Mesoscale modeling reveals formation of an epigenetically driven HOXC gene hub

Gavin D. Bascom[a], Christopher G. Myers[a], and Tamar Schlick[a,b,c,1]

[a]Department of Chemistry, New York University, New York, NY 10003; [b]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; and [c]New York University-East China Normal University Center for Computational Chemistry, New York University Shanghai, 200122 Shanghai, China

Gene expression is orchestrated at the structural level by nucleosome positioning, histone tail acetylation, and linker histone (LH) binding. Here, we integrate available data on nucleosome positioning, nucleosome-free regions (NFRs), acetylation islands, and LH binding sites to "fold" in silico the 55-kb HOXC gene cluster and investigate the role of each feature on the gene's folding. The gene cluster spontaneously forms a dynamic connection hub, characterized by hierarchical loops which accommodate multiple contacts simultaneously and decrease the average distance between promoters by ∼100 nm. Contact probability matrices exhibit "stripes" near promoter regions, a feature associated with transcriptional regulation. Interestingly, while LH proteins alone decrease long-range contacts and acetylation alone increases transient contacts, combined LH and acetylation produce long-range contacts. Thus, our work emphasizes how chromatin architecture is coordinated strongly by epigenetic factors and opens the way for nucleosome resolution models incorporating epigenetic modifications to understand and predict gene activity.

chromatin modeling | chromatin folding | gene structure | chromatin loop domains | contact hub

Developmental regulation in eukaryotic organisms depends on chromatin fibers looping across large genomic distances spanning kilobases to megabases (1). Such looping depends on epigenetic factors that are tightly regulated during development and differentiation stages of the cell (2, 3). Chromatin fibers are made up of nucleosomes, where each nucleosome consists of ∼147 bp of DNA wrapped tightly around eight core histone proteins (two copies each of H2A, H2B, H3, and H4) (4–6). Each histone protein has flexible N-terminal tail domains that extend away from the nucleosome into the surrounding solvent (7). Among many epigenetic chromatin modifications, acetylation of histone tails, nucleosome positioning, and linker histone (LH) binding are known to strongly affect chromatin structure.

Histone tail acetylation—the addition of acetyl groups to the lysine residues within N-terminal tails—is one of many reversible, posttranslational modifications (PTMs) that modulate chromatin architecture and play key roles in regulating gene expression. The modified residues can participate in hydrogen bonding, promoting more stably folded secondary structures of the respective tail (9–11). Acetylation of lysine 16 on the H4 tail (H4K16ac) decreases kilobase-range contacts and is associated with linear fiber decondensation (i.e., fiber lengthening along its principal axis) (10, 12). Acetylation in living chromatin is often found in small islands that range from 1 to 5 kb (13).

Nucleosome positions, often measured as nucleosome repeat length (NRL), or the amount of DNA wrapped around the parent nucleosome (147 bp) plus the linker length, are nonuniform in eukaryotic fibers; in mouse embryonic stem (mES) cells, NRL follows a distribution that can be approximated by positions $10n + 5$ (where $n = 0, 1, 2 \ldots 9$), with $n = 4$ or 5 most common (14). Regions near the 5′ end of each gene-encoding region, commonly referred to as transcription start sites (TSSs), are gen-

erally depleted of nucleosomes (also known as nucleosome-free regions, or NFRs) (15). Studies of chromatin fibers reconstituted in vitro and mesoscale modeling show that DNA linker lengths can strongly affect the structure of chromatin fibers, where short linkers (i.e., NRL 162–182 bp) are associated with stiff, straight fibers, and longer linkers are associated with more disordered globules (16, 17).

Besides tail acetylation and nonuniform nucleosome positions, the LH protein strongly affects fiber secondary structure through binding DNA dynamically at various concentrations, on or off the dyad axis formed by entering and exiting DNA (18–24), leading to significant linear compaction (i.e., a shortening of the fiber along its principal axis) (25). Each LH contains a small, uncharged N-terminal domain, a globular head (GH) that binds nucleosomal DNA (19), and a long, positively charged and intrinsically disordered C-terminal domain (CTD) that binds to the entering or exiting linker DNA (26). Chromatin immunoprecipitation (ChIP) combined with sequencing assays (ChIP-Seq) shows that LH localizes to specific regions of chromatin fibers, often anticorrelated with specific PTMs such as acetylation and gene-encoding regions (27). LH knockout studies in mice have shown that while LH silences many genes, it is also necessary to maintain active transcription of some genes (28). LH also has roles in development regulation and diseases (29, 30).

Changes in epigenetic features are part of the regulation of development (31), differentiation (32), and cancer progression (33). Information regarding epigenetic features as a function of genome position is publicly available for many organisms and cell lines [e.g., ENCODE (34)], but combining these features into a unified structural model is not straightforward. Here, we integrate all available data to build and "fold" a mesoscale

---

Significance

The precise role of epigenetic factors in controlling chromatin structure is not well understood. Here, we use publicly available data to build and "fold" a nucleosome-resolution mesoscale model of the developmentally regulated HOXC gene locus that incorporates histone tail acetylation, linker histone binding, and nucleosome occupancy. We discover a spontaneous contact hub that bridges promoters and exon/intron junctions. Our work emphasizes how epigenetic factors are coordinated to influence chromatin architecture and opens the way for nucleosome resolution modeling of epigenetically regulated genes.

---

chromatin model of the 55-kb HOXC gene cluster and systematically investigate the effect of LH, acetylation, NFRs, and DNA linker length distributions on chromatin folding through reference systems that contain different subsets of the gene cluster's epigenetic features. Our model of the HOXC gene locus includes five genes (containing 11 exons), 284 nucleosomes, five acetylation islands, and 55 kb of DNA.

We find that epigenetically modified chromatin (with LH, acetylation, and NFRs) in this context cooperates to compact a gene hub, forming a series of densely packaged hierarchical loops. This increased density is associated with a dynamic kilobase-range contact hub that bridges promoters within the locus, specifically bringing into contact an LH-rich and acetylation-rich region, as well as "stripe" regions in the nucleosome contact maps, indicating sampling around promoter regions. Our results indicate that nucleosome resolution models that explicitly incorporate epigenetic features derived from experimental data can explain and help predict the genetic regulation of various gene loci as a function of epigenetic modifications.

