Long-range genomic loci stochastically assemble into combinatorial forms of chromosome skeleton

Jingyu Zhang[1,*], Siyuan Wang[2], Simon C Watkins[3], Jianhua Xing[1,4,5,*]

[1] Department of Computational and Systems Biology, University of Pittsburgh; Pittsburgh, PA 15232, USA.

[2] Department of Cell Biology, Yale School of Medicine; New Haven, CT 06510, USA.

[3] Department of Cell Biology, University of Pittsburgh; Pittsburgh, PA 15232, USA.

[4] UPMC-Hillman Cancer Center, University of Pittsburgh; Pittsburgh, PA, USA.

[5] Department of Physics and Astronomy, University of Pittsburgh; Pittsburgh, PA 15232, USA.

[*] To whom correspondence should be addressed. Email: xing1@pitt.edu, zhangjy4316@gmail.com

Abstract

One fundamental yet open question is how eukaryotic chromosomes fold into segregated territories, a process essential for gene transcription and cell fate. Through analyzing Hi-C and chromatin-tracing DNA-FISH data, we identify long-range chromo skeleton loop structures that span over 100 Mb, extending beyond the reach of loop extrusion models. Spatial density analyses point to assembly formation independent of major nuclear structures. A subset of genomic loci serves as nucleation centers, driving loop clustering. These complexes are highly stable, as shown by live-cell imaging with sequence-specific fluorescent labeling, and biophysical model analyses reveal a multivalent binding mechanism. Our findings suggest a redundant, distributed cluster mechanism that ensures robustness across cell types and against mutations, guiding both chromosome compaction and the formation of smaller-scale chromosomal structures.

MAIN TEXT

A central problem in structural biology and biological physics is how macromolecules fold into three-dimensional structures. An emerging "local-to-global" mechanism highlights the role of local cooperativity among peptide monomers in guiding the formation of a native protein structure from an expansive conformational space (*1*). In contrast to typical polypeptides with 300 – 400 amino acids, human chromosomes are composed of 48 long-chain polymers, ranging from ~40 to 250 Mb, with a total linear length extending up to two meters. Given that the 3D structures play a central role in regulating cell identity and gene expressions (*2*), an intriguing question is how these linear DNA molecules compact into segregated chromatin territories within a cell nucleus, typically only about 10 μm in diameter.

Chromatin, unlike a polypeptide, is often considered to lack a stable, ordered three-dimensional structure due to its significantly greater length. When modeled as a linear polymer chain without

structural supports such as protein bridges or long non-coding RNA networks, theoretical analyses predict that the average contact frequency (*f*) between two chromosomal fragments depends on their genomic separation (*l*) and follows a power-law decay, $f \propto l^{\alpha}$. Both Hi-C and DNA FISH measurements confirmed this relationship, with an exponent $\alpha \sim -1$ within the 1 – 5 Mb range (Fig. 1A, Fig. S1A-C). This result aligns with polymer models describing chromatin as a crumpled polymer (*3-5*) or emphasizing the excluded volume effect (*6*).

However, at genomic distances ~100 – 200 kb, deviations from the power-law behavior (Fig. 1A, Fig. S1A-C) suggest the presence of stable chromatin loops, potentially formed through a loop-extrusion mechanism (*7*). Beyond 5 Mb, Hi-C data reveal further deviations, where the *f-l* curves level off (Fig. 1A blue line, Fig S1A&B). This phenomenon has been attributed to polymer models incorporating random-sized loop formation between chromatin segments (*8*). Additional factors, such as A/B compartment segregation and nuclear structure tethering, also contribute to deviations from the crumpled polymer model at large genomic distances (*4, 9*).

Therefore, previous studies suggest that chromosomes form functionally important local structures such as topologically associated domains (TADs), typically spanning ~ 1 Mb. In contrast, locus pairs separated by more than 10 Mb generally lack stable structures with specific interactions. In large genomes like the human genome, such long-range proximities are considered rare, though scattered instances of locus pairs interacting across distances exceeding 20 Mb have been reported (*4, 9, 10*). In this study, we investigate the presence and conservation of spatial proximity between two well-separated loci on the same human chromosome across various cell lines and explore the underlying physical mechanisms that may explain such stable contacts.

Hi-C data reveals the prevalence of long-range genomic locus colocalization beyond 50 Mb

We analyzed Hi-C data for chromosome 2 (chr2) and 14 (chr14) in MCF10A cells. While most genomic pairs followed the expected Hi-C *f-l* curve, certain pairs (e.g., red box in Fig. 1A & Fig. S1A) showed significant deviations, with contact frequencies comparable to the median value of 1-Mb pairs despite their much larger genomic separation. Similar high-interaction locus pairs were also observed in chr2 and chr14 of IMR90 (Fig. S1B). Extending this analysis to 13 additional human cell lines revealed that such deviations are prevalent among genomic locus pairs separated by even more than 50 Mb (Fig. S2), suggesting their stable three-dimensional proximity in at least a subpopulation of chromatin. Notably, these long-range interactions extend beyond typical intrachromosomal contacts such as enhancer-promoter interactions, super-enhancer, and TAD (*11, 12*). Given the entropic cost of maintaining substantial colocalization probability over such long genomic distances, we sought to explore their underlying mechanisms and potential functional significances.

While the above statistical analyses identified long-range locus pairs with spatial proximity, detecting them quantitatively from noisy Hi-C data remains challenging, as strong signals from short-ranged locus pairs often overshadow weaker signals from long-range ones. To address this,

we developed a computational pipeline, Chromosome Long-range colocalization identifier through Outlier Detection (CLOD) (Fig. 1B, details in Methods). The pipeline operates on the premise that colocalized locus pairs increase the contact frequency of their neighboring loci along the genome. To enhance the signal-to-noise ratio, CLOD employs a sliding-window averaging approach, allowing for the distinction of colocalization signals from background noise. It then compares locus pairs with similar genomic distances and identifies outliers exhibiting abnormally high contact frequencies. Specifically, outliers are defined as locus pairs whose signals exceed two times the interquartile range above the third quantile of all pairs with similar genomic distances ($\frac{Signal-Q3}{Q3-Q1} > 2$).

To assess CLOD's performance in detecting experimentally verified colocalized locus pairs, we first analyzed an MCF7 Hi-C dataset (*13*). CLOD successfully identified a 150 kb loop (Fig. S1D), previously validated as a promoter-enhancer interaction of *PRIP1* using a 3C assay (*14*). We then applied CLOD to a Hi-C dataset from the HCT116 cell line (*15*) and detected a colocalized genomic pair separated by 27 Mb (Fig. 1C). This locus pair was independently confirmed by fluorescence in situ hybridization (FISH) (*10*). Collectively, these results demonstrate that the CLOD pipeline can effectively detect colocalized locus pairs across both short and long genomic distances.

We then applied CLOD to Hi-C datasets from a panel of four normal differentiated human cell lines, five cancer cell lines, and four embryonic cell lines (Table S1). Across these datasets, we identified multiple colocalized locus pairs separated by 50 Mb or further in chr2 and chr14 (Fig. 1D&E and Fig. S3). Notably, some of these long-range pairs were consistently detected across multiple cell lines of different types (Fig 1F&G).

Specifically, CLOD identified a colocalized genomic pair on chr14 separated by approximately 80 Mb (Fig. 1H). To determine whether this long-range colocalization was transient or persistent, we labeled the loci with two-color CRISPR-dCas9 guided fluorescent proteins in HEK293 cells (Fig. 1I, Methods, Table S2). Live-cell imaging confirmed the sustained proximity and coordinated movement of green and red puncta, with their co-fluctuations persisting throughout the observed 2-10 hours in some cases (Fig. 1J &K, and movie S1-S2).

MERFISH chromatin tracing data reveals prevalence and conservation of colocalized loci > 100 Mb in human chr2.

While bulk Hi-C provides an ensemble-averaged view of chromosome conformations, it lacks the resolution to capture individual chromosome structures. It cannot distinguish whether colocalized pairs originate from the same or different homologous chromosomes, nor can it determine whether observed contact frequencies reflect widespread interactions or rare sub-populations. To cross-validate the prevalence of long-range pair colocalization and assess their distribution at the single chromatin level, we analyzed previously published IMR90 chromatin tracing data obtained using DNA MERFISH (*16*).

112 As a technical control, we first re-analyzed the 3D distances (d_1) of MERFISH-labeled neighboring
113 genomic loci across chr21 with a genomic separation $l_1 = 50\ kb$. The median distance for
114 chr21:32.45 - 33.35 Mb (Fig. S4A) was found to be 256 nm, consistent with the previous report
115 (*16*). The overall distance distribution of neighboring genomic loci across chr21 followed a log-
116 normal distribution, with a median distance of 325 nm (Fig. S4B, left).

117 Next, we calculated the spatial distances of all neighboring genomic locus pairs across 2991 copies
118 of chr2 that were fluorescently traced with $l_1 = 250\ kb$ (Fig 2A, green line). The distribution also
119 exhibited a log-normal pattern, with a median distance 462 nm and a first quantile distance d_r =
120 299 nm (Fig. 2B, Fig. S4B, right). For subsequent analyses, we adopted d_r as an intrinsic distance
121 reference, classifying a genomic locus pair as in spatial proximity if their measured spatial distance
122 $d \leq d_r$. This criterion aligns with previous studies using super-resolution DNA FISH data at
123 approximately 30 kb resolution, which employed a cutoff distance of ~ 200 nm to define chromatin
124 interactions (*17*), and 500 nm for the chr21 data with 50 kb resolution to identify genomic pairs in
125 proximity (*18*).

126 Most previously studied short-range chromosome interactions, such as enhancer-promoter
127 interactions, occur within genomic distances of ≤ 200 kb, typically in the 20 – 50 kb range (*19,*
128 *20*). However, few studies have specifically focused on long-range chromatin interactions. Here,
129 we calculated the spatial distances between pairs of loci with genomic separations of $l_2 \in$
130 $(50, 100]$ Mb and $l_3 > 100$ Mb, respectively (Fig. 2A, purple and orange lines). Compared to d_1,
131 the distributions of corresponding d_2 and d_3 shifted toward larger values, with median distances
132 of 1.86 μm and 2.08 μm, respectively (Fig. 2B). Interestingly, even at such large genomic
133 separations, certain locus pairs exhibited spatial distances < d_r (Fig. 2B, arrow). In the following
134 analyses, we focused on the genomic locus pairs with $l_3 > 100$ Mb and $d_3 < d_r$, denoting them as
135 long-range colocalization (LRC).

136 Individual chr2 copies show diverse conformations (see examples shown in Fig. 2C, Fig. S4C&D,
137 Movie S3-S5), consistent with previous studies (*17*). A negative correlation exists between the
138 total number of LRCs of an individual chromosome copy and its conformation measured by
139 convex hull volumes (Fig. 2D), suggesting a potential relationship between LRC and chromosome
140 conformation. Notably, in these structures LRCs form either scattered small clusters or join into
141 larger ones, along with varying compactness of the chromatin structures.

