nature structural & molecular biology

ANALYSIS

https://doi.org/10.1038/s41594-021-00678-3

Check for updates

Three-dimensional genome organization via triplex-forming RNAs

Irene Farabella[1]✉, Marco Di Stefano[1,5], Paula Soler-Vila[1,6], Maria Marti-Marimon[1] and Marc A. Marti-Renom[1,2,3,4]✉

An increasing number of long noncoding RNAs (lncRNAs) have been proposed to act as nuclear organization factors during interphase. Direct RNA-DNA interactions can be achieved by the formation of triplex helix structures where a single-stranded RNA molecule hybridizes by complementarity into the major groove of double-stranded DNA. However, whether and how these direct RNA-DNA associations influence genome structure in interphase chromosomes remain poorly understood. Here we theorize that RNA organizes the genome in space via a triplex-forming mechanism. To test this theory, we apply a computational modeling approach of chromosomes that combines restraint-based modeling with polymer physics. Our models suggest that colocalization of triplex hotspots targeted by lncRNAs could contribute to large-scale chromosome compartmentalization cooperating, rather than competing, with architectural transcription factors such as CTCF.

In the cell nucleus, chromatin is organized in space at multiple scales, from nucleosome positioning to chromosome territories. At a scale of tens to hundreds of kilobases, chromosomes fold into chromatin domains, also referred to as topologically associating domains (TADs)[1–3], defined by boundaries occupied by architectural chromatin-binding proteins (for example, CTCF and cohesin SMC complex) or noncoding RNAs (ncRNAs)[4–6]. At a larger scale, chromosomes spatially segregate into regions of preferential long-range interaction, consisting predominantly of transcriptionally active (A) or inactive (B) chromatin compartments[7]. The proposed mechanisms underlying compartmentalization are generally distinct from those forming TADs and loops[6,8]. Indeed, the multiple level of chromosome organization has been shown to be the result of various coexisting mechanisms, some of which involve protein macromolecular complexes, including polycomb-group proteins, cohesin, CTCF and Mediator complex[9–11]. Others involve repetitive elements, the most common being simple sequence repeats that comprise up to half of the human genome[12–15].

Additionally, increasing evidence suggests that nuclear-retained RNA has a role as a genome structure regulator at various levels of genome organization[16–22]. For example, nuclear-retained RNA is an integral component of chromatin[23], constituting up to 10% of chromatin mass[24] and can affect local structural organization of chromatin[25], as well as serve as a guide to facilitate the recruitment of chromatin factors[26]. Within the nuclear space, RNA also forms an intricate network of compartments at varying RNA density, with high-density RNA regions being associated with transcriptionally active chromatin and preferentially located in the interface with inactive ones[16]. Finally, RNA molecules have been shown to play an active role in governing phase-separation effects related to spatial organization of the nucleus[16,27–29].

An increasing number of lncRNAs have been proposed to act as nuclear organization factors in interphase and mitotic cells[30,31]. For instance, some lncRNAs such as NEAT1 or MALAT1 are fundamental components of nuclear bodies[32] while others have chromatin-related functions that may influence chromosomal three-dimensional (3D) structure. For example, Xist silences the X chromosome by altering chromatin compaction via a 3D proximity-guided search[33]. ASAR15 plays a role in coordinating the synchronous replication of homologous autosomal chromosomes and in maintaining their stability[34]. Other ncRNAs are stably associated with interphase chromosome territories[35,36], and their disruption leads to chromatin condensation[36]. Further, lncRNAs could influence changes in gene expression or chromosome activity through various mechanisms such as recruitment of chromatin-modifying proteins via direct DNA binding acting in *cis* or in *trans*[37–39].

However, given the structural complexity of RNAs organized into combinations of discrete functional domains[40], the precise mechanisms by which RNAs associate with chromatin and modulate its overall structural organization are largely unknown. On the one hand, the association of RNA with chromatin can be mediated by the formation of ribonuclear-protein assemblies (indirect association), as in the case of SAF-A–RNA interactions that promote the formation of a chromatin mash responsive to transcription and which regulates local chromatin structure in an ATP-depended manner[41], or as the locus-specific targeting of CTCF mediated by RNA interaction[42]. Moreover, several techniques that localize RNAs on chromatin have recently confirmed that a substantial proportion of RNA transcripts is directly associated with chromatin[18,21,22]. Direct RNA-DNA interactions can occur via the formation of triplex helix structures (single-strand RNA hybridizing via noncanonical base pairing with the major groove of double-stranded DNA), which permits RNA to bind DNA in a sequence-specific manner with no requirement for disruption of the double-helix complementary base pairing[43]. RNA–DNA triplex formation has also been shown to affect chromatin state through the recruitment of epigenetic modifiers, particularly when the interacting RNA in the triplex structure is a lncRNA. This is the case for MEG3, HOTAIR and PARTICLE lncRNAs that directly interact with GA-rich chromatin motifs to

[1]CNAG-CRG, Centre for Genomic Regulation, Barcelona Institute of Science and Technology, Barcelona, Spain. [2]Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation, Barcelona, Spain. [3]Universitat Pompeu Fabra, Barcelona, Spain. [4]ICREA, Barcelona, Spain. [5]Present address: Institute of Human Genetics, Centre National de la Recherche Scientifique, University of Montpellier, Montpellier, France. [6]Present address: Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain. ✉e-mail: irene.farabella@cnag.crg.eu; martirenom@cnag.crg.eu

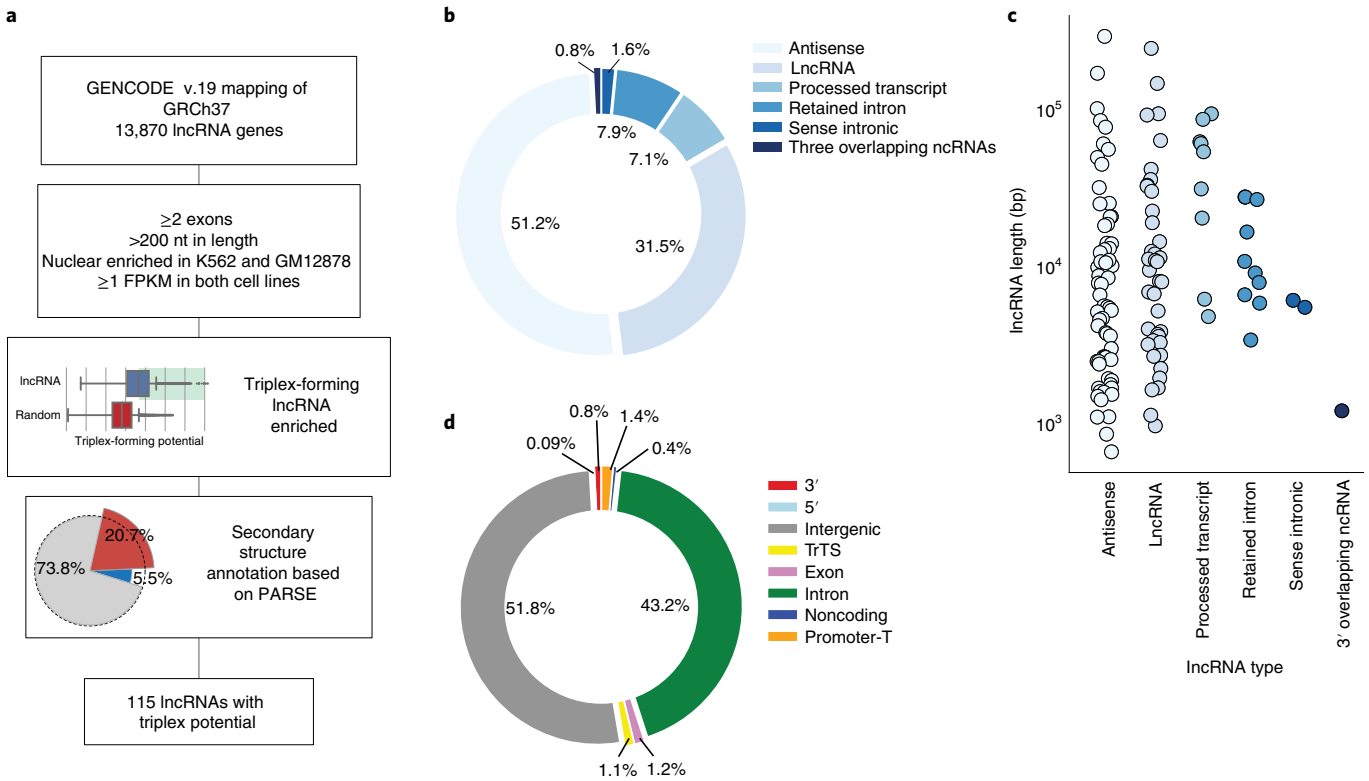


Fig. 1 | Characterization of lncRNAs with triplex-forming potential. a, Flowchart of lncRNA selection criteria. **b**, Classification of lncRNAs engaged in triplex structures based on transcript type. **c**, Length distribution of lncRNAs with triplex-forming potential, classified based on the categories shown in **b**. **d**, Genetic sequence features of TrTS engaging in triplex formation with lncRNAs.

facilitate recruitment of polycomb-repressive complex 2 to target gene promoters[37–39].

Here, we hypothesize that direct RNA-DNA interaction via the formation of triplex helix structures could play a role in spatial structuring of the genome. To test this hypothesis, we used an integrative modeling protocol, in line with other modeling methods that have been already used to provide insight into the multiple levels of chromosome organization[44–49]. Our modeling protocol, similarly to other methods[46,47], spatially brings close pairs of genomic loci enriched with predicted triplex-forming sequences targeted by selected lncRNAs, with the aim of reconstructing 3D models of entire chromosomes. The resulting structural models revealed that genomic triplex target sites targeted by selected lncRNAs correlate with experimentally derived Hi-C maps and promote spatial segregation between compartment types, possibly by creating an RNA-rich interface between A and B compartments. Our models also suggest that this large-scale chromatin structure segregation effect may complement the local action of architectural proteins such as CTCF.

Results

lncRNAs with triplex-forming potential. Long noncoding RNA transcripts (that is, multiexonic RNA transcripts >200 nucleotides (nt) with triplex-forming potential were selected from the list of lncRNAs available in GENCODE v.19 (ref. [50] for the human genome (GRCh37)). The selected lncRNAs were filtered based on three criteria (Fig. 1a). First, because it was necessary that lncRNA be cell type unspecific, the selected transcripts were nuclear enriched with an expression level ≥1 fragment per kilobase per million (FPKM) in cell lines K562 and GM12878. Second, it was required that lncRNAs result in high length-normalized RNA–DNA triplex-forming potential[43] compared to a random selection of transcripts ($P < 0.00001$, Kolmogorov–Smirnov two-sample test). And third, lncRNA had to harbor putative triplex-forming oligos (TFOs) in a non-double-stranded portion of the transcript based on experimentally available RNA secondary structure profiles[51]. These three filters ensured that the selected lncRNAs were expressed in several cell lines and had the potential to form triplex in a single-stranded part of the transcript. In total, we collected a set of 115 lncRNAs that passed these selection criteria and, as such, could potentially interact directly with DNA (Supplementary Table 1).

Consistent with experimentally detected RNAs engaged in triplex structures[22], the selected lncRNAs were mainly antisense to either a known protein-coding gene RNA or long intergenic noncoding RNA (51.6% and 31.5%, respectively; Fig. 1b) and were variable in length, ranging from 6.2 to 285.9 kb (Fig. 1c). Distribution along the human genome of triplex target sites (TrTSs) engaging in triplex formation with lncRNAs indicated that the latter target mainly intergenic regions (51.8%) or introns of protein-coding genes (43.2%) (Fig. 1d).