Results

In addition to the HOXC gene model shown in Fig. 1 *B* and *C*, which contains life-like linker length distribution with LH, acetylation islands, and NFRs, we define five other reference systems that contain a subset of the HOXC gene cluster's factors: **Uniform, Uniform+NFR, Life-Like, Life-Like+Ac**, and **Life-Like+LH** (*Materials and Methods* and *SI Appendix*). The **Uniform** fiber has uniform linker lengths without NFRs. The **Uniform+NFR** fiber has uniform linker lengths and NFRs. **Life-Like** fibers have a distribution of linker lengths modeled

after living systems. We use this term as in prior work (35) to distinguish these from fibers that have uniform or heterogeneous nucleosome positions. **Life-Like+Ac** fibers have the same life-like linker length distribution plus acetylation islands with acetylated histone tails. Finally, **Life-Like+LH** fibers contain a life-like linker length distribution with LH placed in intergenic regions.

Contact Probability Matrices Show Increased Contact Densities in HOXC Fibers. In Fig. 2, we show the 2D contact probability matrix for the **HOXC** system compared with **Life-Like+Ac** and **Life-Like** reference systems. **HOXC** contact maps show a significant increase in contact density and coverage, with two discernible contact domains: LH-rich (top left of contact map) and acetylation-rich region (bottom right of contact map). The latter region is more dense than the former. In fact, stripes or lines running horizontally or vertically, originating near NFRs, demarcate these two subdomains (also see *SI Appendix*, Fig. S1). Two stripes originating at the HOXC8 promoter indicate contact between this promoter and all other portions of the fiber. This is also evident in a fiber image shown in Fig. 2. The **Life-Like+Ac** (Fig. 2*B*) and **Life-Like** (Fig. 2*C*) reference fibers show homogeneously distributed contacts where acetylation leads to increased contacts of low intensity, but do not form stripes, nor do they show the separation of two domains. **Life-Like** fibers show the least amount of long-range contacts, with a uniformly distributed contact map.

The epigenetic features also emerge from a close-up of the 30- to 55-kb region (acetylation-rich) (*SI Appendix*, Fig. S2). This subregion shows increased contacts bridging HOXC5 and HOXC6 genes within the acetylated region. Contact matrices for

Fig. 1. Mesoscale model and HOXC system setup. (*A*) The repeating unit of the mesoscale model consists of a rigid nucleosome core with wrapped DNA represented by 300 point charges and beads for linker DNA, LH, and the flexible histone tails. See details in ref. 8. NCP, nucleosome core particle. (*B*) A rendering of the HOXC gene cluster model demonstrates gene positions and epigenetic elements in a 3D model. Acetylated tails are red, and WT tails are blue. Intron DNA is drawn in dark blue, exon DNA is colored light blue, and intergenic DNA is dark red. (*C*) DNA linker lengths used in the HOXC model are plotted against the nucleosome index, along with annotated gene locations and epigenetic features, with acetylation islands in red, LH in teal, and NFRs in green.
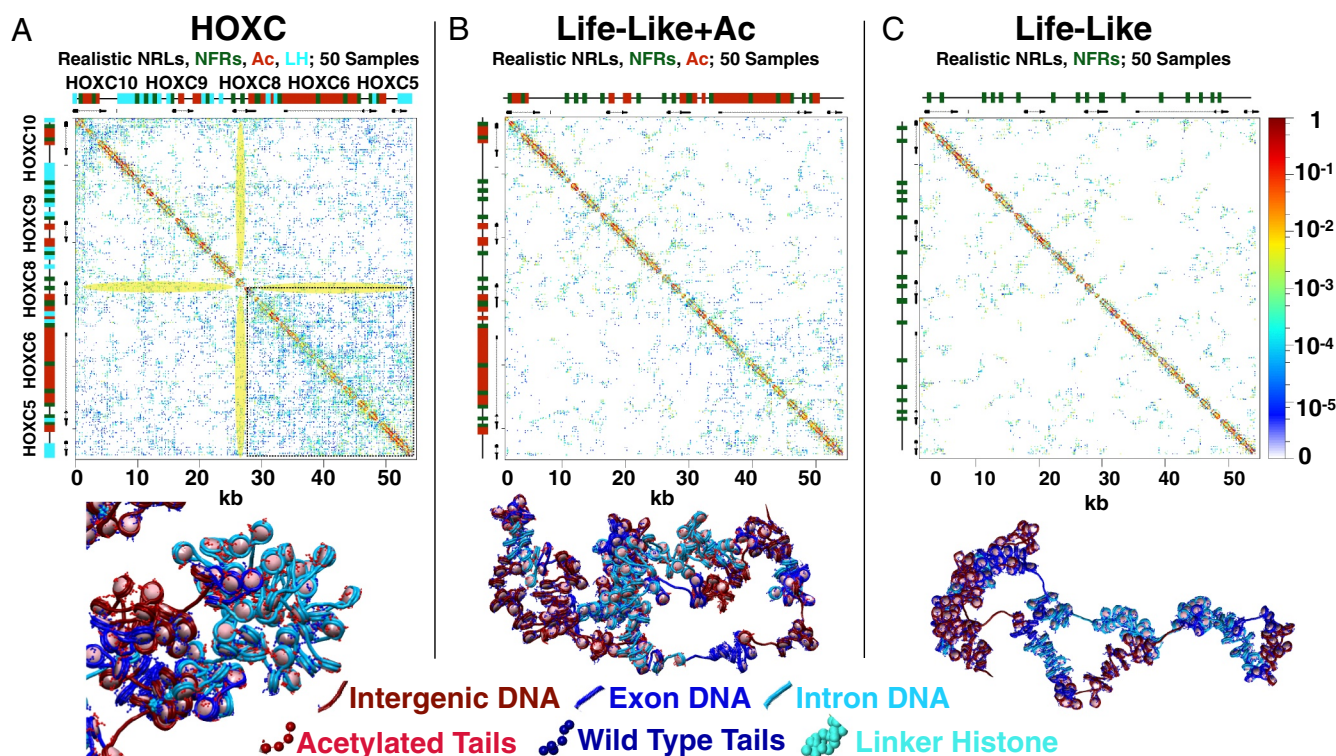
Fig. 2. Fiber renderings and contact probability matrices for **HOXC** and reference fibers. Black bars (top and sides) show the positions of HOXC genes. (*A*) HOXC system. (*B*) Reference system with life-like nucleosome positions, NFRs, and acetylation marks. (*C*) Reference system with life-like nucleosome positions and NFRs.

additional reference systems are shown in *SI Appendix*, Figs. S4 and S5.

HOXC Gene System Forms Dense Globules with Extensive Hierarchical Looping. Previously, we have shown that compact chromatin fibers at metaphase tend to fold into hierarchical loops, leading to a compact 3D structure where loops stack in space like rope flaking (8, 36, 37). While we showed that these loops can reproduce contacts observed in chromatin conformation capture (3C) experiments (8), we had not included the effects of life-like nucleosome placement, acetylation, NFRs, or LH.