142 LRC counts per chr2 copy range from fewer than 10 to over 10,000, with a peak at ~600 per copy
143 (Fig. S4E). These LRCs can be classed into two groups: transient contacts, likely due to
144 fluctuations, and stable structures. We hypothesized that transient LRCs appear infrequently,
145 whereas stable ones recur in subpopulations of chr2 copies. To test this, we analyzed the
146 occurrence frequency of 155,009 unique LRCs across 2,991 chr2 copies (Fig. 2E). We identified
147 7,366 LRCs with occurrence frequencies beyond the 95% population interval, defined them as
148 stable LRCs (sLRCs) (Fig. 2E & F). Each sLRC appeared in at least 23 chr2 copies (~0.8% of the
149 dataset), with some found in over 10% of samples. That is, sLRC exhibited a significantly higher
150 probability of occurring within 299 nm compared to other LRC (Fig. 2G).

151 We then compared sLRC detected via DNA FISH with CLOD-identified colocalized pairs (> 100
152 Mb separation) from Hi-C. Most Hi-C outliers were also classified as sLRCs (Fig. 2H). Notably,
153 Hi-C and DNA FISH detect proximity differently: Hi-C captures direct contact, whereas DNA
154 FISH assesses spatial distances, using a 299 nm threshold here. Technical limitations also
155 contribute discrepancies — Hi-C may miss repetitive or low-accessibility regions, whereas its
156 much larger sample size (~1 million cells vs. ~3,000 chr2 copies in DNA FISH) could detect rare
157 configurations potentially overlooked by DNA FISH. Despite these differences, the strong
158 agreement between Hi-C and DNA FISH reveals that LRCs are prevalent in chromatin structure,
159 with a subset forming stable structural features.

160 **A subset of genomic sites serves as nucleation centers (NCs) for clustering sLRCs.**

161 Our structural analyses revealed that sLRC-associated genomic loci form intertwined networks
162 with sLRCs acting as links (Fig. 2C, Fig. S4C&D, Movie S6-S9). Linkage maps from both Hi-C
163 and DNA FISH data show that certain genomic loci colocalize with multiple distant loci (Fig.
164 2F&G). These loci likely function as NCs, linking multiple sLRCs within the network (Fig. 3A).
165 The genomic locations of NCs, including regions such as ~28 – 40 Mb and the ends of chr2, are
166 evident in sLRC distribution plots (Fig. 3B & C). Notably, locus 242 Mb (NC 242 Mb) exhibited
167 significantly more sLRCs and was analyzed separately.

168 Statistical analyses (Fig. 3B, right) showed that 26% of sLRC-associated loci formed only one pair
169 of sLRC, while 50% established fewer than five. However, a subset of loci participated in over
170 150 sLRCs (Fig. 2B, right). Even within individual chr2 copies, some loci paired with up to ~ 90
171 others to form sLRCs (Fig. 3D). A striking example is NC 242 Mb, which paired with loci across
172 the chromosome (Fig. 3E). At the population level, 89.8% of the 2,991 observed chromosomes
173 had at least one sLRC partner at NC 242 Mb. Other NCs also showed frequent long-range
174 interactions, such as NC 34.25 Mb, found in 38% of chr2 copies (Fig. 3F). Table S3 lists all loci
175 associated with at least two sLRCs in chr2 copies.

176 Modeling sLRCs as networks (Fig. 3G) revealed that NCs coordinated sLRCs into distinct clusters
177 of varying sizes, with chr2 copies containing 1 – 48 clusters (Fig. 3H). While most clusters (59.2%)
178 contained only 2 – 5 loci, 5.4% had over 100 loci as nodes (Fig. 3I). Notably, large clusters
179 frequently contained NC 242 Mb and its associated sLRCs. Detailed analyses revealed distinct
180 structural motifs (Fig. 3J): some NCs formed isolated single-NC clusters, while others coalesced
181 into multi-NC hubs via direct colocalization or shared pairing loci, forming clusters with 100 –
182 780 nodes.

183 These findings suggest that NCs link multiple genomic regions, assembling hub-like domains that
184 may regulate chromosome conformation and territory segregation.

185 **Neighboring LRCs cooperate to form stable complexes through multivalent binding**

186 DNA FISH data revealed the widespread presence of stable LRC ($p \geq 0.8\%$) in chr2 copies, with
187 some occurring in 5 – 10% of all copies (Fig. 2E). However, consider an isolated sLRC pair that
188 stochastically transits between bound and unbound states, with association (α) and dissociation (γ)

rate constants (Fig. 4A left). After dissociation, the loci diffuse apart due to thermal fluctuations, making re-encounters highly unlikely ($\alpha << 1$). Achieving $p = \frac{\alpha}{\alpha+\gamma} \geq 0.8\%$ would require an exceptionally low γ, implying the need for strong and specific interactions, or alternative mechanisms stabilizing spatial proximity.

A potential clue lies in Fig. 2F, which shows sLRC clustering along the genome. Statistical analysis confirmed that most sLRC-associated loci have their nearest neighboring genomic loci within 500 kb (Fig. S5A). We hypothesized that neighboring sLRCs stabilize each other via a multivalent binding mechanism — where an unbound sLRC pair remains spatially close due to adjacent bound sLRCs. This leads to an increased effective association rate constant $\alpha + \delta$, where δ represents the contribution from neighboring sLRCs, consequently the sLRC cluster becomes an effective two-conformation system with no or at least one bound sLRC, respectively (Fig. 4A middle & right).

To test this model, we first defined the genomic proximity between two pairs of sLRCs. Given pairs a (loci a_1, a_2) and b (loci at b_1, b_2), we defined their genomic proximity distance as $l_{ab} = \max\left(\min(|a_1 - b_1|, |a_1 - b_2|), \min(|a_2 - b_1|, |a_2 - b_2|)\right)$ (Fig 4B, Fig. S5Ba). The sLRC proximity distribution exhibited three peaks with 28% of pairs within 12 Mb (Fig. 4B). To detect multivalent sLRCs, we used a more stringent threshold, grouping sLRCs into metaLRCs, where each sLRC had at least one other sLRC with 250 kb at both ends (Fig 4C). This process grouped 72% sLRCs into 736 metaLRCs, each containing 2 to 40+ sLRCs (Fig. S5C).

The multivalent binding model predicts that within a metaLRC, binding of one sLRC enhances the binding of others, leading to positively correlated binding states (Fig 4A, middle). Fig. 4E and Fig. S5D&E depicted three representative metaLRCs. Correlation analyses over MERFISH data confirmed this expectation for three representative metaLRCs (Fig 4E, Fig. S5D&E, middle).

To quantify the cooperative effect, we constructed a minimal model treating all sLRCs within a metaLRC equally (Supplemental Text). Assumes a constant increase in α by δ due to neighboring bound sLRCs (Fig. 4A, middle), we obtained gave $\frac{\alpha}{\gamma} = \frac{p_1}{Np_0}$, where N is the number of sLRCs in the metaLRC, p_0 and p_1, the frequency of observing zero or one bound sLRC, respectively. The effective δ was determined by fitting the distribution of bound sLRCs, p_i vs. i for i > 1, which significantly exceeded predictions from a non-cooperative model ($\delta = 0$) (Fig. 4F).

Analysis of individual chr2 copies (Fig. 4G, bound vs. free forms, Movies S10-13) further supported the model. When one or more sLRCs were bound, the two genomic regions within the metaLRC were held closer (Fig. 4H), though additional bound sLRCs (> 1) did not significantly alter spatial distance. The observed non-monotonic p_i v.s. i curves for certain metaLRCs (e.g., 33.75-35.5 Mb & 146-146.25 Mb; 119.25-123.75 Mb & 242 Mb) suggested even stronger cooperativity than the minimal model predicted (see Supplemental Text) (Fig. S5D&E).

For all metaLRCs of chr2 with N > 2, α/γ values from FISH data were < 0.007 (Fig. 4I), indicating that associating two distant loci is entropically unfavorable. However, the corresponding ($\alpha +$

226 $\delta)/\gamma$ values were up to 10 – 100× higher, making the probability of observing a bound metaLRC
227 over 1 – 10% (Fig. S5F).

228 In conclusion, pairs of distant loci can remain stably proximate within a chromosome via neighbor-
229 assisted multivalent binding, significantly enhancing structural stability and long-range genomic
230 interactions.

231 **Structural analyses reveal multiple candidate mechanisms for sLRC formation.**

232 Colocalization of distant cis-loci had been intermittently observed, with three general mechanisms
233 proposed to explain LRC formation (Fig. 5A, I-III). (I) A/B compartment, or loose/compact region,
234 segregation may bring distant B compartment loci in transient proximity. (II) Transcriptionally
235 active loci, chromatin around which are typically loose and accessible, may colocalize via
236 attaching to shared nuclear structures like speckles, or nucleoli. (III) The nuclear membrane may
237 recruit multiple loci into spatial proximity. An alternative mechanism (IV) suggests stable sLRC
238 formation independent of major nuclear structures or compartment segregation. While the first
239 three have been proposed previously (*21*), sLRC assemblies formed through Mechanism IV were
240 considered rare.

241 Examination of A/B compartment identities among sLRC-forming loci revealed no clear pattern
242 or correlation between the two loci of each sLRC pair (Fig. S6A-C), while the majority of sLRCs
243 have two loose ends (Fig. S6B, bottom), which is consistent with the observation from a previous
244 study (*18*). Note that the A/B compartment was calculated from bulk level data. That is, an
245 ensemble of chromosomes may be composed with subpopulations with and without the presence
246 of a sLRC under study, thus conclusions from such bulk-averaged data are not definite. Bulk
247 averaging also prevented us from concluding whether sLRCs or the loci linked to sLRCs have
248 specific epigenetic patterns (Fig. S6 D-E).

249 Using MERFISH data, we quantified local geometric structures within 299 nm of a locus by
250 defining regional density of short-ranged (< 10 Mb) neighbors (RD-SN) and regional density
251 heterogeneity (RDH), which measures the uniformity of local spatial distributions (Fig. 5B). Note
252 that an A compartment has higher probability to has a loose neighborhood than a B compartment
253 has (Fig. S6F). Then one expects that Mechanism I and III lead to high RD-SN for both loci
254 associated with a sLRC, Mechanism II leads to low-low RD-SN, while the pattern for Mechanism
255 IV is undetermined. Furthermore, a locus next to a nuclear structure (Mechanism II and III) due to
256 the excluded volume of the latter. Therefore, a combined signature of mixed low/high RD-SN and
257 low RDH only comes from Mechanism IV. Analyses of the sLRCs revealed broad distributions of
258 these two quantities, implying possible contribution from diverse mechanisms on sLRC formation
259 (Fig. 5C&D), as also reflected from examining representative structures (Fig. 5E).