TrTS hotpots in the 3D genome. To examine whether TrTSs form 3D clusters, we investigated the presence of colocalizing TrTS (co-TrTS) hotspots, defined as highly interacting sites on GM12878 intrachromosomal Hi-C maps[5] at 50-kb resolution (Methods). Depending on the selected lncRNA and regardless of its length (Extended Data Fig. 1), a median of 18.5% (range 1.4–95.4%) of TrTSs significantly interacted. The vast majority of co-TrTS interacting sites were between distal intrachromosomal loci (~90% >5 Mb and ~2% <1 Mb apart; Fig. 2a). This is in line with previous observations that lncRNA transcripts form direct RNA-DNA interactions at long genomic distances[18]. These co-TrTSs were rich in purine nucleotides (Extended Data Fig. 1b) and depleted in DNase I hypersensitive sites probably devoid of nucleosomes (as DNA

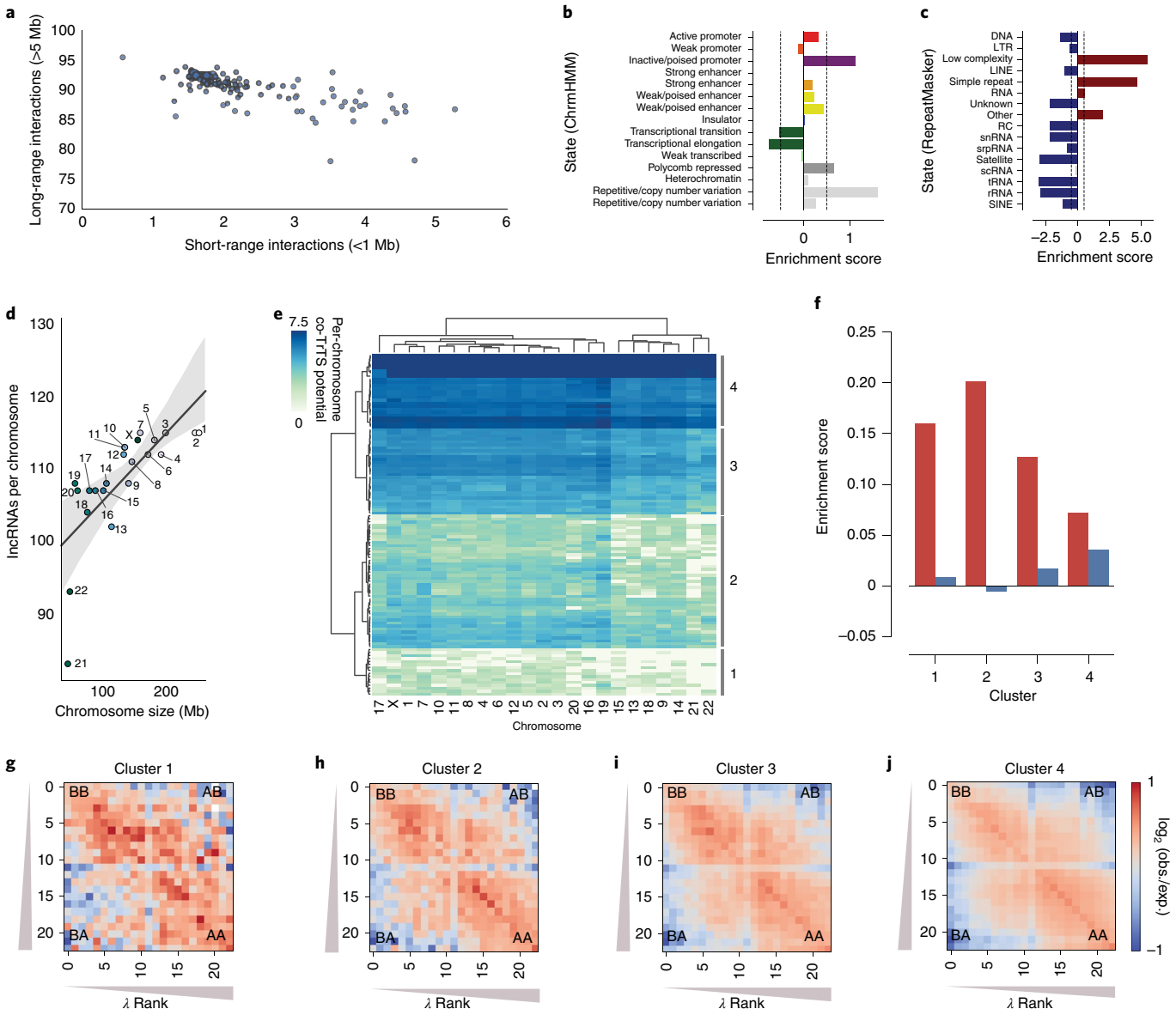


Fig. 2 | Characterization of co-TrTS hotspots in the genome. a, Percentage of co-TrTS hotspots engaging in long-range (>5 Mb) and short-range (<1 Mb) interactions. **b**, Enrichment of co-TrTS hotspots in 15 different chromatin states, as defined by the Roadmap Epigenomics project[94]. **c**, Enrichment of co-TrTS hotspots in different classes of repeat elements, as classified by RepeatMasker[95]. **d**, Amount of lncRNAs with triplex potential targeting per chromosome depending on chromosome size. **e**, Heatmap diagram of two-way hierarchical clustering of 115 triplex-forming lncRNAs (rows) over the 23 chromosomes (columns) based on co-TrTS potential (Methods). **f**, A/B compartment enrichment score of co-TrTS hotspots per cluster identified in **e**. **g**–**j**, Extent of compartmental interactions across different triplex-forming lncRNA clusters (1, **g**; 2, **h**; 3, **i**; 4, **j**) in *cis*. Saddle plots were calculated using the first eigenvector obtained from Hi-C data[5] at 100-kb resolution. Upper left and lower right: homotypic interactions between B-B and A-A compartments, respectively; top right and lower left: heterotypic A-B interactions. λ ranking denotes eigenvector components. Obs, observed; exp, expected.

linker) (Extended Data Fig. 1c), which is consistent with in vivo observations[22,52]. Indeed, in vivo nucleosomes specifically stabilized triple single-stranded RNA–double-stranded DNA helix structures located in the DNA linker at the nucleosomal entry-exit site. Most linker DNA in eukaryotes is torsionally relaxed[53], and thus devoid of local positive supercoiling that can disrupt Hoogsteen H-bonds disfavoring triplex interaction[54].

Furthermore, based on a 15-chromatin-state ChromHMM model specific for GM12878 (ref. [55]), co-TrTSs were depleted of transcriptionally related active states (such as transcription elongation and transition chromatin states, with −0.75- and −0.52-fold change, respectively; Fig. 2b). Interestingly, co-TrTSs were generally enriched in repetitive elements (1.6-fold change) and poised chromatin (poised-promoter and polycomb-repressed chromatin states, with 1.12- and 0.66-fold change, respectively; Fig. 2b). Moreover, co-TrTSs associated significantly more with simple-repeat and low-complexity regions (4.7- and 5.5-fold change, respectively; Fig. 2c), which suggests that these lncRNAs could target multiple repetitive genomic sequences. Finally, co-TrTSs for multiple lncRNAs were found in all chromosomes, with the longest harboring larger numbers of lncRNA co-TrTSs (Fig. 2d) independently of their gene density (Extended Data Fig. 1d).

To further investigate whether different lncRNAs have varying preference for a specific subset of chromosomes, we calculated a

lncRNA co-TrTS potential for each chromosome (Methods). Some lncRNAs showed preference for only a specific subset of chromosomes while others had a widespread per-chromosome co-TrTS potential. A total of four different clusters of lncRNA were identified from more (cluster 1) to less (cluster 4) chromosome-specific groups (Fig. 2e and Extended Data Fig. 1e). The lncRNA chromosome co-TrTS potential also resulted in two major groups of chromosomes being clustered in their propensity to host co-TrTS with the selected 115 lncRNAs (Fig. 2e). The first cluster includes chromosomes such as 9 and 18, as well as all acrocentric chromosomes (that is, 13, 14, 15, 21 and 22) that have high co-TrTS potential with cluster 3 and 4 lncRNAs. The second cluster includes all remaining chromosomes, which have high co-TrTS independently of lncRNAs clusters (Fig. 2e).

Next, we surveyed whether co-TrTSs involving different lncRNA clusters were preferentially located in chromatin compartment A (open) or B (closed), as obtained from GM12878 Hi-C interaction maps[5] at 100-kb resolution (Methods). Compared to genome-wide compartmentalization, bins in the genome containing lncRNA co-TrTSs were slightly enriched in differential compartments depending on their cluster (Fig. 2f). Bins of sites targeted by lncRNAs belonging to clusters 1 and 3 were enriched in the A compartment, with no clear enrichment in B. Cluster 4 was slightly enriched in both A and B. To assess whether those lncRNA co-TrTS genomic bins might be involved in the colocalization of compartments, we next plotted interactions between those bins ranked by their component in the first eigenvector (Fig. 2g–j). Interestingly, although to varying degrees, in all four clusters co-TrTS interactions were enriched in heterotypic (A-B) and depleted in homotypic (A-A or B-B) pairings (Fig. 2g–j) as compared to the genome-wide trend (Extended Data Fig. 1f). This result suggests that the selected lncRNAs could play a mechanistic role in compartment segregation, especially at the interphase between compartments A and B.

TrTSs as drivers of chromosomal conformation. To test whether the identified lncRNAs could act as 3D genome organizers by bringing spatially closer co-TrTSs, we focused on lncRNAs in cluster 4 with the highest per-chromosome co-TrTS potential genome wide. In particular, simulations were carried out for a total of seven lncRNAs, with ENST00000541775.1 lncRNA presented as an example case in all figures. Briefly, the identified pairs of bins enclosing co-TrTS hotspots for a given lncRNA were restrained to be spatially close with steered molecular dynamics simulations of entire chromosomes, using a protocol similar to one previously published[46,47]. Each chromosome was modeled with a homogeneous beads-on-a-string representation, in which each bead represents 50 kb of chromatin. At this coarse graining, the number of imposed restraints between co-TrTSs was <2.8% of all significant pairwise interactions in a Hi-C intrachromosomal matrix (Methods and Supplementary Table 2) that is, a very small percentage of all significant interactions in a chromosome that can potentially drive chromosome folding. The steering process was repeated independently 1,000 times for each of the 23 chromosomes. For each of the 1,000 runs only those conformations best satisfying the imposed restraints were retained to constitute the final ensemble of 3D chromosome models for further analysis. To validate our approach, we next compared the experimentally derived Hi-C interaction map to a contact map generated from the reconstructed ensemble of 3D models at 50-kb resolution (Methods, Fig. 3 and Extended Data Fig. 2). For example, the chr19 Hi-C interaction map displays the typical plaid pattern representing chromatin compartmentalization (Fig. 3a). Interestingly, this plaid pattern was observed in the contact map generated from the ENST00000541775.1 ensemble of models for chr19 (Fig. 3b and Supplementary Video 1). To assess the significance of this observation, we defined a positive control where 3D models were reconstructed to impose restraints between pairs of significantly interacting loci that contained chromatin immunoprecipitation sequencing (ChIP–seq) peaks of CTCF transcription factor, a well-known architectural protein involved in TAD organization[6,56] (Methods and Fig. 3c). As a negative control, 3D models were also reconstructed using randomly selected significant Hi-C interacting pairs of loci (Methods and Extended Data Fig. 2a). The compartmentalization-derived plaid pattern appeared less pronounced for CTCF-derived models and was totally absent from the ensemble of models based on randomly selected interacting pairs. Indeed, as measured by mean diagonal cross-correlation coefficient (dCCC; Methods), 3D models reconstructed by imposing ENST00000541775.1 co-TrTS colocalization resulted in contact patterns significantly more similar to the Hi-C than 3D models reconstructed with the CTCF or random pairs of sites (mean dCCC = 0.58, 0.41 and 0.22 for ENST00000541775.1, CTCF and Random, respectively; Mann–Whitney rank test, $^{***}P < 10^{-3}$; Fig. 3d and Extended Data Fig. 2c–e). Additionally, we derived a control where 3D models were reconstructed imposing restraints between pairs of significantly interacting loci enclosing low-complexity sites, to test the specific forming ability of co-TrTS (Methods and Extended Data Fig. 2b). Three-dimensional models reconstructed using ENST00000541775.1 co-TrTS colocalization resulted in contact patterns significantly more similar to the Hi-C than the 3D models reconstructed with low-complexity pairs of sites (mean dCCC = 0.58 and 0.50 for ENST00000541775.1 and low-complexity models, respectively; Mann–Whitney rank test, $^{***}P < 10^{-3}$; Fig. 3d and Extended Data Fig. 2c,f). Given that the co-TrTSs targeted by ENST00000541775.1 overlap consistently (63%) with low-complexity pairs of bins, this result highlights that interactions involving co-TrTSs targeted by lncRNA suggest a more accurate reconstruction of chr19 than all the repetitive elements together. This piece of evidence reinforces the fact that co-TrTSs specifically targeted by lncRNAs could be driving compartmentalization.

The results obtained for the ENST00000541775.1 ensemble of models for chr19 were consistent along the entire genome (Fig. 3e), as measured by the element-wise Spearman cross-correlation coefficient (spCCC; Methods). All chromosomes resulted in higher spCCC than the CTCF-driven models, with the exception of chromosomes 1 and 17. Moreover, all but chromosomes 1, 2 and 9 resulted in contact maps with spCCC > 0.5 (Fig. 3e). Notably, 3D models of acrocentric chromosomes (13, 14, 15, 21 and 22), as well as of chr19, resulted in the highest agreement with experimental Hi-C maps (Fig. 3f). It is important to note that the results obtained from simulations are independent of the percentage of satisfied restraints in the models. Indeed, on average, CTCF-imposed restraints were equally or even better satisfied than those derived from lncRNA (Extended Data Fig. 3a), which rules out the possibility that improved reconstruction can be simply accounted for by greater compliance of the steering process.

