Despite the extent of experimental information available on chromatin organization around transcription start sites (TSS)—protein localization from ChIP-seq and chemical mapping studies [19–21], cryoEM structures of ternary complexes of the transcriptional machinery [22, 23], hierarchical organization of protein factors [18], and Micro-C informed interaction between CREs [8]—a molecular-level understanding of the chromatin organization and the mechanistic basis of such organization is still lacking.

The transcription regulation of pluripotency factors in mouse embryonic stem cells (mESCs) is well characterized in the literature, supplemented with a plethora of chromatin interaction maps and CRE annotations, representing an ideal system to investigate the interplay between 3D genome organization and transcription at the molecular level. In this work, we develop a multiscale modeling pipeline to build a comprehensive model of mESC gene loci at a near-atomic resolution, integrating several experimental chromatin interaction maps and organization features into a consistent model. Such a model acts as an *in silico* nanoscope and is suitable for subsequent molecular dynamics (MD) simulations to study mechanistic details of the chromatin organization and the intertwined transcription regulation [24]. The molecular modeling of gene loci poses two major challenges: (i) localizing protein complexes and factors along a three-dimensionally organized DNA with characteristic local and global structural features and (ii) delineating the information from ensemble-averaged chromatin interaction maps (e.g. Micro-C experiments) to generate biologically relevant molecular models. We overcome these challenges by developing a multiscale modeling pipeline that encompasses the ensemble nature of these experiments.

Here, we describe our data-driven modeling pipeline by building comprehensive models of the 200 kb *Nanog* gene locus from mESCs, which contains three enhancers at various distances from the *Nanog* promoter (−45, −5, and +60 kb), and is considered a model system for the understanding of enhancer–promoter (E-P) interactions and transcription regulation [25]. Our models integrate nucleosome chemical mapping, ChIP-seq data on protein–DNA binding and epigenetic modifications, and Micro-C genome contact frequencies into an experimentally consistent model of the *Nanog* locus at near-atomic resolution, providing a realistic molecular-level picture of chromatin organization consistent with *in vivo* single-gene imaging studies. The mapping of nucleosomes, linker histones, and transcription (co-)factors, together with the spatial chromatin organization observed from the high-

resolution molecular model, reveals the distinct organization principles at CREs. The chromatin is locally expanded at CREs, strategically designed by weak nucleosome positioning signals (NPS) and a longer and asymmetric entry/exit linker DNA that favors local chromatin accessibility for transcription (co)factor binding and recruitment of RNA polymerase II–mediator complex. The contrasting nucleosome positioning features along the genome guide the mutually excluded localization of linker histones and transcription machinery and suggest the DNA-sequence-guided interplay between chromatin organization and transcription regulation. The expanded chromatin segments at CREs, owing to the increased capture radius, facilitate nonlocal internucleosome interactions, forming a road map for communication between CREs. The models are an excellent starting point for MD simulations to study the molecular basis of chromatin organization and the role of protein factors, as well as test E-P communication models. Overall, the models are expected to be a valuable asset in exploring the structure–function relationships of the genome.

Materials and methods

Integrative modeling pipeline

The genome contact frequencies are used to build a mesoscopic ensemble of 3D chromatin conformations at a coarse resolution (1 kb) using Bayesian polymer simulation [26]. Based on chemical mapping and ChIP-seq experimental data, we employ data-driven Monte Carlo simulations to generate an ensemble representation of nucleosome positioning and protein localization consistent with the experiments. The representative protein positional maps and coarse chromatin conformation are combined using a backmapping pipeline to generate the molecular model of the *Nanog* locus at near-atomic resolution. The individual protocols of the pipeline are further explained in the following sections.

Generating a mesoscopic model of the *Nanog* locus

The 3D conformational ensemble of chromatin in the *Nanog* locus was based on mESC Micro-C data [8] at 1 kb resolution, normalized by Juicebox [27]. To generate the conformations, a 128-replica Hi-C metainference MD simulation [26, 28] is performed using a prior 1 kb resolution chromatin model defined by harmonic bond, angle, and Lennard–Jones potentials, with the bead size of 22 nm and other parameters based on higher-resolution chromatin simulations with the 1CPN model [29]. Hi-C metainference performs replica polymer simulations based on the prior model and introduces an additional Bayesian energy score to ensure the agreement between experimental and replica-averaged contact frequencies calculated using a distance-dependent forward model (Supplementary Methods). The *Nanog* locus is then simulated as a polymer made of 200 1 kb beads for 1 million MD steps, and the final half of the replica trajectories are accumulated and further analyzed. The conformational ensemble is clustered based on pairwise distances between the CREs into five conformational clusters with distinct combinations of *cis*-interactions (Supplementary Methods). Representative conformations from each cluster with E-P and enhancer–enhancer (E-E) distances close to the cluster average are selected for generating high-resolution molecular models.

For the analysis, a more detailed time-lagged independent component analysis [30] (tICA) is performed over pairwise genomic contacts at 5 kb resolution using the PyEMMA python package [31]. The relative TF accessible surface area (r. TF-ASA) is calculated using an 11 nm bead and 3 nm probe radius and normalized by the surface area of the beads. The distance autocorrelation function is calculated for bead pairs and fitted to the single-exponential functions to measure their decay rate ($1/\tau_r$) and hence their reconfiguration time τ_r . The protein occupancy/epigenetic markers are mapped to the mesoscopic beads at 1 kb resolution and are considered enriched if the Z-value = $(x - \mu)/\sigma$ is > 1 , where μ and σ are the mean and standard deviation calculated over the 200 kb segment, respectively, and treated as background if $Z < 1$.

Positional maps of protein factors and complexes

The DNA sequence of the *Nanog* gene locus (Chr6: 122 600–122 800 kb) and its annotations are obtained from the reference mouse genome [32] (accession ID: GCA_000001305.2; mm10; release date: 9 January 2012). The occupancy and epigenetic profiles in mESC obtained in various formats (Supplementary Table S1; Supplementary Methods) are converted into bedGraph format using the BEDOPS package [33]. The mouse genome assembly, GRCm39 (mm10; release date: 9 January 2012), is used as the reference, and the genome coordinates from various assemblies are appropriately converted using the LiftOver tool (<http://genome.ucsc.edu>). The occupancy profiles at the region of interest are isolated, and the genomic positions of proteins are subsequently mapped.

The positions of RNA polymerase II–mediator complexes (transcription pre-initiation complex; PIC) along the target locus are determined based on the ChIP-seq occupancy profile, and the orientation of the PIC is determined by the relative signal of the positive and negative strands from the GRO-seq experiment [34]. The nucleosomes are mapped on DNA using simulated annealing Monte Carlo simulations with protein association, dissociation, and translocation moves. The energy function is proportional to the nucleosome occupancy measured by *in vivo* chemical mapping experiments [20], and the energy constant is optimized to achieve saturated nucleosome association. Similarly, linker histones and BRD4 (a transcriptional cofactor) are mapped onto the previously mapped nucleosome positions to reproduce the *in vivo* H1/Nuc ratio and the number of BRD4 molecules observed by microscopy in the proximity of *Nanog* locus [18, 19, 35] (Supplementary Methods).

Twenty copies of each TF (SOX2, OCT4, NANOG, and KLF4) are added, assuming a uniform 1 μ M concentration inside the nucleus [36, 37] at positions ranked based on their ChIP-seq signal and the strength of their cognate DNA sequences (Supplementary Methods). Due to the low copy number of P300 in cells [16], three copies of P300 are added to the model in the proximity of transcription factor clusters at super-enhancer (SE) regions.

The generated positional maps are consistent with the mm10 mouse genome assembly, and the sequences in the 5' to 3' direction, corresponding to the protein localization, are extracted from the corresponding genomic locations for further analysis. The A/T (C/G) probability is the probability of A or T (C or G) at a specific position (−98 to 98 bp) relative to the dyad (set to 0) calculated from the mapped nucleosome positions. Similarly, (A/T)₅ probability is the frequency of finding

a 5-mer composed solely of combinations of A and T (e.g. ATATA, ATTTT, TTAAA, etc.) centered at a given position. A₅/T₅ probability is the chance of finding 5-mer composed of A or T alone (e.g. AAAAA and TTTTT) centered at a given position. The A/T fraction (A/T%) is calculated as the fraction of A and T in any given DNA segment.

Backmapping mesoscopic model to near-atomistic model

Based on the positional maps of nucleosomes and the associated proteins (i.e. nucleosome—DNA wound around core histones, Nuc-H1—nucleosome bound to H1, Nuc-BRD4—H3-acetylated nucleosome bound to BRD4; hereafter called nucleosome modules), the *Nanog* locus is deconstructed into fragments of nucleosome and PIC modules connected by linker DNA of various length (collectively referred to as fiber modules). The fiber modules at atomistic resolution are modeled using MODELLER [38] if the structural template of homologous protein complexes is available—nucleosome: 1KX5, nucleosome bound to H1: 7K5Y, nucleosome bound to BRD4: 2WP1, and PIC: 7ENC and 6W1S (Supplementary Methods). The intrinsically disordered regions (IDRs) in PIC lacking structural templates are selectively modeled for subunits that are known to participate in liquid–liquid phase separation: MED1, MED14-15, Pol II subunits, TFIID3-5, and TFIID11 [39, 40]. The structural models for the IDRs are generated using AlphaFold [41] and are added to the PIC model using MODELLER to account for any residual structures. About 83 out of 203 IDRs in the remaining subunits are trimmed (see Supplementary Methods) to avoid the risk of topological loops and knots formed while modeling IDRs at the interface of protein subunits. The all-atom structures of the other modules are modeled using MODELLER—SOX2: 1GT0, OCT4: 3L1P, NANOG: 2VI6, KLF4: 4M9E or obtained from the AlphaFold protein structure database [42]—BRD4: Q9ESU6, and P300: B2RWS6. The molecular models are coarse-grained to residue level, and the disordered tails are artificially compacted using GENESIS to reduce the chances of clashes during backmapping (Supplementary Methods).

Using representative structures from the 1 kb metainference mesoscopic model as a reference, we grow the chromatin fiber of the *Nanog* locus from one end to the other by adding one nucleosome module (+PIC module, if PIC is at the consecutive position) at a time by generating structural ensembles on the fly (Supplementary Methods; Supplementary Videos S1–S5). The chromatin fiber grows by adding the nucleosome modules that vary in nucleosomal DNA unwrapping, orientation, and linker DNA bending to the current fiber (Supplementary Methods). The fiber modules are joined by aligning the phosphates of the three terminal bp of DNA. At each step, the structure of the growing chromatin fiber is sampled using a Monte Carlo-guided procedure attempting to avoid the clashes between the newly added nucleosome module and the structure generated until the previous step, and at the same time, minimize the distance from the reference 1 kb mesoscopic bead and the distance between $i-1$ and $i+1$ nucleosomes to promote a compact local nucleosome organization and decrease the chances of topological knots. The protocol also ensures that the growing fiber is knot-free by rejecting knotted structures on the fly at each iteration [43, 44]. The longer linker DNA is compacted by introducing decoy nucleosomes to avoid topological loops during the modeling, and

the core histones are removed later to leave behind bare linker DNA. After chromatin fiber backmapping is complete for the reference mesoscopic conformations, the DNA component of the generated *Nanog* locus molecular models is coarse-grained to 25 bp resolution and analyzed for topological knots using KymoKnot [43], and the backmapping procedure is repeated if knots are identified.

The performance of the backmapping protocol is evaluated by generating 25 nucleosome chromatin fibers with varying nucleosome repeat lengths (NRL = 167–207) and reference polymer structures with five beads. The distance from the reference bead and the distance between i and $i+2$ nucleosomes are calculated as the distance between the center of mass (COM) of the nucleosome. The sedimentation coefficient of the nucleosome fibers is calculated as before [45]:

$$S_{20,w} = S_0 \times \left[1 + \frac{2R}{N} \left(\sum_i \sum_j 1/r_{ij} \right) \right] \quad (1)$$

here, $S_0 = 11.1 S$ is the sedimentation coefficient of a mononucleosome, $R = 54.5 \text{ \AA}$ is the spherical radius of a mononucleosome, N is the total number of nucleosomes, and r_{ij} is the distance between the COM of nucleosomes i and j .

Finally, the compacted TFs and P300 models are added in the proximity of mapped genomic locations by iteratively sampling the translational and rotational moves to decrease steric clashes while retaining the binding orientation and proximity to the mapped genomic location (Supplementary Methods).

