

UC San Diego

UC San Diego Previously Published Works

Title

Direct DNA crosslinking with CAP-C uncovers transcription-dependent chromatin organization at high resolution

Permalink

<https://escholarship.org/uc/item/8zc5z24m>

Journal

Nature Biotechnology, 39(2)

ISSN

1087-0156

Authors

You, Qiancheng
Cheng, Anthony Youzhi
Gu, Xi
[et al.](#)

Publication Date

2021-02-01

DOI

10.1038/s41587-020-0643-8

Peer reviewed



Published in final edited form as:

Nat Biotechnol. 2021 February ; 39(2): 225–235. doi:10.1038/s41587-020-0643-8.

Direct DNA crosslinking with CAP-C uncovers transcription-dependent chromatin organization at high resolution

Qiancheng You^{1,2,9}, Anthony Youzhi Cheng^{3,4,9}, Xi Gu^{1,2,9}, Bryan T. Harada^{1,2}, Miao Yu⁵, Tong Wu^{1,2}, Bing Ren^{5,6,∞}, Zhengqing Ouyang^{3,4,7,∞}, Chuan He^{1,2,8,∞}

¹Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, Chicago, IL, USA.

²Howard Hughes Medical Institute, Chicago, IL, USA.

³The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA.

⁴Department of Genetics and Genome Sciences and Institute for Systems Genomics, University of Connecticut, Farmington, CT, USA.

⁵Ludwig Institute for Cancer Research, La Jolla, CA, USA.

⁶Center for Epigenomics, Department of Cellular and Molecular Medicine, Institute of Genomic Medicine, and Moores Cancer Center, University of California, San Diego School of Medicine, La Jolla, CA, USA.

⁷Department of Biostatistics and Epidemiology, School of Public Health and Health Sciences, University of Massachusetts, Amherst, MA, USA.

⁸Department of Biochemistry and Molecular Biology, The University of Chicago, Chicago, IL, USA.

⁹These authors contributed equally: Qiancheng You, Anthony Youzhi Cheng, Xi Gu.

Abstract

Determining the spatial organization of chromatin in cells mainly relies on crosslinking-based chromosome conformation capture techniques, but resolution and signal-to-noise ratio of these

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

Correspondence and requests for materials should be addressed to B.R., Z.O. or C.H. biren@ucsd.edu; ouyang@schoolph.umass.edu; chuanhe@uchicago.edu.

Author contributions

C.H. and Q.Y. conceived the original idea. Q.Y., A.Y.C., X.G., Z.O. and C.H. designed the experiments. Q.Y., X.G., T.W. and M.Y., performed the experiments, A.Y.C. and B.L. performed analysis of the sequencing data. B.T.H. aided in the analysis of the FISH data. A.Y.C., Q.Y., X.G., B.R., Z.O. and C.H. analyzed the data and interpreted the finding. Q.Y., X.G., B.T.H. and A.Y.C. wrote the manuscript with input from B.R., Z.O. and C.H.

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/s41587-020-0643-8>.

Competing interests

C.H. is a scientific founder and a member of the scientific advisory board of Accent Therapeutics, Inc. and a shareholder of Epican Genetech. B.R. is a co-founder and a member of the scientific advisory board of Arima Genomics Inc.

Supplementary information is available for this paper at <https://doi.org/10.1038/s41587-020-0643-8>.

approaches is limited by interference from DNA-bound proteins. Here we introduce chemical-crosslinking assisted proximity capture (CAP-C), a method that uses multifunctional chemical crosslinkers with defined sizes to capture chromatin contacts. CAP-C generates chromatin contact maps at subkilobase (sub-kb) resolution with low background noise. We applied CAP-C to formaldehyde prefixed mouse embryonic stem cells (mESCs) and investigated loop domains (median size of 200 kb) and nonloop domains (median size of 9 kb). Transcription inhibition caused a greater loss of contacts in nonloop domains than loop domains. We uncovered conserved, transcription-state-dependent chromatin compartmentalization at high resolution that is shared from *Drosophila* to human, and a transcription-initiation-dependent nuclear subcompartment that brings multiple nonloop domains in close proximity. We also showed that CAP-C could be used to detect native chromatin conformation without formaldehyde prefixing.

The organization of DNA into chromatin in the nucleus of eukaryotic cells influences transcription, DNA replication and other nuclear processes^{1–3}. Chromosome conformation capture approaches (such as 3C and Hi-C)^{4–7} have been instrumental in elucidating the principles of chromatin folding. These techniques use formaldehyde-mediated crosslinking, in situ enzymatic fragmentation and proximity ligation to acquire contact frequency read-outs that are used to infer spatial relationships between different genomic loci. The acquisition of highly resolved contact maps in different species has led to the discovery of general structural features of genome organization, such as chromosome territories, compartments⁷, topologically associating domains (TADs)⁸, subtopologically associating domain (sub-TADs)⁹, insulated neighborhoods¹⁰, chromatin loops¹¹ and stripes³.

Most current methods to probe chromatin organization depend on extensive in situ crosslinking of protein (for example, transcription factors (TFs) and histones) with genomic DNA, which allows probing of the proximity between genomic loci mediated through the interacting proteins bound at each locus. However, the presence of these proteins masks enzymatic digestion sites and hinders proximal ligation of digested DNA fragments in these procedures. These limitations lead to substantially reduced resolution and increased background noise. The resolution gap between three-dimensional (3D) (Hi-C) and one-dimensional (1D) (that is, chromatin immunoprecipitation–sequencing (ChIP–seq), and assay for transposase-accessible chromatin–sequencing (ATAC–seq)) genomic maps limits our understanding of how transcription affects chromatin organization. Specifically, the typical kilobase (kb) resolution of Hi-C experiments in mammalian cells has limited the characterization of chromatin structure at gene level. Given that 77% of mouse protein-coding genes are less than 50 kb in length (median, 28 kb), new technologies that can assess chromatin conformation at subkilobase resolution may precisely reveal the relationships between chromatin organization and transcription.

To circumvent potential fragmentation and ligation limitations caused by interference from DNA-bound proteins and to increase the resolution and sensitivity of 3D genomic maps, we developed a crosslinking approach relying on a multifunctional chemical platform that captures proximal DNA loci through direct crosslinking on ultraviolet (UV) irradiation. We term this approach chemical-crosslinking assisted proximity capture (CAP-C). The crosslinked complexes can be isolated, purified and stripped of DNA-bound proteins, which

allows homogeneous DNA fragmentation into ~50–200 bp length around captured proximal loci followed by ligation. The absence of DNA-bound proteins dramatically reduces background noise commonly observed in in situ methods such as Hi-C. Applying CAP-C to formaldehyde prefixed mESCs we achieved subkilobase resolution of 3D genomic maps with high sensitivity and low background. Compared with another high-resolution Hi-C method (DNase Hi-C)¹², CAP-C captured more valid chromatin interactions at similar sequencing depth. From our high-resolution CAP-C contact maps, we found that local chromatin organization is closely related to transcription, revealed a critical role of active transcription on domain insulation and observed subnuclear compartments that are dependent on transcription initiation. We further applied CAP-C to investigate the native chromatin conformation without formaldehyde prefixing. We observed generally similar 3D genomic maps between native and formaldehyde prefixed chromatins, with native chromatins displaying more dynamic chromatin contacts than the fixed ones.

Results

CAP-C: a new crosslinking strategy to study chromatin architecture.

To establish CAP-C, an approach that captures proximal chromatin contacts without relying on protein–DNA crosslinking, we used a new type of crosslinker: multifunctional poly(amidoamine) (PAMAM) dendrimers¹³. The size of each generation dendrimer can be precisely tuned by controlling the number of repeated branching cycles that are performed during dendrimer synthesis¹⁴. We install tens of crosslinking groups on the surface of these spherical polymers with diameters ranging from 3 nm for generation 3 (G3) dendrimers to 11 nm for generation 9 (G9) dendrimers, which could potentially capture interactions in both closed and open chromatin conformations.

We used psoralen, which crosslinks to double-stranded DNA on UV irradiation¹⁵, to functionalize approximately half of the surface amine branches on PAMAM dendrimers. The remaining amine branches were masked with acetyl groups, making them inert to cellular interactions (Supplementary Fig. 1). To enhance the ligation between proximal loci, an azide handle was attached to the dendrimer surface that could react with a bifunctional bridge linker oligonucleotide for proper two-step ligation¹⁶. These azide- and psoralen-functionalized dendrimers will be referred to as dendrimers throughout the manuscript.