Next, to assess whether the observed correlations between contact maps from the ensemble of models with the experimental Hi-C also hold true for lncRNAs apart from ENST00000541775.1, we reconstructed 3D models for six additional lncRNAs from cluster 4: ENST00000561611.2, ENST00000547963.1, ENST00000540866.2, ENST00000421202.1, ENST00000449111.1 and ENST00000434346.1. Overall, the calculated contact frequency maps from these ensembles of models highly correlated (spCCC > 0.5) with experimental Hi-C data. Similarly to ENST00000541775.1, the 3D structures of acrocentric chromosomes and chr19 were better reconstructed than larger chromosomes such as 1 or 2 (Fig. 3f). Interestingly, despite a similar trend, some chromosomes were better reconstructed by different lncRNAs (Fig. 3f), although the triplex-forming potential of the selected lncRNAs did not correlate with the capacity to reconstruct genomic structure (Extended Data Fig. 3b). We also reconstructed 3D models of chr22 using as restraints all pairs of bins enclosing co-TrTS

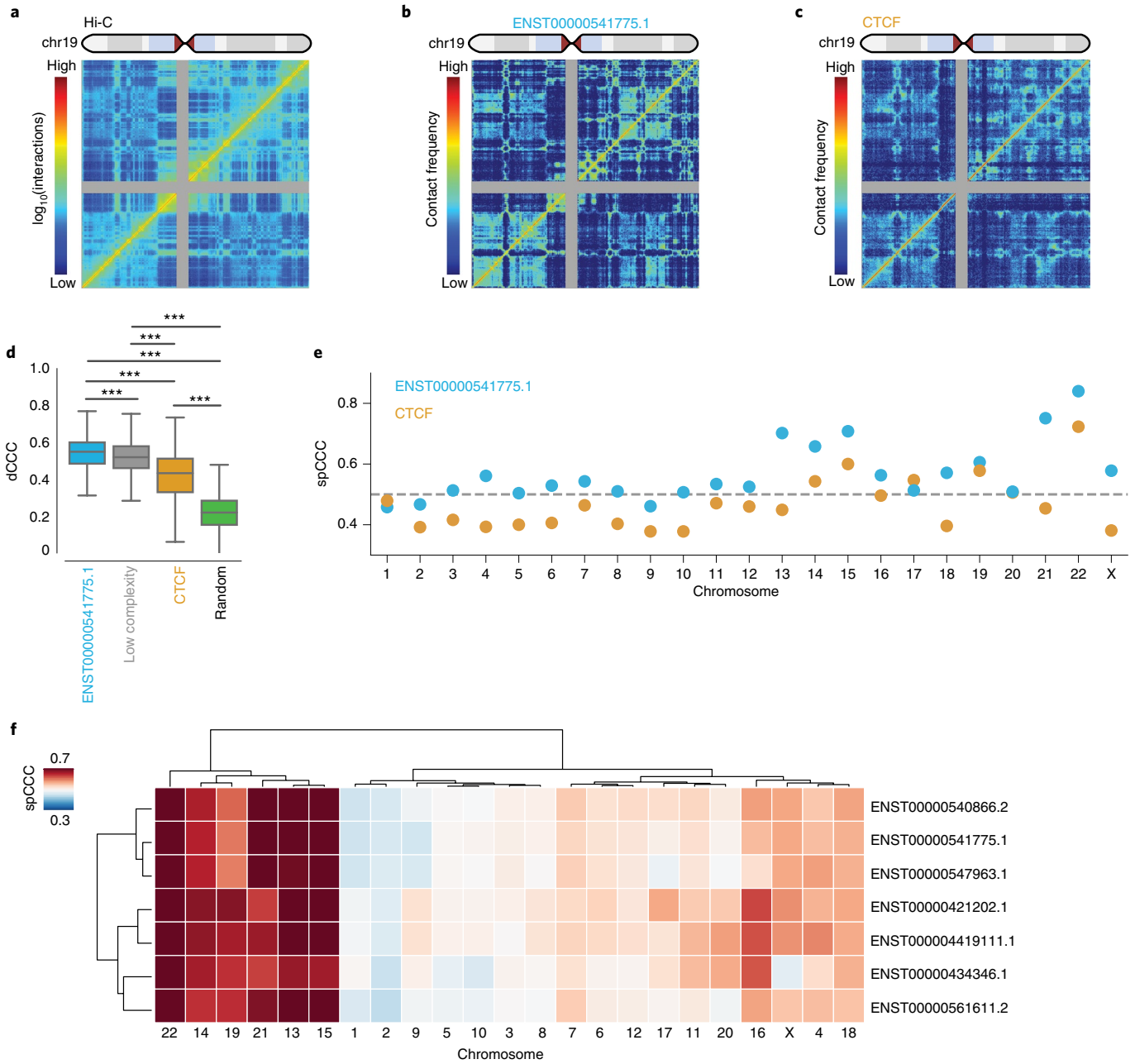


Fig. 3 | Co-TrTSs as drivers of chromosomal conformation. a, Hi-C normalized interaction matrix at 50-kb resolution for human chr19 (ref. [5]). **b**,**c**, Contact maps derived from the ensemble of 3D models generated using colocalizing pairs of loci driven by ENST00000541775.1 co-TrTS hotspots (**b**) and colocalizing CTCF-enriched genomic site pairs (**c**). **d**, Distribution of ddCCC (Methods) of the contact maps derived from the ensemble of 3D models for chr19. Box boundaries represent first and third quartiles, the middle line represents the median and whiskers extend to 1.5× interquartile range (two-sided Mann–Whitney rank test using python default parameters, $^{***}P < 10^{-3}$; $n = 1{,}182$ beads in chr19). **e**, Element-wise spCCC between Hi-C and model contact maps. **f**, spCCC between experimental Hi-C contact map and that derived from the ensemble of 3D models generated using colocalizing pairs of loci driven by the co-TrTS hotspots of seven representative lncRNAs with triplex potential belonging to cluster 4.

hotspots for lncRNAs with genome-wide high per-chromosome co-TrTS potential (that is, from cluster 4). On chr22, on average, a TrTS hotspot was targeted by ten lncRNAs suggesting that co-TrTSs could be shared by many lncRNAs. Three-dimensional models reconstructed using all pairs of bins enclosing co-TrTS hotspots for cluster 4 as restraints resulted in significantly more similar contact maps to the experimental Hi-C datasets compared to models reconstructed based on all but one of the lncRNAs alone (Extended Data Fig. 4a). Overall, our simulations on chr22 indicate that combining different lncRNAs results in better reconstruction than the majority (6/7) of individual lncRNAs, but no better than the best lncRNA for that specific chromosome, suggesting that there may be competition between individual lncRNAs and that there might be no synergy by combining the restraints detected for each lncRNA.

In summary, colocalizion of TrTSs targeted by specific lncRNAs guided an accurate reconstructional 3D chromosomal organization based on polymer physics models, driven by a very limited number of restraints (<2.8% of the total significant interaction in a chromosome). Importantly, this effect was mostly observed for chromosome compartmentalization.

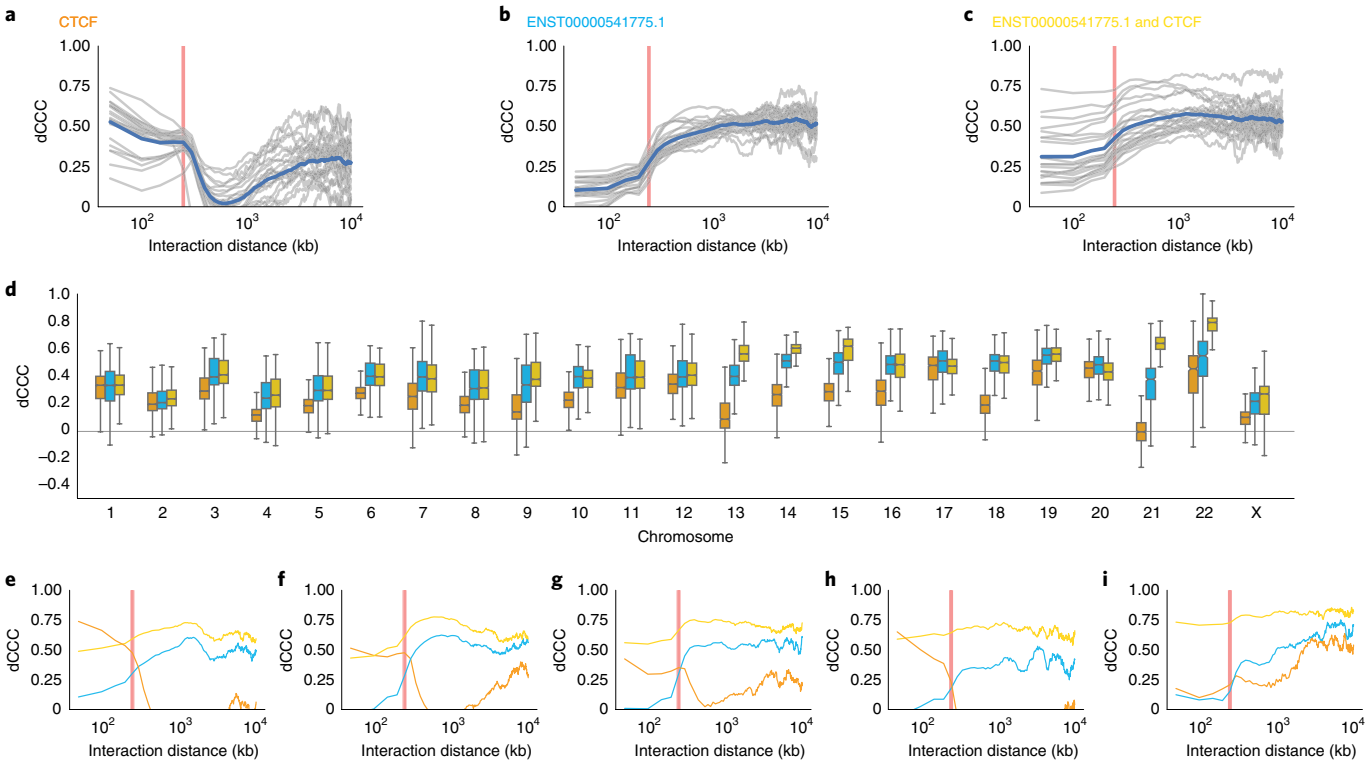


Fig. 4 | Cooperative effects of CTCF and ENST00000541775.1 on 3D genome structure. a–c, dCCC values for ≤10 Mb of genomic distance between experimental Hi-C contact map and contact maps derived from the ensemble of 3D models generated for each of the 23 chromosomes (light gray) and genome-wide average (blue). Vertical red bars denote 250 kb, the median genomic separation between convergent CTCF sites[61]. Models based on CTCF-enriched genomic loci (**a**), ENST00000541775.1 co-TrTS hotspots (**b**) and CTCF- and ENST00000541775.1-enriched genomic loci (**c**). **d**, dCCC distribution for each chromosome of contact maps derived from the ensemble of 3D models generated using ENST00000541775.1 co-TrTS hotspots (light blue), CTCF-enriched genomic loci (orange) and CTCF- and ENST00000541775.1-enriched genomic loci (yellow). Box boundaries represent the first and third quartiles, the middle line represents the median and whiskers extend to 1.5× interquartile range; *n*, number of beads in each chromosome. **e–i**, dCCC values for ≤10 Mb of genomic distance between the experimental Hi-C contact map and contact maps derived from the ensemble 3D models from CTCF-enriched genomic loci (orange), ENST00000541775.1 co-TrTS hotspots (light blue), and CTCF- and ENST00000541775.1-enriched genomic loci (yellow) for acrocentric chromosomes, chr13 (**e**), chr14 (**f**), chr15 (**g**), chr21 (**h**) and chr22 (**i**). Vertical red bars denote 250 kb.

Cooperative effect on 3D genome structure of CTCF and lncRNAs. The processes underlying the formation of TADs and loops are distinct from those forming compartments[6,8,57]. In fact, it has been proposed that TADs and loops are formed by the loop-extrusion model involving CTCF and cohesin complexes[58–60]. However, the molecular factors that drive compartmentalization in vivo remain unknown. Our results suggest that the identified lncRNAs, through the formation of triplex with DNA, could drive compartment formation. Next, we examined the interplay between the identified lncRNAs driving compartment formation and the formation of local structures driven by CTCF-CTCF interactions. The 3D models restrained by pairs of enriched CTCF sites resulted in high correlation at short genomic distances (Fig. 4a), with a significant dCCC decay at about 250 kb of genomic distance, which is the median between convergent CTCF loops in the human genome[61]. These results confirmed that models restrained by a small percentage of CTCF pairwise sites were sufficient to reconstruct genome organization at short genomic distances, but could not recover Hi-C interactions at larger genomic distances. In contrast, 3D models restrained by lncRNA co-TrTSs resulted in contact maps that correlated best at large genomic distances (>1 Mb; Fig. 4b for lncRNA ENST00000541775.1 and Extended Data Fig. 5a–f for other simulated lncRNAs).

Next, we asked whether these two driving forces act cooperatively to ensure chromosomal organization. To test this hypothesis, we generated an ensemble of models using as restraints both CTCF and ENST00000541775.1 co-TrTS pairs of loci. The resulting 3D models had an overall higher dCCC for all chromosomes at both local and non-local genomic distance (Fig. 4c). Strikingly, the CTCF- and ENST00000541775.1-driven models for the majority of chromosomes generally agreed better with the experimental data (Extended Data Fig. 3g–i), with an increase in overall correlation for acrocentric chromosomes (Fig. 4d). Interestingly, for these chromosomes we observed a constant high correlation both at local- and non-local-scale structures (Fig. 4e–i and Extended Data Fig. 5g–i), with the ensemble of models driven by CTCF- and lncRNA-mediated interactions correlated as much as 0.8 with the experimental data for chr22. Overall, our results support the premise that the proposed local and non-local mechanisms of genome structure (that is, CTCF- and triplex-forming lncRNA, respectively) could synergize in the organization of the genome in space.