Fig. 3 shows fiber renderings of all systems studied, with average volumes, radii of gyration (R_g^2), and sedimentation coefficients ($S_{w,20}$). **HOXC** fibers form the most dense globules with extensive hierarchical looping (8). **Uniform** and **Uniform+NFR** fibers display similarly condensed fibers with extensive hierarchical looping, where NFRs increase the amount of DNA exposed to solvent. **Life-Like+LH** fibers behave similarly to **Life-Like** fibers, with some increased linear compaction in LH-rich regions and LH/NFR looping, but LH does not significantly affect the overall volume when compared with **Life-Like** fibers. **Life-Like+Ac** fibers form slightly more compact globules that **Life-Like** fibers, but occupy a significantly larger volume than **HOXC** fibers. **Life-Like** fibers, which contain ∼20% short linkers (where NRL = 173 bp), produce stiff fiber segments near NFRs, forming nucleosome "clutches" and intraclutch short-range interactions similar to those observed in mES cells (38). These fibers produce the largest volume of those studied (Fig. 3).

Acetylation and LH Cooperate to Form a Dynamic Hub that Bridges Gene Promoters. To demonstrate the cooperative nature of LH and acetylation in forming kilobase-range contacts and illustrate the various folding motifs, we analyzed the total interaction prob-

ability as a function of genome position in Fig. 4. The largest peak corresponds to next-neighboring nucleosomes at ∼250 bp, indicative of zigzag fiber folding. Experimental evidence for zigzag morphology has been observed in situ in human cells via electron microscopy-assisted nucleosome interaction capture (EMANIC) (36) and genome-wide probing of short-range inter-nucleosomal interactions with ionizing radiation (39), as well as in situ yeast chromatin via Micro-C data (40). Several near-neighbor nucleosome peaks in the 1- to 5-kb range arise from intraclutch interactions, observed in living cells by EMANIC (36). Next, contacts in the 5- to 10-kb range arise from local looping and near-neighbor clutch interactions, and interactions in the 10- to 20-kb range arise from large loops or hierarchical loops. All contacts are clearly elevated for **HOXC** fibers (green curve) compared with the reference fibers. **Life-Like** fibers (with no LH or acetylation; gray) show the least amount of contacts in the 1- to 20-kb range.

Each Epigenetic Feature Imparts Distinct Aggregation Tendencies to the Fiber. To further distinguish the nature of contacts observed in contact maps, we plotted the contact count of six distinct contact types as a function of genomic position. Fig. 5*A* shows the normalized contact count observed in **HOXC** fibers between NFRs and LH regions (NFR/LH), NFRs and acetylation regions (NFR/Ac), LH regions (LH/LH), acetylation and LH regions (Ac/LH), and acetylation/acetylation regions (Ac/Ac). Interactions involving unmodified elements (WT) make up the majority of contacts within the hub. Among epigenetically modified elements, however, the strongest interactions are between LH and LH-dense regions, forming loops between intergenic regions. The second-strongest loop type formed is between acetylation and LH regions, particularly in the intergenic region between HOXC8 and HOXC6 genes, where acetylation islands are found immediately on either side of the LH.

A

HOXC　　**Uniform**　　**Uniform+NFR**　**Life-Like+Ac**　**Life-Like+LH**　**Life-Like**
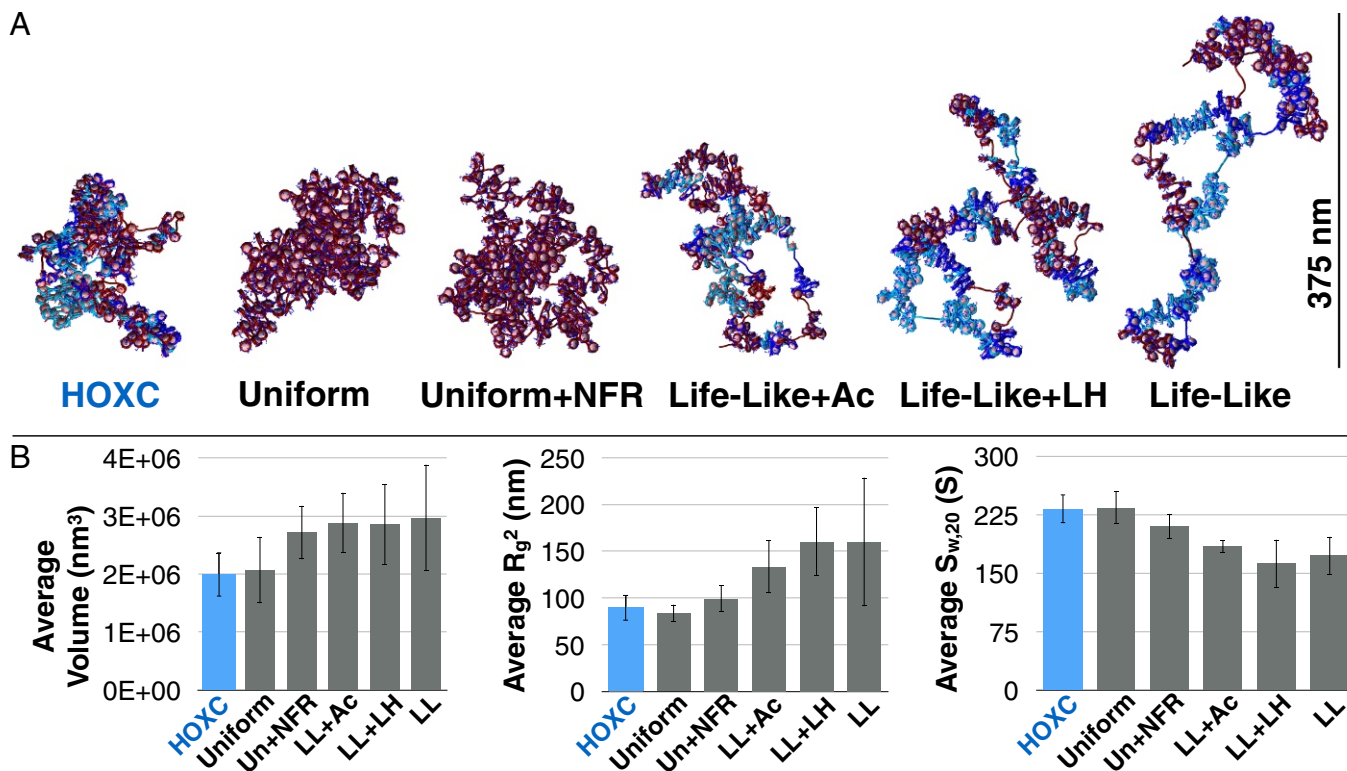
375 nm

B

Fig. 3. Epigenetic factors control fiber morphology and density. (*A*) Fiber renderings of **HOXC**, **Uniform**, **Uniform+NFR**, **Life-Like+Ac**, **Life-Like+LH**, and **Life-Like** systems. (*B*) The volume, R_g^2, and $S_{w,20}$ measurements as averaged across all 50 trajectories.