260 Among the 50 most frequent sLRCs, most exhibited low-low RD-SN patterns, consistent with
261 previous studies (*18*), followed by high-low patterns, while high-high patterns were rare (Fig. 5F).
262 Notably, 41 out of the top 50 involved locus 242 Mb, consistent with its role as the leading NC.

Analyses of the sLRCs associated with the top 20 NCs revealed similar RD-SN patterns (Fig. 5G). These observations exclude Mechanism I as the dominant mechanism for sLRC formation.

Locus 242 Mb exhibited a more compact neighborhood than typical A compartment loci, with disproportionately high numbers of neighbors spanning >10 Mb (Fig. 5I&H). Sequence analysis identified unique repetitive elements at 242 Mb-242.01 Mb (Table S4), supporting Mechanism IV. These findings suggest an additional, previously underappreciated mechanism contributing to stable sLRC formation.

Together, all the epigenetic, positional, and spatial distribution results suggest an additional, previously underappreciated mechanism contributing to stable sLRC formation.

DISCUSSION

Our analyses of Hi-C and DNA FISH chromosome tracing data reveal widespread colocalization of genomic loci separated by distances far exceeding a typical TAD. Statistically, the probability of multiple genomic loci with $l \geq 100$ Mb being in physical proximity without stable structural mechanisms would be negligible. While the exact molecular mechanism remains unclear, some sLRCs may form through anchoring to nuclear structures such as nucleoli, speckles, or the nuclear envelope. However, the presence of many sLRCs and NCs away from these structures suggests additional nuclear structure-independent mechanisms, such as molecular assembly or condensates. Notably, in contrast to assumptions in random loop models (*8*), some large-sized loops appear sequence-specific, implying potential roles of non-coding DNA elements. For instance, the 242 Mb locus, enriched in repetitive sequences, may serve as a selective binding site for partnering loci, either directly or through bridging molecular factors.

Beyond their varied formation mechanism, these sLRC structures likely play diverse functional roles, including transcriptional regulation and broader chromosome 3D organization. Proper DNA folding into spatially segregated territories with minimal entanglement and knot formation is essential for eukaryotic genome organization and gene regulation. This process must be robust across all the cell types, including those with abnormal karyotypes, and must function despite cell type-specific euchromatin/heterochromatin partitioning, folding stochasticity, and mutations. From an engineering perspective, achieving this through a single centralized mechanism requiring precise tuning would be challenging. Instead, our findings suggest a redundant, distributed component mechanism that facilitates chromosome folding into compact structures (Fig. 5J). The genome harbors an extensive repertoire of NCs and partnering loci separated by long genomic distances, from which only a subset is available for a given cell type. At the individual chromosome level, subsets of these loci stochastically form sLRCs and larger structures, contributing to the observed heterogeneity of chromosome structures (*22*). The structures provide a "divide-and-conquer" mechanism to facilitate forming chromosome structures at smaller length scales, e.g., by forming segregated globally compact chromosome configurations and close-end boundary

conditions for simultaneous crumpled folding at multiple regions (Fig S4D) (*5*), and enhanced contact frequencies for forming loops at smaller scales.

It is important to note that our study employed stringent criteria to identify representative examples rather than exhaustively cataloging all stable long-range colocalizations. While we focused on loci separated by >100 Mb, similar structures likely exist at shorter genomic distances, particularly in smaller chromosomes such as chr21 (~48 Mb).

A key unsolved question is how sLRC loci locate and associate with each other. One plausible mechanism is that, during post-mitotic DNA decondensation, some genomically distant loci may be in spatial proximity and associate stochastically, leading to subsequent formation of additional sLRCs and stabilization of the assembled structures.

In summary, our analyses of chromosome tracing and Hi-C data uncover prevalent, conserved large-sized loops and their clustering into stable structures. These findings raise important questions regarding the molecular mechanisms, temporal dynamics, and functional significances of these structures on chromosome configurations, gene transcriptional activity, and cell type regulation.

Abbreviations and glossary:

LRC: Long-range colocalization for genomic locus pairs > 100 Mb and within 299 nm in spatial distance.

sLRC: Stable long-range colocalization for LRC pairs with occurrence frequency > 4% in the IMR90 chr2 MERFISH dataset.

metaLRC: a group of sLRCs, in which for each sLRC there was at least one other sLRC with 250 kb for both ends.

NC: Nucleation center referring to locus having multiple sLRC partner loci.

RD-SN: Regional density of short-ranged neighbors (< 10 Mb) within a 299 nm range of a tagged locus.

RDH: Reginal density heterogeneity.

Acknowledgments: We thank Tom Mistelli, Harinder Singh, Jing Chen, Jian Liu for helpful discussions.

Funding:

National Institute of General Medical Sciences R01GM148525 (JX)

The Charles E. Kaufman Fund of the Pittsburgh Foundation KA2018-98550 (JX)

Author contributions:

Conceptualization: JZ, JX

334 Methodology: JZ, SW, SCW, JX

335 Investigation: JZ, JX

336 Visualization: JZ, JX

337 Funding acquisition: JX

338 Project administration: JX

339 Supervision: JX

340 Writing – original draft: JX

341 Writing – review & editing: JZ, SW, SCW, JX

342

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

344 **Supplementary Materials**

345 Materials and Methods

346 Supplementary Text

347 Fig. S1 CLOD pipeline

348 Fig. S2 boxplot of chr2 and chr14 in 13 human cell lines

349 Fig. S3 linkage map of chr2 and chr14 in 13 human cell lines

350 Fig. S4 statistic results and two additional examples of individual chr2 from DNA FISH

351 Fig. S5 metaLRC

352 Fig. S6 Features of loci associated with sLRC

353 Table S1 Cell lines

354 Table S2 Summary of sgRNAs

355 Table S3 List of all loci associated with at least two sLRCs in chr2 copies

356 Table S4 Repetitive sequences identified within the 242 Mb region of human chr2

357

358 Movies S1-S2 two CRISPR knock-in FL microscopy imaging.

359 Movie S3-S5 3D structure of representative chr2 copies.

360 Movies S6-S9 3D structure of representative chr2 copies shown nucleation centers.

361 Movies S10-S13 comparison of 3D structures of example chr2 copies with or without metaLRC.

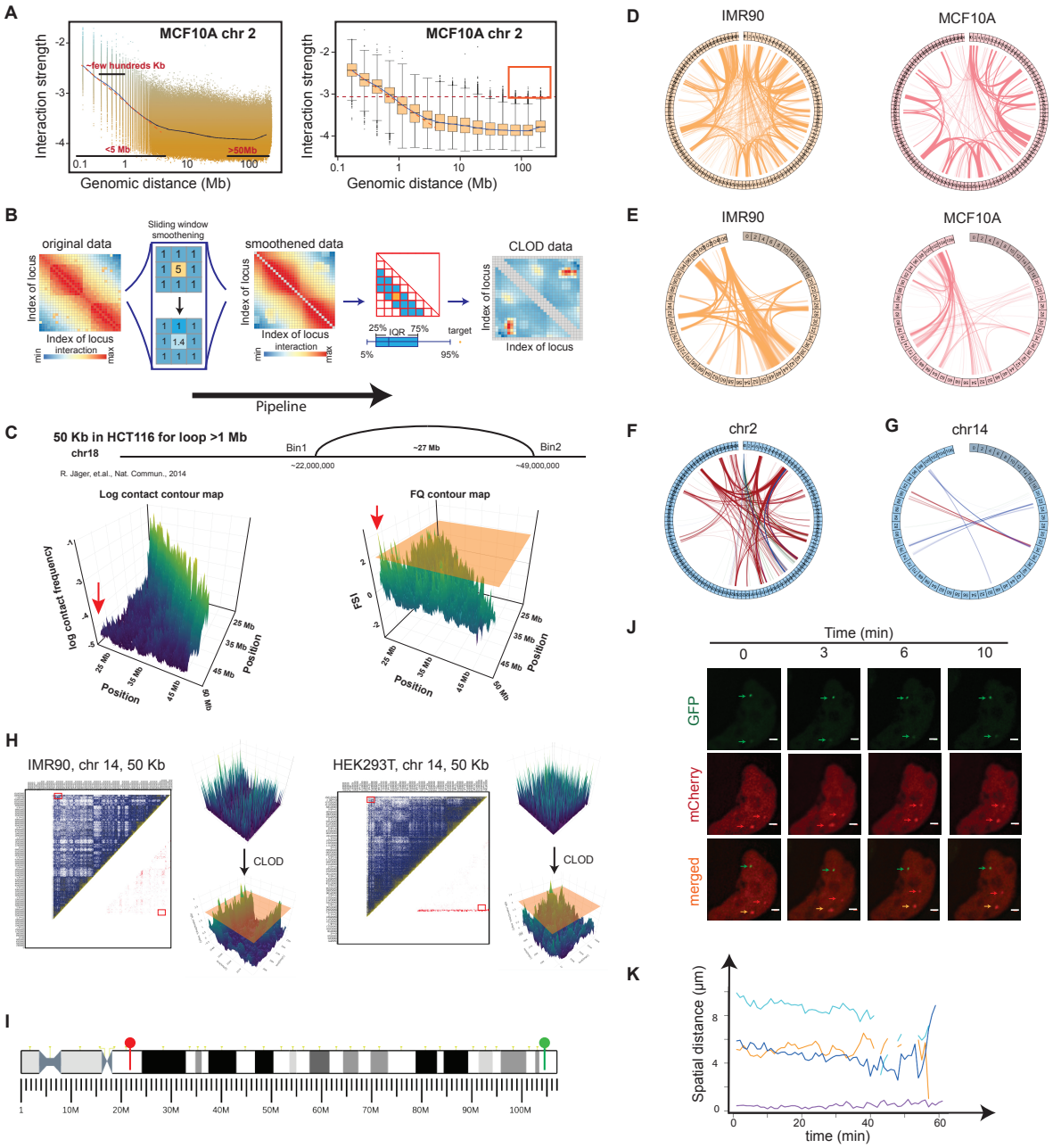


Fig. 1. Hi-C data reveals prevalent existence of colocalized pairs of long-range genomic loci. (**A**) Scatter plot (left) and boxplot (right) showing the decay trend (blue line) of contact frequency with increasing genomic distance. Red boxed points represent locus pairs > 50 Mb apart with contact frequency exceeding the third quartile of those observed for ~1 Mb pairs. (**B**) Schematic of the CLOD pipeline. (**C**) Experimental validation of colocalized pairs predicted by CLOD using Hi-C data from HCT116 cells at 50 kb resolution. (**D-E**) Representative linkage maps of colocalized locus pairs > 50 Mb on chr2 and chr14 in two different cell lines. (**F-G**) Linkage maps of colocalized pairs > 50 Mb apart, shared across at least two of the following cell types: four

372 embryonic cell lines (brown), four normal differentiated cell lines (green), and four cancer cell
373 lines (blue). (**H**) CLOD-based identification of colocalized pairs on chr14 from Hi-C data of
374 IMR90 and HEK293T cells at 50 kb resolution. (**I**) CRISPR-dCas9 two-color labeling sites on
375 chr14. (**J**) Live-cell imaging snapshots of a HEK293T cell labeled with RFP and GFP. Scale bar:
376 1 μm. (**K**) Time-resolved trajectories of distances between blue-red puncta pairs.