Compartment segregation via co-TrTSs. To characterize the relationship between long-range interactions involving co-TrTS hotspots and the 3D architecture of chromosomes, we investigated which compartments were more accurately reconstructed by our models. To do so, we decomposed the Hi-C experimental map by separating homotypic and heterotypic interactions and calculating the spCCC for each type of interaction (Methods). For the majority of chromosomes (18/23), homotypic B-type interactions were reconstructed more accurately by lncRNA co-TrTS-driven models than other types of compartmental interactions (Fig. 5a), suggesting that co-TrTSs are involved in

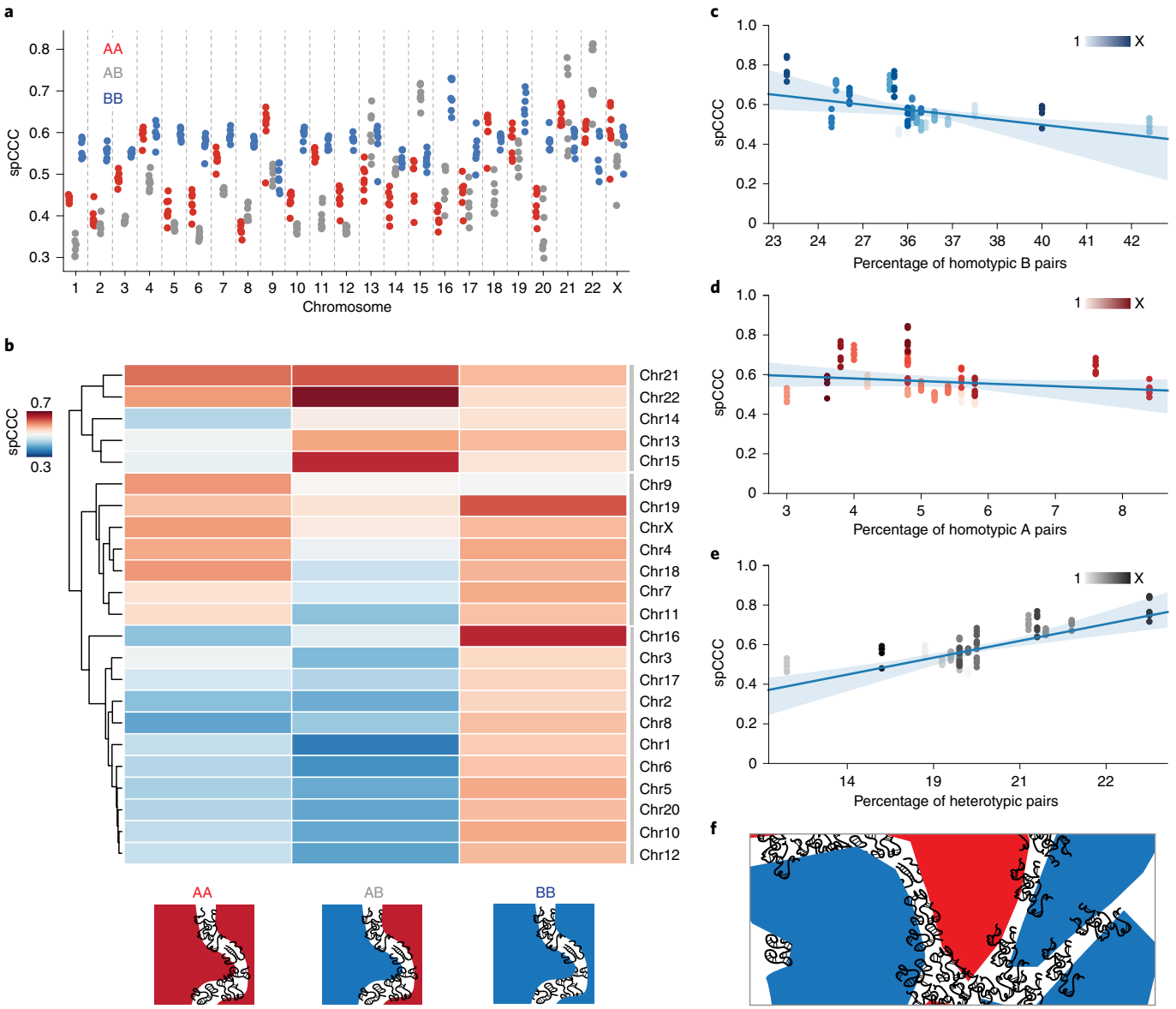


Fig. 5 | Compartment segregation mediated by co-TrTSs. a, Element-wise spCCC values between homotypic and heterotypic contacts in the experimental Hi-C contact map and contact maps derived from the ensemble of 3D models generated using colocalizing pairs of loci driven by the co-TrTS hotspots of seven representative lncRNAs from cluster 4. Red, homotypic A contacts; blue, homotypic B contacts; gray, heterotypic contacts. **b**, One-way hierarchical clustering of each chromosome (rows) based on mean spCCC over interaction type (columns). **c**–**e**, spCCC values between the experimental Hi-C contact map and that derived from the ensemble of 3D models generated using colocalizing pairs of loci driven by the co-TrTS hotspots of seven representative lncRNAs from cluster 4; percentages of homotypic B pairs (**c**), homotypic A pairs (**d**) and heterotypic pairs (**e**). The blue line represents a linear regression model fit and the transparent shade is its 95% confidence interval. **f**, Illustration of the proposed model by which co-TrTSs targeted by lncRNA (black) promote compartment segregation, mainly creating an interface between compartments A (red) and B (blue).

maintaining heterochromatin. However, in a few chromosomes (13, 15, 21 and 22) the lncRNA co-TrTS-driven ensembles more accurately reconstructed the 3D organization of heterotypic interactions or, in the case of chr9, homotypic A interactions. Clustering by mean correlation with homotypic and heterotypic interactions revealed the existence of three main groups of chromosomes (Fig. 5b): (1) one in which lncRNA co-TrTS-driven ensembles reconstructed better homotypic B-type interactions (including chromosomes 1, 2, 3, 5, 6, 8, 10, 12, 16, 17 and 20); (2) a second group in which both homotypic interactions and, in a lower proportion, heterotypic interactions were better reconstructed (chromosomes 4, 7, 9, 11, 18, 19 and X); and (3) a third cluster with heterotypic interactions accurately reconstructed (chromosomes 13, 14, 15, 21 and 22). Interestingly, the observed trend was dependent on the proportion of homotypic B-type and heterotypic interactions present in a given chromosome, while it was independent of the proportion of homotypic A-type interactions (Fig. 5c–e). Indeed, chromosome models that better recapitulated the experimental Hi-C map (higher spCCC) were those with a greater extent of heterotypic interactions, as is the case for acrocentric chromosomes. Based on this evidence, we propose a model where lncRNA co-TrTS contributes to genome compartment segregation by creating an interface between compartments A and B (Fig. 5f).

Discussion

Here, we have theoretically tested the hypothesis that triplex-forming lncRNAs contribute to the 3D organization of nuclear chromatin,

by performing steered molecular dynamics simulations of polymer models of chromosomes. Collectively, the findings from this study have several implications.

First, we predict the existence of lncRNAs with triplex-forming potential that do not randomly position themselves along the genome, but instead establish widespread preferential contacts between active and inactive chromatin regions. Their triplex-forming sequences are enriched in contacts with GA-rich repetitive elements, as simple-repeat and low-complexity regions that are over-represented in both prokaryotic and eukaryotic genomes[62], and are mainly located in repressed loci of the genome. Interestingly, it has been shown in vivo that both short and long purine-rich sequences can engage in triplexes[63] and that a G-rich triplex target site leads to a more stable[63] and more specific pH-independent triplex interaction[64]. Many thousands of uncharacterized lncRNAs are related to interspersed repeat elements[65,66] and show prolonged half-lives[67], which opens up the possibility of the existence of widespread lncRNA-enriched hotspots that stably form patches located in multiple distal genomic locations. Some of the identified lncRNAs with triplex-forming potential have preference only for a specific subset of chromosomes, while others show a widespread binding potential with all chromosomes. The latter appears to be the only subgroup targeting acrocentric chromosomes.

Second, triplex target hotspots could act as 'anchor palates' that can be used as restraints during modeling to accurately reconstruct the 3D organization of entire chromosomes, reproducing the plaid pattern characteristic of Hi-C interaction maps. Specifically, we found that acrocentric chromosomes and chr19, both located in the nuclear center[68], can be better reconstructed than larger chromosomes.

Third, we found that lncRNAs could be drivers for long-range interactions in contrast to more local genomic structures, such as TADs and loops, that often involve architectural proteins such as CTCF, and that these two driving forces might act cooperatively. These findings support the model in which chromatin organization is driven by the interplay of loop extrusion and compartmental segregation[8], and highlight that cooperativity between these two regimes, rather than competition, is what synergistically may govern overall 3D organization. Most importantly, our models support that colocalizing triplex sites targeted by lncRNAs could be key molecular factors driving compartmental segregation. These sites are shared by many lncRNAs, and multiple lncRNAs can target a chromosome dependently of chromosome size, suggesting that this could be a widespread mechanism. This is well in line with experiments showing that perturbation of RNAs by RNase A treatment which, together with other effects, can also disrupt triplex structures, induces changes in the nuclear organization of both euchromatin[30] and heterochromatin[69]. Interestingly, chromosome conformation experiments on RNase A-treated cells displayed subtle perturbation in long-range interactions between B-type compartments, even though TAD structures were largely intact[70]. Conversely, transcriptional activity perturbation led to weakening of TAD boundaries without drastic alterations in genomic compartmentalization[70]. These results suggest that pre-existing RNAs may play a role in genomic compartmentalization rather than newly transcribed transcripts.

Fourth, in our simulations, homotypic B interactions are generally better reconstructed than other types of interaction, suggesting that co-TrTSs are involved in maintaining heterochromatin. Interestingly, B-type compartments at single cells, although located in close proximity to each other, tend to intermingle less than the A type[45]. We hypothesized that coating of heterochromatin with lncRNAs could prevent intermixing of these compartments. Additionally, our models can more accurately reconstruct the structure of chromosomes with a higher presence of heterotypic interactions, suggesting possibly that coating of lncRNAs could also actively promote segregation between compartment types. Interestingly, RNA-dense compartments, composed of RNA with high connectivity and associated with active chromatin regions, are preferentially located spatially closer to inactive ones[16], which would be in line with our hypothesis of accumulation of lncRNAs at the compartment interface. Notably, acrocentric chromosomes, which are better reconstructed by our models, carry nucleolar organizer regions on their short arms[71] and are associated with the nucleolus[72]. Indeed, the nucleolus is another nuclear compartment where RNAs with high connectivity are found[16], further sustaining our hypothesis. Interestingly, and also in agreement with our hypothesis, other evidence shows that in mouse embryos, L1 transcripts, another class of triplex-forming potential RNAs[73], are required to seed homotypic B interaction clustering[15] while other mechanisms may further reinforce and stabilize compartmentalization as the attachment of repeat DNA sequences to other subnuclear structures such as the nucleolus[74].

Fifth, from a physics viewpoint, compartmentalization has been proposed to result from phase separation promoted by attraction among compartments of the same or similar chromatin state[75–77]. It has been hypothesized that these attractions may result from coassociation of domains with subnuclear bodies that, themselves, appear to be formed as condensates mediated by weak multivalent self-attraction between proteins such as heterochromatin protein 1 (ref. [76]), scaffold attachment factor A[41] or polycomb-repressive complexes[78]. These observations are in line with mechanistic modeling studies[8,79–82] in which compartmentalization has been proposed to result from the attraction among chromatin particles of similar compartments. Our modeling approach suggests an additional mechanism where the possible coating of lncRNAs could create an interface between compartments A and B and promote their mutual segregation, while also preventing the intermixing of B chromatin. This could be due to the formation of lncRNA foci at TrTS-enriched sites on chromatin that, due to the capacity of RNAs for multivalent binding with various RNA-binding proteins, act as 'sticky patches' for the recruitment of protein complexes facilitating their phase separation. Additionally, one could also speculate that lncRNA foci directly influence specific chromosomal interactions, especially between chromatin patches enriched in sequence repeats (simple-repeat and low-complexity regions). The direct binding of triplex-forming lncRNAs at widespread repeat patches in the genome could mediate the proximity between multiple loci of similar sequence motifs, thus promoting the formation of a chromatin network regulating large-scale chromatin structure. These two models of action are not mutually exclusive, and could coexist in maintaining A/B compartmentalization. This is supported by new evidence showing that, in condensates with high RNA density, RNA molecules are localized at the surface, limiting condensate growth and coalescence[83], and, we speculate, possibly leaving RNAs to interact with macromolecules outside the condensate.

To dissect both proposed plausible mechanisms for maintaining genome compartmentalization and to shed light on their coexistence, further microscopy and molecular experiments will be critical. To this end, it will be necessary to overcome various limitations in conducting experiments that take into account the repetitive nature of the identified lncRNAs. Imaging multicopy targets of specific repetitive DNA and messenger RNAs has been achieved using oligoprobes comprising between one and several dozen individually designed oligos[84–87]. However, targeting simple-repeat and low-complexity regions would lead presumably to a substantial amount of off-target binding that needs to be taken into account during the experimental design. Visualization of foci enriched in triplex-forming lncRNAs requires simultaneous imaging, at super-resolution (combining, for example oligoSTORM[45,87] and immuno-STORM[88]), repetitive triplex-forming lncRNAs, the double-stranded DNA–single-stranded RNA

triplex with specific antibody labeling[89–91] and the surrounding genomic regions. This challenging experiment will help to reveal the locations of triplex hotspots at the interface of compartments, and of heterochromatin coated with the identified lncRNAs and, importantly, to quantify the variation that arises at the single-cell level. Alternatively, to confirm the hypothesis driven by our integrative modeling approach one could perform knockdown experiments of triplex-forming lncRNAs[92]. However, in designing such experiments one should consider that here we predicted that multiple lncRNAs can target the same chromosome, and therefore the interference of only one lncRNA could easily be compensated by the presence of others. Indeed, experiments that broadly perturb triplex formation, or the strength of RNA binding with the duplex[63,64] changing the propensity of lncRNAs to perform triplex, will be necessary although extremely challenging. Alternatively, the association of the triplex forming lncRNAs at the predicted chromatin triplex hotspot could be tested via genome-wide RNA-chromatin proximity ligation approaches, for example using RADICL-seq[93], that is based on an improved protocol for the detection of intergenic transcripts and ncRNAs and for investigating long-range RNA-chromatin associations, or RNA-associated chromosome conformation experiments[94]. However, ideally, these experiments should be expanded to account for long-read sequencing to identify with more confidence repeat-element-rich lncRNAs.