The molecular models of the *Nanog* locus (~2.5 million CG particles) are energy-minimized and subsequently simulated at 150 mM ionic strength and 300 K using GENESIS CGDYN [46, 47] on the Fugaku supercomputer. The simulations employed the AICG2+ model [48] for proteins, with a statistical flexible and solvation potential for disordered residues [49, 50], and the 3SPN2.C model [51] for DNA with a -1.0 charge on phosphate groups. The interactions between the DNA and proteins are modeled as before [52, 53]. Additional details on the protein sequences, reference structures used for globular domains, description of disordered segments, and CG model parameters for MD simulation are available in the Supplementary Methods. The simulations are conducted for 100 ns (10^7 MD steps with 10 fs integration timesteps) to relax the artificially compacted fiber modules used in the backmapping procedure and to demonstrate the stability of the generated molecular model for MD simulations.

Results

CRE interactions at *Nanog* gene locus

Our approach integrates three kinds of experimental data to build an experimentally consistent molecular model of the *Nanog* gene locus: Micro-C contact frequency maps, chemical mapping/ChIP-seq data on protein localization, and *in situ* microscopy on the local concentration of proteins and spatial organization (Fig. 1A). The *Nanog* gene locus is particularly interesting as the 200 kb segment (Chr6: 122 600–122 800 kb) comprises four genes (*Gdf3*, *Dppa3*, *Nanog*, and *Slc2a3*), a partial gene at the 5' end (*Apobec1*), and three SE elements, -45SE , -5SE , and $+60\text{SE}$, located at -45 , -5 , and 60 kb relative to the *Nanog* promoter (Fig. 1B), respectively [4, 32]. The interaction of the gene promoters with the three SEs

is evident from the virtual-4C interaction maps [8] from the viewpoints of promoters (Fig. 1C; note the spike in contact frequency of promoters with other CREs). The contribution of SEs to the transcriptional output of *Nanog* and *Dppa3* [4, 11, 25] and the hierarchical organization of TFs at the *Nanog* gene [18] are well characterized in mESC.

3D modeling of the *Nanog* gene Locus

Most chromosome conformation capture methods characterize the 3D genome organization in terms of ensemble-averaged pairwise interaction frequency maps and mask the conformational heterogeneity of chromatin [54]. The replica-based Bayesian approach Hi-C metainference [26, 28] addresses this by reconstructing chromatin conformational ensembles from experimental contact frequencies and prior models and has been validated against synthetic and *in vivo* data [26]. Using this protocol (Fig. 2A; see “Materials and methods” section), we modeled the 200 kb *Nanog* locus with 128 replicas based on mESC Micro-C data and a 1 kb resolution chromatin model. The replica-averaged pairwise interaction frequencies match the ensemble-averaged Micro-C data (Pearson’s correlation coefficient $r = 0.90$, Fig. 2A and B).

We estimate the chromatin ensemble size by rescaling spatial coordinates to the 1 kb bead size of 22 nm, derived from nucleosome-resolution 1CPN model simulations [26, 29]. The average radius of gyration of the *Nanog* locus is $122 \pm 26 \text{ nm}$, accounting for $\sim 0.01\%$ of the nuclear volume (Fig. 2C). Assuming a uniform density of ~ 5 billion bp in the diploid mouse genome within a $10 \times 10 \times 5 \text{ \mu m}$ ellipsoidal nucleus [55], the expected R_g is $\sim 136 \text{ nm}$, slightly larger than our estimate but within the observed range (Fig. 2C). We observe a compact organization with a $1/3$ power-law scaling of end-to-end distances with segment length [26] (Supplementary Fig. S1C), characteristic of fractal globule architecture [56, 57], influenced by the bias toward Micro-C data.

The H3K9Ac signal, an epigenetic marker enriched at active enhancers and promoters, shows a linear dependence with average (local and nonlocal) contact density ($r = -0.43$; Fig. 2B and Supplementary Fig. S1A) and relative TF-accessible surface area (a proxy for accessibility of chromatin to TFs; $r = 0.48$; Supplementary Fig. S1B; see “Materials and methods” section). This indicates that the CREs enriched with histone tail acetylation are weakly packed and are preferentially accessible for TFs on the surface of the mesoscopic chromatin ensemble (Supplementary Fig. S1H), which is otherwise compact. This observation reflects the increased DNA accessibility of the acetylated regions probed by the MNase-seq [58]. The interactions between the H3K9Ac enriched beads are similar to nonspecific chromatin interactions indicated by their comparable reconfiguration timescales (Supplementary Fig. S1D and E). The CTCF- (and, similarly, SMC1/SMC3-) bound sites show slow reconfiguration, also evident from the slowest independent components from the tICA [30, 31] analysis of the ensemble (Supplementary Figs S1D and E and S4D and E), indicative of cohesin’s action as a topological constraint.

ChIP-seq signals of H3K27Ac (active enhancers) and transcriptional cofactor (BRD4) decrease exponentially with distance from enhancers (Fig. 2D; $r = -0.53$ for purple circles), suggesting that histone acetylation and cofactor recruitment

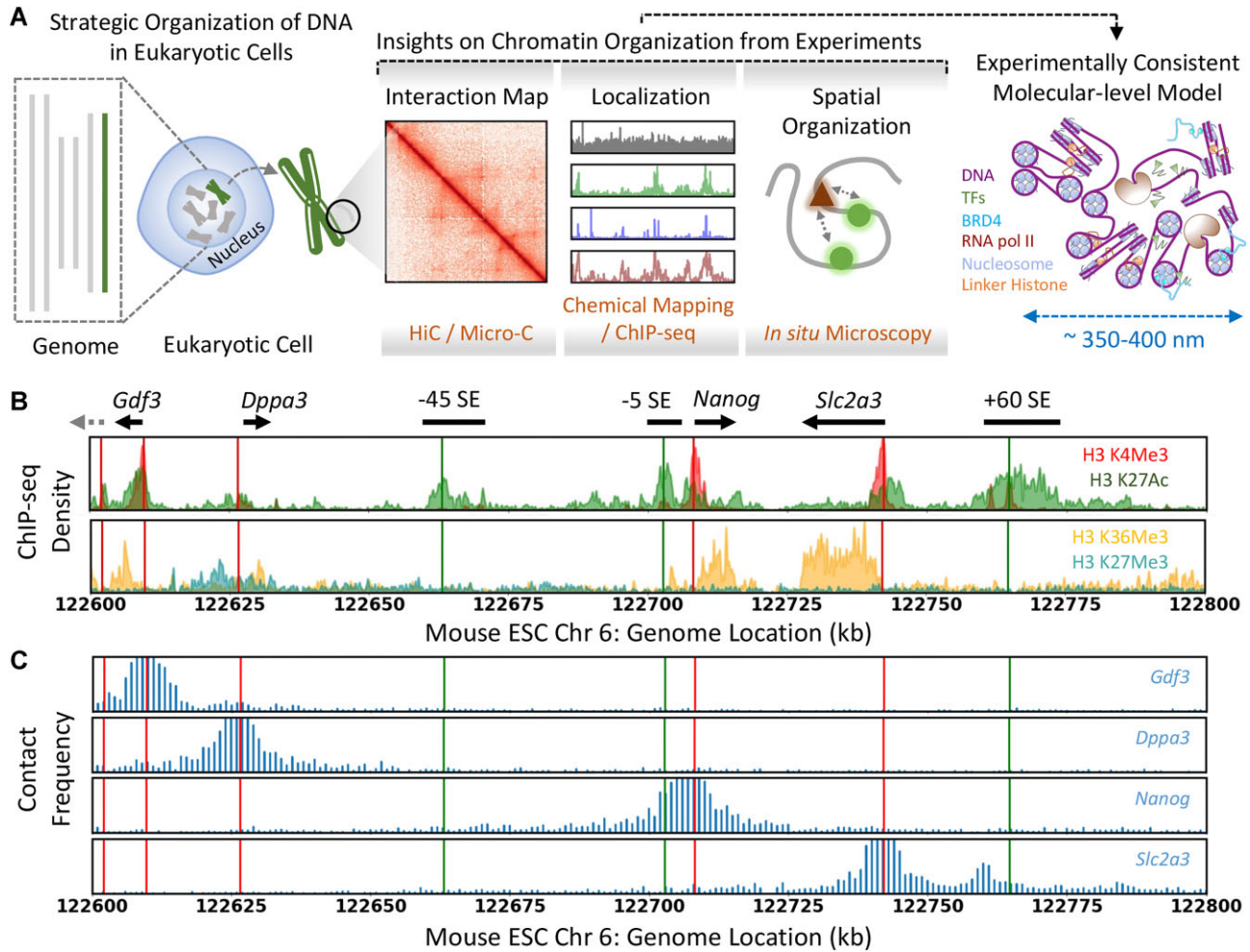


Figure 1. (A) Molecular modeling of the mESC *Nanog* gene locus combining a range of experiments reporting on the distinct features of chromatin organization. (B) Epigenetic signatures highlight the active enhancers (green), promoters (red), transcriptionally active (orange), and inactive (cyan) genes comprising the *Nanog* locus. The annotations are at the top, and the arrows indicate the orientation of the genes. The dashed gray arrow denotes the location of the partial *Apobec1* gene present inside the modeled chromatin segment. (C) Virtual 4C interaction maps reconstructed from the Micro-C data at 1 kb resolution from the viewpoint of gene promoters show the selective *cis*-interactions between promoters and three SE elements. As in panel (B), the red and green vertical lines denote the location of promoters and enhancers, respectively. The y-axis scale is uniform for direct comparison across panels.

depend on physical proximity to the three SE elements. Transcription signals extend ~ 100 nm from enhancers, with background noise up to ~ 200 nm, matching transcriptional condensate sizes observed in microscopy studies [14]. Genomic proximity alone cannot explain this, as the H3K27Ac signal extends up to ~ 400 nm (~ 18 kb) without spatial context (Supplementary Fig. S1F; $r = -0.31$ for purple circles). This suggests that the spatial proximity of CREs is crucial for E-P communication. The 50 kb region upstream of -45 SE does not follow this trend due to the weak dependence of *Apobec1* and *Gdf3* on the three SEs [4] (Fig. 2A) and may rely on upstream enhancers (Supplementary Fig. S1G).

We applied *k*-means clustering of the mesoscale ensemble based on the pairwise distance between 6 CREs (*Gdf3*, *Nanog*, and *Slc2a3* promoters and three SEs; Supplementary Figs S2 and S3; see “Materials and methods” section) to understand the distinct organization of CREs. The optimal number of clusters is determined using the elbow method, resulting in five representative conformational clusters with distinct organization of CREs (Fig. 2E, Supplementary Fig. S4 and

Supplementary Table S3). The projection of the mesoscopic chromatin ensemble over the first two tICA components reveals a heterogeneous conformational state within each cluster separated by a fuzzy boundary (Supplementary Fig. S4F–H).

The pairwise correlation map suggests three micro-domains with enhanced correlated motions indicating preferential intra-domain contacts (Supplementary Fig. S4K), each spanning ~ 65 kb and bound by cohesin acting as a topological anchor: (i) segment containing *Apobec1*, *Gdf3*, and *Dppa3*, (ii) -45SE to *Nanog* segment, and (iii) *Slc2a3* to +60SE segment, and the relative association of these domains defines the CRE interactions across the clusters (Fig. 2E and Supplementary Fig. S4I–L). The contact probabilities (Supplementary Methods) among the CREs (probed by distinct epigenetic markers on histones) and the cohesin-enriched beads (based on SMC3, SMC1, and CTCF ChIP-seq signals) hint at the distinct topological anchors characterizing the clusters (Supplementary Fig. S5B). We observe a preferential dissociation of acetylated beads compared to nonspecific chromatin interaction with identical pairwise sequence