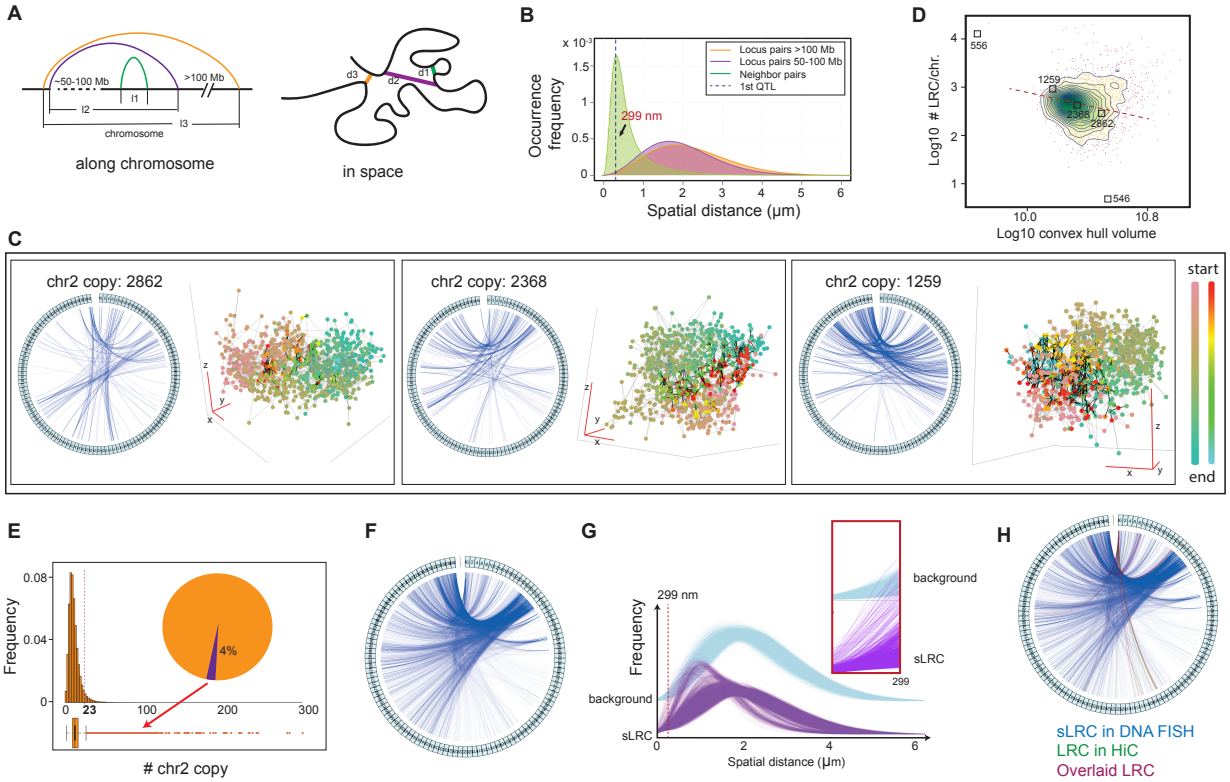


Fig. 2. MERFISH chromatin tracing reveals widespread long-range colocalization (LRC). (**A**) Schematic illustrating genomic pairs and their genomic versus spatial distances. (**B**) Density curves of spatial distance distributions for loci pairs with genomic separations of 250 kb, 50-100 Mb, and >100 Mb. The dashed line represents the first quantile of neighboring locus pairs. (**C**) LRC linkage maps and 3D structures of 3 representative chromosomes. (**D**) Scatter plot showing the relationship between chromosome compaction, quantified by the convex hull volume, and the number of LRCs. (**E**) Histogram (top) and boxplot (bottom) showing the distribution of chr2 copy counts (n = 2991) containing LRCs. LRCs present in >23 copies (red dash line) are classified as stable LRCs (sLRCs). (**F**) Identification of chr2 sLRCs from DNA FISH data. (**G**) Spatial distance distributions of locus pairs with genomic separation >100 Mb. Purple lines (n = 6443): sLRCs; blue lines (n = 6443): randomly selected non-sLRCs. The red dashed line represents the first quantile of spatial distances between neighbor loci on chr2 (299 nm). (**H**) Comparison of colocalized chr2 locus pairs (>100 Mb apart) identified in IMR90 cells using DNA FISH (blue lines) and Hi-C (pink lines) data. Overlapping links are highlighted in purple.

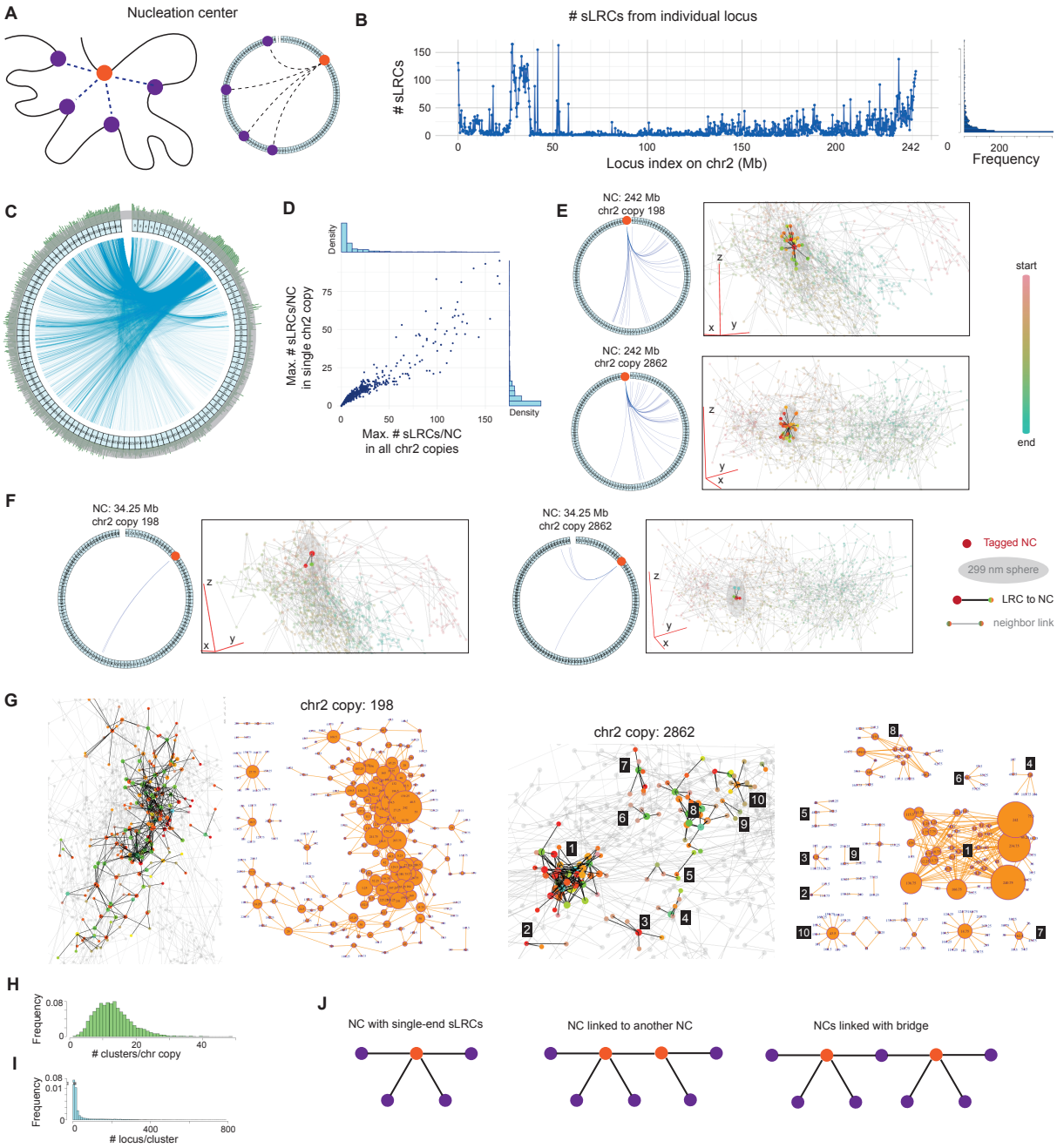


Fig. 3. Nucleation centers (NC) coordinate clustering of sLRCs. (**A**) Left: Schematic illustration of a nucleation center (NC) (orange) forming spatial proximity with multiple partner LRC loci (purple). Right: Ring plot depicting the same interaction network. (**B**) Left: Line plot showing the number of sLRCs originating from a given locus, based on 2,991 chr2 copies. Right Histogram illustrating the distribution of loci (x-axis) connecting to a given number of sLRC (y-axis). (**C**) sLRC linkage map with corresponding numbers of potential sLRCs (green bars outside the ring) originating from specific loci in the chr2 copies. (**D**) Scatter plot comparing the total number of potential sLRC from targeted loci across the analyzed chr2 population to the maximum number of sLRCs from a single targeted locus in an individual chr2 copy. Top and right panels:

Corresponding distributions. (**E-F**) Linkage maps and 3D structures highlighting two representative NCs (NC 242 Mb and NC 34.25 Mb) on chr2, along with their associated sLRCs. (**G**) 3D structural representations and network maps showing connections among NC loci in two representative chr2 copies. (**H**) Histogram displaying the number of locus clusters — defined as sets of loci connected to at least one other locus — per individual chr2 copy. (**I**) Histogram depicting the number of loci per locus-cluster, as identified in H. (**J**) Diagrams illustrating single NC elements and various types of connections between NCs.