The CAP-C procedure includes diffusing dendrimers into either native (termed nCAP-C) or formaldehyde prefixed nuclei (termed pfCAP-C) of mouse embryonic stem cells (mESCs), followed by photocrosslinking to capture DNA in proximity to each dendrimer. For pfCAP-C, formaldehyde fixation was reversed after UV-induced dendrimer crosslinking. To expose all the DNA sequences for fragmentation, we used proteases to remove proteins from all DNA fragments crosslinked to dendrimers. The dendrimer–DNA complexes were then purified and digested using DNase I to ensure universal fragmentation of genomic DNA into 50 to 200 bp pieces with no bias toward specific chromatin regions (Supplementary Fig. 2). Fragmented DNAs were subjected to end-polishing and A-tailing. Next, bifunctional bridge linkers were added and attached onto the dendrimer via ‘click chemistry’¹⁷. Excess linkers not conjugated to dendrimers were removed through size selection. Proximal DNA fragments captured on the same dendrimer were then joined via the bridge linker by using a

two-step ligation¹⁶; this step reduces random DNA collisions and increases the specificity for intramolecular ligation. The ligated products were further enriched through capture of the biotin on the bridge linker and subjected to high-throughput sequencing (Fig. 1a). For each CAP-C experiment, dendrimers of a defined size were used as the chemical crosslinkers. We specify which dendrimer was used for each CAP-C experiment (for example, CAP-C conducted with dendrimer G3 is labeled as CAP-C (G3)).

To validate the feasibility of this approach, we first confirmed effective photocrosslinking between the modified dendrimers and genomic DNA (Supplementary Fig. 3a). As expected, capturing chromatin contacts requires both the dendrimer and UV irradiation (Supplementary Fig. 3b,c). pfCAP-C contact maps generated by each sized dendrimer (G3, G5 and G7) show high reproducibility between biological replicates (Supplementary Fig. 4).

CAP-C enables capture of native chromatin conformation.

We next conducted nCAP-C (G3, G5, G7 and G9) on mESC cells (see Methods) and successfully captured native mESC chromatin conformation (Supplementary Fig. 5a,b). We discovered that the size of the dendrimer plays an important role when capturing interactions in native chromatin. For instance, G3, the small-sized dendrimer with ~3.6 nm diameter, effectively captures interactions in fixed chromatin but is less effective on native chromatin (Supplementary Fig. 5a). In general, nCAP-C (G3, G5, G7 and G9) detected fewer chromatin loops and showed reduced contact signals on the contact matrix compared to pfCAP-C (G3, G5 and G7, Supplementary Fig. 5c,d). These differences could be due to native chromatin being more dynamic than fixed chromatin, such that the small-sized dendrimers might not be able to simultaneously contact adjacent DNA loci in a dynamic native chromatin environment upon UV irradiation. Nevertheless, nCAP-C (G5, G7 and G9) maps displayed high similarity to those of pfCAP-C (G3, G5 and G7) maps (Supplementary Fig. 5b). Compared to previous Hi-C results, CAP-C conducted on either native or fixed mESCs detect reported chromatin organizations, such as compartments, TADs, contact domains, loops and stripes (Supplementary Fig. 5e,f). As the 3D genomic maps show high similarity between native chromatin and formaldehyde prefixed ones, we performed pfCAP-C (G3, G5 and G7), which we will subsequently refer to as CAP-C, for the rest of the manuscript.

CAP-C resolves short-length-scale local chromatin structures with substantially reduced noise.

To assess the effectiveness of CAP-C in enriching short-range chromatin contacts, we examined the genome-wide average interaction frequency decay curves. These plots revealed a six- to eightfold increase in short-range (1–20 kb) interactions detected by CAP-C over in situ Hi-C libraries (Fig. 1d and Supplementary Fig. 6a). We further validated that the removal of DNA-bound proteins is critical to the enrichment of short-range contacts in CAP-C libraries (Supplementary Figs. 3d and 6c). Compared to in situ Hi-C at similar sequencing depth, contact maps plotted for CAP-C recapitulated higher-order chromatin features, such as compartments, TADs, contact domains, loops and stripes (Fig. 1b and Supplementary Fig. 7a,b). Most of the TAD boundaries, domains and loops detected by CAP-C were also reported in in situ Hi-C (Supplementary Fig. 8a–e), suggesting that CAP-C can identify

more short-range contacts while maintaining the ability to detect reported higher-order chromatin structures.

Compared to in situ Hi-C contact maps at 500 bp resolution, CAP-C maps were visually clearer and sharper with lower background noise (Fig. 1c and Supplementary Fig. 9, Supplementary Fig. 10a,b and see Methods), which enabled us to identify more small-sized contact domains (10–40 kb in length). These domains tend to encapsulate single genes, with domain boundaries that overlap active histone marks (H3K4me3 and H3K27ac) as well as various architectural proteins CTCF, MED12 and SMC1. DNase Hi-C has also been reported to achieve high resolution by enriching short-range contacts. However, compared with CAP-C at similar sequencing depth, DNase Hi-C captured notably fewer chromatin interactions (Supplementary Fig. 6b,d–f).

Next, we took advantage of the higher resolution of CAP-C over in situ Hi-C to obtain higher resolution maps of contact domains and loops genome-wide. To systematically characterize contact domains, we used Arrowhead¹¹ (see Methods) to obtain a total of 28,381 (CAP-C) and 29,028 (in situ Hi-C) unique contact domains across multiple resolutions. At 500 bp resolution, CAP-C identified at least ~1,000 more contact domains than in situ Hi-C (Fig. 1e). These additional contact domains appear to be functionally relevant as they are enriched with active histone marks. The overall size distribution of these domains was substantially smaller (median 9 kb) than previously characterized TADs (median 880 kb) in mESCs or contact domains (median 185 kb) previously identified in GM12878 cells, suggesting that CAP-C resolves additional short-length structures that were not previously appreciated.

To systematically characterize loops, we used HiCCUPS¹¹ (see Methods) to call enriched pixels (or loops) at multiple resolutions. CAP-C identified ~10,000 more loops than in situ Hi-C across all length scales (Fig. 1f). We demonstrated that enhanced signal-to-background ratio on the contact matrix allows CAP-C to identify more loops (Supplementary Figs. 11, 12a,b and 13a and Supplementary Note 1). To confirm the reliability of loops identified by CAP-C, we designed DNA fluorescent in situ hybridization (FISH) probes targeting loci around the loop anchors. We indeed discovered that these loci are in closer proximity than the negative controls (Supplementary Figs. 12c–j and 13b–e). In addition, loops called specifically in CAP-C showed much stronger correlation with active histone marks, CTCF and cohesin compared with loops identified by in situ Hi-C (Fig. 1f), indicating that CAP-C is a precise and sensitive method for identifying loops genome-wide.

Two types of chromatin domains with distinct structural and genomic properties.

Previous studies¹¹ reported that 40% of contact domains have loops demarcating their boundaries, which were termed loop domains^{10,18}. Unlike loop domains, contact domains that may not possess clearly visible loops at the intersection corners (we term this a ‘nonloop domain’) have been less characterized (Fig. 2a, left). Within the CAP-C mESC datasets, we categorized 12,882 nonnested contact domains into loop domains (20.1%) or nonloop domains (79.9%). Both types of domains showed no preference for a specific compartment (Fig. 2a, right). However, nonloop domains are significantly smaller (Fig. 2b), encapsulate

fewer protein-coding genes per domain (Fig. 2c) and lack CTCF motif orientation preference at the boundaries (Fig. 2a, middle and Supplementary Fig. 14a–c).

To investigate the genomic properties of the bodies and boundaries of the loop and nonloop domains, we used 18 ChIP–seq datasets, including different transcription factors and histone modifications from mESCs, to examine whether these features correlate differently with the two domain types (see Methods). By first correlating ChIP–seq signals with domain boundaries, we observed stronger enrichment of H3K4me3, RNA polymerase II (RNA Pol II) and H3K27ac at the boundaries of nonloop domains over loop domains versus slightly stronger enrichment of CTCF levels at the boundaries of loop domains over nonloop domains (Fig. 2d and Supplementary Fig. 15a). Fitting a logistic regression model and performing fivefold cross-validation on seven individually selected features, we observed that individual features, especially CTCF (area under curve (AUC) of the receiver operating characteristic curve 0.67) or SMC3 (AUC 0.67), are moderately predictive of domain type, suggesting CTCF or cohesin alone cannot adequately explain the two domain types. A combined model of all seven features (H3K4me3, RNA Pol II and H3K27ac, MED1, YY1, CTCF and SMC3) improved accuracy to a reasonable level (AUC 0.77). Standardized regression coefficients indicate that H3K4me3, RNA Pol II and H3K27ac can mostly explain nonloop domains, whereas CTCF and SMC3 can explain loop domains (Fig. 2e). Furthermore, using precision nuclear run-on sequencing (PRO-seq) data, we discovered that genes within the nonloop domains clearly show higher transcription rates than those in loop domains (Fig. 2f), suggesting that transcription might promote the formation of nonloop domains more than of loop domains.

In summary, we showed that chromatin domains can be categorized into at least two distinct classes: loop domains and nonloop domains. Loop domains are large domains spanned by convergently oriented CTCF binding and characterized by a relatively modest transcription rate, whereas nonloop domains tend to be smaller, contain actively transcribed genes and show boundaries enriched with active promoters and enhancers (Supplementary Fig. 15b,c).