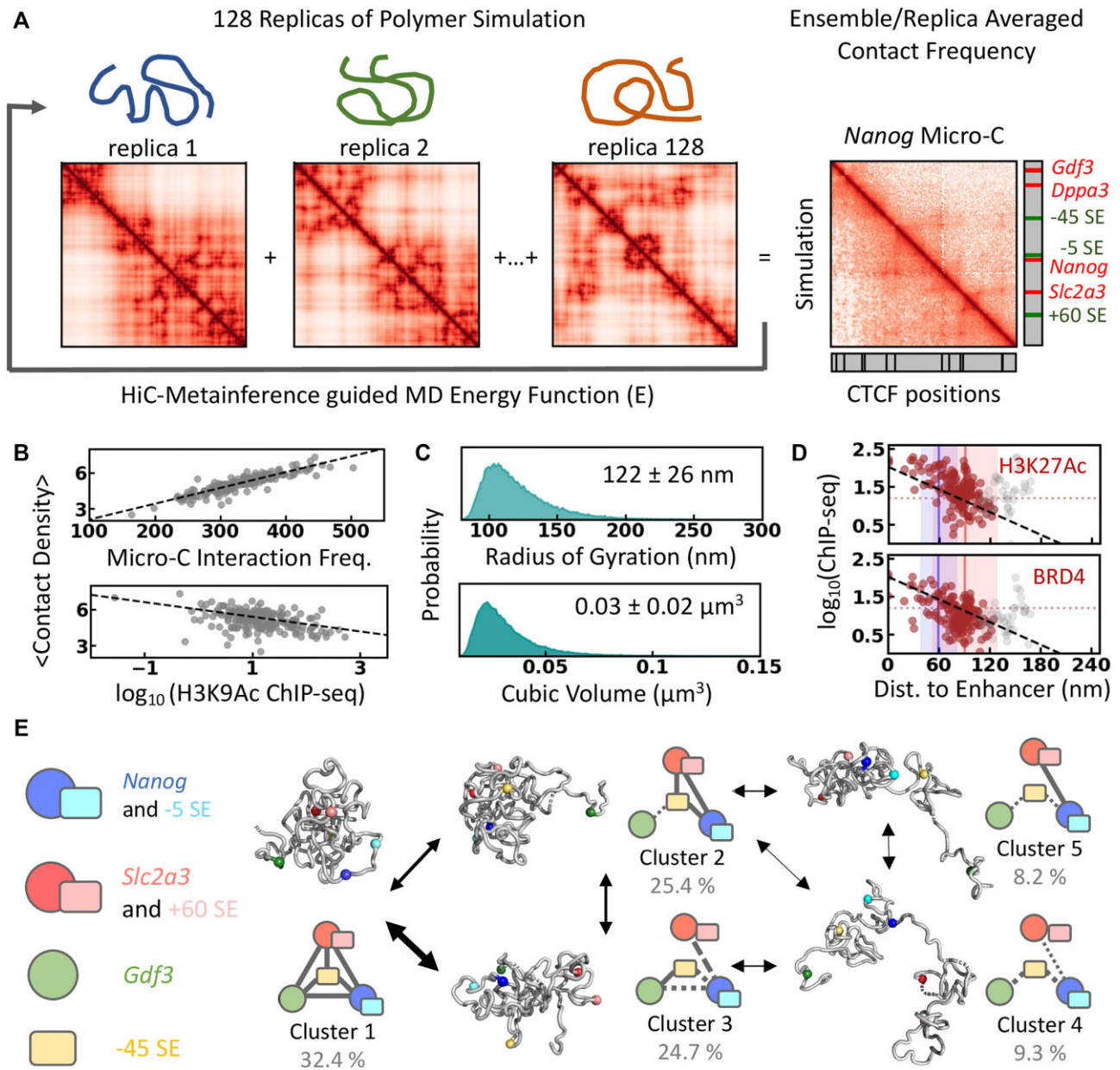


Figure 2. Mesoscopic model of the *Nanog* locus. **(A)** Schematic representation of the Hi-C metainference protocol. 128 replicas of polymer simulations at 1 kb resolution (1 bead = 1 kb) are performed. Hi-C metainference uses Bayesian inference to introduce energy bias such that the replica-averaged pairwise interaction frequency (bottom left panel on the right) quantitatively resembles experimental Micro-C data (top-right panel on the right). **(B)** The average contact density positively correlates with Micro-C contact frequency and shows weak negative dependence on the H3K9 acetylation signal. **(C)** Distribution of radius of gyration and volume of the generated mesoscopic ensemble. **(D)** The log-transformed ChIP-seq signals as the function of ensemble-averaged minimum distance from enhancers (−45, −5, and +60 SE) show a linear dependence. The colored and gray circles correspond to the 160 kb segment at the 3' end (122 640–122 800 kb) and the remaining 40 kb, respectively. The blue and red vertical lines mark the average distance of the *Nanog* and *Slc2a3* promoters from the enhancers, respectively, and the shaded area corresponds to 1 standard deviation from the mean. The dotted horizontal line marks the mean ChIP-seq signal. **(E)** Graphical representation of the conformational clusters obtained from *k*-means clustering and the observed transitions between them. Representative conformations from each cluster used for further modeling are shown in gray, and spherical beads represent the positions of CREs.

separation (Supplementary Fig. S5A). This suggests CRE interactions are not particularly favored but can instead be easily disrupted, possibly by a mechanism similar to the RNA-mediated disruption of transcriptional condensates [59] (Supplementary Fig. S5A). The representative conformations from each cluster with E-P and E-E distances close to the cluster average (Fig. 2E and Supplementary Fig. S6) are used for further modeling.

Mapping PIC and nucleosomes

We mapped the transcription PIC, comprising RNA polymerase II, mediator, and general TFs, based on the overlapping RNA polymerase II and H3K9Ac ChIP-seq peaks (Fig. 3A), resulting in 7 PICs at the CREs (see “Materials and methods” section). The number of PICs is close to the lower limit of the RNA polymerase II complexes estimated at the *Nanog* locus [18]. The RNA polymerases II–mediator complexes are not

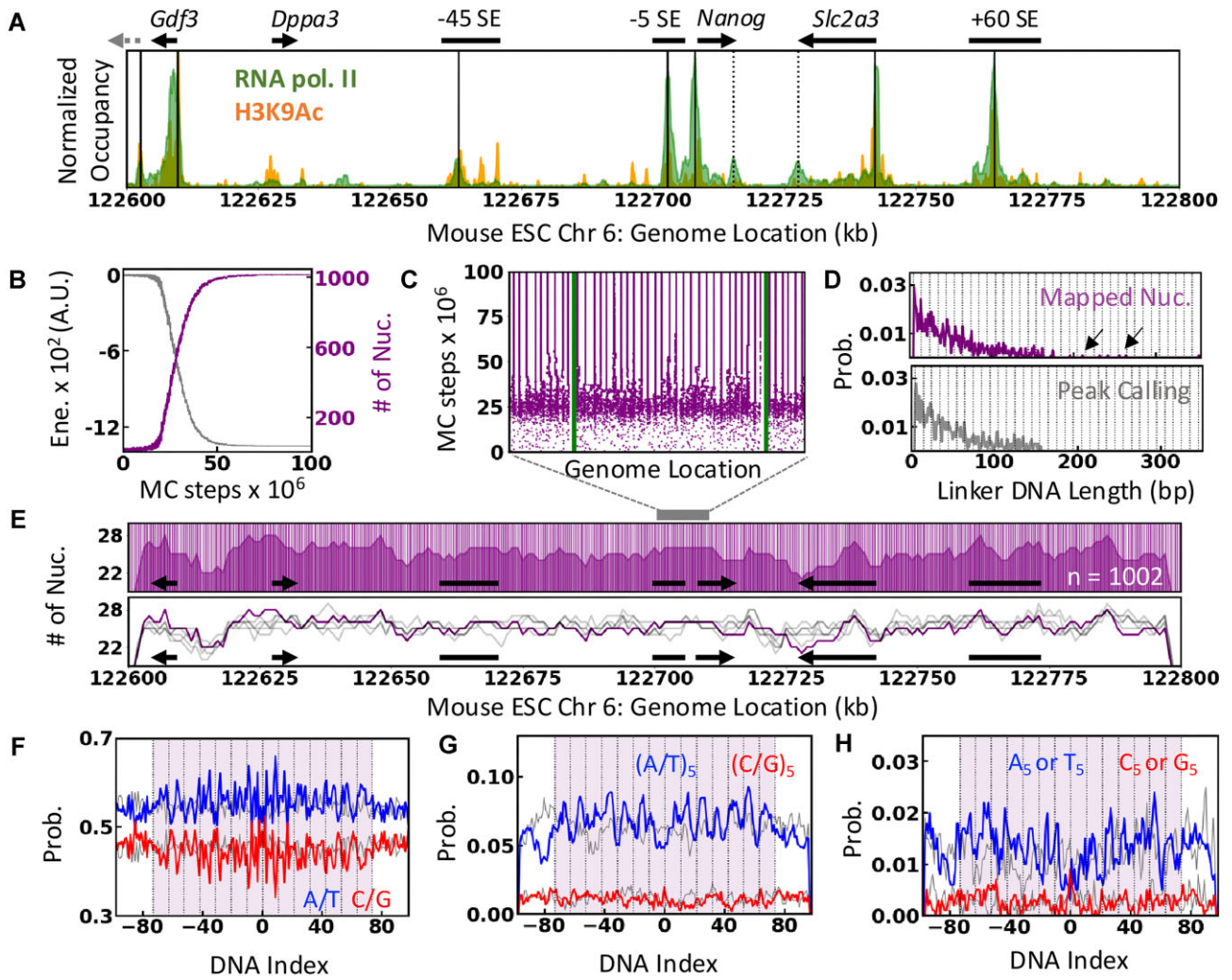


Figure 3. Mapping PIC and nucleosome positions. **(A)** Mapped positions of PIC (solid black line) in the background of RNA pol II (green) and H3K9Ac (orange) ChIP-seq profiles. The CREs and gene annotations are shown at the top for reference. Dashed black lines denote the RNA pol II peaks at the gene ends. **(B and C)** Converging the number of nucleosomes, total energy, and nucleosome positions as the function MC steps. **(D)** Distribution of DNA linker length obtained from the representative nucleosome map (top) and the peak-calling method (bottom). The arrows highlight the nucleosome-free regions (>185 bp), and the vertical dotted line indicates the $(10n + 5)$ periodicity. **(E)** Nucleosome positions (vertical lines; $n = 1002$) and the number of nucleosomes per 5 kb (shaded area) of the selected nucleosome position map (top). Stochastic variation in the nucleosome positions generated by independent MC simulations (gray) compared to the selected positional map for further modeling (purple). **(F–H)** The probability of the sequence features calculated as the function of the DNA index for the mapped nucleosome positions (colored) and randomly chosen 197 bp DNA segments (gray). X_5 in the legend indicates 5mers of X. The dotted vertical lines indicate the position of SHL (10.5 bp) with the dyad position at 0. The shaded area marks the mapped nucleosome position.

mapped to the *Nanog* and *Slc2a3* gene termini, as the experimental model of the transcriptional machinery at the termination sites is unavailable.

The nucleosomes are mapped using Monte Carlo (MC) simulations guided by the nucleosome center positioning scores (NCPS) obtained from experimental chemical mapping [20] (Supplementary Fig. S7A). The simulated annealing protocol converges to well-defined nucleosome positions with strong NCPS scores compared to the random mapping of nucleosomes (Fig. 3B and C, and Supplementary Fig. S7B). Consistent with the genome-wide estimate of mESC, the typical NRL in the *Nanog* locus ranges between 185 and 217 bp [60], and the linker length distribution shows a periodicity of $10n + 5$ [20] (Fig. 3D). The NCPS score used for mapping is influenced by various *in vivo* factors that disrupt the nucleosome positioning, such as chromatin remodelers and transcriptional

machinery. Although we attempted to densely pack the 200 kb segment with nucleosomes (Supplementary Fig. S7A), our protocol detects nucleosome-free regions (>185 bp) around the PIC-loaded regions (Fig. 3D and Supplementary Fig. S7D), reminiscent of consecutively loaded RNA Pol-II molecules at CREs, that are not apparent in the conventional peak-calling methods [20].

Individual MC runs generate slightly different nucleosome positional maps with a periodicity of $10n$ shift in nucleosome positions compared to a representative nucleosome position map (Supplementary Fig. S7C). This is expected in an ensemble view of nucleosome positions (Fig. 3E), but the individual runs retain a narrow distribution of nucleosome density measured over a 5 kb sliding window, i.e. 25 ± 2 nucleosomes per 5 kb (Fig. 3E). Despite the stochastic nature of the mapping resulting in subtle differences in

the nucleosome positions, ~14% of the nucleosomes are mapped to unique positions in the 50 independent runs and are roughly uniformly distributed in the 200 kb segment (Supplementary Fig. S7F). Specifically, nucleosomes in the proximity of TSS along the direction of the gene are uniquely mapped as estimated from the 50 independent runs (Supplementary Fig. S7E).

We selected a tentative nucleosome map (1002 nucleosomes) for further modeling and analysis. The mapped nucleosome positions reveal strong A/T phasing patterns with peaks corresponding to the DNA minor grooves facing the histone octamer ($SHL \pm 0.5, \pm 1.5$, etc.; Fig. 3F), that becomes prominent for (A/T)₅ 5-mers (Fig. 3G; e.g. ATATA, TTTTA, AAAAA, etc.). The phasing (A/T)₅ patterning is consistent with the analysis of whole-genome studies using chemical mapping [20] and MNase-based ChIP-seq [61], and corresponds to the strong nucleosome positioning sequences (NPS) [20, 62]. The (A/T)₅ phasing is not apparent at the linker DNA and the randomly mapped positions (gray in Fig. 3F–H), validating the nucleosome positions (Fig. 3F and G). The poly-A and poly-T stretches (A₅ - AAAAA and T₅ - TTTTT) are depleted at the dyad but populate further away from the dyad and at the linker DNA (Fig. 3H). A₅ and T₅ stretches are reported to decrease the stability of DNA wrapping at the ends and increase the accessibility of nucleosome-wrapped DNA [63]. Meanwhile, an apparent patterning of G/C probability is observed as expected for mono- or di-nucleotide G/C stretches [64, 65] but the feature disappears for longer stretches of G/C (Fig. 3F–H).