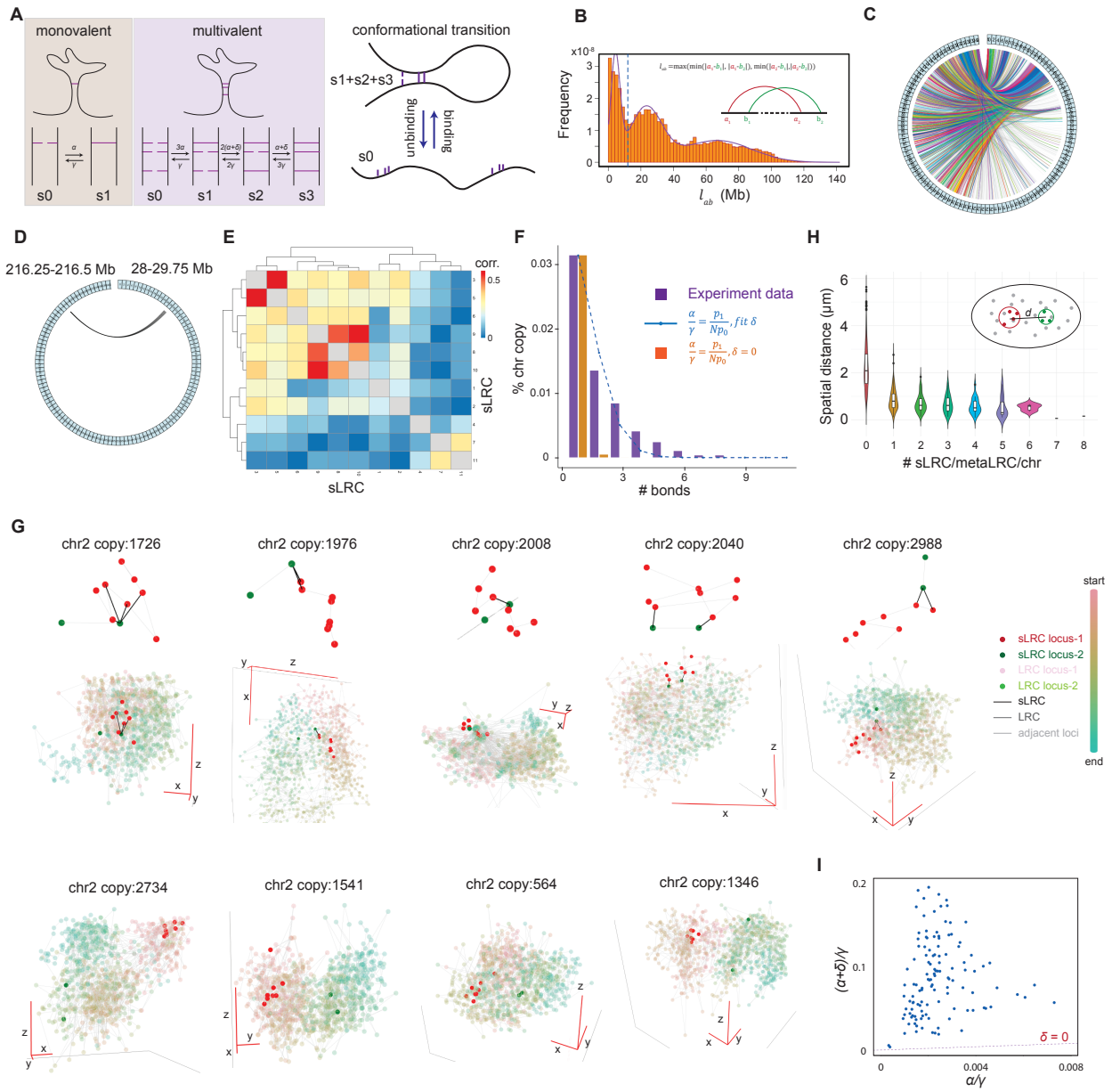


Fig. 4. DNA FISH data analysis reveals that multivalent binding stabilizes sLRC clusters. (**A**) Left: Schematic illustration of monovalent binding vs. multivalent binding models. α: association rate constant of a single pair in the absence of other bound sLRCs; γ: dissociation rate constant of a single pair; δ: association rate enhancement due to neighboring bound sLRCs. Right: A multivalent site is coarse-grained into a two-state system — either free or bound (with one or more bound sLRCs) two-state system. (**B**) Frequency distribution of the genomic separation between two LRCs. (**C**) sLRC and metaLRC linkage map. The outer circle represents chr2, and each line denotes a sLRC. Gray lines: isolated sLRC; colored lines (except gray): sLRCs belonging to a metaLRC. (**D**) A representative metaLRC. (**E**) Heatmap showing correlation between sLRCs within the metaLRC from D. (**F**) Comparison of the probabilities of observing various numbers of bound sLRCs within the metaLRC (purple bar) vs. predictions based on an independent sLRC

binding model (orange bar), and a multivalent binding model (blue dashed line). p_0 is not shown. (**G**) Representative 3D structures of chr2, highlighting the metaLRC from D. Top: Structures with at least one bound sLRC within the metaLRC. Bottom: Structures with no bound sLRC within the metaLRC. (**H**) Violin plot showing the relationship between 3D distances (between centers of minimal-sized bounding spheres of the two genomic regions forming the metaLRC from D) and the number of bound sLRCs per individual chr2 copy. (**I**) Multivalent binding model parameters for metaLRCs, obtained through fitting DNA FISH data.

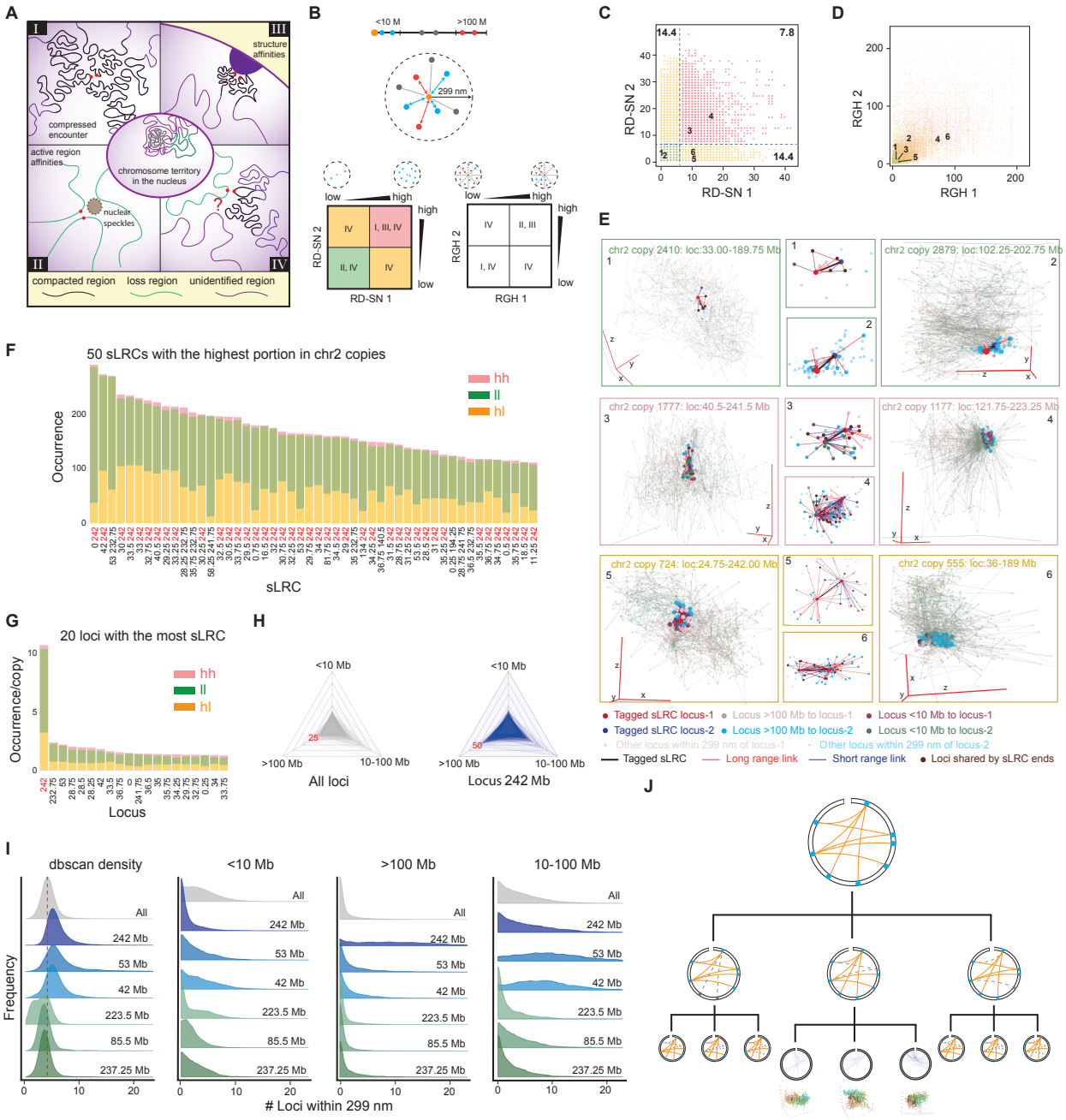


Fig. 5 Possible mechanisms of sLRC formation and functional roles on chromosome folding. (**A**) Schematic of four possible mechanisms for sLRC formation. I & II: A/B compartment segregation leads to compressed encounters of genomically distant loci in the B and A compartments, respectively. III: Colocalization of transcriptionally inactive, genomically distant loci interacting with common nuclear membrane structures. IV: Specific interactions between genomically distant loci form an assembly without the involvement of large nuclear structures. (**B**) Definition of regional density of short-ranged neighbors (RD-SN) and regional density heterogeneity (RDH) to characterize structural properties of sLRCs formed via mechanisms in (A). (**C-D**) Scatter plots of RD-SN and RDH around both ends of individual sLRCs in each chr2 copy.

Each dot represents one sLRC in a single chr2 copy. (**E**) 3D structures of six representative chr2 copies highlighted in (C-D). (**F**) Bar plot of the occurrence numbers (grouped into three RD-SN patterns) for the top 50 most populated sLRCs in the 2,991 chr2 copies. Colors match those in (C). (**G**) Bar plot of the numbers of NCs with the highest number of associated sLRCs in the 2,991 chr2 copies. Colors match those in (C). (**H**) Radar plots showing the spatial distribution of loci within 299 nm of a tagged locus, sampled over all loci (left) and a specific locus at 242 Mb (right). (**I**) Occurrence frequency vs. the number of sLRCs within 299 nm, analyzed across all loci and selected loci. (**J**) Schematic illustration of the proposed excessively redundant distributed component mechanism for chromosome folding.

Supplementary Text

Derivation of the minimal multivalent binding model

Consider a metaLRC that contains N pairs of sLRCs. Given the observed abnormally high contact frequency of a sLRC pair as compared to non-sLRC pairs separated with similar genomic distance, it is likely that the two loci are held together through some direct interactions or mediated by other molecular components. Thus, we can assume that an sLRC pair i exists in one of the two possible configurations, $\sigma_i = 1$ if the pair are bound (with spatial distance $d_r <= 299$ nm), and 0 otherwise. We assume this metaLRC is genomically distant from other metaLRCs and sLRCs, so for a good approximation there is no need to explicitly consider the influence of the latter on the association-dissociation dynamics of sLRCs in this tagged metaLRC.