CTCF and transcription both contribute to domain organization.

We then investigated contributions of CTCF and transcription to the formation of loop and nonloop domains by performing CAP-C in mESCs with either CTCF depletion or transcription inhibition (Supplementary Fig. 16a,b and Supplementary Note 2). To quantitatively measure the average changes in intradomain interactions between untreated and perturbed samples, we performed meta-domain analysis for loop domains and nonloop domains (Fig. 2i and Supplementary Fig. 16c, and see Methods). Inhibition of RNA Pol II elongation caused a greater loss of intradomain contacts in the body of nonloop domains than loop domains (Fig. 2i). As expected, greater contact losses were observed in the body of nonloop domains in compartment A than in compartment B. Transcription inhibition also led to reduced loop contacts within-loop domains in compartment B (Fig. 2i, Supplementary Fig. 17 and Supplementary Note 3).

Regardless of compartment status, acute depletion of CTCF decreased chromatin contacts between loop anchors in loop domains (Fig. 2i), which affected large loop domains more than small ones (Supplementary Fig. 16d). In nonloop domains, CTCF removal also resulted

in observable contact losses, but affected smaller nonloop domains more than larger ones, which is the opposite trend as observed in loop domains (Fig. 2i and Supplementary Fig. 16d). The patterns of intradomain contact loss between the two types of domain on CTCF removal are different as well (Fig. 2i), with loop domains displaying more long-range contact losses between loop anchors, whereas nonloop domain showing more decreases of short-range interactions near diagonal regions, suggesting that CTCF affects intradomain interactions differently in loop domains and nonloop domains.

Simultaneous depletion of CTCF and inhibition of transcription elongation resulted in greater loss of intradomain contacts than each individual treatment for all domains (Fig. 2i and Supplementary Fig. 16d), suggesting that transcription and CTCF binding make complementary contributions toward mediating interactions within loop and nonloop domains. Loop domain and nonloop domain classified using loops called from CTCF ChIA-PET also recapitulated the same observations (Supplementary Fig. 16c). In summary, the higher resolution CAP-C maps revealed that CTCF promotes intradomain interactions for both loop domains and nonloop domains but likely through different mechanisms¹⁹.

Induction of transcription initiation establishes weak chromatin insulation.

We next investigated whether transcription and CTCF binding contribute to domain boundary strength. Similar to previous observations¹⁹, removal of CTCF led to a near complete loss in the boundary strength of loop domains (Supplementary Fig. 18a). However, transcription inhibition only gave rise to moderate boundary losses on loop domains (Supplementary Fig. 18a) without changing CTCF or cohesin levels (Fig. 2g,h), suggesting that transcription elongation itself might only facilitate but is not responsible for establishing insulation of loop domain boundaries. In contrast, we discovered that both CTCF depletion and transcription inhibition weakened boundaries of nonloop domains (Supplementary Fig. 18b), indicating that both CTCF binding and transcription elongation are partially responsible for the insulation of nonloop domains.

We further discovered that genes with alternative promoter usage²⁰ can form chromatin boundaries on their promoter sites without the presence of CTCF (Fig. 3a–d and Supplementary Fig. 19a–d). Classification of alternative promoters based on their transcription status suggested that only loci with active promoter showed clear boundaries (Fig. 3c,d). In addition, the formation of these boundaries is not dependent on RNA Pol II binding (Supplementary Figs. 19e and 20a–d). Furthermore, we found that transcription induction on silenced chromatin is sufficient to create weak chromatin insulation, but recruiting RNA Pol II binding alone may not be sufficient to create strong boundaries (Fig. 4a–e and Supplementary Fig. 18c). Therefore, in addition to RNA Pol II binding, the establishment of active transcription complexes around stably transcribed transcription start sites (TSSs) is likely required to cause clear chromatin insulation (Supplementary Note 4).

We also observed that compartmentalization changed after transcription induction. We found gross changes in the Pearson correlation maps (Fig. 4g,h) and a longer periodicity in the autocorrelation plot of compartment eigenvector tracks (Fig. 4f, see Methods), suggesting a decrease in compartmentalization after transcription induction. We found that transcription

initiation could be involved in switching loci from compartment B to compartment A (Fig. 4i,j, Supplementary Fig. 18d,e and Supplementary Note 5).

CAP-C identifies transcription-state-dependent small compartments.

Compartments have been shown as fine-scale structures (10 kb resolution) in *Drosophila* genome^{21,22}. Together with our observations that local changes in transcription state can cause short-length changes in compartmentalization (100 kb), we hypothesized that transcription-state-dependent compartmentalization may also exist in mouse and human genomes at kilobase resolution. Testing this hypothesis requires a way to detect compartments at high resolution. However, computing Hi-C or CAP-C ‘compartment eigenvectors’ by eigendecomposition of the Pearson’s correlation matrix for resolution less than 25 kb requires substantial time and computer memory. In our CAP-C data, we observed that contact maps derived from large (G5 and G7) and small (G3) dendrimers were slightly different, with larger dendrimers preferring A over B compartment (Supplementary Fig. 21a). To determine whether this feature allows us to resolve compartments at high resolution, we performed singular value decomposition (SVD) on the CAP-C mean-centered data matrix and generated a 1D ‘CAP-C eigenvector’ and two-dimensional (2D) ‘dendrimer map’ (Supplementary Fig. 21b and see Methods). Dendrimer maps were reproducible between biological replicates (Supplementary Fig. 21c).

Dendrimer maps at multiple resolutions visually reproduced the ‘plaid-like’ patterns observed in Pearson correlation matrices computed from Hi-C contact map data. The CAP-C eigenvector displayed relatively high Pearson’s correlation coefficient with the Hi-C compartment eigenvector ($R = 0.896$, 500 kb; $R = 0.855$, 50 kb) (Fig. 5a), suggesting that CAP-C eigenvector might be an alternative way to identify compartments. The CAP-C eigenvector could be generated quickly and easily for up to the maximum map resolution (5 kb in this study) at any given chromatin loci, allowing us to determine compartments in large genomes, such as mouse and human, at high resolution. The CAP-C eigenvector identified small (25–100 kb) compartment intervals that were missed by the compartment eigenvectors previously derived for mouse and human genomes (Fig. 5b–d and Supplementary Fig. 21d–g). The CAP-C eigenvector displays positive correlation with active histone marks H3K36me3 (Spearman’s correlation coefficient = 0.491; 5 kb) and assay for transposase-accessible chromatin peaks (Spearman’s correlation coefficient = 0.193; 5 kb), and shows negative correlation with H3K9me2 (Spearman’s correlation coefficient = -0.161 , 5 kb). Therefore, these small compartments share similar features with the broad compartments observed at lower resolutions, suggesting that the finer-scale compartmentalization observed by our CAP-C eigenvector is real.

Therefore, our CAP-C eigenvector analysis uncovered that, similar to organisms with much smaller genomes (for example, *Drosophila*), mammalian chromatin exhibits fine-scale segregation into A and B compartments corresponding to open active and compact inactive chromatin, respectively.

CAP-C detects transcription-initiation-dependent chromatin subcompartments.

On transcription inhibition in our flavopiridol-treated samples, we frequently identified regions showing increased interdomain interactions compared with wild type (WT) controls (Fig. 6a–c). It was validated that the increased interdomain interactions represent statistically significant increases in chromatin contacts instead of noises on the contact maps (Fig. 6g and Supplementary Fig. 22a,b). A total of 491 regions with increased interdomain interactions were detected genome-wide in CAP-C contact maps, of which 81% overlap at least one protein-coding gene (see Methods). A total of 365 gene pairs were subsequently mapped from these regions, and used as the basis for defining 228 gene clusters (or connected component subgraphs, see Methods). These clusters have two or more genes interacting with each other (Fig. 6a–c,e and Supplementary Fig. 23b) and resemble higher-order nuclear subcompartments^{5,23,24}; they also occur more frequently in nonloop domains than loop domains (Fig. 6d and Supplementary Fig. 23a). These gene clusters generally have higher Pol II-S5P levels and transcription rates than a randomized set of gene pairs (Fig. 6h,f). Multiple actively transcribed DNA regions are thought to cluster together to form higher-order chromosomal interactions, which can exhibit crucial effects on proper gene expression regulation^{25,26}. Because flavopiridol treatment inhibits elongation and causes the accumulation of initiating Pol II, we hypothesized that, at the gene clusters whose interactions are strengthened by flavopiridol treatment, transcription initiation might contribute to the clustering of these genes by bringing multiple active promoters in close proximity. Therefore, we tentatively refer to these regions as transcription-initiation clusters (TICs).