Mapping nucleosome-associated protein factors

The linker histone and transcriptional cofactor BRD4 directly bind nucleosomes at dyad and acetylated histone tails, respectively [66, 67]. Their positions are mapped using a similar Monte Carlo-based approach with the energy terms proportional to the ChIP-seq signals averaged over the mapped nucleosome positions (Supplementary Methods). The energy terms are tuned to reproduce the *in vivo* H1/Nuc ratio (0.36–0.46) [19, 35] and the number of BRD4 measured in the proximity of *Nanog* locus [18] (“Materials and methods” section), resulting in 418 linker histones (H1/Nuc \approx 0.42) and 20 BRD4 molecules mapped onto the predetermined nucleosome positions (Fig. 4A). The number density of H1 per 5 kb shows a broad distribution, ranging from 3 to 20 ($\sim 11 \pm 3$) H1 per 5 kb, unlike the near-uniform distribution of nucleosomes (Fig. 4B).

The number density of H1 calculated over the 5 kb sliding window shows a positive correlation ($r = 0.52$) with the local contact frequency from Micro-C data (orange shaded region and black line in Fig. 4A, respectively), indicating compact local organization mediated by the linker histones. The mapped BRD4 positions largely overlap with the BRD4 and H3K9Ac ChIP-seq peaks at the CREs (Fig. 4A), consistent with the expectation that BRD4 binds acetylated nucleosomes [68] and hence, BRD4 bound nucleosomes (Nuc-BRD4) can be used as a proxy for acetylated nucleosomes. Interestingly, the average ChIP-seq frequency of BRD4 (and H3K9Ac), calculated over 5 kb sliding windows, exponentially decays with increasing H1/nucleosome ratio (Fig. 4A and C). This suggests that the association of linker histone and BRD4 is mutually exclusive, and the balance between H1 association and acetylation (as BRD4 binds acetylated

histones) results in distinct local compaction of chromatin fibers.

The collective mapping of nucleosome, H1 and BRD4 provides a unique opportunity to identify sequence features that may assist the regulated association of the accessory proteins with nucleosome. The three classes of nucleosomes—Nuc-H1 (H1 bound nucleosome), Nuc-BRD4 (BRD4 bound nucleosome or acetylated nucleosome), and Nuc (other free nucleosomes)—show the distinct distribution of A/T fraction (A/T%, Fig. 4D), with H1-bound nucleosomes relatively more enriched in A/T% compared to free nucleosomes and Nuc-BRD4 (P -value < 0.001 , using two-sample Kolmogorov–Smirnov test). The A/T% of the nucleosomal DNA increases with the increasing H1 binding probability (Fig. 4E, top panel), which could be explained by the previously observed higher affinity of linker histones for A/T-rich sequences [69] (Fig. 4E). The phasing of (A/T)₅ motifs is retained for all classes of nucleosomes (Fig. 4E and F) with a strong enrichment of (A/T)₅ tracks specifically at half-integer SHL for Nuc-H1 and gradually decreases for Nuc followed by Nuc-BRD4. The sequence trend in Nuc-BRD4 is not apparent due to the small number of nucleosomes in this class.

Mapping transcription factors and cofactors

We mapped the binding sites of four pioneer TFs (SOX2, OCT4, NANOG, and KLF4) based on the ChIP-seq density and their cognate sequence bias, assuming a uniform nuclear concentration of 1 μ M based on the available experimental estimates [36, 37] (Fig. 5A; Supplementary Methods). Most TF-binding sites are mapped at CREs, and the number of each TF mapped is equivalent to the number of SOX2 estimated at the *Nanog* gene locus *in vivo* [18]. We mapped three P300 molecules based on the theoretical estimate of one P300 per active gene [16], guided by the ChIP-seq density (Fig. 5A).

The occupancy of the TFs, cofactors, and acetylation markers, measured using the ChIP-seq density averaged over the mapped nucleosome positions as a proxy, are strongly depleted at Nuc-H1 positions equivalent to the background (i.e. mean ChIP-seq density of the 200 kb segment; Fig. 5B and Supplementary Fig. S8A). However, they are prominently enriched at linker DNA $<$ free nucleosomes $<$ Nuc-BRD4 [70]. Based on the strong localization bias of the transcriptional signal, we hypothesize that the sequence-regulated chromatin accessibility, where H1-depleted regions spanning CREs are more accessible for TFs, leads to the cascading cycle of protein recruitment and acetylation required for transcription regulation. In line with these expectations, out of ~75% of the TFs binding sites mapped to the nucleosome, only ~20% are mapped to H1-Nuc (Fig. 5C). For better statistics, we randomly map an equal number of TF-binding sites with a uniform sequence preference for the 200 kb DNA. The randomly mapped TF-binding sites and the TF-binding sites identified based on the ChIP-Seq density and cognate sequence bias show a similar preference for the nucleosome over the linker DNA except for a subtle decrease in the preference for H1-Nuc in the latter (Fig. 5C).

Of the four TFs, SOX2, NANOG, and KLF4 bind the minor groove, and OCT4 binds the major groove of the DNA. Experiments and simulations have demonstrated that SOX2 can better recognize the exposed target sites and induce nucleosome sliding to expose target sites for other TFs [70] or change the activation domain accessibility for further recruitment of tran-

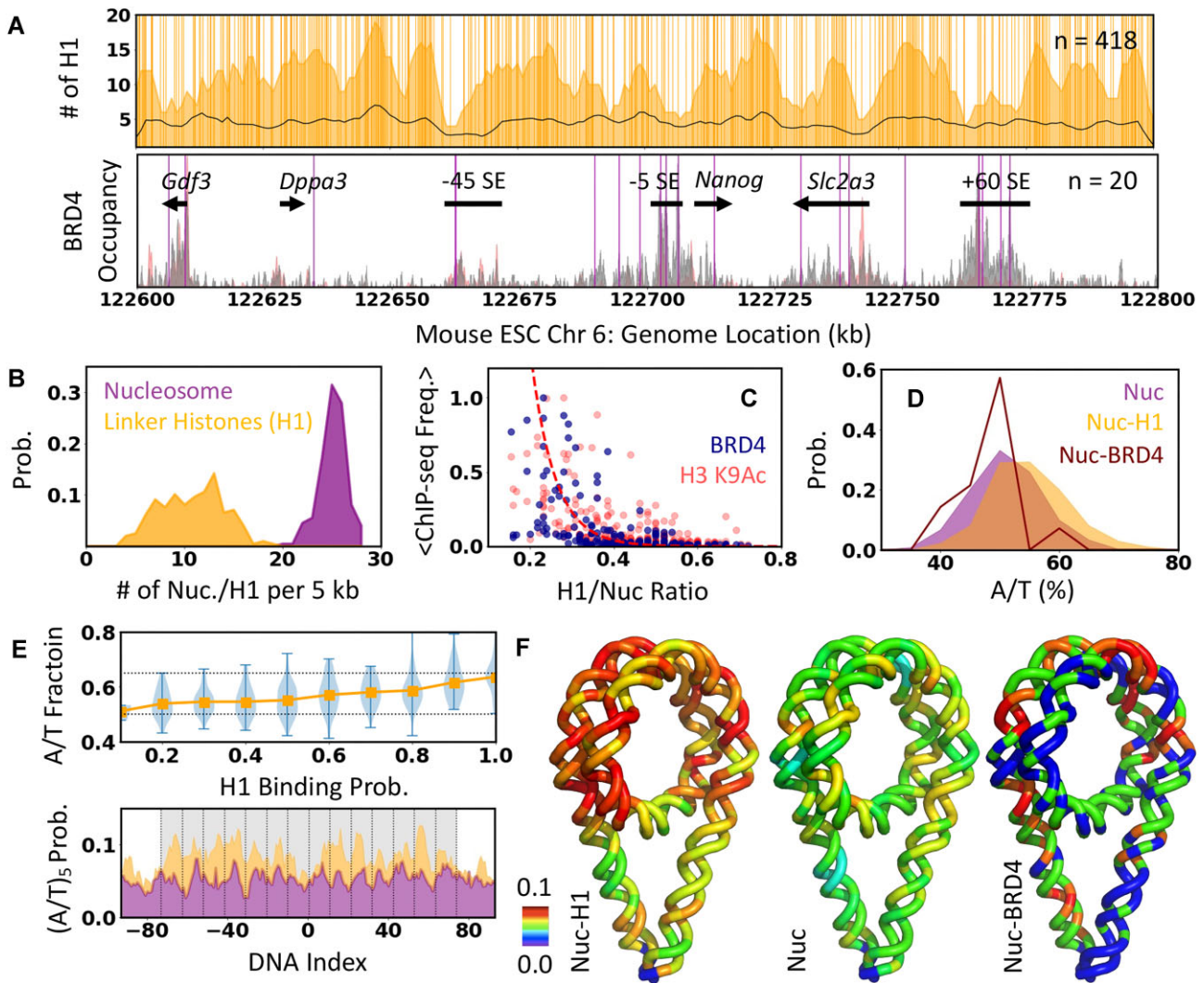


Figure 4. Mapping nucleosome-associated accessory proteins. **(A)** Top: Mapped genomic locations of linker histone (vertical lines, $n = 418$), the number of linker histones mapped per 5 kb (orange shaded region), and the local contact frequency estimated from Micro-C data (black line). Bottom: Mapped positions of BRD4 (vertical lines, $n = 20$) compared against BRD4 and H3K9Ac ChIP-seq profile (gray and red shaded region, respectively). The annotations are shown in top. **(B)** Distribution of the number of nucleosomes (purple) and H1 (orange) per 5 kb shown in panel (A) and Fig. 3E. **(C)** Comparison of normalized mean BRD4 and H3K9Ac ChIP-seq frequency and H1/nuc ratio per 5 kb sliding window. The red dashed line is an exponential curve to visualize the apparent trend. **(D)** Distribution of A/T% of different nucleosome classes based on the generated positional maps. **(E)** A/T% as the function of H1 binding probability estimated from the experimental data (orange). The distribution of A/T% in each bin is shown in blue. Black dotted lines at 0.5 and 0.7 highlight the linear increase in the A/T%. The bottom panel shows the phasing poly A/T feature of Nuc-H1 (orange) and Nuc (purple). The gray-shaded area marks the position of the nucleosome. **(F)** The $(A/T)_5$ probability shown in panel (E) is mapped on the nucleosomal DNA for easy visualization.

scription (co-)factors [71]. Consistent with this observation, we note that the mapped SOX2-binding sites populate the exposed minor grooves (SHL -4 to 4; dashed lines on Fig. 5D and E and Supplementary Fig. S8D), and the OCT4-binding sites populate the major grooves facing the histone core. The mapped binding sites of the NANOG and KLF4 also show a subtle preference for the exposed minor grooves of the nucleosomal DNA (Fig. 5D and E and Supplementary Fig. S8D). About 10% of OCT4 and 35% of NANOG-binding sites are colocalized with SOX2-binding sites on the same nucleosome (Supplementary Fig. S8B). Importantly, $\sim 50\%$ of each TF-binding site is in close genomic proximity (< 500 bp or ~ 2 nucleosomes away) to all the other TFs, compared to $< 0.05\%$ for the randomly-mapped TFs, suggesting a strong colocalization of the TFs (Supplementary Fig. S8C).

The combined positional maps reveal that the BRD4s are mapped near the independently mapped TF-binding sites (Fig. 5F and G). Specifically, $\sim 50\%$ – 60% of the mapped TF binding sites are < 1 kb genomic distance from the dyad position of BRD4-Nuc. In terms of nucleosome separation, $\sim 50(70)\%$ of the TF-bound nucleosomes are $< 4(8)$ nucleosomes away from the BRD4-Nuc with preferential $i-i\pm 2n$ separation. When randomly mapped, TF binding sites are almost always 10 nucleosomes away from the BRD4-Nuc (Fig. 5G). We also note that $\sim 20\%$ – 50% of the mapped TF-binding sites are within 1 kb of the three P300 molecules. The spatial proximity of the TFs, P300, and BRD4 likely reflects an underlying molecular mechanism where the TF-recruited acetylation factors (possibly P300) modify the histone tails of proximal nucleosomes [16].