Suppose that without influence from other sLRCs within the tagged metaLRC, an sLRC psair has a bare association rate α and a dissociation rate γ. Then at kinetic equilibrium the probability in the bound state is $\frac{\alpha}{\alpha+\gamma}$, and from the FISH data we expect that $\alpha \ll \gamma$. Here for simplicity we assume that each sLRC pair has the same association-dissociation kinetic parameters. Let's focus on one pair i. When there is one or more other bound sLRC pairs within the metaLRC, after dissociation the two loci of pair i are constrained by the nearby bound sLRCs and cannot diffuse far away from each other before rebind. Note that the genomic distance between these two loci of a sLRC ≥ 100 Mb, and by definition the genomic distance between two sLRCs i and j within a metaLRC $l_{ij} << 100$ Mb. Consequently, at the presence of other bound sLRC pair(s), the two loci of a dissociated sLRC have an increased probability to confront each other and rebind, so the effective association constant changes to $\alpha + \delta$, and we expect that $\delta \gg \alpha$. For simplicity, given $l_{ij} << 100$ Mb let us assume that each sLRC pair has the same effect on any other one, and the effect is not additive but δ remains the same value at the presence of one or more bound sLRC pairs. Given the genomic distance between two loci belonging to two different sLRCs, we further assume that there is no direct interaction between two sLRC pairs to simplify the analyses. That is, we assume that the dissociation constant of a sLRC pair, γ, is not affected by the presence of other bound sLRC pairs.

With the above model, one can write down a set of master equation for finding that the metaLRC exists with 0, 1,…, N pairs of bound sLRCs,

$$\frac{dp_0}{dt} = -Nap_0 + \gamma p_1,$$

$$\frac{dp_1}{dt} = -\big((N-1)(a+\delta)+\gamma\big)p_1 + 2\gamma p_2 + Nap_0,$$

$$\frac{dp_2}{dt} = -\big((N-2)(a+\delta)+2\gamma\big)p_2 + 3\gamma p_3 + (N-1)(a+\delta)p_1,$$

$$\ldots$$

$$\frac{dp_{N-1}}{dt} = -\big((a+\delta)+(N-1)\gamma\big)p_{N-1} + N\gamma p_N + 2(a+\delta)p_{N-2} \quad ,$$

$\frac{dp_N}{dt} = -N\gamma p_N + (a+\delta)p_{N-1}$.

Then at steady state,

$p_1 = \frac{Na}{\gamma} p_0$,

$p_2 = \frac{(N-1)(a+\delta)}{2\gamma} p_1$,

$p_3 = \frac{(N-2)(a+\delta)}{3\gamma} p_2$,

...

$p_N = \frac{a+\delta}{N\gamma} p_{N-1}$.

Define the probability of observing presence of one or more pairs as

$p = 1 - p_0 = \sum_{i=1,\dots,N} p_i$.

One has,

$$p = \left(\frac{N!\,(a+\delta)^{N-1}}{N!\,\gamma^{N-1}} + \cdots + \frac{(N-1)N(a+\delta)}{2\gamma} + N\right)\frac{a}{\gamma} p_0$$

$$= \left(\frac{N!\,(a+\delta)^{N}}{N!\,\gamma^{N}} + \cdots + \frac{(N-1)N(a+\delta)^2}{2\gamma^2} + N\frac{(a+\delta)}{\gamma}\right)\frac{a}{a+\delta} p_0$$

$$= \left(\left(1+\frac{a+\delta}{\gamma}\right)^N - 1\right)\frac{a}{a+\delta} p_0 = \frac{\left(\left(1+\frac{a+\delta}{\gamma}\right)^N - 1\right)\frac{a}{a+\delta}}{\left(\left(1+\frac{a+\delta}{\gamma}\right)^N - 1\right)\frac{a}{a+\delta} + 1}.$$

In the limit $\alpha \ll \delta \ll \gamma$, the above expression can be further simplified as,

$$p \approx \frac{N\left(1+\frac{(N-1)}{2}\frac{a+\delta}{\gamma}\right)\frac{a}{\gamma}}{N\left(1+\frac{(N-1)}{2}\frac{a+\delta}{\gamma}\right)\frac{a}{\gamma}+1} \approx \frac{N\left(1+\frac{(N-1)}{2}\frac{\delta}{\gamma}\right)\frac{a}{\gamma}}{N\left(1+\frac{(N-1)}{2}\frac{\delta}{\gamma}\right)\frac{a}{\gamma}+1}.$$

Several mechanisms can further increase the cooperativity beyond what described by a constant δ value in the above model. Presence of additional bound sLRC pair may further constrain the diffusion of the two loci of a dissociated sLRC, leading to further increase of the value of δ. That is, the value of δ may increase with the number of bound sLRC pairs. The bound sLRC pairs may physically contact and stabilize each other with an increased γ value. Indeed, we observed non-monotonic histograms of p_i for some metaLRCs (Fig. S5D&E). With a thermodynamic equilibrium model, one can introduce some binding free energies to describe the cooperativity as in allosteric effect and a previous cooperative enhancer binding model(*23*). For simplicity in this work we restricted the analyses to the minimal model discussed above.

508 Note that all the loci belonging to a metaLRC can be clustered as two clumped genomic regions
509 that are genomically distant (i.e., ~ 100 Mb or further). Subsequent analyses further coarse-grained
510 the binding state of a metaLRC as $s = 1$ if $(\sum_{i=1}^{N} \sigma_i) \geq 1$and 0 otherwise.

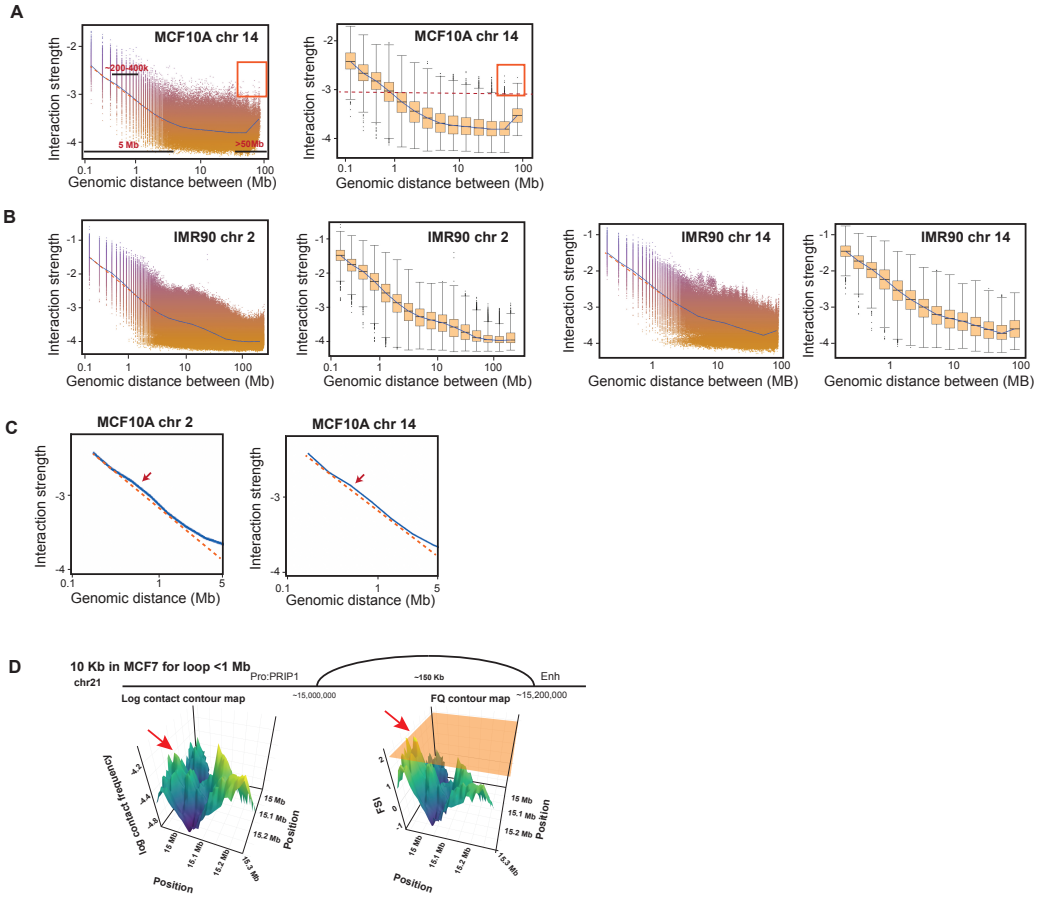


Fig. S1. Decay trends of contact frequency in additional cell lines and CLOD pipeline. (**A**) Scatter plot (left) and boxplot (right) showing the relationship between contact frequency and genomic distance for chr14 in MCF10A cells (blue line). Red-boxed points represent locus pairs separated by > 50 Mb, exhibiting contact frequencies higher than the third quarter of pairs of those separated by ~1 Mb. (**B**) Same as panel A, but for chr2 and chr14 of IMR90 cells. (**C**) Zoomed-in view of the contact frequency vs. genomic distance trend, derived from Fig. 1A and panel A. (**D**) Experimental validation of colocalized pairs identified by CLOD, using MCF7 cell Hi-C data at 10 kb resolution.