There are also moderate interactions between acetylation regions and one NFR in the region between HOXC9 and HOXC10, where an LH-dense region is immediately surrounded by NFRs. Finally, acetylation islands form weakly interacting regions near most of the TSSs. Regions with high amounts of acetylation and nucleosome depletion show some weak Ac/NFR interactions, particularly along the HOXC6 exon region. NFRs interacting with NFRs define the weakest form of epigenetic interaction.

The Contact Hub Decreases Average Distance Between Promoters. To characterize the nature of connections between gene pairs, we calculated the Cartesian distance between pairs of +1 nucleosome cores flanking TSSs between genes in the system. Fig. 5*B* shows the average pairwise distance counts between all promoter pairs for **HOXC** fibers, **Life-Like** fibers, and **Life-Like+Ac** fibers, with a rendering of the DNA and promoter regions for each system. The average distance between gene promoters decreased from ~150 to ~50 nm, indicating the increased proximity between these sites and the formation of a contact hub.

Hierarchical Looping Allows for Many Dynamic Connections Within the Hub. Finally, we show in Fig. 6 that the strongly interacting regions of the connection hub within **HOXC** fibers are dynamic, where hierarchical looping leads to the formation of multiple contacts simultaneously. Fig. 6*A* shows an example where an individual folded HOXC gene fiber forms four simultaneous contacts, involving intragene, contiguous-gene, and noncontiguous gene contacts. The corresponding contact matrix is annotated by contact locations. Note how this folding motif allows for nonexclusive contacts between neighboring pairs, next-neighbor pairs, or nonadjacent pairs of genes without tangling the overall fiber.

These dynamic connections can also be seen from Fig. 6*B*, which shows the distance between gene-promoter pairs for HOXC10/HOXC5, HOXC10/HOXC9, HOXC9/HOXC8, and HOXC9/HOXC5 genes as a function of simulation step, with three accompanying fiber snapshots. All four gene pairs, which represent both neighboring gene segments and gene pairs on the opposite side of the gene neighborhood, transiently form contacts where their distance is close to 50 nm, but no single pair forms a contact that persists throughout the simulation.

Discussion

It is well known that epigenetic features lead to distinct chromatin morphologies both in vitro (41) and in vivo (42), suggesting that active, euchromatin states may spontaneously segregate from inactive, heterochromatin states when modified. Our mesoscale model allows for nucleosome resolution study of the folding and formation of one such activation hub, providing unique insight into the mechanics of gene regulation.

Nucleosome Resolution Epigenetic Models Are Now Possible. Previously, we have shown that chromatin fibers spanning 10 kb can form hierarchical loops while maintaining zigzag topology, in good agreement with EMANIC cross-linking data of interphase and metaphase fibers in HeLa cells (36). We also showed that this effect can accommodate looping at the 80-kb scale found by 3C experiments at developmentally regulated genes such as the GATA4 gene locus (8). Although sophisticated models have been developed that use epigenetic profiles to predict global chromatin-folding features (43), computational models that represent epigenetic factors such as acetylation and LH explicitly are not available, to our knowledge. Our results indicate that epigenetic factors affect the kilobase-scale chromatin folding through modulating the formation and dynamics of hierarchical loops, providing a structural interpretation of the role of epigenetic
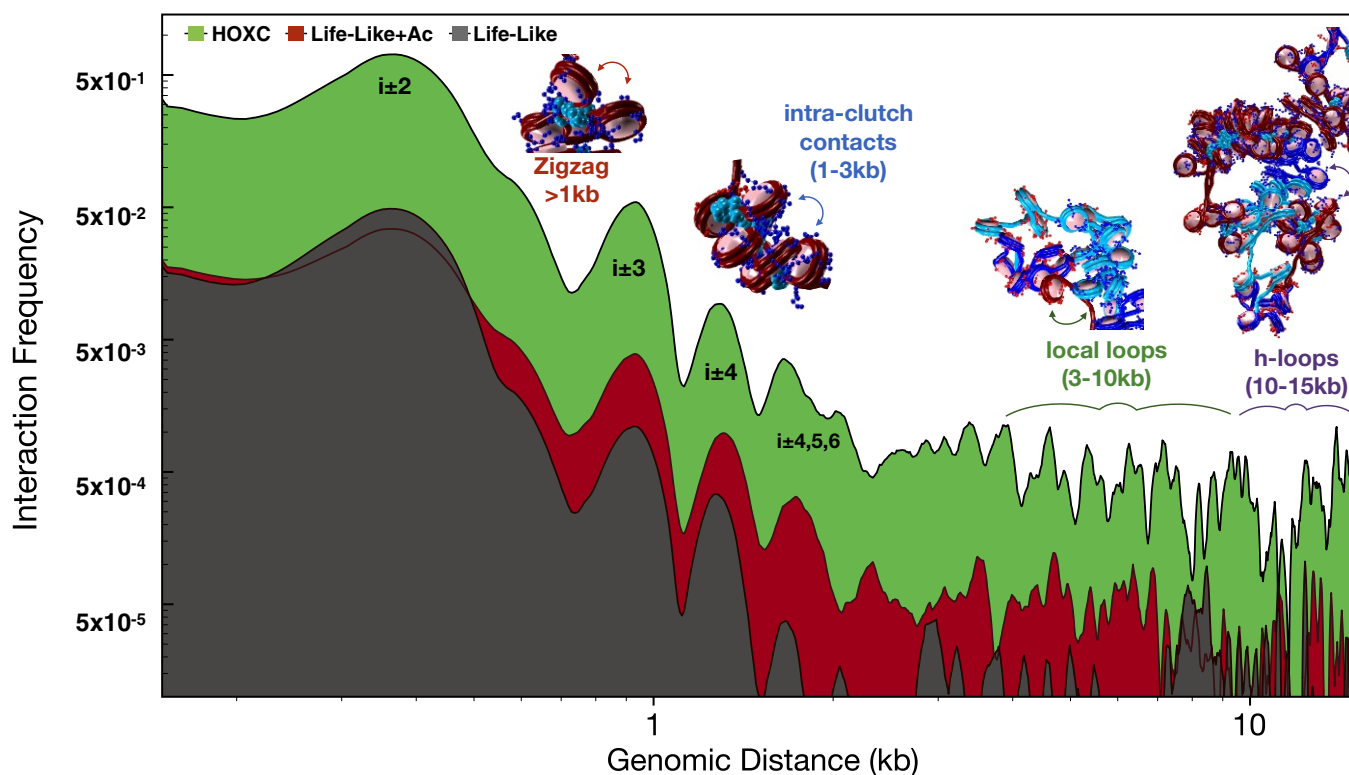
Fig. 4. Analysis of internucleosome interaction contacts for **HOXC** and reference systems. We annotate structural peaks as follows. Short-range contacts (>1 kb) measure next-neighbor interactions common in zigzag fibers. Contacts in the 1- to 2-kb range arise from intraclutch interactions. Chromatin loops between neighboring clutches account for 5- to 10-kb contacts, and hierarchical loops account for 10- to 20-kb contacts.

features on fiber contact morphology. Additionally, ChIP with paired-end tags data (44), circular chromosome conformation capture (4C), and a high-resolution Hi-C variant, Bridge Linker-Hi-C (BL-Hi-C), developed by Liang et al. (45) show looping among HOXC10, HOXC9, and HOXC6 genes. Such contacts are evident in our **HOXC** fibers.