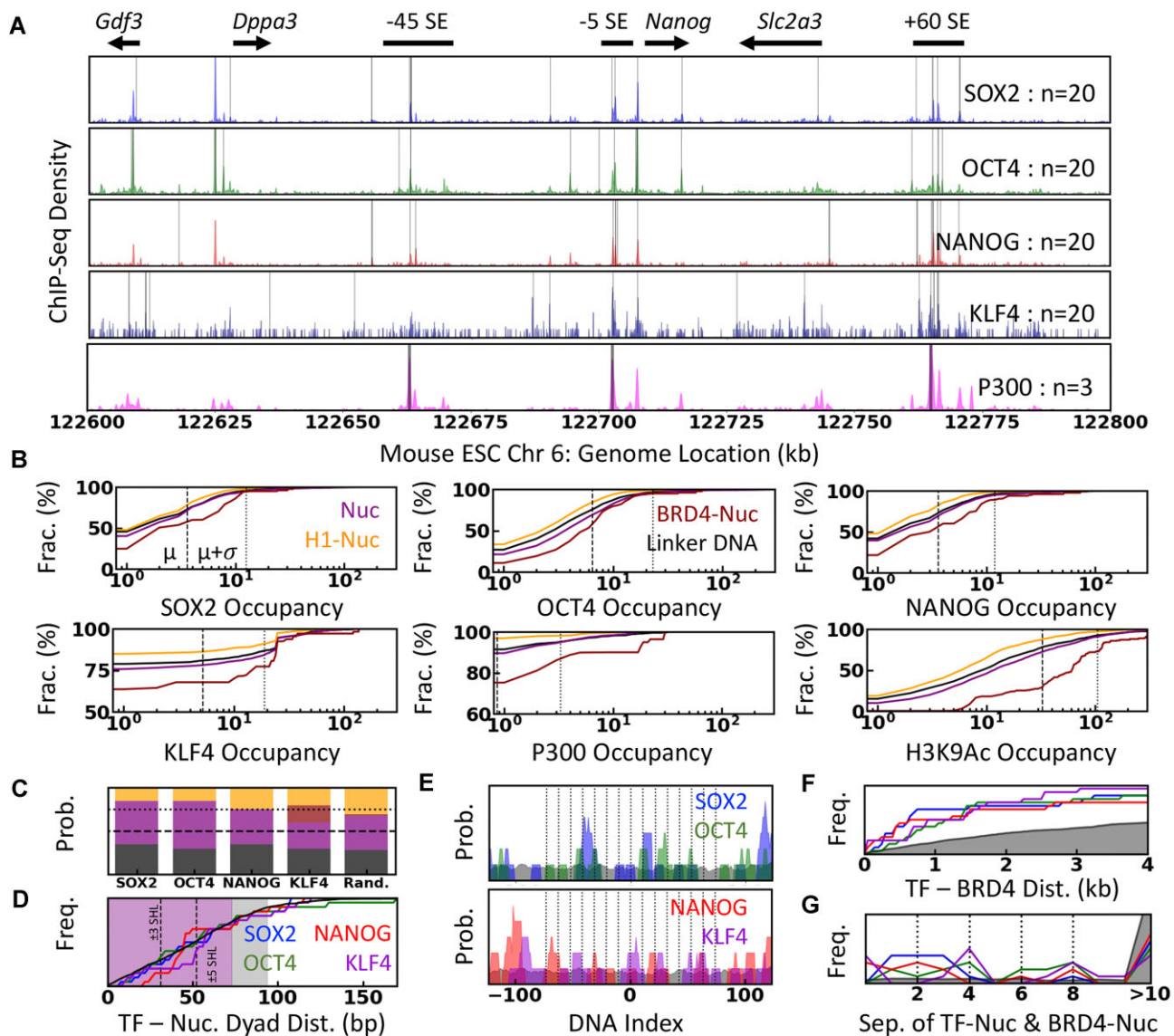


Figure 5. Mapping transcription (co)factors. **(A)** Mapped genomic position of four pioneer TFs (gray vertical bar, $n = 20$) and P300 (gray vertical bar; $n = 3$) alongside the ChIP-seq density (shaded area). The number of molecules mapped is denoted on the right, and the annotations are shown at the top. **(B)** Cumulative distribution of the nucleosome occupancy of TFs, cofactors, and acetylation markers, using mean ChIP-seq density over mapped nucleosome position as a proxy for each nucleosome class and linker DNA, as indicated in the first two panels. The dashed and dotted vertical lines indicate the mean (μ) and one standard deviation from the mean ($\mu + \sigma$) calculated for each data within the *Nanog* locus, respectively. **(C)** The probability of TFs mapped to different nucleosome classes following the color code as in panel (B) and the distribution expected of randomly mapped TF-binding sites with uniform binding preference to the 200 kb DNA. The dashed and dotted lines indicate the 0.5 and 0.75 probability, respectively. **(D)** Cumulative distribution of sequence separation between the center of the TF-binding site and the dyad of proximal nucleosome. The purple and gray shaded regions correspond to the nucleosome and 20 bp linker DNA, respectively. The black line indicates the cumulative frequency expected for a randomly mapped TF. **(E)** Probability of mapped TFs binding sites as the function of nucleosome DNA index. The dyad position is indicated as 0, and the dashed black lines denote the nucleosome SHL. The probability expected for a randomly mapped TF is shown as a black-shaded area. **(F)** Cumulative frequency of the minimum genomic distance between the center of the TF-binding site and the dyad of BRD4-Nuc following the color code in panel (D). The black-shaded region represents the distribution expected for a randomly mapped TF. **(G)** Distribution of shortest nucleosome separation between TF- and BRD4-bound nucleosomes following the same color code in panel (F). Dotted black lines indicate the $(i \pm 2n)$ nucleosome separation.

Backmapping mesoscopic to near-atomistic resolution

We combine the positional maps generated by the ChIP-seq/Chemical-mapping guided Monte Carlo simulations and the mesoscopic conformations generated by Hi-C meta-inference simulations using our backmapping protocol (Fig. 6A and [Supplementary Fig. S10A–E](#)). This protocol recognizes distinct DNA–protein complexes and generates the chromatin fiber by alternatively connecting the fiber

modules—coarse-grained models of nucleosomes (free nucleosome, nucleosome bound to H1 and/or BRD4), PIC, and linker DNA with artificially compacted disordered regions ([Supplementary Figs S9 and S10B](#))—and minimizes steric clashes, distance to the reference mesoscopic fiber and distance between i -i+2 (D_{i-i+2}) nucleosome by a Monte Carlo-like sampling of module conformations at each step (“Materials and methods” section; [Supplementary Fig. S10A–E](#) and [Supplementary Videos S1–S5](#)). The backmapping protocol

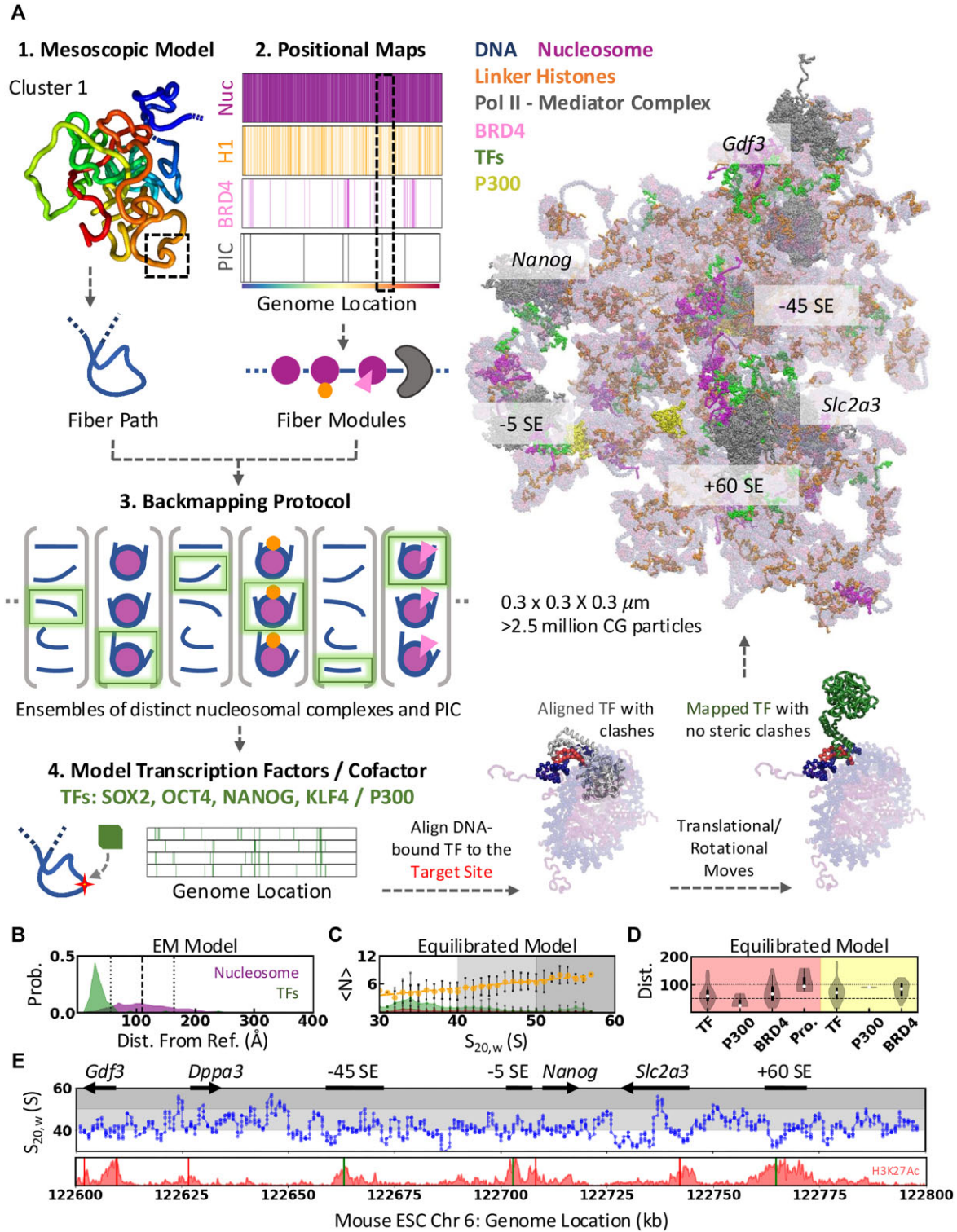


Figure 6. (A) Schematic representation of the integrative modeling pipeline. Our semi-automated modeling pipeline combines multiscale modeling strategies to generate *in vivo* relevant chromatin models at a nanoscopic resolution consistent with various experiments. **(B)** Distribution of distance of the modeled nucleosomes (purple) and TFs (green) from their reference mesoscopic bead and DNA-binding site, respectively, calculated from the energy minimized (EM) model. The radius of the mesoscopic bead (11 nm, black dashed line) \pm radius of the nucleosome (5.5 nm, black dotted lines) are shown as references. **(C)** The number of linker histones (orange), TFs (green), and BRD4 (red) mapped to the 12-nucleosome chromatin segments as the function of their sedimentation coefficient ($S_{20,w}$) from the equilibrated cluster 1 model. The error bars are calculated as the standard deviation in each bin. The shaded backgrounds indicate the broadly categorized expanded (white), moderately compact (light gray), and compact (dark gray) chromatin segments. **(D)** The distribution of the COM-COM distance of four TFs, P300, BRD4, and promoters to the spatially proximal enhancers (red shaded) and promoters (yellow shaded). The promoter and enhancer positions are defined by the DNA segment bound to the PIC at their proximity, which also overlaps with their respective epigenetic markers. **(E)** The $S_{20,w}$ of the 12 nucleosome segments as the function of genomic position. The shaded background is the same as that in panel (C). The annotations are given at the top, and the H3K27Ac ChIP-seq signals are shown at the bottom panel for reference.

accommodates large nucleosome-free regions and simulates an apparent decrease in sedimentation coefficient ($S_{20,w}$) with increasing NRL, qualitatively resembling the experiments [71] (Supplementary Fig. S10F). Each run produces a structurally different chromatin model with a consistent local arrangement of nucleosomes.

We generate two complete models for cluster 1 (Model 1 and 1*) to evaluate the reproducibility of the backmapping protocol and one each for the remaining mesoscopic chromatin structure cluster (Figs 2E and 6A; Supplementary Figs S11 and S12). The RMSD between the two independent models of cluster 1 based on the COM of nucleosomes is 20.3 nm, less than the resolution of the mesoscopic reference structure (22 nm). Across the six models, our backmapping procedure can model ~45(80)% of the nucleosomes within 11(17) nm from the reference mesoscopic bead (Fig. 6B, and Supplementary Figs S10F and S13A). TFs and P300 are iteratively placed within $\sim 4 \pm 1$ and $\sim 20 \pm 10$ nm (COM–COM distance; Fig. 6B and Supplementary Fig. S13A), respectively, from their target DNA-binding site by algorithmically orienting them to avoid steric clashes, as illustrated in Fig. 6A.