To further characterize TICs, we designed DNA–fluorescence in situ hybridization (–FISH) probes targeting TIC loci and performed DNA–FISH combined with immunofluorescence imaging. On inhibition of transcription elongation, we observed increased contacts on CAP-C contact maps around TIC loci (Fig. 6g), accompanied with elevated Pol II-S5P occupancy (Fig. 6h). Similarly, inhibition of transcription elongation caused Pol II-S5P to cluster near TICs (Supplementary Fig. 26a) and induced gene pairs within TIC loci to closer proximity (Supplementary Figs. 22c,d and 25d), suggesting that Pol II-S5P might be responsible for mediating interactions within TICs. We then performed proximity ligation-assisted chromatin immunoprecipitation (PLAC-seq)²⁷ on Pol II-S5P under the same conditions to detect interactions (loops) mediated by Pol II-S5P (see Methods). We found that Pol II-S5P loops increased notably after inhibition of transcription elongation around TICs (Fig. 6i). Besides Pol II-S5P, CYCT1 and BRD4 were also found to cluster around TICs on flavopiridol treatment (Supplementary Figs. 24a,b and 26b,c), suggesting that TIC formation could be mediated by multiple transcription factors. To rule out other factors such as CTCF looping that may also contribute to TIC formation, we analyzed TIC interactions for cells treated with both inhibition of transcription elongation and CTCF removal. We found that the interactions among TIC not only remained enriched on the contact map of CAP-C but also showed increases compared to the unperturbed control (Fig. 6g). We also observed corresponding increases of the Pol II-S5P level (Fig. 6h) and Pol II-S5P-mediated loops (Fig. 6i), suggesting that TICs are formed with no or less dependence on CTCF or CTCF looping.

To confirm that TIC formation is dependent on transcription initiation, we performed CAP-C and DNA-FISH on mESCs treated with triptolide. On inhibition of transcription initiation, gene pairs in TICs are no longer in proximity and no clear Pol II-S5P clusters were observed nearby TIC loci as well (Supplementary Figs. 25b–d and 26a). Similarly, no TIC contacts were found on the CAP-C contact matrix (Supplementary Fig. 25a), suggesting that TIC formation is dependent on transcription initiation.

Discussion

Current methods for investigating 3D genome structures mostly rely on in situ protein–DNA crosslinking. The presence and extensive crosslinking of DNA-bound proteins (for example, TFs and histones) masks enzymatic digestion sites and hinders proximal ligation in Hi-C procedures. The partial digestion of the genome from an ensemble cell population can lead to heterogeneous lengths of chromatin fragments, which increase the propensity for background noise on 3D genomic maps and prevent the sensitive identification of proximal chromatin interactions. Digestion of protein–DNA-crosslinked complexes with MNase or DNase I have been applied, but these approaches are still subject to interference from extensive protein–DNA crosslinking and show high preference to ligate loci at nucleosome-depleted sites or open chromatin sites. CAP-C represents a new, unbiased strategy for studying chromatin architecture. It uses multifunctional chemical crosslinkers that substitute the DNA-bound proteins to crosslink proximal genomic DNA loci directly. The varied sizes allow these probes to access both open and closed compartments. Because DNA fragments are covalently linked to dendrimers, DNA-bound proteins can be removed, allowing the DNA to be homogeneously fragmented to ~50–200 bp length. This more uniform fragmentation reduces background noise on the 3D genomic maps by reducing false or heterogeneous ligation. The technique enriches short-range chromatin interactions, increasing resolution and sensitivity. Collectively, CAP-C enables resolving more functionally relevant loops, domains and subnuclear compartments with high sensitivity at subkilobase resolution.

Using CAP-C, our study draws a closer link between transcription and chromatin organization. The reduced false ligation rate and substantially reduced background noise in CAP-C enabled us to identify more small-sized nonloop domains at high resolution, the boundaries of which are frequently enriched with active promoters and enhancers and encompass highly transcribed genes. Such features are in stark contrast to loop domains characterized previously, which tend to be large, contain multiple modestly transcribed genes and are spanned by convergently oriented CTCF loops. Hsieh et al.²⁸ also reported similar discoveries of small-sized, transcription-associated domains named as micro-TADs. Future investigations are required to study the relationship between micro-TADs and nonloop domains.

Our results indicate that transcription mainly contributes to local chromatin organization at a short-length range and promotes contacts within nonloop domains more than loop domains. Similar observations were also recently reported that transcription inhibition mainly decreases chromatin interactions at the gene scale²⁸. As nonloop domain boundaries are enriched for active promoters, we found that transcription initiation instead of transcription

elongation is key to chromatin insulation. In fact, previous Hi-C studies also revealed that chromatin insulation is independent of transcription elongation during *Drosophila* and mouse early embryo development^{29,30}. We suggest that upon transcription initiation, active transcription protein complexes (including RNA Pol II) established around the TSS, rather than only RNA Pol II binding, are largely responsible for chromatin insulation. This conclusion is consistent with previous in situ Hi-C results that RNA Pol II recruitment alone is insufficient to cause chromatin insulation changes³¹.

Taking advantage of the three dendrimer-based crosslinkers with different abilities to access chromatin structures at various length scales, we introduced the CAP-C eigenvector to resolve chromatin compartmentalization patterns at high resolution. Our analyses revealed the presence of the transcription-state-dependent, fine-scale compartments within mouse and human genomes. The high sensitivity of CAP-C further enabled characterization of a unique subnuclear compartment, the TIC, which brings multiple genes with high transcription rates from nonloop domains into close proximity. We found that TICs are dynamic and affected by the transcription dynamics. Transcription initiation helps to cluster several actively transcribed genes. Transcription factors might then enrich at corresponding loci to phosphorylate Pol II for proper transcription elongation. Thus, phosphorylated RNA Pol II might constantly condensate and distribute to transcribed genes within TICs, which could help maintain the high expression levels of these genes. Inhibiting transcription elongation attenuates release of phosphorylated RNA Pol II from TICs, and as a result, condensates of RNA Pol II form around TICs. We suspect that RNA Pol II clustering around TICs supplies sufficient active RNA Pol II to TSSs of multiple genes to maintain their high-levels of expression.

RNA Pol II was recently shown to cluster in vivo on inhibition of transcription elongation using live-cell, single-molecule super-resolution imaging^{32,33}. Clustering of RNA Pol II was further shown to be a result of phase separation^{32,34}. As three of the TIC-related proteins (CYCT1, BRD4, Pol II-S5P) were demonstrated to form phase separated condensates in vivo^{35–37}, we hypothesize that TIC may form through phase separation. We did show that inhibition of phase separation using 1,6-hexanediol leads to disruption of TIC; however, 1,6-hexanediol lacks specificity to address the potential requirement of phase separation for TIC formation (Supplementary Fig. 27a–c). Further work is required to reveal the mechanism of TIC formation.

Methods such as split-pool recognition of interactions by tag extension (SPRITE) allow genome-wide mapping of higher-order chromatin interactions²⁶. SPRITE overcomes the limitation of proximity ligation between two chromatin loci, enabling detecting multiple spatial proximal chromatin contacts within nucleus. In the future, the crosslinked DNA–dendrimer complexes generated in CAP-C, which preserve intact chromatin organization information, could be purified and coupled with other downstream methods such as SPRITE or super-resolution fluorescent imaging to assay multiple chromatin interactions simultaneously. Finally, CAP-C can detect native chromatin conformation, which could potentially enable investigating the dynamics of enhancer–promoter interactions in the future.

Methods

Cell culture.

F123 mouse embryonic stem cells were grown on gamma-irradiated mouse embryonic fibroblast cells (Thermo A34180) under standard conditions with 85% DMEM (Gibco), 15% Knockout Serum Replacement (Thermo Fisher Scientific), 0.1 mM of nonessential amino acids, 0.1 mM of β -mercaptoethanol, 1 mM of glutamine, 500 U ml⁻¹ of LIF, 100 U ml⁻¹ of penicillin and 100 μ g ml⁻¹ of streptomycin. Before collecting for CAP-C, in situ Hi-C and DNase Hi-C, F123 mESCs were passaged onto feeder free 0.2% gelatin-coated plates for at least two passages to rid the culture of feeder cells. The CTCF-AID knock-in mouse embryonic stem cells expressing TIR1-9myc were gifts from B. Ren³⁸, the cells were passaged on 0.1% gelatin-coated plates without mouse embryo fibroblasts (MEFs). We added 1 μ l of 500 mM auxin (Abcam) per 1 ml of medium to deplete CTCF, and changed medium with auxin every 24 h. Cells were gathered 48 h after auxin treatment. Transcription elongation was inhibited by addition of 1 μ M of flavopiridol (Sigma) to CTCF-AID knock-in mouse embryonic stem cells for 6–8 h. Transcription initiation was inhibited by an addition of 1 μ M of triptolide (Sigma) to mESCs for 6–8 h. Blocking phase separation was performed for mESCs by addition of 5% (v/v) 1,6-hexanediol (Sigma) at room temperature for 10 min or 3% (v/v) 1,6-hexanediol at room temperature for 15 s. *Drosophila* S2 cells were cultured in Schneider's *Drosophila* Medium (Gibco) supplemented with 10% heat-inactivated fetal bovine serum (FBS) (Gibco). Cells were grown at 28 °C without CO₂. HepG2 cells were cultured in DMEM (Gibco) supplemented with 10% (v/v) FBS, 100 U ml⁻¹ of penicillin and 100 μ g ml⁻¹ of streptomycin and grown at 37 °C with 5% CO₂. HCT116 cells were cultured in McCoy's 5A medium (Gibco) supplemented with 10% (v/v) FBS, 2 mM of L-glutamine, 100 U ml⁻¹ of penicillin and 100 μ g ml⁻¹ of streptomycin at 37 °C with 5% CO₂. Induction of nonannotated TSS was performed on HCT116 cells as reported previously. Briefly, HCT116 cells were treated with 500 nM DAC (Sigma) for 72 h. DAC containing medium was refreshed every 24 h.