Epigenetic Features Cooperate to Form a Dynamic Connection Hub Bridging Local Promoters. Our results suggest that each epigenetic feature imparts complex folding properties that cannot easily be generalized at all genomic scales. For example, nucleosome depletion at specific regions separates fiber clutches along the genomic sequence. Life-like linker length distributions (or nucleosome positions) further serve to separate neighboring clutches by stiffening individual clutches, but this stiffening comes at a cost, decreasing kilobase-range contacts and leading to a less dynamic fiber. Histone tail acetylation near such short linkers reduces this stiffness, leading to a more stochastic structure with transient contacts at long distances. LH cooperates with acetylation, acting to compact the total polymer by attracting NFRs neighboring acetylated regions, where the linker DNA polyelectrolyte is less effectively screened by strongly folded acetylated tails. Together, these features serve to condense the fiber into dynamic globules with volumes and densities similar to those observed experimentally, significantly increasing coverage of contacts in the 10- to 20-kb range.

Attraction between the LH- and acetylation-rich regions brings the HOXC8 promoter into contact with all other genes. We also see an NFR-rich region just downstream of HOXC10 forming strong contacts with HOXC9, HOXC8, and HOXC6 genes (Fig. 2). However, these features only become apparent in aggregate ensembles or long time scales. While we observed some simultaneous multiloop contacts in single structures

(Fig. 6A), the increased propensity to bridge multiple promoters arises on the ensemble level, reflecting the cumulative effects of transient contacts forming between promoters and other regions (Fig. 5A). Thus, stripe patterns in the contact maps may help explain how the cell facilitates loading and efficient diffusion of proteins, by forming strong connections between hubs while maintaining local sampling within the hub. This additionally helps explain the stochastic nature of gene expression observed in eukaryotes (46).

The hypothesis that epigenetic marks can act as boundary insulators is thus supported by our results. CHiP-Seq assays show that acetylation is sparse in the region downstream of the HOXC genes, which does not interact with the connection hub. Thus, it is plausible that heavily modified chromatin spontaneously separates from nearby unmodified chromatin. Our data indicate, however, that the boundaries resulting from such a mechanism are dynamic and only arise as a function of aggregate probability across dynamic samples, as opposed to precisely positioned boundary elements along the genome. Of course, other factors not modeled here, such as CTCF (and loop extrusion), long-noncoding RNAs, and transcription factors, likely play similarly complex roles.

Conclusion

A current attractive model for chromatin structure involves phase transitions (i.e., liquid to gas-like states) as a function of elevated transcription, explaining the spontaneous separation of active from inactive compartments, as seen in Hi-C experiments (47–49). However, as the majority of these models incorporate this assumption implicitly, it is important to develop nucleosome-resolution models that produce these patterns from basic physical variables in the living system. Here, we show that epigenetic features modeled based in a physical–chemical model can fold a gene, helping to interpret biological
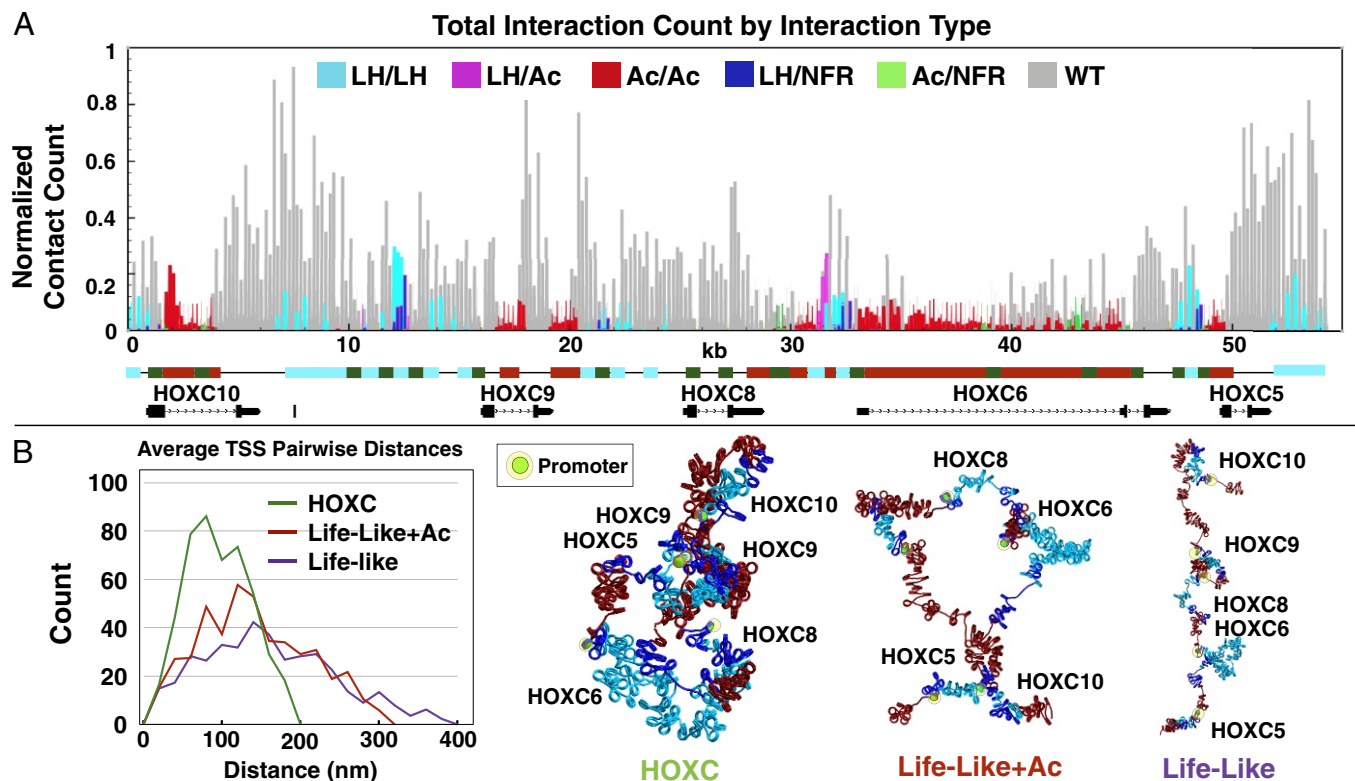
Fig. 5. Formation of dynamic contact hub depends on epigenetic features. (*A*) Number of contacts observed in **HOXC** fiber system between LH/LH (teal), LH/Ac (pink), Ac/Ac (red), LH/NFR (blue), and Ac/NFR (green), and any interaction that involves unmodified regions (WT; gray) as a function of genomic position. (*B*, *Left*) Plot of average pairwise distance count between gene promoters in the **HOXC** system vs. two reference systems. (*B*, *Right*) Renderings of chromatin configurations with promoter contacts (green/yellow highlight).