The challenges for microscopy and molecular biology experiments to investigate highly repetitive triplex-forming lncRNAs highlight the importance of the theoretical approach introduced here in guiding future experiments and in revealing drivers of genome organization that are experimentally challenging to identify. Indeed, our computational modeling enabled us to unveil that direct lncRNA–DNA interaction may have a role in compartment segregation and act cooperatively, rather than competitively, with CTCF-mediated chromatin folding.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-021-00678-3.

Received: 13 October 2020; Accepted: 30 September 2021;
Published online: 10 November 2021

References

1. Nora, E. P. et al. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
2. Dixon, J. R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
3. Sexton, T. et al. Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* **148**, 458–472 (2012).
4. Barutcu, A. R., Maass, P. G., Lewandowski, J. P., Weiner, C. L. & Rinn, J. L. A TAD boundary is preserved upon deletion of the CTCF-rich Firre locus. *Nat. Commun.* **9**, 1444 (2018).
5. Rao, S. S. et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014).
6. Nora, E. P. et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* **169**, 930–944 (2017).
7. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289–293 (2009).
8. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl Acad. Sci. USA* **115**, E6697–E6706 (2018).
9. Rowley, M. J. & Corces, V. G. Organizational principles of 3D genome architecture. *Nat. Rev. Genet.* **19**, 789–800 (2018).
10. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
11. Bantignies, F. & Cavalli, G. Polycomb group proteins: repression in 3D. *Trends Genet.* **27**, 454–464 (2011).
12. Yang, J. & Li, F. Are all repeats created equal? Understanding DNA repeats at an individual level. *Curr. Genet.* **63**, 57–63 (2017).
13. Cournac, A., Koszul, R. & Mozziconacci, J. The 3D folding of metazoan genomes correlates with the association of similar repetitive elements. *Nucleic Acids Res.* **44**, 245–255 (2016).
14. Winter, D. J. et al. Repeat elements organise 3D genome structure and mediate transcription in the filamentous fungus *Epichloe festucae*. *PLoS Genet.* **14**, e1007467 (2018).
15. Lu, J. Y. et al. Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res.* **31**, 613–630 (2021).
16. Morf, J. et al. RNA proximity sequencing reveals the spatial organization of the transcriptome in the nucleus. *Nat. Biotechnol.* **37**, 793–802 (2019).
17. Quinodoz, S. A. et al. RNA promotes the formation of spatial compartments in the nucleus. Preprint at *bioRxiv* https://doi.org/10.1101/2020.08.25.267435 (2020).
18. Bonetti, A. et al. RADICL-seq identifies general and cell type–specific principles of genome-wide RNA-chromatin interactions. *Nat. Commun.* **11**, 1108 (2020).
19. Bell, J. C. et al. Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *eLife* **7**, e27024 (2018).
20. Li, X. et al. GRID-seq reveals the global RNA-chromatin interactome. *Nat. Biotechnol.* **35**, 940–950 (2017).
21. Sridhar, B. et al. Systematic mapping of RNA-chromatin interactions in vivo. *Curr. Biol.* **27**, 602–609 (2017).
22. Senturk Cetin, N. et al. Isolation and genome-wide characterization of cellular DNA:RNA triplex structures. *Nucleic Acids Res.* **47**, 2306–2321 (2019).
23. Nickerson, J. A., Krochmalnic, G., Wan, K. M. & Penman, S. Chromatin architecture and nuclear RNA. *Proc. Natl Acad. Sci. USA* **86**, 177–181 (1989).
24. Holmes, D. S., Mayfield, J. E., Sander, G. & Bonner, J. Chromosomal RNA: its properties. *Science* **177**, 72–74 (1972).
25. Rodriguez-Campos, A. & Azorin, F. RNA is an integral component of chromatin that contributes to its structural organization. *PLoS ONE* **2**, e1182 (2007).
26. Tsai, M. C. et al. Long noncoding RNA as modular scaffold of histone modification complexes. *Science* **329**, 689–693 (2010).
27. Hnisz, D., Shrinivas, K., Young, R. A., Chakraborty, A. K. & Sharp, P. A. A phase separation model for transcriptional control. *Cell* **169**, 13–23 (2017).
28. Michieletto, D. & Gilbert, N. Role of nuclear RNA in regulating chromatin structure and transcription. *Curr. Opin. Cell Biol.* **58**, 120–125 (2019).
29. Frank, L. & Rippe, K. Repetitive RNAs as regulators of chromatin-associated subcompartment formation by phase separation. *J. Mol. Biol.* **432**, 4270–4286 (2020).
30. Caudron-Herger, M. et al. Coding RNAs with a non-coding function: maintenance of open chromatin structure. *Nucleus* **2**, 410–424 (2011).
31. Meng, Y. et al. The non-coding RNA composition of the mitotic chromosome by 5'-tag sequencing. *Nucleic Acids Res.* **44**, 4934–4946 (2016).
32. Chujo, T., Yamazaki, T. & Hirose, T. Architectural RNAs (arcRNAs): a class of long noncoding RNAs that function as the scaffold of nuclear bodies. *Biochim. Biophys. Acta* **1859**, 139–146 (2016).
33. Engreitz, J. M. et al. The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. *Science* **341**, 1237973 (2013).
34. Donley, N., Smith, L. & Thayer, M. J. ASAR15, a cis-acting locus that controls chromosome-wide replication timing and stability of human chromosome 15. *PLoS Genet.* **11**, e1004923 (2015).
35. Fey, E. G., Ornelles, D. A. & Penman, S. Association of RNA with the cytoskeleton and the nuclear matrix. *J. Cell Sci.* **1986**, 99–119 (1986).
36. Hall, L. L. et al. Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* **156**, 907–919 (2014).
37. Kalwa, M. et al. The lncRNA HOTAIR impacts on mesenchymal stem cells via triple helix formation. *Nucleic Acids Res.* **44**, 10631–10643 (2016).
38. Mondal, T. et al. MEG3 long noncoding RNA regulates the TGF-beta pathway genes through formation of RNA-DNA triplex structures. *Nat. Commun.* **6**, 7743 (2015).
39. O'Leary, V. B. et al. PARTICLE, a triplex-forming long ncRNA, regulates locus-specific methylation in response to low-dose irradiation. *Cell Rep.* **11**, 474–485 (2015).
40. Johnson, R. & Guigo, R. The RIDL hypothesis: transposable elements as functional domains of long noncoding RNAs. *RNA* **20**, 959–976 (2014).
41. Nozawa, R. S. et al. SAF-A regulates interphase chromosome structure through oligomerization with chromatin-associated RNAs. *Cell* **169**, 1214–1227 (2017).
42. Saldaña-Meyer, R. et al. RNA interactions are essential for CTCF-mediated genome organization. *Mol. Cell* **76**, 412–422 (2019).

43. Buske, F. A., Bauer, D. C., Mattick, J. S. & Bailey, T. L. Triplexator: detecting nucleic acid triple helices in genomic and transcriptomic data. *Genome Res.* **22**, 1372–1381 (2012).
44. Baù, D. et al. The three-dimensional folding of the alpha-globin gene domain reveals formation of chromatin globules. *Nat. Struct. Mol. Biol.* **18**, 107–114 (2011).
45. Nir, G. et al. Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling. *PLoS Genet.* **14**, e1007872 (2018).
46. Di Stefano, M. et al. Transcriptional activation during cell reprogramming correlates with the formation of 3D open chromatin hubs. *Nat. Commun.* **11**, 2564 (2020).
47. Di Stefano, M., Rosa, A., Belcastro, V., di Bernardo, D. & Micheletti, C. Colocalization of coregulated genes: a steered molecular dynamics study of human chromosome 19. *PLoS Comput. Biol.* **9**, e1003019 (2013).
48. Tjong, H. et al. Population-based 3D genome structure analysis reveals driving forces in spatial genome organization. *Proc. Natl Acad. Sci. USA* **113**, E1663–E1672 (2016).
49. Tiana, G. et al. Structural fluctuations of the chromatin fiber within topologically associating domains. *Biophys. J.* **110**, 1234–1245 (2016).
50. Frankish, A. et al. GENCODE reference annotation for the human and mouse genomes. *Nucleic Acids Res.* **47**, D766–D773 (2019).
51. Wan, Y. et al. Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* **505**, 706–709 (2014).
52. Maldonado, R., Schwartz, U., Silberhorn, E. & Langst, G. Nucleosomes stabilize ssRNA-dsDNA triple helices in human cells. *Mol. Cell* **73**, 1243–1254 (2019).
53. Sinden, R. Torsional tension in the DNA double helix measured with trimethylpsoralen in living *E. coli* cells: analogous measurements in insect and human cells. *Cell* **21**, 773–783 (1980).
54. Pyne, A. L. B. et al. Base-pair resolution analysis of the effect of supercoiling on DNA flexibility and major groove recognition by triplex-forming oligonucleotides. *Nat. Commun.* **12**, 1053 (2021).
55. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
56. Rao, S. S. P. et al. Cohesin loss eliminates all loop domains. *Cell* **171**, 305–320 (2017).
57. Mirny, L. A., Imakaev, M. & Abdennur, N. Two major mechanisms of chromosome organization. *Curr. Opin. Cell Biol.* **58**, 142–152 (2019).
58. Fudenberg, G. et al. Formation of chromosomal domains by loop extrusion. *Cell Rep.* **15**, 2038–2049 (2016).
59. Alipour, E. & Marko, J. F. Self-organization of domain structures by DNA-loop-extruding enzymes. *Nucleic Acids Res.* **40**, 11202–11212 (2012).
60. Sanborn, A. L. et al. Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl Acad. Sci. USA* **112**, E6456–E6465 (2015).
61. Tang, Z. et al. CTCF-mediated human 3D genome architecture reveals chromatin topology for transcription. *Cell* **163**, 1611–1627 (2015).
62. Buske, F. A., Mattick, J. S. & Bailey, T. L. Potential in vivo roles of nucleic acid triple-helices. *RNA Biol.* **8**, 427–439 (2011).
63. Kunkler, C. N. et al. Stability of an RNA•DNA–DNA triple helix depends on base triplet composition and length of the RNA third strand. *Nucleic Acids Res.* **47**, 7213–7222 (2019).
64. Kaufmann, B. et al. Identifying triplex binding rules in vitro leads to creation of a new synthetic regulatory tool in vivo. Preprint at *bioRxiv* https://doi.org/10.1101/2019.12.25.888362 (2019).
65. Hadjiargyrou, M. & Delihas, N. The intertwining of transposable elements and non-coding RNAs. *Int. J. Mol. Sci.* **14**, 13307–13328 (2013).
66. Hoekstra, H. E. et al. Transposable elements are major contributors to the origin, diversification, and regulation of vertebrate long noncoding RNAs. *PLoS Genet.* **9**, e1003470 (2013).
67. Clark, M. B. et al. Genome-wide analysis of long noncoding RNA stability. *Genome Res.* **22**, 885–898 (2012).
68. Boyle, S. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Hum. Mol. Genet.* **10**, 211–219 (2001).
69. Maison, C. et al. Higher-order structure in pericentric heterochromatin involves a distinct pattern of histone modification and an RNA component. *Nat. Genet.* **30**, 329–334 (2002).
70. Barutcu, A. R., Blencowe, B. J. & Rinn, J. L. Differential contribution of steady-state RNA and active transcription in chromatin organization. *EMBO Rep.* **20**, e48068 (2019).
71. Henderson, A. S., Warburton, D. & Atwood, K. C. Location of ribosomal DNA in the human chromosome complement. *Proc. Natl Acad. Sci. USA* **69**, 3394–3398 (1972).
72. Misteli, T. et al. Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase Rosettes. *PLoS Biol.* **3**, e157 (2005).
73. Fadloun, A. et al. Chromatin signatures and retrotransposon profiling in mouse embryos reveal regulation of LINE-1 by RNA. *Nat. Struct. Mol. Biol.* **20**, 332–338 (2013).
74. Quinodoz, S. A. et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell* **174**, 744–757 (2018).
75. Strom, A. R. et al. Phase separation drives heterochromatin domain formation. *Nature* **547**, 241–245 (2017).
76. Larson, A. G. et al. Liquid droplet formation by HP1alpha suggests a role for phase separation in heterochromatin. *Nature* **547**, 236–240 (2017).
77. Hult, C. et al. Enrichment of dynamic chromosomal crosslinks drive phase separation of the nucleolus. *Nucleic Acids Res.* **45**, 11159–11173 (2017).
78. Plys, A. J. et al. Phase separation of Polycomb-repressive complex 1 is governed by a charged disordered region of CBX2. *Genes Dev.* **33**, 799–813 (2019).
79. Jost, D., Carrivain, P., Cavalli, G. & Vaillant, C. Modeling epigenome folding: formation and dynamics of topologically associated chromatin domains. *Nucleic Acids Res.* **42**, 9553–9561 (2014).
80. Di Pierro, M., Zhang, B., Aiden, E. L., Wolynes, P. G. & Onuchic, J. N. Transferable model for chromosome architecture. *Proc. Natl Acad. Sci. USA* **113**, 12168–12173 (2016).
81. Brackley, C. A., Johnson, J., Kelly, S., Cook, P. R. & Marenduzzo, D. Simulated binding of transcription factors to active and inactive regions folds human chromosomes into loops, rosettes and topological domains. *Nucleic Acids Res.* **44**, 3503–3512 (2016).
82. Di Stefano, M., Nützmann, H.-W., Marti-Renom, Marc, A. & Jost, D. Polymer modelling unveils the roles of heterochromatin and nucleolar organizing regions in shaping 3D genome organization in *Arabidopsis thaliana*. *Nucleic Acids Res.* **49**, 1840–1858 (2021).
83. Cochard, A. et al. RNA at the surface of phase-separated condensates impacts their size and number. Preprint at *bioRxiv* https://doi.org/10.1101/2021.06.22.449254 (2021).
84. Beliveau, B. J. et al. Versatile design and synthesis platform for visualizing genomes with Oligopaint FISH probes. *Proc. Natl Acad. Sci. USA* **109**, 21301–21306 (2012).
85. Femino, A. M. Visualization of single RNA transcripts in situ. *Science* **280**, 585–590 (1998).
86. Matera, A. G. & Ward, D. C. Oligonucleotide probes for the analysis of specific repetitive DNA sequences by fluorescence in situ hybridization. *Hum. Mol. Genet.* **1**, 535–539 (1992).
87. Chang, C. H. et al. Islands of retroelements are major components of *Drosophila* centromeres. *PLoS Biol.* **17**, e3000241 (2019).
88. Ricci, M. A., Manzo, C., Garcia-Parajo, M. F., Lakadamyali, M. & Cosma, M. P. Chromatin fibers are formed by heterogeneous groups of nucleosomes in vivo. *Cell* **160**, 1145–1158 (2015).
89. Stollar, B. D. & Raso, V. Antibodies recognise specific structures of triple-helical polynucleotides built on poly(A) or poly(dA). *Nature* **250**, 231–234 (1974).
90. Ohno, M., Fukagawa, T., Lee, J. S. & Ikemura, T. Triplex-forming DNAs in the human interphase nucleus visualized in situ by polypurine/polypyrimidine DNA probes and antitriplex antibodies. *Chromosoma* **111**, 201–213 (2002).
91. Gorab, E., Amabis, J. M., Stocker, A. J., Drummond, L. & Stollar, B. D. Potential sites of triple-helical nucleic acid formation in chromosomes of *Rhynchosciara* (Diptera: Sciaridae) and *Drosophila melanogaster*. *Chromosome Res.* **17**, 821–832 (2009).
92. Kuo, C.-C. et al. Detection of RNA–DNA binding sites in long noncoding RNAs. *Nucleic Acids Res.* **47**, e32 (2019).
93. Mumbach, M. R. et al. HiChIRP reveals RNA-associated chromosome conformation. *Nat. Methods* **16**, 489–492 (2019).
94. Ernst, J. et al. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**, 43–49 (2011).
95. Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16**, 418–420 (2000).