Synthesis of psoralen, azide functionalized PAMAM dendrimer.

To synthesize each functional dendrimer crosslinker, 1.54 μ mol of PAMAM dendrimer G3 (Sigma), G5 (Sigma) or G7 (Sigma) was dissolved in 2 ml of methanol, respectively. To modify half of the amine branches on the surface of each dendrimer, 24.64, 98.56 or 394.24 μ mol of SPB (Thermo Fisher Scientific) was added in the solution containing dendrimer G3, G5 or G7, respectively (Total available amines on the surface of each dendrimer are 32, 128 and 512 for G3, G5 and G7). To generate a site to attach bridge linker, 1.54 μ mol of NHS-PEG₄-Azide (Thermo Fisher Scientific) was added to each dendrimer reaction solution. Then 5 μ l of Et₃N (Sigma) was added to each reaction solution to catalyze the reaction. The reaction mixture was stirred overnight at room temperature before the addition of 100 μ l of Ac₂O (Sigma) to modify the rest unreacted amine branches with acetyl groups. Each reaction mixture was stirred for another 24 h at room temperature before the addition of 3 ml of water to neutralize the reaction. Each dendrimer solution was purified and concentrated with a 3 KDa Amicon Ultra Centrifugal Filter Unit (Millipore) by centrifugation at 14,000g for 10 min at room temperature.

Characterization of psoralen, azide-modified PAMAM dendrimer.

Psoralen has a characteristic of UV absorbance at 300 nm. Unfunctionalized dendrimers have no UV absorbance at 300 nm. Thus, we could characterize the psoralen-modified dendrimers by measuring their UV absorbance at 300 nm. A series of SPB solutions with known concentrations were prepared as the standard solutions. UV absorbance at 300 nm of each standard solution was measured by UV-visible light spectrometry. The calibration curve was plotted with the values of UV absorbance on the vertical axis and the concentrations of standard solutions on the horizontal axis. UV absorbance at 300 nm was then measured for each psoralen-modified dendrimer. The concentration of each psoralen-modified dendrimer was determined using the calibration curve.

CAP-C.

In this study, we developed two types of CAP-C; the restriction enzyme fragmentation-based CAP-C and the DNase I fragmentation-based CAP-C. For each type, we performed experiments for native chromatin (termed nCAP-C) or chromatin prefixed with formaldehyde (termed pfCAP-C). The detailed procedures used for each type of experiments are listed in the following subsections.

Crosslinking dendrimer with chromatin.

For formaldehyde fixed cells.: Grow five million cells under recommended culture conditions. Detach adherent cells by centrifugation at 300g for 5 min. Resuspend cells in fresh medium at 1 million cells per 1 ml of medium. Add 16% formaldehyde solution (Thermo Fisher Scientific) to a final concentration of 1% (v/v). Incubate at room temperature for 5 min on rotating rocker. Add fresh prepared 2.5 M of glycine solution to a final concentration of 0.2 M to quench the reaction. Incubate at room temperature for 5 min on a rotating rocker. Centrifuge for 5 min at 300g at 4 °C. Discard the supernatant. Resuspend cells in 1 ml of cold 1× PBS and spin for 5 min at 300g at 4 °C. Discard the supernatant and flash-freeze cell pellets in liquid nitrogen (can be stored in –80 °C for up to 1 year). Combine 250 µl of ice-cold CAP-C lysis buffer (10 mM of Tris-HCl, pH 8.0, 10 mM of NaCl, 0.2% Igepal CA-630) with 50 µl of protease inhibitors (Sigma). Add to formaldehyde fixed cell pellets. Incubate cell suspension on ice for 20 min. Centrifuge at 2,500g for 5 min at 4 °C. Discard the supernatant. Wash pelleted nuclei once with 500 µl of ice-cold CAP-C lysis buffer. Centrifuge and discard the supernatant. Resuspend the cell pellet in 500 µl of 50 µM dendrimer in methanol. Incubate at 4 °C on a rocker with rotation for 10 min. Photocrosslink the nuclei by irradiating under 365 nm of UV for 15 min at 4 °C. The nuclei should be allowed to rest on ice for 5 min before another 15 min of 365 nm of UV irradiation at 4 °C. Centrifuge for 5 min at 2,500g at 4 °C. Discard the supernatant. Wash pelleted nuclei twice with 500 µl of ice-cold CAP-C lysis buffer. Centrifuge and discard the supernatant. Resuspend the pellet in proteinase K digestion buffer (420 µl of CAP-C lysis buffer, 50 µl of 10% SDS, 30 µl of 20 mg ml⁻¹ proteinase K). Incubate the resuspension at 65 °C overnight on a thermomixer at 800 r.p.m.

For nonfixing cells.: Combine 250 µl of ice-cold nucleus lysis buffer (10 mM of Tris-HCl, pH 7.5, 10 mM of NaCl, 3 mM of MgCl₂, 0.5% NP-40, 0.15 mM of spermine (Sigma), 0.5

mM of spermidine (Sigma)) with 50 μ l of protease inhibitors (Sigma). Add to 5 million cells without fixing. Incubate cell suspension on ice for 5 min. Centrifuge at 500g for 5 min at 4 °C. Discard the supernatant. Wash pelleted nuclei once with 500 μ l of nucleus resuspension buffer (10 mM of Tris-HCl pH 7.4, 15 mM of NaCl, 60 mM of KCl, 0.15 mM of spermine (Sigma), 0.5 mM of spermidine (Sigma)). Centrifuge at 500g for 5 min at 4 °C and discard the supernatant. Resuspend the cell pellet in 500 μ l of 50 μ M dendrimer in methanol. Incubate at 4 °C on a rocker with rotation for 10 min. Photocrosslink the nuclei by irradiating under 365 nm of UV for 15 min at 4 °C. The nuclei should be allowed to rest on ice for 5 min before another 15 min of 365 nm of UV irradiation at 4 °C. Centrifuge for 5 min at 2,500g at 4 °C. Discard the supernatant. Wash pelleted nuclei twice with 500 μ l of nucleus resuspension buffer. Centrifuge and discard the supernatant. Resuspend the pellet in proteinase K digestion buffer (420 μ l of nucleus resuspension buffer, 50 μ l of 10% SDS, 30 μ l of 20 mg ml⁻¹ proteinase K). Incubate the resuspension at 65 °C overnight on a thermomixer at 800 r.p.m.

Purify UV crosslinked DNA–dendrimer complexes.—Extract the DNA with 500 μ l of phenol:chloroform (Sigma). Centrifuge at the maximum for 10 min at room temperature. Transfer the upper layer to a new tube. Add 800 μ l of EtOH and 50 μ l of 3 M NaOAc (pH 5.5, Thermo Fisher Scientific). Incubate at –80 °C for 1–2 h. Centrifuge at 14,000g for 15 min at 4 °C. Discard the supernatant. Wash the pellet twice with 800 μ l of ice-cold 80% EtOH. Centrifuge at the maximum for 5 min at 4 °C. Discard the supernatant. Resuspend the DNA pellet in 50 μ l of water.

Fragment genome and proximity ligation.

For restriction enzyme fragmentation-based CAP-C.: Digest DNA–dendrimer complexes with MboI:

Component	Amount (μ l)
DNA-dendrimer complex	50
10 \times NEB buffer 2	20
5 U μ l ⁻¹ MboI (NEB)	20
H ₂ O	110
Total	200

Incubate at 37 °C overnight on a thermomixer at 800 r.p.m. The next day, inactivate MboI by incubating at 65 °C for 20 min on a thermomixer at 800 r.p.m. Mark the DNA ends with biotin as detailed:

Component	Amount (μ l)
dCTP (10 mM)	1.5
dGTP (10 mM)	1.5

Component	Amount (μ l)
dTTP (10 mM)	1.5
Biotin-14-dATP (0.4 mM, Thermo Fisher Scientific)	37.5
5 U μ l ⁻¹ DNA polymerase I, Large (Klenow) Fragment (NEB)	8
Total	50

Incubate at 37 °C for 1 h on a rotating rocker. Inactivate the Klenow by incubating at 65 °C for 30 min on a thermomixer at 800 r.p.m. Proximity ligation in the ultradiluted solution is conducted as follows:

Component	Amount (μ l)
10 \times NEB T4 DNA ligase buffer	500
10 mg ml ⁻¹ BSA (Thermo Fisher Scientific)	12
400 U μ l ⁻¹ T4 DNA ligase (NEB)	20
H ₂ O	6,200
Total	6,732

Incubate at 16 °C for 8 h or overnight on a rotating rocker.