function of the HOXC gene locus. Specifically, LH, histone tail acetylation, and nucleosome placement cooperate to control the dynamic folding landscape of chromatin. In particular, that LH and acetylation reverse the trend that each factor imparts alone to promote and target kilobase-range contacts is both surprising and indicative of the complex orchestration of many factors on the structural and biophysical levels of the genome.

While our HOXC fibers provide a model system for studying the structure of genes in mammals, many more structural epigenetic features need to be investigated. In particular, the DNA CpG methylation, transcription factor binding, long noncoding RNAs, and CTCF factors missing from our current model likely play important roles in modulating structure. In addition, because epigenetic genomic data based on DNA sequencing generally represent ensembles which are averaged across heterogenous cell populations, and some nucleosome positions can have poor resolution depending on the methodology and ensemble averaging involved (50), more work on proper interpretation of such data for incorporation into structural models is needed. Both improvements in modeling and experiments and in their joint interpretation will be important for moving toward higher-resolution views of genes and genome architecture. Nonetheless, our current mesoscale approach represents a link between structural experiments and the functions of genes in living tissues, while retaining predictive power of the underlying models. As we have shown here, incorporating publicly available epigenetic profiles into in silico structural models makes it feasible now to fold into 3D and investigate the structure and function of various genes in model organisms or cell lines and, ultimately, to obtain a precise description of the dynamic 3D topology of the genomic code.

Materials and Methods

Mesoscale Model. Our mesoscale chromatin model (16, 51, 52) is composed of four distinct components: a modified worm-like chain polymer to represent linker DNA as beads, partially charged beads representing the electrostatically charged surface of the nucleosome determined by our Discrete Surface Charge Optimization algorithm (53), coarse-grained beads for histone tails, and coarse-grained beads for LH residues (see ref. 8 and *SI Appendix* for full details of the model and parameters).

Building the HOXC Gene Locus. To obtain realistic nucleosome positions, acetylation island positions, and LH distributions, we used various experimental data to model each parameter and produce the model of nucleosome positions, acetylation islands, and LH binding regions shown in Fig. 1 *B* and *C*. The University of California, Santa Cruz (UCSC) genome browser was used to annotate gene locations and acetylation peak locations (by H3K27ac Chip-Seq) (54). NFRs were modeled based on micrococcal nuclease sequencing (MNase-Seq) data of H9 embryonic stem cells from Yazdi et al. (55), downloaded from the Gene Expression Omnibus (accession code GSM1194221) (56) and visualized by using the Integrated Genome Browser (57). Using these data, we identified the locations of 17 NFRs, which we classified as being either "long" or "short." The former group has 4 NFRs with 162-bp linkers, and the latter has 13 NFRs with 108-bp linkers. To model realistic nucleosome positions, we used the distribution of linker lengths from chemical mapping in mES cells (14), where we placed nucleosomes such that the overall distribution observed by chemical mapping matched that of our model (Fig. 1*C*). LHs were placed to emulate trends seen in mES cells (27), where they are found in intergenic and nonacetylated regions only.

Simulation Parameters and Data Analysis. In addition to the HOXC gene model shown in Fig. 1 *B* and *C*, we defined five other reference systems of ∼55 kb in length that we called **Uniform, Uniform+NFR, Life-Like, Life-Like+LH,** and **Life-Like+Ac** (*SI Appendix*). These systems were designed to test for the effect of each parameter in comparison with the real system.
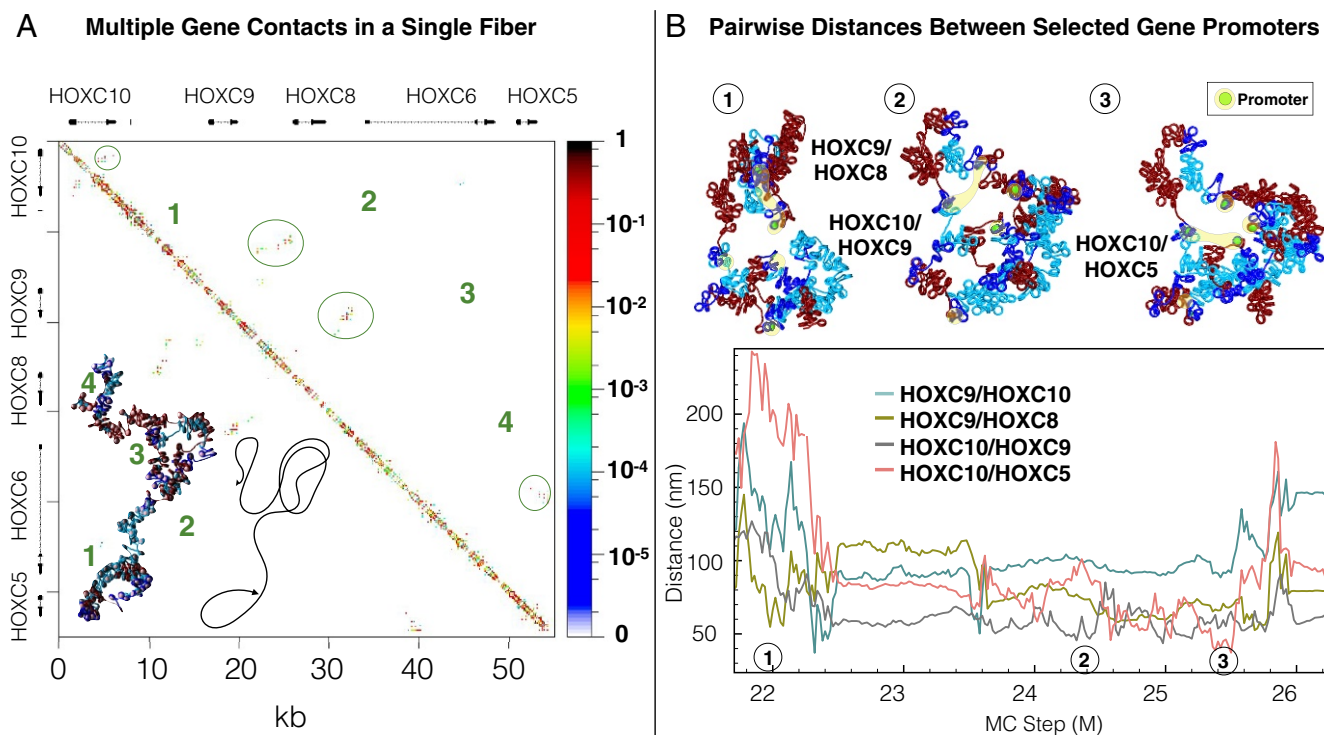
Fig. 6. Multiple dynamic contacts form within the hub. (*A*) Hierarchical looping accommodates four simultaneous loops in an individual fiber. One folded **HOXC** gene cluster fiber is shown with a sketch demonstrating the fiber folding and the genes and fiber-folding pattern and the corresponding contact probability matrix. (*B*) Cartesian distances between promoter regions of four gene pairs are plotted as a function of simulation length in a single trajectory. Fiber snapshots, with the promoters highlighted in green and yellow, demonstrate the formation of transient connections between promoters. MC, Monte Carlo.