The chromatin conformations at near atomistic resolution are energy minimized and equilibrated for 100 ns using GENESIS CGDYN [47] to remove the modeling artifacts due to the artificially compacted fiber modules used in the backmapping procedure and to evaluate the stability of the generated gene locus models. The backmapping protocol favors compact local structure for convenient modeling, and accordingly, we observe local relaxation and subtle rearrangements favoring multivalent inter-nucleosome interactions after equilibration (RMSD = ~ 15.5 nm; Supplementary Fig. S14). The TFs and P300 molecules diffuse freely or by interacting with spatially proximal molecules, including sequence-specific interaction with DNA as demonstrated in the previous work [53, 72], during the equilibrium process (supplementary methods), resulting in subtle differences in the intermolecular distances measured from the energy-minimized and equilibrated model (Supplementary Fig. S13B and C). We treat the final snapshot from the short equilibrium run for each cluster as the representative near-atomistic structures of the 200 kb *Nanog* locus.

Chromatin organization at the near-atomistic resolution

The local chromatin organization in our high-resolution models is compatible with the protein localization and mesoscopic ensemble generated using the Micro-C data, as we observe a mutually correlated trend in the number of mapped linker histones, local contact frequency measured from the Micro-C data, contact density measured from the mesoscopic ensemble, and the local compaction measured from the high-resolution models (Figs 2B, 4A, and 6C). The sedimentation coefficient ($S_{20,w}$) calculated for the 12-nucleosome sliding window highlights that the compact chromatin segments have a high H1/Nuc ratio (orange circles in Fig. 6C; $r = 0.95$). This aligns with our previous observation that local contact frequencies calculated from the Micro-C data are weakly correlated to the H1 density (Fig. 4A) and several experimental evidence of H1-mediated chromatin compaction [71, 73–75]. We also observe that chromatin segments mapped with TFs and BRD4 are highly expanded and accessible, as indicated by the low $S_{20,w}$ values (green in Fig. 6C, and Supplementary Figs S12 and S15A). Specifically, the chromatin accessibility is higher at

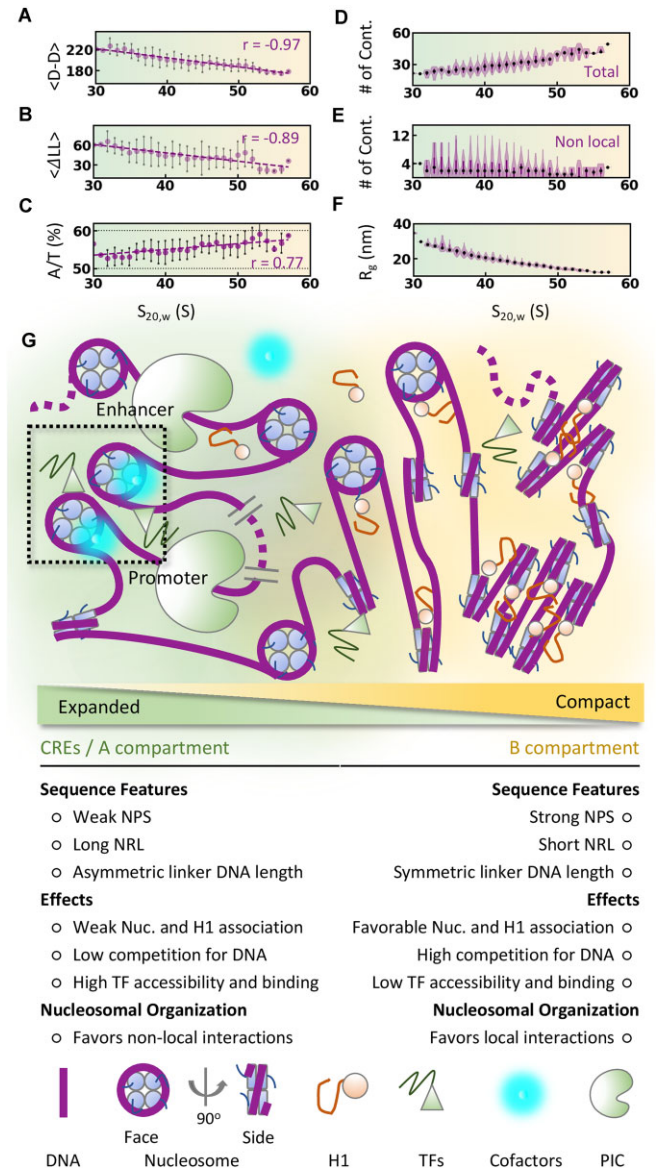


Figure 7. (A) The mean dyad–dyad distance in bp ($\langle D-D \rangle$; proxy for $i-i+1$ nucleosome separation). (B) Asymmetry in entry/exit linker DNA length (ΔLL ; proxy for $i-i+2$ nucleosome interaction). (C) A/T fraction (A/T %; proxy for nucleosome positioning signal). (D and E) Total and nonlocal ($|i-j| > 10$) nucleosome–nucleosome contacts. (F) Radius of gyration as the function of the sedimentation coefficient ($S_{20,w}$) of 12-nucleosome chromatin segments. The solid lines are linear fits, and Pearson's correlation coefficients (r) are given in panel (A–C). The panels are colored in a gradient from green to orange, representing expanded to compact chromatin segments. (G) Cartoon representation of the sequence and nucleosome organization principles guiding the distinct local organization of the chromatin. The CREs are locally expanded, and the increased accessibility facilitates binding regulatory protein complexes. The transient nonlocal interactions between these expanded segments (dotted black box)—“Nucleosomal Handshake”—possibly form the roadmap for *cis*-interactions.

CREs (Fig. 6E, and Supplementary Figs S1B and S16), and we also observe an apparent increase in the H3K27Ac ChIP-seq signal with the increased chromatin accessibility in the mesoscopic ensembles (Fig. 2B). The distinct local organization—compact (50–60 S) and expanded (30–40 S)—shows contrasting nucleosome organization and sequence features (Fig. 7A–C and Supplementary Fig. S15B–D). The compact chromatin

segments have relatively short NRL (175 ± 3 bp) and smaller differences in the entry/exit linker DNA length ($\sim 20 \pm 5$ bp) and linearly increase to 220 ± 4 bp ($r = -0.93$) and 65 ± 15 bp ($r = -0.88$), respectively, for the expanded segments. We also observe strong nucleosome positioning signals (high A/T%) at compact chromatin segments, corroborating our previous result that H1 favorably binds DNA regions with strong NPS (Fig. 4F).

The models span $\sim 350 \times 350 \times 350$ nm³ and provide a physiologically realistic representation of the *in vivo* chromatin organization and spatial compartmentalization of transcriptional components. In the cluster 1 model, most transcriptional components are encased within a radius of 125 nm, similar to the experimental estimated size of transcriptional condensates (~ 100 – 200 nm) [14], embedded within a network of molecular interactions (Supplementary Fig. S17). The models do not show apparent phase-separated condensates comprised of transcriptional components (Supplementary Figs S11 and S17). However, we observe a hierarchical spatial organization of transcription regulators from the enhancers, TFs (SOX2, OCT4, NANOG, and KLF4 at ~ 60 nm) < BRD4 (~ 66 nm) < promoters (~ 93 nm), corroborating with the expected molecular mechanism of transcription (co)factor recruitment, and the spatial molecular organization at the *Nanog* locus [14, 16, 18] (Fig. 6D, and Supplementary Fig. S13B and C). Despite the apparent higher median distance of the TFs from the enhancers compared to P300 (~ 24 nm), there are more TFs in the proximity of enhancers compared to P300. We observe a similar hierarchical spatial organization of the TFs and cofactors from the enhancers across all five models, with a subtle increase in median distance with increasing distances between the CREs (Supplementary Figs S13B and C, and S17).

A similar spatial organization is not observed from the viewpoint of promoters, where the distance to the transcription (co)factors is approximately equivalent to or larger than the distance to the enhancers (Fig. 6D, and Supplementary Fig. S13B and C). Despite the differential spatial organization between promoters and enhancers, the average pairwise distance between the RNA polymerases and SOX2/BRD4 (~ 130 – 220 nm, depending on the cluster) agrees with the estimates from experimental image tracking studies at the *Nanog* locus, further attesting to the quality of global organization represented in our models [18].

The experimentally consistent local and global molecular organization in our models reveal a network of transcription (co)factors bridging the enhancers and promoters (Supplementary Fig. S17). These models can suggest potential molecular mechanisms underlying transcription regulation, although the lack of large-scale dynamics will make the discussion mostly speculative. A recent multiscale modeling study highlighted the role of nucleosome plasticity in promoting multivalent nucleosome interactions [76]. Considering the distinct local expansion of chromatin at CREs compared to other regions, we analyzed the nucleosome-nucleosome interactions as the function of chromatin compaction. Interestingly, we identify that despite the relatively low local nucleosome interactions formed by the expanded 12-nucleosome chromatin segments (~ 24 or valency = ~ 2 ; Fig. 7D and Supplementary Fig. S18A), they favor nonlocal interactions (Fig. 7E and Supplementary Fig. S18B) preferably with other expanded segments, compared to the compact segments. The radius of gyration of the expanded 12 nucleosome segments ranges between 20 and 35 nm, facilitating the nonlocal nu-

cleosomal interactions due to the increased accessibility, and can sufficiently bridge the promoters to the nearest enhancers (Figs 6D, and 7E and F; Supplementary Figs S13C and S18B and C). This suggests that the local chromatin organization contains the roadmap for communication between CREs and is encoded by the strategically organized nucleosomes.

Discussion

In this work, we develop a multiscale modeling pipeline to explore the *Nanog* gene locus in mESCs at a near-atomistic resolution. We combine information from experimental ensemble-averaged protein localization, high-resolution pairwise interaction frequencies among genomic loci, cryo-electron microscopy, and *in vivo* single-molecule fluorescent studies using a multiscale approach. The generated model is a first step towards understanding the functional role of chromatin organization in facilitating E-P communication. Thanks to the availability of extensive experimental data, the proposed protocol is easily transferable to model other genomic segments and cell types or the same locus under different developmental or perturbed (e.g. CTCF/cohesin deletions) stages.

In addition to the detailed nanoscopic model, the adopted multiscale modeling methodology also provides insights into the principles of chromatin organization and molecular organization at CREs that are not directly accessible from experimental data of individual chromatin components. We explored how the observed chromatin ensemble informs the interplay between chromatin structure and gene regulation. The graded polyA/T phasing signal for Nuc-H1 > Nuc > Nuc-BRD4 suggests a sequence-regulated chromatin accessibility mechanism (Fig. 4F). The strong NPS may enhance the H1 association via favorable linker DNA geometry resulting from the strong DNA wrapping or decreased nucleosome dissociation and sliding. The resulting compact local nucleosome arrangements inhibit chromatin accessibility for TF/cofactor binding and histone tails for acetylation factors (Figs 4C and 5). Chromatin compaction is further ensured by smaller NRL and symmetric entry/exit linker DNA length of the nucleosomes [71] (Figs 6C–E, 7A–C, and Supplementary Fig. S15).

The transcriptional status of the genes is regulated by dynamic modification of the chromatin status at CREs, suggesting regulated access of TFs and cofactors to CREs. Accordingly, CREs are populated with weak NPS that favor the transient local nucleosome-nucleosome interactions due to a relatively high unwrapping rate [76]. The larger NRL and asymmetric entry/exit linker DNA length could disrupt the regular local organization and H1 binding due to increased electrostatic repulsion between the linker DNA and unfavorable linker DNA geometry [71, 74, 75, 77]. The strategically orchestrated chromatin expansion and possibly increased nucleosome turnover ensure the DNA accessibility for protein binding and histone tail modifications. Together with the acetylation-dependent contact density and TF-ASA observed in the Hi-C meta-inference simulations (Fig. 2B, and Supplementary Fig. S1A and B), the results reveal an organization principle at CREs favoring the accessibility of DNA for transcriptional (co)factors and the resulting acetylation of histone-tails can further increase their accessibility for other regulatory proteins. The contrasting chromatin organization at CREs and elsewhere reflects the preferential localization of CREs into kilobase-scale A compartments, observed in the recent *in situ* Hi-C studies [78]. We hypothe-

size that the recruited TFs and the subsequent histone modifications propagate to the proximal chromatin segments in a distance-dependent manner, as observed in the mesoscopic chromatin ensemble (Fig. 2D), resulting in the hierarchical organization of the transcription (co)factors (Fig. 6C and Supplementary Fig. S13C).