Methods

Identification of triplex-forming lncRNAs. *Selection of nuclear-enriched lncRNAs.* The GENCODE v.19 (ref. [50]) lncRNA set was downloaded from https://www.gencodegenes.org/human/release_19.html, and transcripts with ≥200 nucleotides, multiexonic and nuclear-enriched (expression level >1 read per kilobase million) in cell types GM12878 and K562 (based on RNA-sequencing (RNA-seq) expression data from ENCODE[96], IDs: ENCSR000CPO and ENCSR000CQF for GM12878, and ENCSR000CQM, ENCSR530NHO and ENCSR000CPS for K562, respectively), were defined as nuclear-enriched lncRNAs, numbering 1,120 in total (Fig. 1a).

Prediction of TFOs and TrTSs. Putative triplex sites were identified using Triplexator[43], which defines a TFO in lncRNAs as well as their TrTSs in the target genome (here, the human GRCh37 assembly). Triplexator analysis was performed with the following default parameters: 10% error rate (number of mismatches), 25% guanine content, minimum triplex length of 15 base pairs (bp) and using canonical triplex-formation rules for the purine motif, which have been suggested as being more physiologically relevant[97], do not require acidic pH and are stabilized mainly by bivalent cations such as Mg^{2+} (ref. [98]). As a control, we generated an in silico set of RNA molecules of identical length and base-composition distributions of the selected lncRNAs. This set was then used to obtain background distribution of Triplexator triplex-forming potential scores. Only those lncRNAs with a score exceeding the upper quartile level in the background distribution were then retained, which resulted in a total of 339 lncRNAs (Fig. 1a).

lncRNA secondary structure profile. Parallel analysis of RNA structure (PARS) scores were computed for the selected transcripts as previously described[51] (GEO ID: GSE50676). Transcripts for which the putative TFOs were not localized in double-stranded parts of the lncRNA were selected. A total of 115 lncRNAs contained TFOs in single-stranded RNA sites (Fig. 1a).

TrTS and lncRNA annotation. The transcript and genomic characteristics of TrTSs were annotated with HOMER using the script *annotatePeaks.pl*, with default parameters[99].

3D colocalization of TrTSs. *Hi-C interaction maps.* Individual chromatin-chromatin contact frequencies from GM12878 cells at 50- and 100-kb resolution were obtained from the original publication (GEO ID: GSE63525) and were normalized as previously reported[5].

Intrachromosomal interaction calling. To identify significant interactions between pairs of TrTS hotspots, we implemented an intrachromosomal peak calling similarly to one previously published[100]. First, we generated an all-versus-all TrTS pairs set using as reference the previously generated TrTS profile. Second, the set of significant TrTS pairs for each chromosome, *c*, was identified by comparing distribution of intrachromosomal Hi-C interaction frequencies between TrTS pairs at genomic distance *d* with that of intrachromosomal Hi-C interaction frequencies with comparable sequence-based distance. Next, the minimum number of observations required in distribution was set equal to the number of observations present in the diagonal of the Hi-C interaction matrix of each chromosome. For this, Hi-C interaction frequencies between bins of neighboring sequence-based distance in the distribution were concatenated. The set of co-TrTS pairs was defined as TrTS pairs exceeding the upper quartile level. Finally, 3D colocalization of the co-TrTS set was quantified by calculating an enrichment score, given by $\log_2$ of the ratio between observed colocalization and that expected by chance at a specific genomic distance.

Analysis of co-TrTSs. *Motif analysis.* De novo motif analysis was performed using the findMotifsGenome.pl Perl script in HOMER, with default parameters[99].

Enrichment analysis. Enrichments between the 'test' and 'query' sets were calculated from the $\log_2$(observed/expected) ratio of direct overlaps, using bedtools intersect (with '–f 0.9 –wo' flags); for the randomly shuffled 'query' set we used bedtools shuffle (with '–chrom' flag). The 'query' sets are represented by the co-TrTS. Different test sets were used: (1) chromatin state profile (chromHMM 15 state profile[94]) for GM12878 from the UCSC table browser (http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeBroadHmm); (2) repeat elements profile using RepeatMasker[95] annotations from the UCSC table browser (https://genome.ucsc.edu/cgi-bin/hgTrackUi?g=rmsk); (3) DNAse sequencing in GM12878 cells from the ENCODE project[96] (ENCODEid: ENCFF097LEF, ENCFF273MVV and ENCFF804BNU); and (4) chromatin compartment (Compartment analysis).

Per-chromosome co-TrTS potential. Per-chromosome co-TrTS potential (CPcoTrTS) is the fraction of nucleotide subsequences (nt_TTS) participating in co-TrTS formation per kilobase in a specific chromosome:

$$\mathrm{CPcoTrTS} = \log_{10}\left(\frac{100 \times \mathrm{nt\_TTS}}{l_{\mathrm{kb}}}\right)$$

where l_{kb} is chromosome length in kilobases.

Compartment analysis. To segment the genome into compartments A and B, normalized Hi-C matrices at 100-kb resolution were transformed into correlation matrices using Pearson product–moment correlation. The first eigenvector of a principal component (ev1) on each of these matrices was used as a quantitative measure of compartmentalization, and H3k4me1 and CG content data were used to assign negative and positive ev1 categories to the correct compartments (using the rich_in_A option in TADbit). If necessary, the sign of ev1 (which is arbitrary) was inverted so that positive ev1 values corresponded to compartment A regions, and vice versa for B. To measure inter- and intracompartment strength, we followed a previously reported strategy[101]. Briefly, for each chromosome we calculated observed/expected Hi-C matrices at 100 kb and then sorted the bin of observed/expected according to their ev1 values (from lowest to highest). The resulting map was coarse grained into a 25×25 matrix. Finally, all intrachromosomal interactions with similar ev1 values were aggregated to obtain compartmentalization saddle plots as the average of these coarse-grained maps over all chromosomes.

3D colocalization of CTCF sites. CTCF ChIP–seq data for GM12878 cells were obtained from the ENCODE project[96] (ENCODEid: ENCSR000AKB). Colocalized CTCF peaks were defined as the subset of selected interactions (Intrachromosomal interaction calling) that contained in their base at least one CTCF peak.

3D colocalization of low-complexity sites. Colocalization of low-complexity sites was defined as the subset of selected interactions (Intrachromosomal interaction calling) that enclose low-complexity repetitive elements as listed in RepeatMasker[95] annotations from the UCSC table browser (https://genome.ucsc.edu/cgi-bin/hgTrackUi?g=rmsk).

Chromosome ensemble modeling. *Chromosome representation.* Each chromosome was described with a beads-on-a-string model based on a previously implemented protocol[46,47]. Thus, a chromosome is represented by *n* spherical beads with diameter σ containing 50 kb of chromatin, which determines the genomic unit length of each model.

System setup for molecular dynamics. All simulations were done using the LAMMPS package[102]. The potential energy of each system comprises the terms of the Kremer–Grest polymer model[103], including chain connectivity (finitely extensible nonlinear elastic, FENE, with $K = 300.0\, k_B T/\sigma^2$, where K is the attractive force strength, k_B is the Boltzmann constant and T is the temperature equal to 300 K; and $R_0 = 1.5\sigma$, where R_0 is the maximum extend of the bond) and excluded-volume (purely repulsive Lennard-Jones with $\varepsilon = 10.0$ for nearest-neighbor beads, 1.0 otherwise and cutoff$= \sqrt[6]{2}\sigma$; where ε is the interaction energy) interactions. Each model chromosome was initially prepared in an elongated, solenoidal-like configuration resulting from stacked rosette patterns[104], placed randomly inside a cubic simulation box of size 300σ centered at the origin of the Cartesian axis $O = (0.0, 0.0, 0.0)$, tethered at the center of the box using a harmonic (spring constant $K_t = 50.0\, k_B T/\sigma^2$ and equilibrium distance $d_{eq} = 0.0\sigma$, where σ is the particle diameter) to avoid any border effect, and energy minimized using a short run of the Polak–Ribiere version of the conjugate gradient algorithm (LAMMPS command minimize 1.0e-4 1.0e-6 100000 100000)[102].

Equilibration procedure. Each simulation was carried in the canonical ensemble (constant temperature, constant volume) ($T = 1.0\, k_B T$). The integration time step was set equal to $0.006\, \tau_{LJ}$, where τ_{LJ} is the internal time unit and $m = 1.0$ is the bead mass, which was set equal to the LAMMPS default value. Each system was relaxed with a run of unrestrained Langevin dynamic for $6{,}000\, \tau_{LJ}$. After this relaxation phase, production runs were performed with a steered molecular dynamics protocol using the COLVARS LAMMPS-plugin[105].

Steered molecular dynamics protocol. The steered molecular dynamics protocol was used to progressively favor the spatial proximity of the identified colocalizing loci (Intrachromosomal interaction calling) using harmonic restraints between target pairs of beads. For each restraint, the equilibrium distance was set to one bead diameter (σ). The spring constant $k(L,t)$ was weighted with the sequence-separation L between restrained beads[47] to ensure that the steering process was not dominated by target pairs at the largest sequence separation. $k(L,t)$ was progressively increased during the steering phase in 1,000 discrete steps (s_i), each lasting $60\, \tau_{LJ}$, by multiplying the sequence-separation weighted value of k by s_i. The initial values of k ranged between 0.002 and $0.006\, k_B T/\sigma^2$ and the final ones between 2.2315 and $6.7714\, k_B T/\sigma^2$. To sample a reliable 3D ensemble structure, the steering process was repeated independently 1,000 times for each chromosome and, in each individual run, only a random set including 10% of potential restraints was applied. For each of the 1,000 replicate runs, the conformation satisfying the majority of imposed restraints within a radius of 2σ was retained to form the ensemble of 1,000 3D models used for further analysis.

3D chromosome ensemble analysis. *Contact map generation.* A contact map (C) was calculated at 50-kb resolution to visualize the frequency of contacts from the

ensemble of chromosomal models. Two beads were considered as constituting a contact when their Euclidean distance was $<2\sigma$ cutoff.

Comparison of contact maps derived from models and Hi-C interaction maps. The degree of similarity between two maps was computed by TADbit[106] using different metrics: (1) element-wise spCCC; (2) off-dCCC, where spCCC was calculated between corresponding off-diagonals for both matrices, thus generating a genomic distance profile of intrachromosomal interaction correlation; and (3) eigenvector correlation, with Pearson correlation between the first eigenvectors of each matrix.

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

Data availability

RNA-seq datasets were downloaded from ENCODE with accession nos. ENCSR000CPO, ENCSR000CQF, ENCSR000CQM, ENCSR530NHO and ENCSR000CPS. The PARS dataset was downloaded from Gene Expression Omnibus (GEO) with accession no. GSE50676. Hi-C was downloaded from GEO with accession no. GSE63525. CTCF ChIP–seq was downloaded from ENCODE with accession no. ENCSR000AKB. DNAse sequencing was downloaded from ENCODE with accession nos. ENCFF097LEF, ENCFF273MVV and ENCFF804BNU. The GENCODE v.19 lncRNA set was downloaded from https://www.gencodegenes.org/human/release_19.html.