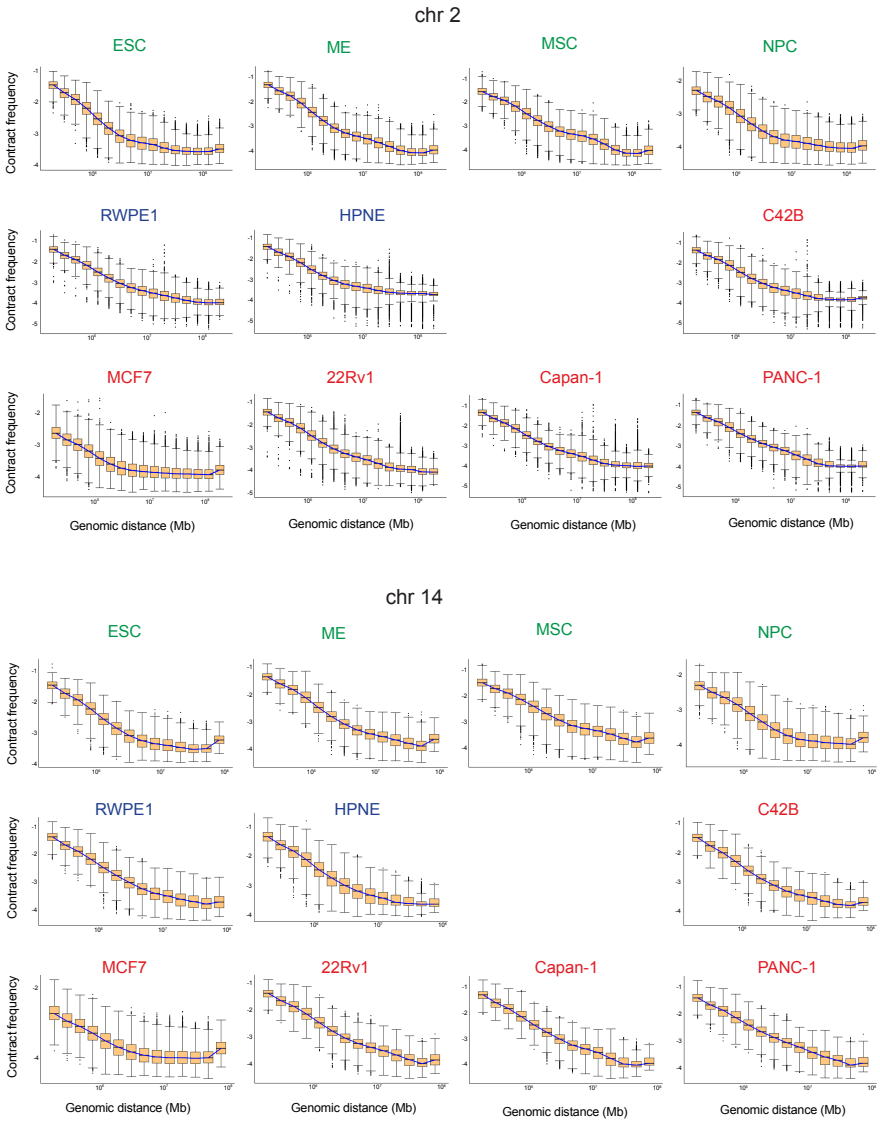
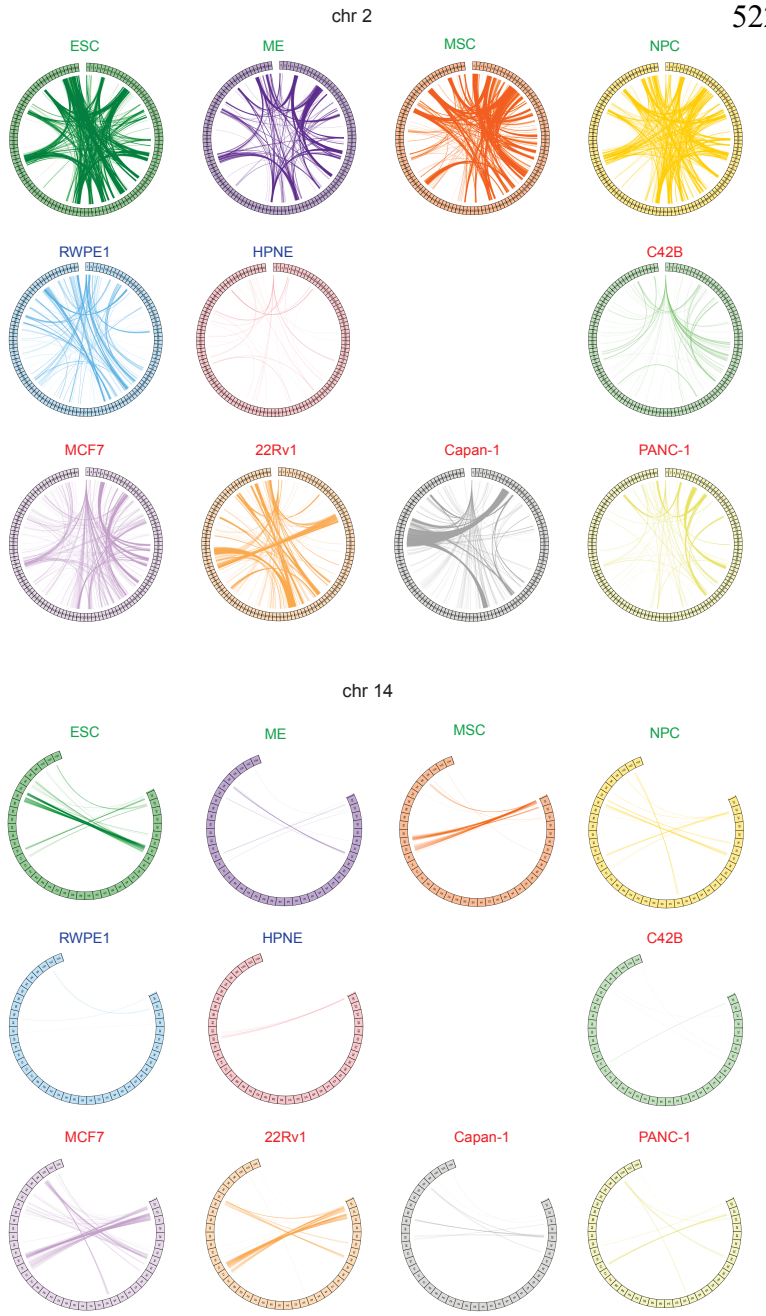


Fig. S2. Prevalent of abnormally high contact frequencies in additional Hi-C datasets. Box plots of 11 additional Hi-C datasets show the widespread presence of genomic locus pairs separated by > 50 Mb in chr2 and chr14, exhibiting abnormally high contact frequencies.

522



523

524 **Fig. S3. Chromosome linkage maps of CLOD-identified long-range colocalization loci.**
525 Chromosome linkage maps depicting CLOD-identified genomic locus pairs (>50 Mb apart) in
526 chr2 and chr14 that exhibit abnormally high contact frequencies across 11 additional cell lines**.**

527

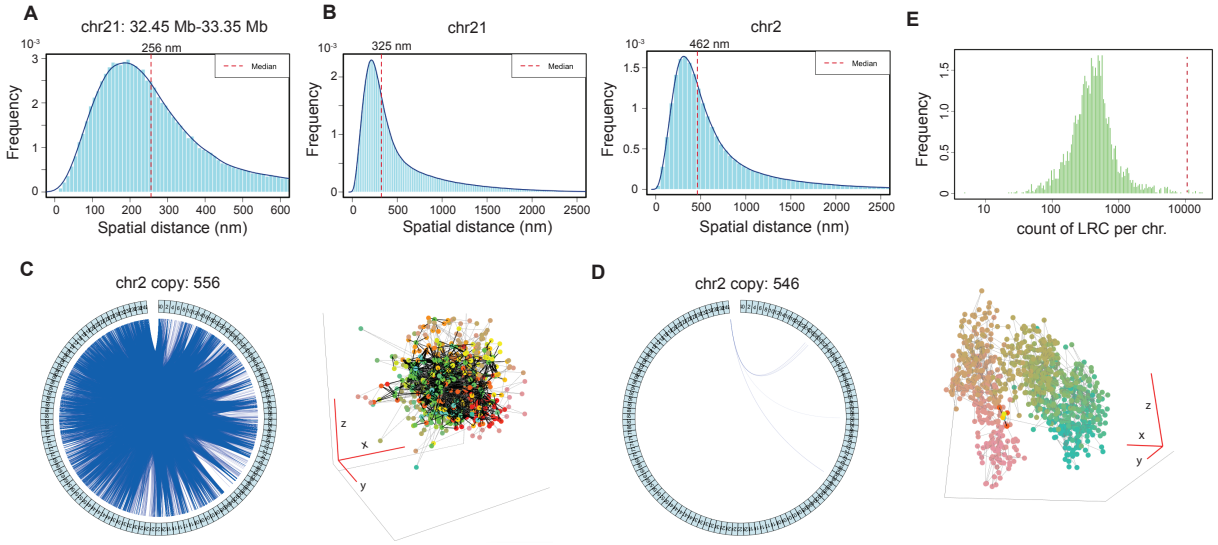


Fig. S4 Statistic analysis and representative examples of individual chr2 from MERFISH chromatin tracing data. (**A**) Histogram and density curve illustrating the 3D distance distribution of adjacent labeled loci in chr21:32.45-33.35 Mb, with the red dashed line indicating the median distance. (**B**) Histogram and density curves showing 3D distance distributions of adjacent labeled loci in chr2 (left) and chr21 (right), with red dashed lines representing the median distance between adjacent bins for each chromatin. **(C-D)** Linkage maps of LRCs and 3D structures of representative chr2 copies, where black lines denote LRCs and gray lines indicate neighbor links. **(E)** Histogram depicting the distribution of LRC counts per individual chr2 copy.

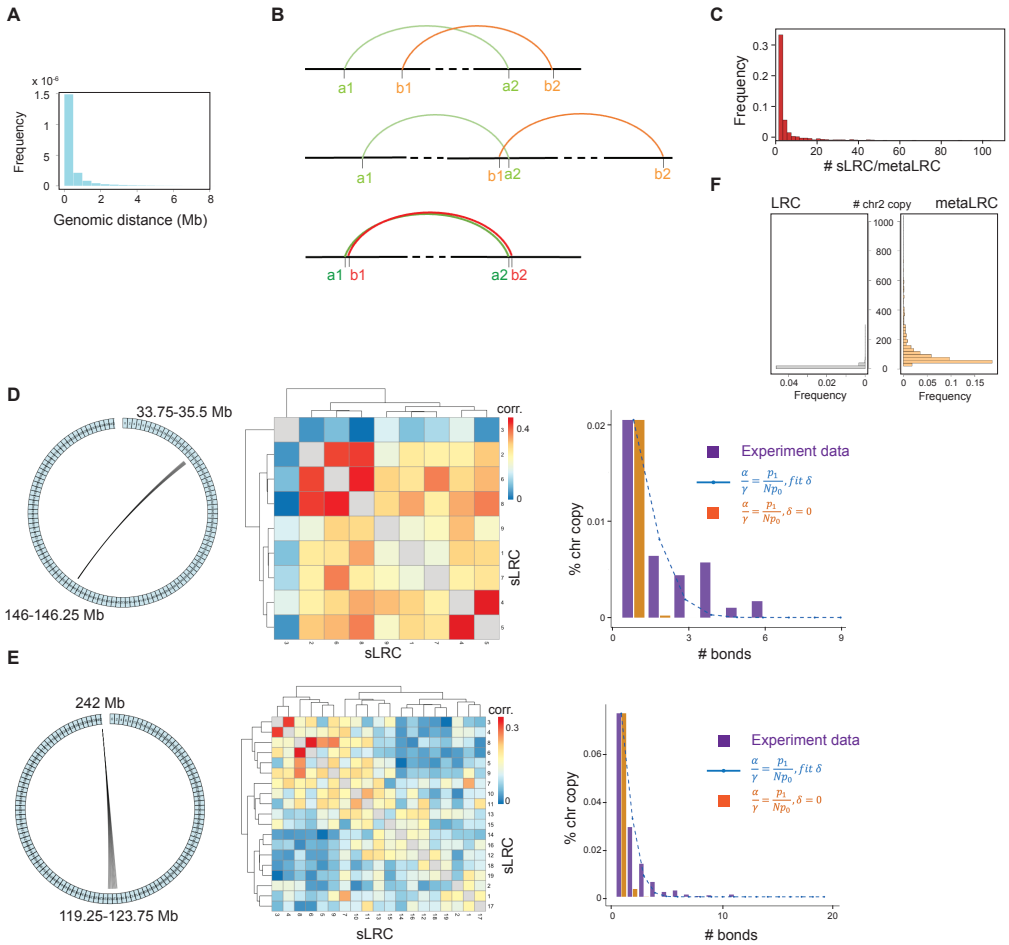


Fig. S5. Additional results for multivalence binding model analyses. (**A**) Histogram depicting the distribution of genomic distances between two adjacent loci involved in sLRC. (**B**) Sketch plots illustrating three typical types of genomic relationships between sLRC a and sLRC b. (**C**) Density distribution of the number of sLRC pairs per metaLRC. (**D-E**) Linkage maps (left) of two representative metaLRCs, their heatmaps of correlation between sLRCs (middle), and probability comparisons (right) of observing various numbers of bound sLRCs within the metaLRC in panel E. The comparison includes experimentally observed frequencies (purple bars), predicted independent sLRC binding (orange bar), and the multivalent binding model (blue dashed line). p_0 is not shown. (**F**) Occurrence frequency of individual sLRCs (in gray) and metaLRCs (in orange) among the 2,991 copies of chr2.