The **Uniform** fiber system has a uniform linker length of 53 bp without NFRs, whereas the **Uniform+NFR** fiber system has a uniform linker length of 53 bp interrupted by NFRs, modeled either as a 108- or 612-bp linker length. The positions of these NFRs were placed as above, and the percentages of 108-bp vs. 162-bp NFRs were placed as above. **Life-Like** fibers have a distribution of linker lengths (*SI Appendix*) modeled after living mammalian genome nucleosome positions by chemical mapping with NFRs (14). **Life-Like+Ac** fibers have a life-like linker length distribution with acetylation peaks placed according to data taken from the UCSC genome browser (58). **Life-Like+LH** fibers contain a life-like linker length distribution with LH placed in intergenic regions. The positions of NFRs, acetylation, and LH are given for all reference systems in *SI Appendix*. An idealized zigzag starting structure was used for all simulations. See details in *SI Appendix*. Our LH model was based on the rat H1d LH structure predicted by Bharath et al. through fold recognition and molecular modeling (59). LH was rigidly fixed and placed on the dyad axis separated by a distance of 2.6 nm. We showed previously that, despite LH positioning in the dyad axis, the asymmetric structure of the GH and nucleosome interaction leads to an asymmetric chromatosome organization (60).

All systems were simulated for 60 million Monte Carlo steps or more. The Uniform, Uniform+NFR, and Life-Like+LH systems were simulated with 10 independent trajectories, and the other systems were simulated with 50 independent trajectories. Contact maps were normalized across each trajectory for each system and summed for the final contact map. Volume measurements were calculated from a convex bounding surface enclosing all elements of the fiber. A sample surface is shown in *SI Appendix*, Fig. S3. For a detailed description of calculating fiber volume, radius of gyration (R_g^2), and sedimentation coefficient ($S_{w,20}$), see *SI Appendix*.

1. Vernimmen D, De Gobbi M, Sloane-Stanley JA, Wood WG, Higgs DR (2007) Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression. *EMBO J* 26:2041–2051.
2. Heintzman ND, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Gen* 39:311–318.
3. Ong C-T, Corces VG (2011) Enhancer function: New insights into the regulation of tissue-specific gene expression. *Nat Rev Gen* 12:283–293.
4. Finch JT, et al. (1977) Structure of nucleosome core particles of chromatin. *Nature* 269:29–36.
5. Kornberg RD, Lorch Y (1999) Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome. *Cell* 98:285–294.
6. Davey CA, Sargent DF, Luger K, Maeder AW, Richmond TJ (2002) Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J Mol Biol* 319:1097–1113.
7. Luger K, Richmond TJ (1998) The histone tails of the nucleosome. *Curr Opin Gen Dev* 8:140–146.
8. Bascom GD, Sanbonmatsu KY, Schlick T (2016) Mesoscale modeling reveals hierarchical looping of chromatin fibers near gene regulatory elements. *J Phys Chem B* 120:8642–8653.

9. Potoyan DA, Papoian GA (2011) Energy landscape analyses of disordered histone tails reveal special organization of their conformational dynamics. *J Am Chem Soc* 133:7405–7415.
10. Collepardo-Guevara R, et al. (2015) Chromatin unfolding by epigenetic modifications explained by dramatic impairment of internucleosome interactions: A multiscale computational study. *J Am Chem Soc* 137:10205–10215.
11. Chen Q, Yang R, Korolev N, Liu CF, Nordenskiöld L (2017) Regulation of nucleosome stacking and chromatin compaction by the histone H4 N-terminal tail–H2A acidic patch interaction. *J Mol Biol* 14:2075–2096.
12. Zhang R, Erler J, Langowski J (2017) Histone acetylation regulates chromatin accessibility: Role of H4K16 in inter-nucleosome interaction. *Biophys J* 112:450–459.
13. Roh T, Wei G, Farrell CM, Zhao K (2007) Genome-wide prediction of conserved and nonconserved enhancers by histone acetylation patterns. *Genome Res* 17:74–81.
14. Voong LN, et al. (2016) Insights into nucleosome organization in mouse embryonic stem cells through chemical mapping. *Cell* 67:1555–1570.e15.
15. Lu B, Charvin G, Siggia ED, Cross FR (2010) Nucleosome-depleted regions in cell-cycle-regulated promoters ensure reliable gene expression in every cell cycle. *Dev Cell* 18:544–555.
16. Perišić O, Collepardo-Guevara R, Schlick T (2010) Modeling studies of chromatin fiber structure as a function of DNA linker length. *J Mol Biol* 403:777–802.