Code availability

Custom scripts used to identify colocalizing loci and perform restraint-based modeling are available at http://sgt.cnag.cat/3dg/datasets/. The custom python scripts for analysis used the libraries Numpy[107], Pandas[108], Seaborn and Matplotlib[109], Scipy[110] and Multiprocessing[111].

References

96. Davis, C. A. et al. The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res.* **46**, D794–D801 (2018).
97. Ayel, E. & Escudé, C. In vitro selection of oligonucleotides that bind double-stranded DNA in the presence of triplex-stabilizing agents. *Nucleic Acids Res.* **38**, e31 (2010).
98. Barsh, G. S., Bacolla, A., Wang, G. & Vasquez, K. M. New perspectives on DNA and RNA triplexes as effectors of biological activity. *PLOS Genet.* **11**, e1005696 (2015).
99. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
100. Paulsen, J. et al. Handling realistic assumptions in hypothesis testing of 3D co-localization of genomic elements. *Nucleic Acids Res.* **41**, 5164–5174 (2013).
101. Imakaev, M. et al. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
102. Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **117**, 1–19 (1995).
103. Kremer, K. & Grest, G. S. Dynamics of entangled linear polymer melts: a molecular-dynamics simulation. *J. Chem. Phys.* **92**, 5057–5086 (1990).
104. Rosa, A. & Everaers, R. Structure and dynamics of interphase chromosomes. *PLoS Comput. Biol.* **4**, e1000153 (2008).
105. Fiorin, G., Klein, M. L. & Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **111**, 3345–3362 (2013).
106. Serra, F. et al. Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors. *PLoS Comput. Biol.* **13**, e1005665 (2017).
107. Harris, C. R. et al. Array programming with NumPy. *Nature* **585**, 357–362 (2020).
108. McKinney, W. Data structures for statistical computing in Python. In *Proc. 9th Python in Science Conference* (eds van der Walt, S. & Millman, J.) 56–61 (SciPy.org, 2010); https://doi.org/10.25080/Majora-92bf1922-00a
109. Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **9**, 90–95 (2007).
110. Virtanen, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods* **17**, 261–272 (2020).
111. McKerns, M. M. et al. Building a framework for predictive science. Preprint at https://arxiv.org/abs/1202.1056 (2012).
112. Lejeune, J. et al. A PROPOSED standard system of nomenclature of human mitotic chromosomes. *Lancet* **275**, 1063–1065 (1960).

Acknowledgements

We thank all current and past members of the Marti-Renom laboratory for their continuous discussions and support; H. Y. Chang, R. A. Flynn and K. Qu for help with PARS data analysis; J. Morf for fruitful discussions; M. Dabad and A. Esteve-Codina of the Functional Genomics Team at CNAG for initial RNA-seq analysis; and C.T. Wu and members of the Wu laboratory for their support. This work was supported by the European Research Council under the 7th Framework Program FP7/2007-2013 (ERC grant agreement no. 609989 to M.A.M.-R.) and the Spanish Ministerio de Ciencia, Innovación y Universidades through nos. IJCI-2015-23352 to I.F. and BFU2017-85926-P and PID2020-115696RB-I00 to M.A.M.-R. CRG acknowledges support from 'Centro de Excelencia Severo Ochoa 2013-2017', SEV-2012-0208 and the CERCA Program/Generalitat de Catalunya, as well as support from the Spanish Ministry of Science and Innovation through the Instituto de Salud Carlos III and the EMBL partnership, the Generalitat de Catalunya through Departament de Salut and Departament d'Empresa i Coneixement, and cofinancing with funds from the European Regional Development Fund by the Spanish Ministry of Science and Innovation corresponding to the Programa Opertaivo FEDER Plurirregional de España 2014–2020 and by the Secretaria d'Universitats i Recerca, Departament d'Empresa i Coneixement of the Generalitat de Catalunya corresponding to the program Operatiu FEDER Catalunya 2014–2020 and the NIH (to C.T. Wu no. R01HD091797 for supporting I.F.).

Author contributions

I.F. and M.A.M.-R. conceived the study. I.F. and M.D.S. performed modeling. P.S.-V. and M.M.-M. supported modeling protocol development and implementation. I.F. wrote the manuscript with M.D.S., P.S.-V., M.M.-M. and M.A.M.-R. M.A.M.-R. oversaw the project.

Competing interests

The authors declare no competing interests.

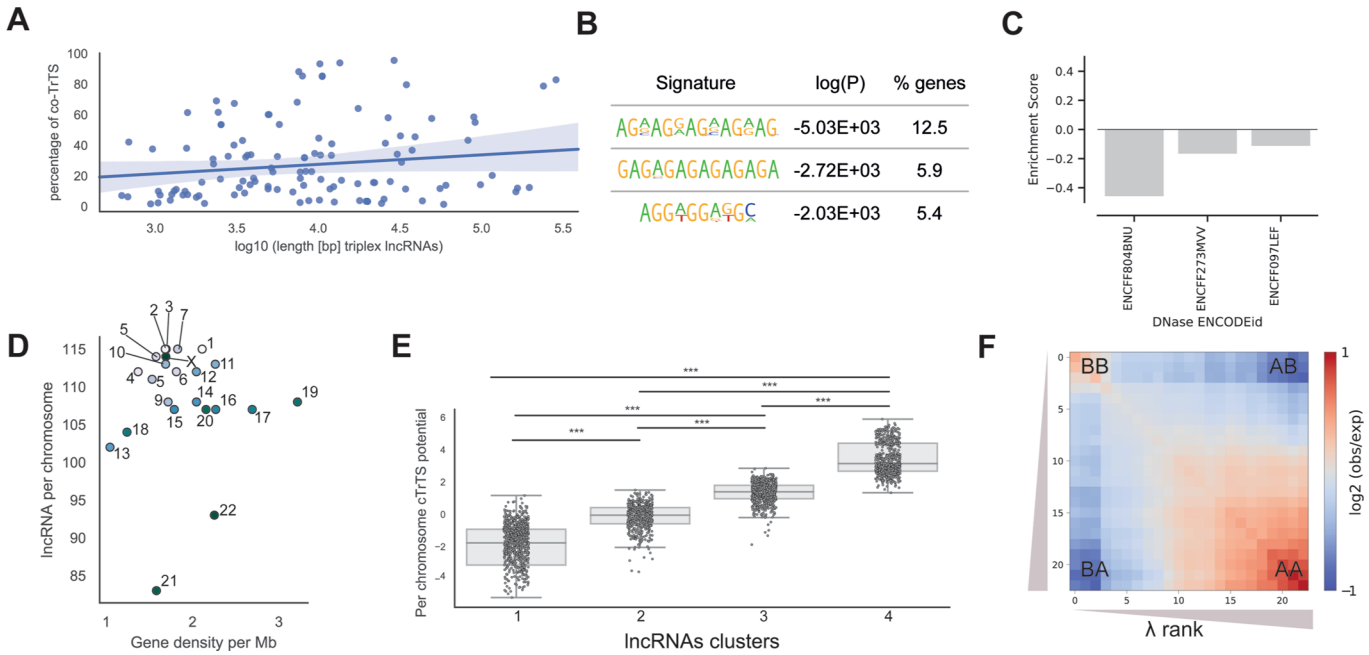
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41594-021-00678-3.

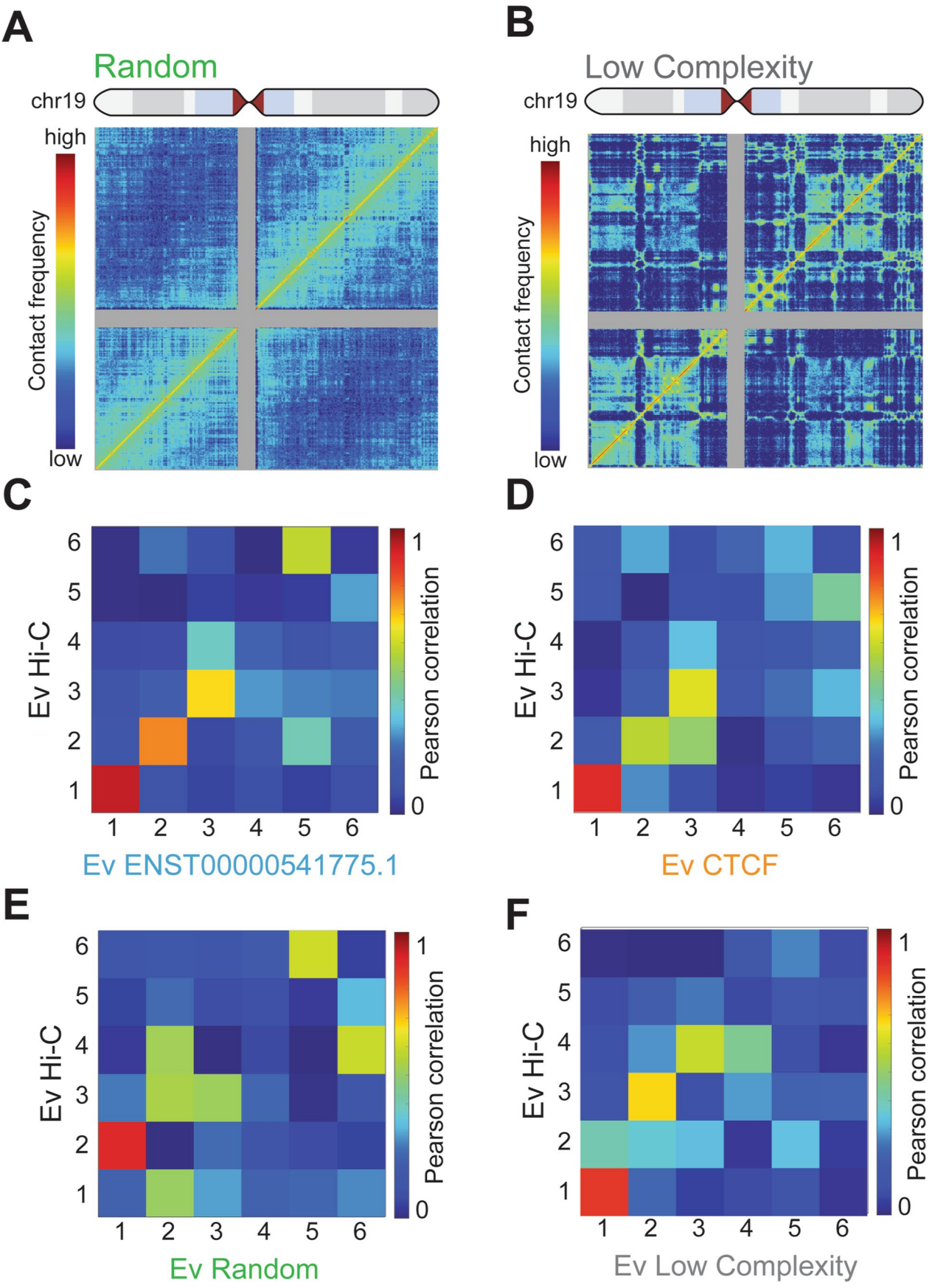
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-021-00678-3.

Correspondence and requests for materials should be addressed to Irene Farabella or Marc A. Marti-Renom.

Peer review information *Nature Structural & Molecular Biology* thanks Sarah Harris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Carolina Perdigoto and Beth Moorefield were the primary editors on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

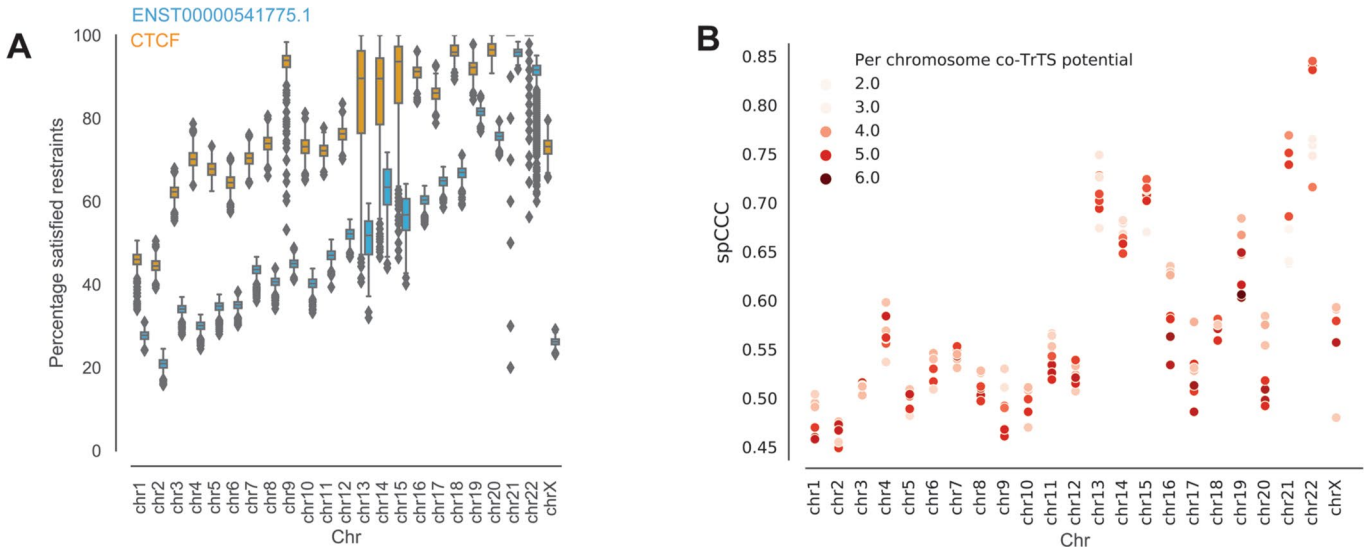


Extended Data Fig. 1 | Genomic features of co-localized Triplex Target hotspot. (a) Percentage of TrTS that co-localize genome-wide versus the length of the identified lncRNAs with triplex forming potential. The blue line represents a linear regression model fit and the transparent shade is the 95% confidence interval. **(b)** Top three enriched motifs (Methods) for co-localizing Triplex Target Sites over background obtained with HOMER[99]. **(c)** Enrichment of co-TrTS hotspots in DNase I hypersensitive sites. **(d)** Number of lncRNAs with triplex forming potential with respect to chromosomal gene density. **(e)** Co-TrTS potential distribution per-chromosome for the four clusters defined in Fig. 2e. Box boundaries represent 1st and 3rd quartiles, middle line represents median, and whiskers extend to 1.5 times the interquartile range (two-sided Mann-Whitney rank test, using python default parameters, ***: $p < 10^{-3}$; n = 897, 506, 667, and 575 for cluster 1, 2, 3, and 4, respectively. **(f)** Compartmentalization saddle plot (Methods) of all intra-chromosomal interactions in GM12878 cell line.