The population distribution of the clusters from mesoscopic simulations suggests that the cluster with the most CRE interactions, cluster 1, is also the most populated compared to the clusters with weak interactions among CREs (Fig. 2E). However, CREs show contact frequencies and distance reconfiguration comparable to the other mesoscopic beads, suggesting that the CRE interactions are comparable to non-specific chromatin interactions, unlike the strong association between CTCF-bound beads acting as topological restraints (Supplementary Figs S1D and E, S4J, and S5). This could possibly result from using a uniform bead radius for 1 kb chromatin segments in the HiC-metainference simulations, whereas in reality, CREs are likely to be relatively expanded. Based on the molecular model, we also observe that the expanded chromatin segments favor transient nonlocal inter-nucleosome interactions with other expanded segments (Fig. 7D–G and Supplementary Fig. S18). We propose that such transient nonlocal inter-nucleosome contacts—“Nucleosomal Handshake”—form the roadmap for *cis*-regulatory interactions. Overall, the detailed analysis of our 3D models suggests that the design principle for such a roadmap is encoded by the nucleosome organization along the genome (Fig. 7G).

The high-resolution *Nanog* model provides a realistic molecular-level picture of chromatin organization consistent with *in vivo* single-gene imaging studies [18]. The model is also an excellent starting point for MD simulations to study the molecular basis of interplay between chromatin organization and communication between CREs, as well as the mechanistic details of the molecular interactions at CREs [14, 18, 79, 80]. Past chromatin simulations established that the local nucleosome organization could sufficiently drive transient interactions between CREs and extensively studied the roles of NRLs, NDR, H1 association, and acetylation in local chromatin organization [26, 81–88]. More recently, nucleosome-resolution integrative modeling and simulations of several 50–100 kb gene loci integrated the Micro-C and nucleosome positioning data to generate *in situ* relevant chromatin fiber models [89]. In our work, we integrate the site-specific protein association data and an ensemble view of chromatin organization to reveal molecular organization principles consistent with *in vivo* experiments. Importantly, our near-atomic large-scale models of the *Nanog* locus bridge the existing understanding of local chromatin organization with that of large-scale 3D genome architecture, revealing the interplay between chromatin structure and transcription regulation. While much work has explored the effect of protein association and histone tail modifications on chromatin [72, 81, 83, 85, 87, 88, 90, 91], our work provides insights into the features driving the orchestrated protein association along the genome.

A current limitation of the molecular models of the *Nanog* locus presented here is that they are simplistic regarding the constituting protein factors and include only the most fundamental components required for chromatin organization and enhancer-mediated transcription regulation. Our current method does not yet include important protein complexes such as cohesin, chromatin remodelers, topoisomerases, several hundred other TFs, the complete epigenetic status of chro-

matin, and the effect of prevalent histone variants. However, the protocol demonstrated here can accommodate the additional details depending on the purpose of the study and the spatial motions accessible within the timescales of the simulations. For example, while the current pipeline only provides an integrated high-resolution view of chromatin organization in mESC, combined with complementing models in differentiated cells, it could serve as a starting point to predict key functional effects, such as cell type-dependent transcription regulation. Finally, the proposed model is a step toward building an experimentally consistent atomistic molecular model of eukaryotic cells, expanding previous advances to build cellular-scale models of cytoplasm and synaptic vesicles [92–94].

Acknowledgements

We thank Diego Ugarte La Torre, Computational Biophysics Research Team, RIKEN Center for Computational Science, Kobe, Hyogo, Japan, for his help in modeling the mouse pre-initiation complex.

Author contributions: Soundhararajan Gopi (Project administration, Data curation, Formal analysis, Methodology, Software, Validation, Resources, Writing—original draft), Giovanni B. Brandani (Data curation, Methodology, Software, Writing—review & editing), Cheng Tan (Methodology, Software, Writing—review & editing), Jaewoon Jung (Software, Writing—review & editing), Chenyang Gu (Methodology, Software, Writing—review & editing), Azuki Mizutani (Methodology, Writing—review & editing), Hiroshi Ochiai (Supervision, Writing—review & editing), Yuji Sugita (Supervision, Resources, Writing—review & editing), and Shoji Takada (Conceptualization, Supervision, Data curation, Methodology, Software, Funding acquisition, Resources, Writing—review & editing)

Supplementary data

Supplementary data is available at NAR online.

Conflict of interest

None declared.

Funding

This work was supported by the Ministry of Education, Culture, Sports, Science and Technology [JPMXP1020200101 to S.T. and Y.S.; JPMXP1020230119 to S.T.], and the Japan Society for the Promotion of Science KAKENHI grants [20H05934, 21H02441, and 24K01991 to S.T.]. This work used computational resources of the supercomputer Fugaku provided by RIKEN through the HPCI System Research Project [hp230095 to S.G.]. Funding to pay the Open Access publication charges for this article was provided by the Japan Society for the Promotion of Science KAKENHI grant [24K01991 to S.T.].

Data availability

All processed input files, Code to perform Hi-C metainference simulations, Hi-C metainference trajectory, all-atom and coarse-grained models of the fiber modules, 200 kb CG *Nanog* models, GENESIS CGDYN setup files, and modeling scripts

used in this study are included in the Zenodo repository <https://doi.org/10.5281/zenodo.13958906>.

References

- Banigan EJ, Tang W, van den Berg AA *et al.* Transcription shapes 3D chromatin organization by interacting with loop extrusion. *Proc Natl Acad Sci USA* 2023;120:e2210480120. <https://doi.org/10.1073/pnas.2210480120>
- Chacin E, Reusswig K-U, Furtmeier J *et al.* Establishment and function of chromatin organization at replication origins. *Nature* 2023;616:836–42. <https://doi.org/10.1038/s41586-023-05926-8>
- Dixon JR, Jung I, Selvaraj S *et al.* Chromatin architecture reorganization during stem cell differentiation. *Nature* 2015;518:331–6. <https://doi.org/10.1038/nature14222>
- Blinka S, Reimer MH, Pulakanti K *et al.* Super-enhancers at the nanog locus differentially regulate neighboring pluripotency-associated genes. *Cell Rep* 2016;17:19–28. <https://doi.org/10.1016/j.celrep.2016.09.002>
- Lieberman-Aiden E, van Berkum NL, Williams L *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the Human genome. *Science* (1979) 2009;326:289–93.
- Nora EP, Lajoie BR, Schulz EG *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 2012;485:381–5. <https://doi.org/10.1038/nature11049>
- Hsieh T-HS, Weiner A, Lajoie B *et al.* Mapping nucleosome resolution chromosome folding in yeast by micro-C. *Cell* 2015;162:108–19. <https://doi.org/10.1016/j.cell.2015.05.048>
- Hsieh T-HS, Cattoglio C, Slobodyanyuk E *et al.* Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol Cell* 2020;78:539–53. <https://doi.org/10.1016/j.molcel.2020.03.002>
- Batur PJ, Bing XY, Sisco Z *et al.* Genome organization controls transcriptional dynamics during development. *Science* (1979) 2022;375:566–70.
- Benabdallah NS, Williamson I, Illingworth RS *et al.* Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol Cell* 2019;76:473–84. <https://doi.org/10.1016/j.molcel.2019.07.038>
- Agrawal P, Rao S. Super-enhancers and CTCF in early embryonic cell fate decisions. *Front Cell Dev Biol* 2021;9:653669. <https://doi.org/10.3389/fcell.2021.653669>
- Lee R, Kang M-K, Kim Y-J *et al.* CTCF-mediated chromatin looping provides a topological framework for the formation of phase-separated transcriptional condensates. *Nucleic Acids Res* 2022;50:207–26. <https://doi.org/10.1093/nar/gkab1242>
- Alexander JM, Guan J, Li B *et al.* Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife* 2019;8:e41769. <https://doi.org/10.7554/eLife.41769>
- Sabari BR, Dall'Agnese A, Boija A *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 2018;361:eaar3958.
- Richter WF, Nayak S, Iwasa J *et al.* The mediator complex as a master regulator of transcription by RNA polymerase II. *Nat Rev Mol Cell Biol* 2022;23:732–49. <https://doi.org/10.1038/s41580-022-00498-3>
- Karr JP, Ferrie JJ, Tjian R *et al.* The transcription factor activity gradient (TAG) model: contemplating a contact-independent mechanism for enhancer-promoter communication. *Genes Dev* 2022;36:7–16. <https://doi.org/10.1101/gad.349160.121>
- Su J-H, Zheng P, Kinrot SS *et al.* Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* 2020;182:1641–59. <https://doi.org/10.1016/j.cell.2020.07.032>
- Li J, Dong A, Saydaminova K *et al.* Single-molecule nanoscopy elucidates RNA polymerase II transcription at single genes in live cells. *Cell* 2019;178:491–506. <https://doi.org/10.1016/j.cell.2019.05.029>
- Cao K, Lailier N, Zhang Y *et al.* High-resolution mapping of h1 linker histone variants in embryonic stem cells. *PLoS Genet* 2013;9:e1003417. <https://doi.org/10.1371/journal.pgen.1003417>
- Voong LN, Xi L, Sebeson AC *et al.* Insights into nucleosome organization in mouse embryonic stem cells through chemical mapping. *Cell* 2016;167:1555–70. <https://doi.org/10.1016/j.cell.2016.10.049>
- Chronis C, Fiziev P, Papp B *et al.* Cooperative binding of transcription factors orchestrates reprogramming. *Cell* 2017;168:442–59. <https://doi.org/10.1016/j.cell.2016.12.016>
- El Khattabi L, Zhao H, Kalchschmidt J *et al.* A pliable mediator acts as a functional rather than an architectural bridge between promoters and enhancers. *Cell* 2019;178:1145–58. <https://doi.org/10.1016/j.cell.2019.07.011>
- Chen X, Yin X, Li J *et al.* Structures of the human mediator and mediator-bound preinitiation complex. *Science* 2021;372:eabg0635.
- Huertas J, Woods EJ, Colleparado-Guevara R. Multiscale modelling of chromatin organisation: resolving nucleosomes at near-atomistic resolution inside genes. *Curr Opin Cell Biol* 2022;75:102067. <https://doi.org/10.1016/j.celb.2022.02.001>
- Blinka S, Rao S. Nanog expression in embryonic stem cells—an ideal model system to dissect enhancer function. *Bioessays* 2017;39:1700086. <https://doi.org/10.1002/bies.201700086>
- Brandani GB, Gu C, Gopi S *et al.* Multiscale bayesian simulations reveal functional chromatin condensation of gene loci. *PNAS Nexus* 2024;3:pgae226. <https://doi.org/10.1093/pnasnexus/pgae226>
- Robinson JT, Turner D, Durand NC *et al.* Juicebox.js provides a cloud-based visualization system for Hi-C data. *Cell Syst* 2018;6:256–8. <https://doi.org/10.1016/j.cels.2018.01.001>
- Bonomi M, Camilloni C, Cavalli A *et al.* MetaInference: a bayesian inference method for heterogeneous systems. *Sci Adv* 2016;2:e1501177. <https://doi.org/10.1126/sciadv.1501177>
- Lequieu J, Córdoba A, Moller J *et al.* 1CPN: a coarse-grained multi-scale model of chromatin. *J Chem Phys* 2019;150:215102. <https://doi.org/10.1063/1.5092976>
- Naritomi Y, Fuchigami S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J Chem Phys* 2011;134:065101. <https://doi.org/10.1063/1.3554380>
- Scherer MK, Trendelkamp-Schroer B, Paul F *et al.* PyEMMA 2: a software package for estimation, validation, and analysis of Markov models. *J Chem Theory Comput* 2015;11:5525–42. <https://doi.org/10.1021/acs.jctc.5b00743>
- Lee BT, Barber GP, Benet-Pagès A *et al.* The UCSC Genome Browser database: 2022 update. *Nucleic Acids Res* 2022;50:D1115–22. <https://doi.org/10.1093/nar/gkab959>
- Neph S, Kuehn MS, Reynolds AP *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* 2012;28:1919–20. <https://doi.org/10.1093/bioinformatics/bts277>
- Williams LH, Fromm G, Gokey NG *et al.* Pausing of RNA polymerase II regulates mammalian developmental potential through control of signaling networks. *Mol Cell* 2015;58:311–22. <https://doi.org/10.1016/j.molcel.2015.02.003>
- Zhang Y, Liu Z, Medrzycki M *et al.* Reduction of hox gene expression by histone H1 depletion. *PLoS One* 2012;7:e38829. <https://doi.org/10.1371/journal.pone.0038829>
- Chen J, Zhang Z, Li L *et al.* Single-molecule dynamics of enhancosome assembly in embryonic stem cells. *Cell* 2014;156:1274–85. <https://doi.org/10.1016/j.cell.2014.01.062>
- Xie L, Torigoe SE, Xiao J *et al.* A dynamic interplay of enhancer elements regulates *Klf4* expression in naïve pluripotency. *Genes Dev* 2017;31:1795–808. <https://doi.org/10.1101/gad.303321.117>
- Webb B, Sali A. Comparative protein structure modeling using MODELLER. *Curr Protoc Bioinform* 2016;54:1–37. <https://doi.org/10.1002/cpbi.3>
- Palacio M, Taatjes DJ. Merging established mechanisms with new insights: condensates, hubs, and the regulation of RNA