For DNase I fragmentation-based CAP-C.: Digest DNA–dendrimer complexes with DNase I:

Component	Amount (μ l)
2 U μ l ⁻¹ DNase I (NEB)	1
10 \times DNase I reaction buffer	5
DNA-dendrimer complex	44
Total	50

Incubate at 37 °C for 5 min then stop the reaction by adding 150 μ l of stop buffer (20 mM of EDTA). Incubate the mixture at 75 °C for 10 min. Purify DNA with ethanol precipitation by adding 800 μ l of EtOH and 50 μ l of 3 M NaOAc (pH 5.5). Incubate at -80 °C for 1 h. Centrifuge at the maximum for 15 min at 4 °C. Discard the supernatant. Wash the pellet twice with 800 μ l of 80% EtOH. Centrifuge at the maximum for 5 min at 4 °C. Discard the supernatant. Resuspend the DNA pellet in 100 μ l of H₂O. Repair DNA ends and add 'A' using the KAPA Hyperplus Kit (KAPA, KK8515):

Component	Amount (μ l)
ER&AT buffer mix	28

Component	Amount (μl)
ER&AT enzyme mix	12
DNA-dendrimer complex	200
Total	240

Incubate at 20 °C for 30 min then 65 °C for 30 min. Purify DNA with ethanol precipitation by adding 500 μl of EtOH and 20 μl of 3 M NaOAc (pH 5.5). Incubate at –80 °C for 1 h. Centrifuge at the maximum for 15 min at 4 °C. Discard the supernatant. Wash the pellet twice with 800 μl of 80% EtOH. Centrifuge at the maximum for 5 min at 4 °C. Discard the supernatant. Resuspend the DNA pellet in 100 μl of H₂O. Attach a biotin linker to the dendrimer by addition of 2 μl of a 100 μM bridge linker (customized with IDT):

Bridge linker_F: /5Phos/GTCAGA/iDBCON/AAGATATCGCGT

Bridge linker_R: /5Phos/CGCGATATC/iBiodT/TATCTGACT

Incubate at 37 °C for 2 h on a thermomixer at 800 r.p.m. Excess of bridge linkers are removed by size selection with Ampure beads. DNA above 100 bp are preserved after size selection. DNA is eluted with 880 μl of H₂O. The two-step ligation was carried out as previously described in diluted solution¹⁶.

Component	Amount (μl)
DNA-dendrimer complex	880
10× NEB T4 DNA ligase buffer	100
400 U μl ⁻¹ T4 DNA ligase (NEB)	20
Total	1,000

The two-step ligation was carried out as previously described in diluted solution¹⁶.

Purify biotinylated ligation product.

For restriction enzyme fragmentation-based CAP-C.: Extract the DNA with 7 ml of phenol:chloroform (Sigma). Centrifuge at 2,000*g* for 10 min at room temperature. Transfer the upper layer to a new tube. Add 17.5 ml of EtOH and 700 μl of 3 M NaOAc (pH 5.5). Incubate at –80 °C for 1 h. Centrifuge at 10,000*g* for 20 min at 4 °C. Discard the supernatant. Resuspend the pellet in 300 μl of water. Shear DNA–dendrimer complexes using Diagenode Bioruptor Pico under the following conditions: Volume of Library, 100 μl in a 0.65 ml Diagenode tube. Program, 30 s on; 30 s off for six cycles.

For DNase I fragmentation-based CAP-C.: Extract the DNA with 1 ml of phenol:chloroform (Sigma). Centrifuge at 2,000*g* for 10 min at room temperature. Transfer the upper layer to a new tube. Add 2.5 ml of EtOH and 700 μl of 3 M NaOAc (pH 5.5). Incubate at –80 °C for 1 h. Centrifuge at 10,000*g* for 20 min at 4 °C. Discard the supernatant. Resuspend the pellet in 264 μl of water. Remove bridge linkers that were only ligated to one DNA:

Component	Amount (μl)
DNA-dendrimer complex	264
10 \times Lambda Exonuclease Buffer (NEB)	30
5 U μl^{-1} Lambda Exonuclease (NEB)	3
20 U μl^{-1} Exonuclease I (NEB)	3
Total	300

Incubate at 37 °C for 30 min. Heat inactivate the enzyme at 75 °C for 10 min. Shear DNA–dendrimer complexes using Diagenode Bioruptor Pico under the following conditions: Volume of Library, 100 μl in a 0.65 ml Diagenode tube. Program, 30 s on; 30 s off for six cycles.

Biotin pull-down and construct library.—Prepare for biotin pull-down by washing 20 μl of 10 mg ml^{-1} Dynabeads MyOne Streptavidin C1 beads (Thermo Fisher Scientific) with 600 μl of 1 \times Tween Washing Buffer (1 \times TWB: 5 mM of Tris-HCl (pH 7.5); 0.5 mM of EDTA; 1 M of NaCl; 0.05% Tween 20). Separate on a magnet and discard the solution. Resuspend the beads in 300 μl of 2 \times Binding Buffer (2 \times BB: 10 mM of Tris-HCl (pH 7.5); 1 mM of EDTA; 2 M of NaCl) and add to the reaction. Incubate at room temperature for 30 min with rotation to bind biotinylated DNA to the streptavidin beads. Separate on a magnet and discard the solution. Wash the beads by adding 600 μl of 1 \times TWB and transferring the mixture to a new tube. Heat the tubes on a thermomixer at 55 °C for 2 min with mixing. Reclaim the beads using a magnet. Discard the supernatant. Repeat wash. Resuspend beads in 100 μl of 1 \times NEB T4 DNA ligase buffer and transfer to a new tube. Reclaim beads and discard the buffer. To repair ends of sheared DNA and remove biotin from unligated ends, resuspend beads in 100 μl of master mix:

Component	Amount (μl)
1 \times NEB T4 DNA ligase buffer	88
25 mM dNTP	2
10 U μl^{-1} T4 PNK (NEB)	5
3 U μl^{-1} T4 DNA polymerase I (NEB)	4
5 U μl^{-1} DNA polymerase I, Large (Klenow) Fragment (NEB)	1
Total	100

Incubate at room temperature for 30 min. Separate on a magnet and discard the solution.

Wash the beads by adding 600 μl of 1 \times TWB and transferring the mixture to a new tube. Heat the tubes on a thermomixer at 55 °C for 2 min with mixing. Reclaim the beads using a magnet. Discard the supernatant. Repeat wash. Resuspend beads in 100 μl of 1 \times NEBuffer 2 and transfer to a new tube. Reclaim beads and discard the buffer. Resuspend beads in 100 μl of dATP master mix:

Component	Amount (μ l)
1 \times NEBuffer 2	90
10 mM dATP	5
5 U μ l ⁻¹ Klenow Fragment (3' \rightarrow 5' exo-) (NEB)	5
Total	100

Incubate at 37 °C for 30 min. Separate on a magnet and discard the solution.

Wash the beads by adding 600 μ l of 1 \times TWB and transferring the mixture to a new tube. Heat the tubes on a thermomixer at 55 °C for 2 min with mixing. Reclaim the beads using a magnet. Discard the supernatant. Repeat wash. Resuspend beads in 100 μ l 1 \times quick ligation reaction buffer (NEB) and transfer to a new tube. Reclaim the beads and discard the buffer. Resuspend beads in 55 μ l of quick ligation master mix:

Component	Amount (μ l)
1 \times Quick ligase reaction buffer (NEB)	50
Quick ligase (NEB)	2
Illumina indexed adapter (Nextflex)	3
Total	55

Incubate at room temperature for 30 min. Separate on a magnet and discard the solution.

Wash the beads by adding 600 μ l of 1 \times TWB and transferring the mixture to a new tube. Heat the tubes on a thermomixer at 55 °C for 2 min with mixing. Reclaim the beads using a magnet. Remove the supernatant and repeat the wash. Wash three times with 100 μ l of water. Reclaim the beads with 23 μ l of water and transfer the beads to an eight-well PCR tube.

PCR amplify 10–12 cycles in 50 μ l of reaction by addition of 25 μ l of 2 \times HiFi PCR Master Mix (KAPA) and 2 μ l of Primer Mix (KAPA) to the 23 μ l of beads under the following conditions:

Initial denaturation	98 °C	30 s
Denaturation	98 °C	15 s
Annealing	60 °C	30 s
Extension	72 °C	30 s
	10–12 cycles (depends on samples)	
Final extension	72 °C	1 min

Purify the libraries with 0.9 \times Ampure beads and elute with 30 μ l of water.