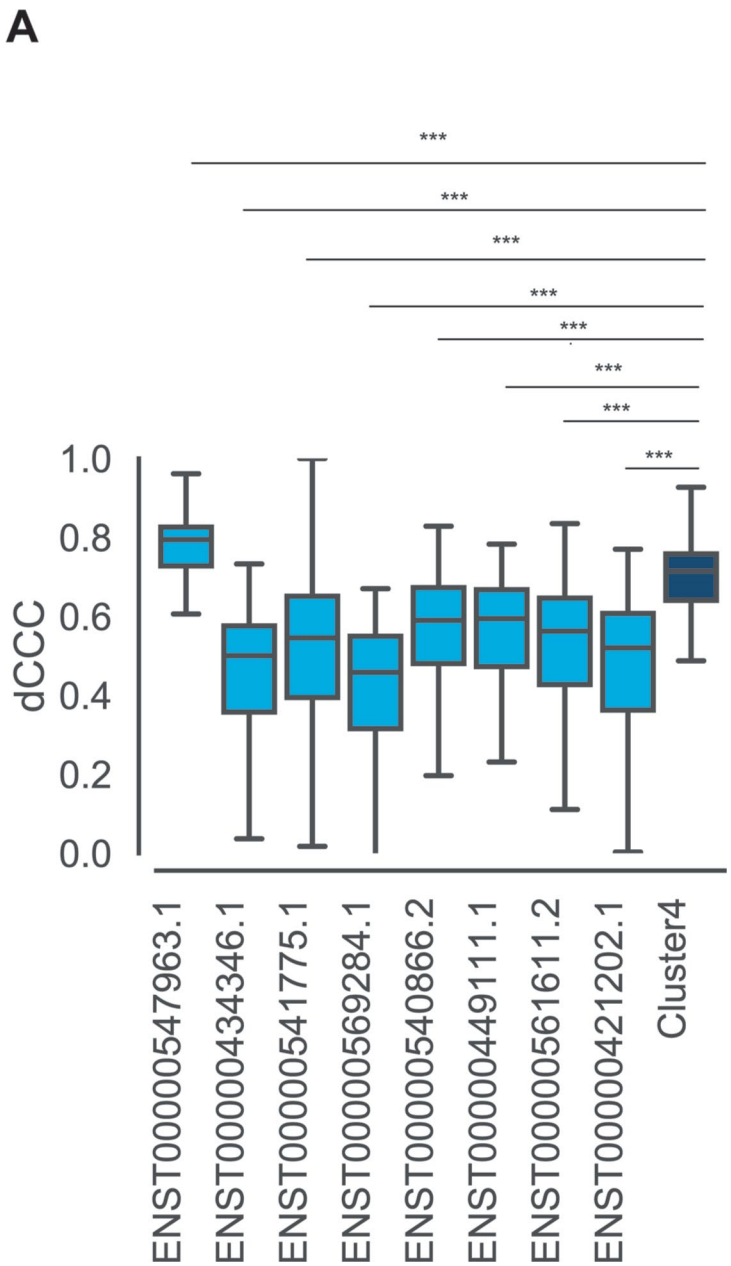


Extended Data Fig. 2 | See next page for caption.

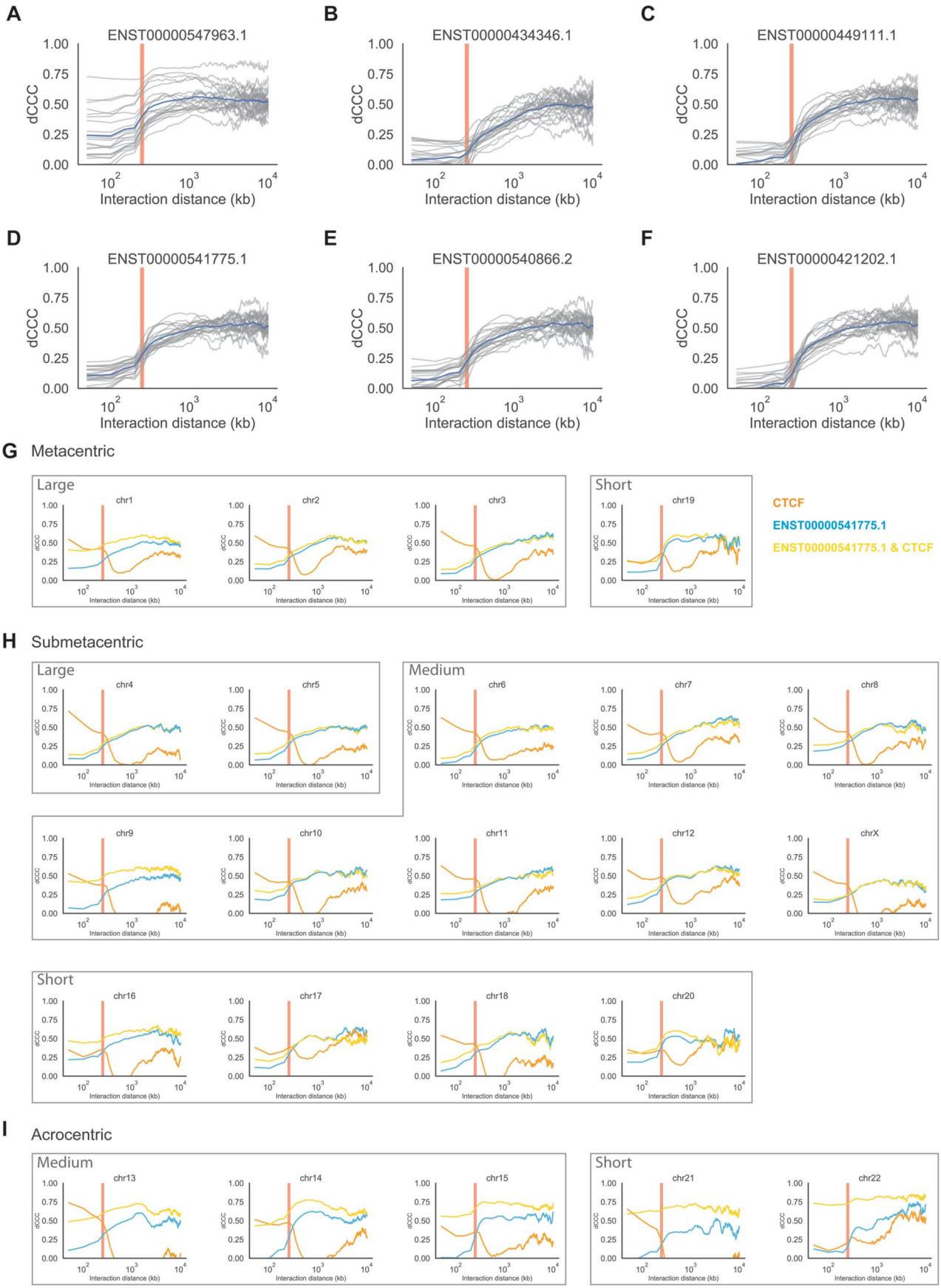
Extended Data Fig. 2 | Restraint-based simulations for human chromosome 19. (**a**,**b**) Contact maps derived from the ensemble of 3D models generated using co-localizing pairs of randomly selected loci **(A)** and low complexity enriched genomic sites (**B**). **(C-F)** Matrices of Pearson cross-correlation coefficients of top six eigenvectors for chromosome 19 of the experimental Hi-C compared to the four simulated datasets (that is, ENST00000541775.1, CTCF, random, and low complexity for **c**, **d**, **e** and **f**, respectively).



Extended Data Fig. 3 | Comparisons of co-localized Triplex Target Sites restraint-based simulations. (a) Distribution of the percentage of satisfied restraints in the ensemble of models using co-localizing pairs of loci driven by the ENST00000541775.1 co-TrTS hotspots and the CTCF enriched sites. Box boundaries represent 1st and 3rd quartiles, middle line represents median, and whiskers extend to 1.5 times the interquartile range; n = 1000 equal to the size of the 3D models ensemble**. (b)** Element-wise Spearman cross-correlation coefficients (spCCC) between the experimental Hi-C contact map and the contact maps derived from the 3D models generated using co-localizing pairs of loci driven by the co-TrTS hotspots of 7 representative lncRNAs with triplex potential belonging to cluster 4.



Extended Data Fig. 4 | Correlation analysis of Hi-C and simulated contact maps for chromosome 22. (a) Distribution of the diagonal cross-correlation coefficients (dCCC) (Methods) in chromosome 22 of the contact maps derived from the ensemble of 3D models with Hi-C. Box boundaries represent 1st and 3rd quartiles, middle line represents median, and whiskers extend to 1.5 times the interquartile range. The statistical significance of the difference between each pair of dCCC distributions has been assessed with the two-sided Mann-Whitney rank test using python default parameters. The p-values are $< 10^{-3}$ unless reported; n = 1026 equal to the number of beads in chromosome 22.



Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Triplex forming lncRNAs govern long-range interactions. (a-f) Diagonal correlations coefficient (dCCC) along the first 10 Mb between the experimental Hi-C contact map and the contact maps derived from the ensemble of the 3D models generated using lncRNAs with triplex potential from cluster 4: (**A**) ENST00000547963.1, (**B**) ENST0000043436.1, (**C**) ENST00000449111.1, (**D**) ENST00000561611.2, (**E**) ENST00000540866.2, and (**F**) ENST00000421202.1 co-TrTS hotspots for each of the 23 chromosomes (grey) and genome-wide average (blue). Vertical red bar marks 250 kb, which is the median length of convergent CTCF loops[61]. **(g-i)** Diagonal correlations coefficient (dCCC) along the first 10 Mb between the experimental Hi-C contact map and the contact maps derived from the ensemble of 3D models generated using CTCF enriched genomic loci (orange), ENST00000541775.1 co-TrTS hotspots (light blue), and CTCF & ENST00000541775.1 enriched genomic loci (yellow) for all (**G**) metacentric, (**H**) submetacentric, and (**I**) acrocentric chromosomes. Chromosomes are classified according to the standard Denver classification[112].

nature research

Corresponding author(s): martirenom@cnag.crg.eu
irene.farabella@gmail.com

Last updated by author(s): Sep 9, 2021

Reporting Summary

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

Statistics

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

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

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	LAMMPS 14May16 COLVARS 2016-05-10
Data analysis	Custom script used for Data analysis are available at http://sgt.cnag.cat/3dg/datasets/ Data analysis: Triplexator (Buske et al., 2012) HOMER (Heinz et al., 2010) TADbit (Serra et al., 2017) BEDTools (Quinlan and Hall, 2010) Python 2.7.16 Python Libraries: Numpy 1.16.4 Pandas 0.24.2 Seaborn 0.9 Scipy 1.2.1 Matplotlib 1.5.1 Multiprocessing 0.7

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

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Custom scripts used to analyze data and generate figures are available at http://sgt.cnag.cat/3dg/datasets/.
RNA-seq datasets were downloaded from ENCODE at accession numbers: ENCSR000CPO, ENCSR000CQF, ENCSR000CQM, ENCSR530NHO, and ENCSR000CPS. PARS dataset was downloaded from Gene Expression Omnibus (GEO) at the accession number GSE50676. Hi-C was downloaded from Gene Expression Omnibus (GEO) at the accession number GSE63525. CTCF ChIP-seq were downloaded from ENCODE at the accession number ENCSR000AKB. DNAse-seq were downloaded from ENCODE at the accession number ENCFF097LEF, ENCFF273MVV, and ENCFF804BNU. The GENCODE v19 lncRNA set was downloaded from https://www.gencodegenes.org/human/release_19.html

Field-specific reporting

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

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

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

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

Sample size	Sample sizes were not predeterminated with a statistical method, but chosen accordingly to starndards in the field. To note, the theoretical approach presented in this work is on the cutting-edge in term of acquired statistics of what is currently achievable in computational physics-based genome-wide studies in human. Indeed, we have simulated a total of 185 chromosomes each in 1,000 independent trajectories that makes this study extremely advanced and robust form the computational point of view.
Data exclusions	No data exclusion were applied in this studies
Replication	We have simulated a total of 185 chromosomes each in 1,000 independent trajectories and no attempts to reproduce failed.
Randomization	No specific randomization method was used.
Blinding	Blinding was not relevant to this study because all matrices were derived from absolute quantitative methods without human subjectivity

Reporting for specific materials, systems and methods

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

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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