17. Correll SJ, Schubert MH, Grigoryev SA (2012) Short nucleosome repeats impose rotational modulations on chromatin fibre folding. *EMBO J* 31:2416–2426.
18. Thoma F, Koller T, Klug A (1979) Involvement of histone H1 in the organization of the nucleosome and of the salt-dependent superstructures of chromatin. *J Cell Biol* 83:403–427.
19. Zhou Y-B, Gerchman SE, Ramakrishnan V, Travers A, Muyldermans S (1998) Position and orientation of the globular domain of linker histone H5 on the nucleosome. *Nature* 395:402–405.
20. Thomas JO (1999) Histone H1: Location and role. *Curr Opin Cell Biol* 11:312–317.
21. Fan Y, et al. (2005) Histone H1 depletion in mammals alters global chromatin structure but causes specific changes in gene regulation. *Cell* 123:1199–1212.
22. Woodcock CL, Skoultchi AI, Fan Y (2006) Role of linker histone in chromatin structure and function: H1 stoichiometry and nucleosome repeat length. *Chromosome Res* 14:17–25.
23. Fyodorov DV, Zhou B-R, Skoultchi AI, Bai Y (2018) Emerging roles of linker histones in regulating chromatin structure and function. *Nat Rev Mol Cell Biol* 19:192–206.
24. Ali Öztürk M, Cojocaru V, Wade RC (2018) Toward an ensemble view of chromatosome structure: A paradigm shift from one to many. *Structure* 26:1050–1057.
25. Bednar J, et al. (1998) Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin. *Proc Natl Acad Sci USA* 95:14173–14178.
26. Caterino TL, Hayes JJ (2010) Structure of the H1 C-terminal domain and function in chromatin condensation. *Biochem Cell Biol* 89:35–44.
27. Cao K, et al. (2013) High-resolution mapping of H1 linker histone variants in embryonic stem cells. *PLoS Genet* 9:e1003417.
28. Geeven G, et al. (2015) Local compartment changes and regulatory landscape alterations in histone H1-depleted cells. *Genome Biol* 16:289.
29. Hu J, et al. (2018) Dynamic placement of the linker histone H1 associated with nucleosome arrangement and gene transcription in early *Drosophila* embryonic development. *Cell Death Dis* 9:765.
30. Popova EY, et al. (2013) Developmentally regulated linker histone H1c promotes heterochromatin condensation and mediates structural integrity of rod photoreceptors in mouse retina. *J Biol Chem* 288:17895–17907.
31. Noordermeer D, et al. (2011) The dynamic architecture of Hox gene clusters. *Science* 334:222–225.
32. Bonev B, et al. (2017) Multiscale 3D genome rewiring during mouse neural development. *Cell* 171:557–572.e24.
33. Salzberg AC, et al. (2017) Genome-wide mapping of histone H3K9me2 in acute myeloid leukemia reveals large chromosomal domains associated with massive gene silencing and sites of genome instability. *PLoS One* 12:e0173723.
34. ENCODE Project Consortium (2004) The ENCODE (ENCyclopedia of DNA elements) project. *Science* 306:636–640.
35. Bascom GD, Kim T, Schlick T (2017) Kilobase pair chromatin fiber contacts promoted by living-system-like DNA linker length distributions and nucleosome depletion. *J Phys Chem B* 121:3882–3894.
36. Grigoryev SA, et al. (2016) Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes. *Proc Natl Acad Sci USA* 113:1238–1243.
37. Gavin B, Schlick T (2017) Linking chromatin fibers to gene folding by hierarchical looping. *Biophys J* 112:434–445.
38. Ricci MA, Manzo C, García-Parajo MF, Lakadamyali M, Cosma MP (2015) Chromatin fibers are formed by heterogeneous groups of nucleosomes *in vivo*. *Cell* 160:1145–1158.
39. Risca VI, Denny SK, Straight AF, Greenleaf WJ (2017) Variable chromatin structure revealed by in situ spatially correlated DNA cleavage mapping. *Nature* 541:237–241.
40. Hsieh TH, et al. (2015) Mapping nucleosome resolution chromosome folding in yeast by Micro-C. *Cell* 162:108–119.
41. Li G, Reinberg D (2011) Chromatin higher-order structures and gene regulation. *Curr Opin Genet Dev* 21:175–186.
42. Boettiger AN, et al. (2016) Super-resolution imaging reveals distinct chromatin folding for different epigenetic states. *Nature* 529:418–422.
43. Di Pierro M, Cheng RR, Lieberman Aiden E, Wolynes PG, Onuchic JN (2017) De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proc Natl Acad Sci USA* 114:12126–12131.
44. Ernst J, Kellis M (2012) ChromHMM: Automating chromatin-state discovery and characterization. *Nat Methods* 9:215–216.
45. Liang Z, et al. (2017) BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. *Nat Commun* 8:1622.
46. Blake WJ, Kærn M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* 422:633–637.
47. Barbieri M, et al. (2012) Complexity of chromatin folding is captured by the strings and binders switch model. *Proc Natl Acad Sci USA* 109:16173–16178.
48. Jost D, Pascal C, Cavalli G, Vaillant C (2014) Modeling epigenome folding: Formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res* 42:9553–9561.
49. Di Pierro M, Zhang B, Aiden EL, Wolynes PG, Onuchic JN (2016) Transferable model for chromosome architecture. *Proc Natl Acad Sci USA* 113:12126–12131.
50. Schöpflin R, et al. (2013) Modeling nucleosome position distributions from experimental nucleosome positioning maps. *Bioinformatics* 29:2380–2386.
51. Collepardo-Guevara R, Schlick T (2014) Chromatin fiber polymorphism triggered by variations of DNA linker lengths. *Proc Natl Acad Sci USA* 111:8061–8066.
52. Arya G, Schlick T (2006) Role of histone tails in chromatin folding revealed by a mesoscopic oligonucleosome model. *Proc Natl Acad Sci USA* 103:16236–16241.
53. Beard DA, Schlick T (2001) Modeling salt-mediated electrostatics of macromolecules: The discrete surface charge optimization algorithm and its application to the nucleosome. *Biopolymers* 58:106–115.
54. Karolchik D, et al. (2003) The UCSC genome browser database. *Nucleic Acids Res* 31:51–54.
55. Yazdi PG, et al. (2015) Increasing nucleosome occupancy is correlated with an increasing mutation rate so long as DNA repair machinery is intact. *PLoS One* 10:e0136574.
56. Edgar R, Domrachev M, Lash AE (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 30:207–210.
57. Nicol JW, Helt GA, Blanchard SG, Jr, Raja A, Loraine AE (2009) The Integrated Genome Browser: Free software for distribution and exploration of genome-scale datasets. *Bioinformatics* 25:2730–2731.
58. Kent WJ, et al. (2002) The human genome browser at UCSC. *Genome Res* 12:996–1006.
59. Bharath MMS, Chandra NR, Rao MRS (2003) Molecular modeling of the chromatosome particle. *Nucleic Acids Res* 31:4264–4274.
60. Luque A, Collepardo-Guevara R, Grigoryev S, Schlick T (2014) Dynamic condensation of linker histone C-terminal domain regulates chromatin structure. *Nucleic Acids Res* 42:7553–7560.