- polymerase II transcription. *J Mol Biol* 2022;434:167216. <https://doi.org/10.1016/j.jmb.2021.167216>
40. Wang X, Zhou X, Yan Q et al. LLPSDB v2.0: an updated database of proteins undergoing liquid–liquid phase separation *in vitro*. *Bioinformatics* 2022;38:2010–4. <https://doi.org/10.1093/bioinformatics/btac026>
 41. Jumper J, Evans R, Pritzel A et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596:583–9. <https://doi.org/10.1038/s41586-021-03819-2>
 42. Varadi M, Anyango S, Deshpande M et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* 2022;50:D439–44. <https://doi.org/10.1093/nar/gkab1061>
 43. Tubiana L, Polles G, Orlandini E et al. KymoKnot: a web server and software package to identify and locate knots in trajectories of linear or circular polymers. *Eur Phys J E* 2018;41:72. <https://doi.org/10.1140/epje/i2018-11681-0>
 44. Dabrowski-Tumanski P, Rubach P, Niemyska W et al. Topoly: python package to analyze topology of polymers. *Brief Bioinform* 2021;22:bbaa196. <https://doi.org/10.1093/bib/bbaa196>
 45. Arya G, Zhang Q, Schlick T. Flexible histone tails in a new mesoscopic oligonucleosome model. *Biophys J* 2006;91:133–50. <https://doi.org/10.1529/biophysj.106.083006>
 46. Tan C, Jung J, Kobayashi C et al. Implementation of residue-level coarse-grained models in GENESIS for large-scale molecular dynamics simulations. *PLoS Comput Biol* 2022;18:e1009578. <https://doi.org/10.1371/journal.pcbi.1009578>
 47. Jung J, Tan C, Sugita Y. GENESIS CGDYN: large-scale coarse-grained MD simulation with dynamic load balancing for heterogeneous biomolecular systems. *Nat Commun* 2024;15:3370. <https://doi.org/10.1038/s41467-024-47654-1>
 48. Li W, Terakawa T, Wang W et al. Energy landscape and multiroute folding of topologically complex proteins adenylate kinase and 2ouf-knot. *Proc Natl Acad Sci USA* 2012;109:17789–94. <https://doi.org/10.1073/pnas.1201807109>
 49. Terakawa T, Takada S. Multiscale ensemble modeling of intrinsically disordered proteins: p53 N-terminal domain. *Biophys J* 2011;101:1450–8. <https://doi.org/10.1016/j.bpj.2011.08.003>
 50. Tesei G, Schulze TK, Crehuet R et al. Accurate model of liquid–liquid phase behavior of intrinsically disordered proteins from optimization of single-chain properties. *Proc Natl Acad Sci USA* 2021;118:e2111696118. <https://doi.org/10.1073/pnas.2111696118>
 51. Freeman GS, Hinckley DM, Lequieu JP et al. Coarse-grained modeling of DNA curvature. *J Chem Phys* 2014;141:165103. <https://doi.org/10.1063/1.4897649>
 52. Niina T, Brandani GB, Tan C et al. Sequence-dependent nucleosome sliding in rotation-coupled and uncoupled modes revealed by molecular simulations. *PLoS Comput Biol* 2017;13:e1005880. <https://doi.org/10.1371/journal.pcbi.1005880>
 53. Tan C, Takada S. Dynamic and structural modeling of the specificity in protein–DNA interactions guided by binding assay and structure data. *J Chem Theory Comput* 2018;14:3877–89. <https://doi.org/10.1021/acs.jctc.8b00299>
 54. Cheng RR, Contessoto VG, Lieberman Aiden E et al. Exploring chromosomal structural heterogeneity across multiple cell lines. *eLife* 2020;9:e60312. <https://doi.org/10.7554/eLife.60312>
 55. Zhou Y, Basu S, Laue E et al. Single cell studies of mouse embryonic stem cell (mESC) differentiation by electrical impedance measurements in a microfluidic device. *Biosens Bioelectron* 2016;81:249–58. <https://doi.org/10.1016/j.bios.2016.02.069>
 56. Grosberg A, Rabin Y, Havlin S et al. Crumpled globule model of the three-dimensional structure of DNA. *Europhys Lett* 1993;23:373–8. <https://doi.org/10.1209/0295-5075/23/5/012>
 57. Mirny LA. The fractal globule as a model of chromatin architecture in the cell. *Chromosome Res* 2011;19:37–51. <https://doi.org/10.1007/s10577-010-9177-0>
 58. Henninger JE, Oksuz O, Shrinivas K et al. RNA-mediated feedback control of transcriptional condensates. *Cell* 2021;184:207–25. <https://doi.org/10.1016/j.cell.2020.11.030>
 59. Fan Y, Nikitina T, Morin-Kensicki EM et al. H1 Linker histones are essential for mouse development and affect nucleosome spacing *in vivo*. *Mol Cell Biol* 2003;23:4559–72. <https://doi.org/10.1128/MCB.23.13.4559-4572.2003>
 60. Segal E, Fondufe-Mittendorf Y, Chen L et al. A genomic code for nucleosome positioning. *Nature* 2006;442:772–8. <https://doi.org/10.1038/nature04979>
 61. Lowary PT, Widom J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J Mol Biol* 1998;276:19–42. <https://doi.org/10.1006/jmbi.1997.1494>
 62. Anderson JD, Widom J. Poly(dA-dT) promoter elements increase the equilibrium accessibility of nucleosomal DNA target sites. *Mol Cell Biol* 2001;21:3830–9. <https://doi.org/10.1128/MCB.21.11.3830-3839.2001>
 63. Kaplan N, Moore IK, Fondufe-Mittendorf Y et al. The DNA-encoded nucleosome organization of a eukaryotic genome. *Nature* 2009;458:362–6. <https://doi.org/10.1038/nature07667>
 64. Yoo J, Park S, Maffeo C et al. DNA sequence and methylation prescribe the inside-out conformational dynamics and bending energetics of DNA minicircles. *Nucleic Acids Res* 2021;49:11459–75. <https://doi.org/10.1093/nar/gkab967>
 65. Filippakopoulos P, Picaud S, Mangos M et al. Histone recognition and large-scale structural analysis of the Human Bromodomain Family. *Cell* 2012;149:214–31. <https://doi.org/10.1016/j.cell.2012.02.013>
 66. Zhou Y-B, Gerchman SE, Ramakrishnan V et al. Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature* 1998;395:402–5. <https://doi.org/10.1038/26521>
 67. Devaiah BN, Case-Borden C, Geggion A et al. BRD4 is a histone acetyltransferase that evicts nucleosomes from chromatin. *Nat Struct Mol Biol* 2016;23:540–8. <https://doi.org/10.1038/nsmb.3228>
 68. Käs E, Izaurralde E, Laemmli UK. Specific inhibition of DNA binding to nuclear scaffolds and histone H1 by distamycin. *J Mol Biol* 1989;210:587–99. [https://doi.org/10.1016/0022-2836\(89\)90134-4](https://doi.org/10.1016/0022-2836(89)90134-4)
 69. Peng Y, Song W, Teif VB et al. Detection of new pioneer transcription factors as cell-type-specific nucleosome binders. *eLife* 2024;12:RP88936. <https://doi.org/10.7554/eLife.88936.4>
 70. Tan C, Takada S. Nucleosome allostery in pioneer transcription factor binding. *Proc Natl Acad Sci USA* 2020;117:20586–96. <https://doi.org/10.1073/pnas.2005500117>
 71. Bjarnason S, McIvor JAP, Prestel A et al. DNA binding redistributes activation domain ensemble and accessibility in pioneer factor Sox2. *Nat Commun* 2024;15:1445. <https://doi.org/10.1038/s41467-024-45847-2>
 72. Correll SJ, Schubert MH, Grigoryev SA. Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J* 2012;31:2416–26. <https://doi.org/10.1038/emboj.2012.80>
 73. Song F, Chen P, Sun D et al. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* (1979) 2014;344:376–80.
 74. Hou Z, Nightingale F, Zhu Y et al. Structure of native chromatin fibres revealed by Cryo-ET *in situ*. *Nat Commun* 2023;14:6324. <https://doi.org/10.1038/s41467-023-42072-1>
 75. Gibson BA, Doolittle LK, Schneider MWG et al. Organization of chromatin by intrinsic and regulated phase separation. *Cell* 2019;179:470–84. <https://doi.org/10.1016/j.cell.2019.08.037>
 76. Farr SE, Woods EJ, Joseph JA et al. Nucleosome plasticity is a critical element of chromatin liquid–liquid phase separation and multivalent nucleosome interactions. *Nat Commun* 2021;12:2883. <https://doi.org/10.1038/s41467-021-23090-3>
 77. Heidarsson PO, Mercadante D, Sottini A et al. Release of linker histone from the nucleosome driven by polyelectrolyte competition

- with a disordered protein. *Nat Chem* 2022;14:224–31. <https://doi.org/10.1038/s41557-021-00839-3>
78. Harris HL, Gu H, Olshansky M *et al*. Chromatin alternates between A and B compartments at kilobase scale for subgenic organization. *Nat Commun* 2023;14:3303. <https://doi.org/10.1038/s41467-023-38429-1>
 79. Kagey MH, Newman JJ, Bilodeau S *et al*. Mediator and cohesin connect gene expression and chromatin architecture. *Nature* 2010;467:430–5. <https://doi.org/10.1038/nature09380>
 80. Shrinivas K, Sabari BR, Coffey EL *et al*. Enhancer features that drive formation of transcriptional condensates. *Mol Cell* 2019;75:549–61. <https://doi.org/10.1016/j.molcel.2019.07.009>
 81. Chang L, Takada S. Histone acetylation dependent energy landscapes in tri-nucleosome revealed by residue-resolved molecular simulations. *Sci Rep* 2016;6:34441.
 82. Kenzaki H, Takada S. Linker DNA length is a key to tri-nucleosome folding. *J Mol Biol* 2021;433:166792. <https://doi.org/10.1016/j.jmb.2020.166792>
 83. Bascom GD, Schlick T. Chromatin fiber folding directed by cooperative histone tail acetylation and linker histone binding. *Biophys J* 2018;114:2376–85. <https://doi.org/10.1016/j.bpj.2018.03.008>
 84. Perišić O, Colleparado-Guevara R, Schlick T. Modeling studies of chromatin fiber structure as a function of DNA linker length. *J Mol Biol* 2010;403:777–802. <https://doi.org/10.1016/j.jmb.2010.07.057>
 85. Bascom GD, Myers CG, Schlick T. Mesoscale modeling reveals formation of an epigenetically driven HOXC gene hub. *Proc Natl Acad Sci USA* 2019;116:4955–62. <https://doi.org/10.1073/pnas.1816424116>
 86. Sridhar A, Farr SE, Portella G *et al*. Emergence of chromatin hierarchical loops from protein disorder and nucleosome asymmetry. *Proc Natl Acad Sci USA* 2020;117:7216–24. <https://doi.org/10.1073/pnas.1910044117>
 87. Portillo-Ledesma S, Tsao LH, Wagley M *et al*. Nucleosome clutches are regulated by chromatin internal parameters. *J Mol Biol* 2021;433:166701. <https://doi.org/10.1016/j.jmb.2020.11.001>
 88. Forte G, Buckle A, Boyle S *et al*. Transcription modulates chromatin dynamics and locus configuration sampling. *Nat Struct Mol Biol* 2023;30:1275–85. <https://doi.org/10.1038/s41594-023-01059-8>
 89. Li Z, Schlick T. Hi-BDiSCO: folding 3D mesoscale genome structures from Hi-C data using brownian dynamics. *Nucleic Acids Res* 2024;52:583–99. <https://doi.org/10.1093/nar/gkad1121>
 90. Colleparado-Guevara R, Portella G, Vendruscolo M *et al*. Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: a multiscale computational study. *J Am Chem Soc* 2015;137:10205–15. <https://doi.org/10.1021/jacs.5b04086>
 91. Watanabe S, Mishima Y, Shimizu M *et al*. Interactions of HP1 bound to H3K9me3 dinucleosome by molecular simulations and biochemical assays. *Biophys J* 2018;114:2336–51. <https://doi.org/10.1016/j.bpj.2018.03.025>
 92. McGuffee SR, Elcock AH. Diffusion, crowding and protein stability in a dynamic molecular model of the bacterial cytoplasm. *PLoS Comput Biol* 2010;6:e1000694.
 93. Wilhelm BG, Mandad S, Truckenbrodt S *et al*. Composition of isolated synaptic boutons reveals the amounts of vesicle trafficking proteins. *Science* 2014;344:1023–8.
 94. Feig M, Harada R, Mori T *et al*. Complete atomistic model of a bacterial cytoplasm for integrating physics, biochemistry, and systems biology. *J Mol Graphics Model* 2015;58:1–9. <https://doi.org/10.1016/j.jmgm.2015.02.004>