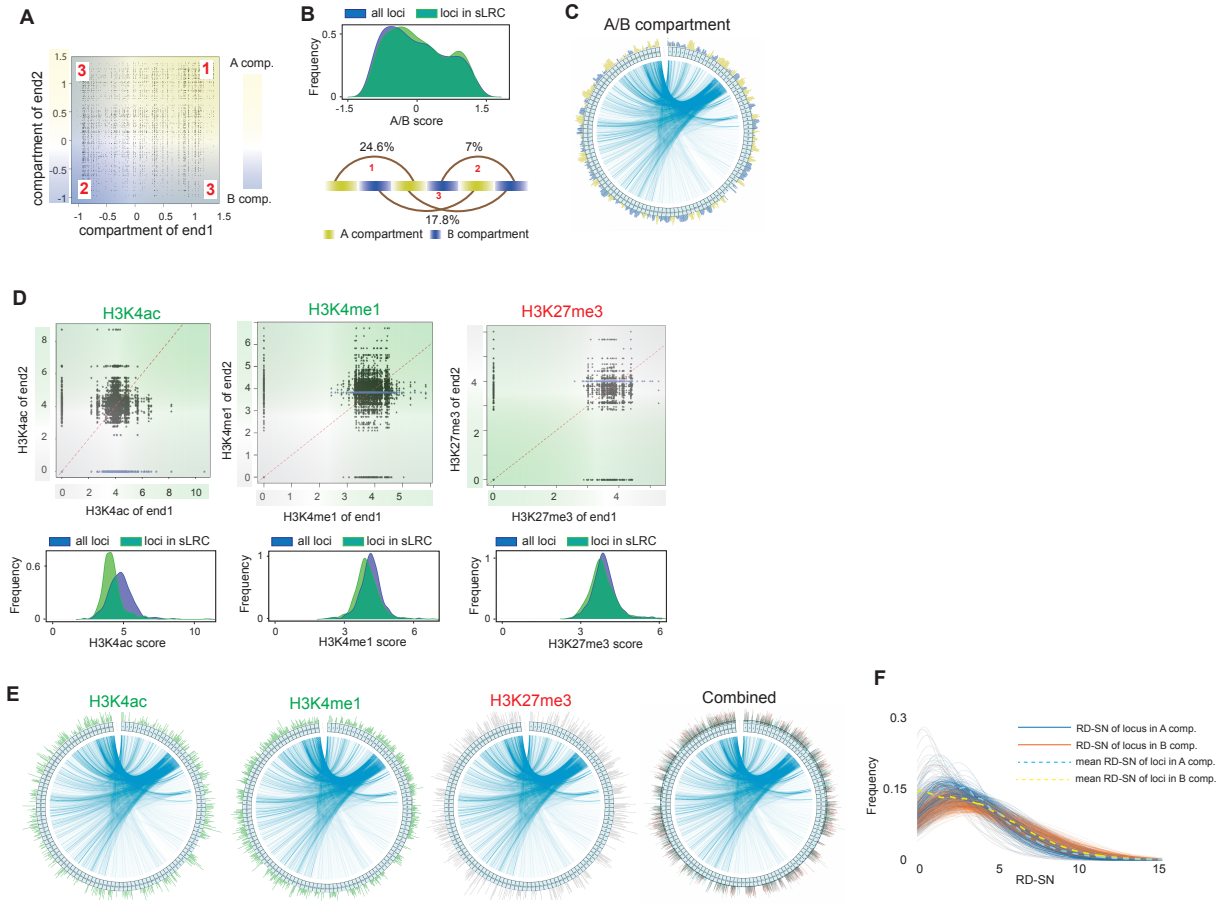


Fig. S6. Characteristics of loci associated with sLRCs. (A) A/B compartment classification of locus pairs involved in sLRCs. **(B)** Top: Comparison of A/B compartment score distributions between all loci and those involved in sLRCs. Bottom: Schematic illustration showing the A/B compartment characteristics of three different types of sLRCs from panel A. **(C)** Genomic distribution of the A/B compartments along chr2. **(D)** Top: Epigenetic characteristics of locus pairs forming sLRCs. Bottom: Comparison of epigenetic scorer distributions between all loci and those involved in sLRCs. **(E)** Genomic distribution of epigenetic characteristics along chr2. **(F)** RD-SN distributions of sLRC loci within A or B compartments, represented in blue and orange, respectively. Mean values for loci in A and B compartments are indicated by dashed lines.

562 ## Table S1 HiC datasets used in CLOD analyses.

Cell type	Dataset source	References
ESC	GSE52457	(*24*)
ME	GSE52457	(*24*)
MSC	GSE52457	(*24*)
NPC	GSE52457	(*24*)
RWRE1	GSE118629	(*25*)
C42B	GSE118629	(*25*)
HPNE	GSE149103	(*26*)
MCF7	GSE66733	(*13*)
MCF10A	GSE66733	(*13*)
22Rv1	GSE118629	(*25*)
Capan-1	GSE149103	(*26*)
PANC-1	GSE149103	(*26*)
IMR90	GSE63525	(*27*)
HEK293T	GSE44267	(*28*)
HCT116	GSE104333	(*15*)

563

564 ## Table S2: Summary of tested sgRNA

FL	**Sequence**	**Spread location**	**# on target**	**# on other sites**
GFP	AGCAACAACCTAGCATCTC AGG	105,229,722 – 105,240,862	107	0
RFP	TGGAGGCCTGTGGAGGCCTG TGG + AGGTCTGCGGGAGCCTGTGG AGG	22,554,200 – 22,554,453 + 22,554,223 – 22,554,476	5 + 5	0

565

566

567 **References and Notes**

568 1. K. A. Dill, J. L. MacCallum, The Protein-Folding Problem, 50 Years On. *Science* **338**,
569 1042 (2012).
570 2. A. Bolzer *et al.*, Three-dimensional maps of all chromosomes in human male fibroblast
571 nuclei and prometaphase rosettes. *PLoS Biol.* **3**, e157 (2005).
572 3. E. Lieberman-Aiden *et al.*, Comprehensive Mapping of Long-Range Interactions Reveals
573 Folding Principles of the Human Genome. *Science* **326**, 289 (2009).
574 4. S. Wang *et al.*, Spatial organization of chromatin domains and compartments in single
575 chromosomes. *Science* **353**, 598 (2016).
576 5. A. Grosberg, Y. Rabin, S. Havlin, A. Neer, Crumpled Globule Model of the Three-
577 Dimensional Structure of DNA. *Europhysics Letters* **23**, 373 (1993).
578 6. G. Gürsoy, Y. Xu, A. L. Kenter, J. Liang, Spatial confinement is a major determinant of
579 the folding landscape of human chromosomes. *Nucleic Acids Research* **42**, 8223-8230
580 (2014).
581 7. K. E. Polovnikov *et al.*, Crumpled Polymer with Loops Recapitulates Key Features of
582 Chromosome Organization. *Physical Review X* **13**, 041029 (2023).
583 8. J. Mateos-Langerak *et al.*, Spatially confined folding of chromatin in the interphase
584 nucleus. *Proceedings of the National Academy of Sciences* **106**, 3812-3817 (2009).
585 9. S. A. Quinodoz *et al.*, Higher-Order Inter-chromosomal Hubs Shape 3D Genome
586 Organization in the Nucleus. *Cell* **174**, 744-757.e724 (2018).
587 10. R. Jäger *et al.*, Capture Hi-C identifies the chromatin interactome of colorectal cancer
588 risk loci. *Nature communications* **6**, 6178 (2015).
589 11. A. Pombo, N. Dillon, Three-dimensional genome architecture: players and mechanisms.
590 *Nature Reviews Molecular Cell Biology* **16**, 245-257 (2015).
591 12. Warren A. Whyte *et al.*, Master Transcription Factors and Mediator Establish Super-
592 Enhancers at Key Cell Identity Genes. *Cell* **153**, 307-319 (2013).
593 13. A. R. Barutcu *et al.*, Chromatin interaction analysis reveals changes in small chromosome
594 and telomere clustering between epithelial and breast cancer cells. *Genome Biology* **16**,
595 214 (2015).
596 14. F. Jin *et al.*, A high-resolution map of the three-dimensional chromatin interactome in
597 human cells. *Nature* **503**, 290-294 (2013).
598 15. S. S. P. Rao *et al.*, Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320.e324
599 (2017).

16. J.-H. Su, P. Zheng, S. S. Kinrot, B. Bintu, X. Zhuang, Genome-scale imaging of the 3D organization and transcriptional activity of chromatin. *Cell* **182**, 1641-1659. e1626 (2020).
17. B. Bintu *et al.*, Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**, eaau1783 (2018).
18. J.-H. Su, P. Zheng, S. S. Kinrot, B. Bintu, X. Zhuang, Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641-1659.e1626 (2020).
19. I. Chepelev, G. Wei, D. Wangsa, Q. Tang, K. Zhao, Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization. *Cell research* **22**, 490-503 (2012).
20. E. E. Furlong, M. Levine, Developmental enhancers and chromosome topology. *Science* **361**, 1341-1345 (2018).
21. J. Dekker, L. A. Mirny, The chromosome folding problem and how cells solve it. *Cell* **187**, 6424-6450 (2024).
22. E. H. Finn *et al.*, Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell* **176**, 1502-1515.e1510 (2019).
23. X. J. Tian, H. Zhang, J. Sannerud, J. Xing, Achieving diverse and monoallelic olfactory receptor selection through dual-objective optimization design. *Proceedings of the National Academy of Sciences of the United States of America* **113**, E2889-2898 (2016).
24. J. R. Dixon *et al.*, Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**, 331-336 (2015).
25. S. K. Rhie *et al.*, A high-resolution 3D epigenomic map reveals insights into the creation of the prostate cancer transcriptome. *Nature Communications* **10**, 4154 (2019).
26. B. Ren *et al.*, High-resolution Hi-C maps highlight multiscale 3D epigenome reprogramming during pancreatic cancer metastasis. *Journal of Hematology & Oncology* **14**, 120 (2021).
27. S. S. Rao *et al.*, A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* **159**, 1665-1680 (2014).
28. J. Zuin *et al.*, Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences* **111**, 996-1001 (2014).