Check the ligation efficiency by aliquot 8 μl of DNA libraries and adding 1 μl of 10 \times CutSmart buffer (NEB), 1 μl of 10 U μl^{-1} BspDI (NEB). Incubate at 37 °C for 1 h. Run a 2% agarose gel with BspDI digested libraries and undigested libraries side by side. A clear shift-down to a small size should be observed with BspDI digested libraries for restriction enzyme fragmentation-based CAP-C.

Libraries are then sequenced on a NextSeq500 pair end for 50 bp or NovaSeq pair end for 150 bp.

In situ Hi-C.

In situ Hi-C experiments were performed as previously described using the MboI restriction enzyme¹¹. We performed in situ Hi-C on mESC and sequenced 2 billion raw reads. The self-made in situ Hi-C was demonstrated to be of high quality and the results similar to published data (Supplementary Fig. 28). Thus, we merged 2 billion reads of in house generated in situ Hi-C with 8 billion reads of in situ Hi-C from Bonev et al.³¹.

DNase Hi-C.

DNase Hi-C experiments were performed as previously described¹².

ChIP-seq.

ChIP-seq experiments was performed as described in ENCODE experimental protocols³⁹ with minor modifications. Cells are crosslinked with 1% formaldehyde for 10 min and quenched with 200 mM of glycine. Five million cells are used for each ChIP sample. Crosslinked cell pellets are resuspended in 1 ml of ice-cold lysis buffer (50 mM of HEPES, pH 7.9; 5 mM of MgCl_2 ; 0.2% Triton X-100; 20% glycerol; 300 mM of NaCl) for 10 min on ice then centrifuge at 500g for 5 min at 4 °C. Remove the supernatant. Resuspend the pellet in 1 ml of 0.1% SDS lysis buffer (50 mM of HEPES, pH 7.5; 1 mM of EDTA; 1% Triton X-100; 0.1% sodium deoxycholate; 0.1% SDS; 150 mM of NaCl) and incubate on ice for 10 min. Shearing of chromatin is performed using Diagenode Bioruptor Pico for sonication using following parameters: 30 s on; 30 s off; 15 cycles at 4 °C. For immunoprecipitation, use 40 μl of Protein G Dynabeads (Thermo Fisher Scientific) and wash them with 1 ml of cold 0.1% SDS lysis buffer three times. The sheared chromatin is added to the beads and incubated on a rotating platform at 4 °C for 2 h to preclear and then save the supernatant. Another 40 μl of Protein G Dynabeads (Thermo Fisher Scientific) are washed with 1 ml of cold 0.1% SDS lysis buffer. After washing, 5 μg of antibodies are added to the beads. Precleared chromatin is then added together and incubated on a rotating platform at 4 °C for 8 h or overnight. After incubation, beads are collected on a magnetic rack and washed twice with ice-cold 1 ml of 0.1% SDS lysis buffer (50 mM of HEPES, pH 7.5; 1 mM of EDTA; 1% Triton X-100; 0.1% sodium deoxycholate; 0.1% SDS; 150 mM of NaCl); followed by washing twice with 1 ml of high salt buffer (50 mM of HEPES, pH 7.5; 1 mM of EDTA; 1% Triton X-100; 0.1% sodium deoxycholate; 0.1% SDS; 350 mM of NaCl); twice with 1 ml of LiCl Wash Buffer (10 mM of Tris-HCl, pH 8.0; 1 mM of EDTA; 0.5% NP-40; 0.5% sodium deoxycholate; 250 mM LiCl) and once with 1 ml of cold TE buffer (10 mM of Tris-HCl, pH 8.0; 1 mM of EDTA; 0.2% Triton X-100). After washing, 100 μl of ChIP elution buffer (50 mM of NaHCO_3 , 10 mM of EDTA, 1% SDS) is added and samples

are incubated at 65 °C for 4 h at 800 r.p.m. on a thermomixer. The beads are discarded using a magnetic rack and 300 mM of NaCl is added to the samples to reverse crosslink by incubating at 65 °C overnight. For input samples, 20 µl of sheared chromatin saved after sonication is added to 80 µl of ChIP Elution Buffer and incubated at 65 °C overnight with the other samples. The input samples are processed in parallel with the ChIP samples from here on. Then, 5 µl of RNase A is added to each sample and incubated at 37 °C for 1 h, and 5 µl of proteinase K is added and incubated at 65 °C for 1 h. The samples are purified using the Zymo DNA Clean & Concentrator kit (Zymo). The purified DNA libraries are prepared using the KAPA Hyperplus Kit (KAPA). Then, PCR amplify 8–14 cycles under following conditions: 98 °C 30 s; 98 °C 15 s; 60 °C 30 s; 72 °C 30 s; repeat for 12 cycles; 72 °C 1 min.

Purify the libraries with 0.9× Ampure beads and elute with 30 µl of water. Libraries can then be sequenced on the HiSeq4000 single end for 50 bp.

PLAC-seq.

PLAC-seq experiments were performed as previously described with minor modifications²⁷. Collect cells by centrifugation at 200g for 5 min at room temperature. Resuspend cell pellets with fresh medium without serum at a concentration of 1×10^6 cells per ml. Add methanol-free formaldehyde to a final concentration of 1% (w/v) and rotate for 15 min at room temperature. Add 2.5 M of glycine solution to a final concentration of 0.2 M to quench crosslinking reaction. Rotate for 5 min at room temperature. Spin at 2,500g for 5 min at 4 °C and discard the supernatant. Wash cell pellets with cold 1× PBS and spin at 2,500g for 5 min at 4 °C. Resuspend up to 5 million crosslinked cells in 300 µl of cold lysis buffer (10 mM of Tris-HCl, pH 8.0; 10 mM of NaCl; 0.2% IGEPAL CA-630) with protease inhibitors (Sigma). Rotate at 4 °C for at least 15 min. Spin at 2,500g for 5 min at 4 °C and remove the supernatant. Resuspend the pellet with 500 µl of cold lysis buffer with protease inhibitor cocktail. Spin at 2,500g for 5 min at 4 °C and then remove the supernatant. Gently resuspend the cell pellet in 50 µl of 0.5% SDS (avoid excess foaming) and incubate for exactly 10 min at 62 °C. Add 135 µl of water and 25 µl of freshly prepared 10% Triton X-100 to quench the SDS. Mix well gently (avoid excess foaming) and incubate for 15 min at 37 °C. Add 25 µl of 10× NEBuffer 2 and 4 µl of 25 U µl⁻¹ MboI. Mix well and digest chromatin for exactly 2 h at 37 °C in a thermomixer, shaking at 900 r.p.m. Inactivate MboI by incubation at 62 °C for 20 min. Cool the reaction to room temperature (~10 min). Add the following reagents to fill in overhangs and mark with biotin:

Component	Amount (µl)
dCTP (10 mM)	1.5
dGTP (10 mM)	1.5
dTTP (10 mM)	1.5
0.4 mM of biotin-14-dATP (Thermo Fisher Scientific)	37.5
5 U µl ⁻¹ DNA polymerase I, Large (Klenow) Fragment (NEB)	8
Total	50

Mix well and incubate for 1 h at 37 °C in a thermomixer, shaking at 900 r.p.m. Prepare the ligation master mix as follows:

Component	Amount (µl)
H ₂ O	664
10× T4 ligase buffer (NEB)	120
10% Triton X-100	100
20 mg ml ⁻¹ BSA (Thermo Fisher Scientific)	6
400 U µl ⁻¹ T4 DNA ligase (NEB)	10
Total	900

Add this ligase master mix to the reaction and mix well by inverting the tube. Rotate at room temperature for 2 h. After ligation, spin down the nuclei at 2,500g for 5 min at 4 °C and discard the supernatant. Resuspend the pellet in 1 ml of 0.1% SDS Lysis Buffer (50 mM of HEPES, pH 7.5; 1 mM of EDTA; 1% Triton X-100; 0.1% sodium deoxycholate; 0.1% SDS; 150 mM of NaCl). Shearing of chromatin is performed using Diagenode Bioruptor Pico for sonication under following parameters: 30 s on; 30 s off; 15 cycles at 4 °C. Transfer 40 µl of resuspended magnetic beads per reaction (for 1–3 million cells) to a 1.7-ml microcentrifuge tube and wash twice with 500 µl of cold 0.1% SDS Lysis Buffer. Place the tube on the magnetic rack for 2 min and remove the supernatant. Add the sheared chromatin to the beads and incubate on a rotating platform at 4 °C for 2 h to preclear and save the supernatant. Wash another 40 µl of Protein G Dynabeads (Thermo Fisher Scientific) with 1 ml of cold 0.1% SDS lysis buffer. After washing, add 5 µg of Pol II-S5P antibodies to the beads. Precleared chromatin is then added together and incubated on a rotating platform at 4 °C for 8 h or overnight. After incubation, beads are collected on a magnetic rack and washed twice with ice-cold 1 ml of 0.1% SDS lysis buffer; followed by washing twice with 1 ml of high salt buffer (50 mM of HEPES, pH 7.5; 1 mM of EDTA; 1% Triton X-100; 0.1% sodium deoxycholate; 0.1% of SDS, 350 mM of NaCl); twice with 1 ml of LiCl wash buffer (10 mM of Tris-HCl, pH 8.0; 1 mM of EDTA; 0.5% NP-40; 0.5% sodium deoxycholate; 250 mM of LiCl) and once with 1 ml cold TE buffer (10 mM of Tris-HCl, pH 8.0; 1 mM of EDTA; 0.2% Triton X-100). After washing, 100 µl of ChIP elution buffer (50 mM of NaHCO₃, 10 mM of EDTA, 1% SDS) is added and samples were incubated at 65 °C for 4 h at 800 r.p.m. on a thermomixer. The beads are discarded using a magnetic rack and 300 mM of NaCl is added to the samples to reverse crosslink by incubating at 65 °C overnight. For input samples, 20 µl of sheared chromatin saved after sonication is added to 80 µl of ChIP Elution Buffer and incubated at 65 °C overnight with the other samples. The input samples are then processed in parallel with the ChIP samples from here on. Next, 5 µl of RNase A is added to each sample and incubated at 37 °C for 1 h, and 5 µl Proteinase K is added and incubated at 65 °C for 1 h. The samples are purified using the Zymo DNA Clean & Concentrator kit (Zymo). Then, wash 20 µl streptavidin C1 beads (Thermo Fisher Scientific) per sample twice with 400 µl of 1× Tween Washing Buffer (1× TWB: 5 mM of Tris-HCl, pH 7.5; 0.5 mM of EDTA; 1 M of NaCl; 0.05% Tween 20) Resuspend the beads in 40 µl of 1× Binding Buffer (1× BB: 5 mM of Tris-HCl, pH 7.5; 0.5 mM of EDTA; 1 M of NaCl).

Combine ChIPed DNA (40 μ l) with C1 beads in 1 \times BB. Incubate at room temperature for 30 min with rotation. Collect the beads on a magnet (1–2 min) and remove the supernatant. Wash the beads twice with 500 μ l of 1 \times TWB. Wash beads once with 100 μ l of 10 mM Tris-HCl (pH 8.0). Resuspend the beads in 43 μ l of 10 mM Tris-HCl (pH 8.0). Add 5 μ l of 10 \times End-Polishing Buffer and 2 μ l of End-Polishing Enzyme Mix (Qiagen Ultralow Input Library Kit, 180492). Mix gently and incubate at 25 $^{\circ}$ C for 30 min, 65 $^{\circ}$ C for 15 min and finally hold at 4 $^{\circ}$ C in a thermomixer. Add the following reagents in the order listed when incubation is done:

Component	Amount (μ l)
Nuclease-free water	18
4 \times Ultralow input ligation buffer (Qiagen Ultralow Input Library Kit, 180492)	25
QIaseq Adapter (1:10 diluted) (Qiagen, 180985)	2
Ultralow Input Ligase (Qiagen Ultralow Input Library Kit, 180492)	5

Mix gently by pipetting 5–6 times and incubate at 25 $^{\circ}$ C for 10 min in a thermomixer. Collect beads on magnet (1–2 min), and remove the supernatant. Wash beads twice with 500 μ l of 1 \times TWB. Wash beads once with 100 μ l of 10 mM Tris-HCl (pH 8.0). Resuspend the beads in 24.5 μ l of 10 mM Tris-HCl (pH 8.0). PCR amplify 12–14 cycles in 50 μ l of reaction by adding 25 μ l of 2 \times HiFi PCR Master Mix (KAPA) and 1.5 μ l of Primer Mix (KAPA) to the 23.5 μ l of library template under following conditions:

Initial denaturation	98 $^{\circ}$C	2 min
Denaturation	98 $^{\circ}$ C	20 s
Annealing	60 $^{\circ}$ C	30 s
Extension	72 $^{\circ}$ C	1 min
	12–14 cycles (depends on samples)	
Final extension	72 $^{\circ}$ C	1 min

Purify the libraries with 0.9 \times Ampure beads and elute with 30 μ l of water. Libraries are then sequenced on a HiSeq4000 pair end for 50 bp.

DNA–FISH combined with immunofluorescence.

DNA–FISH probes were custom designed and generated by Agilent technologies (Supplementary Table 1). MEF feeder cells (Gibco) were cultured on a covered slide glass coated with poly-L-lysine (Sigma) and 0.5% gelatin. mESCs were then cultured on MEF feeder cells for 24 h. mESCs were fixed with 4% PFA solution in PBS (Santa Cruz Biotechnology) for 10 min at room temperature. After three washes with PBS, cells were

permeabilized with 0.5% Triton X-100 (Thermo Fisher Scientific) in 1× PBS for 10 min at room temperature. Following three washes in 1× PBS for 5 min, the cells were blocked with 4% bovine serum albumin (BSA) (Thermo Fisher Scientific) in PBST (0.1% Tween 20 in PBS) for 1 h at room temperature. Cells were then incubated with diluted primary antibodies (anti-BRD4, Abcam ab128874 1:500 dilution; anti-CYCT1, Abcam ab238940, 1:500 dilution; anti-Pol II-S5P Abcam ab5408 1:500 dilution) in 4% BSA in PBST for 2 h at room temperature. The cells were rinsed three times in 1× PBS for 5 min each time, then incubated with the secondary antibodies (goat antirabbit Alexa Fluor 647, Invitrogen A21244, 1:500 dilution) in 4% BSA in 1× PBS at room temperature in the dark. Cells were washed three times with 1× PBS for 5 min each time.

The cells were fixed again with 4% PFA in PBS for 10 min at room temperature then rinsed in 1× PBS three times for 1 min each. Cells were denatured in 0.1 M of HCl solution for 10 min and rinsed in 1× PBS three times for 1 min each. A coplin jar was filled with the denatured solution (28 ml of formamide, 4 ml of 20× SSC (Promega), 8 ml of H₂O) and prewarmed to 82 °C in a water bath. Cells were denatured by immersing the slides with cells in the denaturing solution at 82 °C for 5 min. Cells were incubated in 70% ethanol, 85% ethanol and then 100% ethanol for 1 min at room temperature. While air drying the cells, a hybridization solution was prepared by mixing 7 µl of FISH hybridization buffer with 1 µl of FISH probe 1, 1 µl of FISH probe 2 and 1 µl of H₂O together. The mixture was denatured at 73 °C for 5 min. After addition of 5 µl of hybridization solution on the cells, another coverslip was added and the edges sealed with rubber cement. The slides were transferred to a humidified incubator at 37 °C to incubate overnight, then the cover glasses and rubber cement were removed. Slides with cells were washed in the FISH wash buffer 1 (Agilent technologies, G9401A) at 73 °C for 2 min, followed by washing in the FISH wash buffer 2 (Agilent Technologies, G9402A) at room temperature for 1 min. The cells were air dried at room temperature in the dark, then incubated in Hoechst 33342 (Thermo Fisher Scientific) to stain nuclei for 20 min at room temperature in the dark, before washing in 1× PBS three times for 5 min each. Cells were air dried for 10 min. The slides were mounted with antifade mounting medium (Vectashield) and sealed with nail polish. Images were acquired at the Leica SP5 Tandem Scanner Spectral 2-Photon confocal microscope with a ×100 objective (Integrated Light Microscopy Core Facility, University of Chicago). Images were postprocessed using Fiji is just ImageJ (FIJI). For analysis of distances, images of fiducial beads (Tetraspeck, Thermo Scientific) were used to generate a first degree polynomial function to map the coordinates between the color channels analyzed. To avoid issues with image registration along the *z* axis, maximum intensity projections were analyzed to measure distances in the *x*–*y* plane only. Average immunofluorescence intensities around FISH loci were quantified as previously described³⁶.

Quantification and statistical analysis.

Relative contact frequency and counts per million (CPM) transformation.—We use either the relative contact frequency, *R*, or CPM transformation⁴⁰, *C*, to make comparisons between libraries as nonequal sequencing depths have to be accounted for.

$$R_{ij} = \frac{X_{ij}}{L_c}$$

$$\text{CPM}_{ij} = \frac{X_{ij}}{L_c} 10^6$$

where X refers to intrachromosomal contact matrix, X_{ij} refers to the contact frequency between locus pair i and j , and L_c refers to the sum of all intrachromosomal interactions in a chromosome, c .

ChIP-seq data processing.

All ChIP-seq datasets were aligned to the mm10 or hg19 reference genome using BWA-MEM⁴¹. PCR duplicates were removed using Picard Tools v.2.2.2 (<http://broadinstitute.github.io/picard/>). Peak calling was performed using MACS2 (ref.⁴²) on BAM files using the following parameters: -q 0.05 -f BAM [-g mm|-g hg]. For visualization of all ChIP-seq as tracks, BEDTools v.2.26 (ref.⁴³) was used to create bedGraph files consisting of depth-normalized CPM values. bedGraph files were converted to bigWig files for fast query retrievals using UCSC bedGraphToBigWig (<https://genome.ucsc.edu/util.html>). Programmatic access to bigWig files was done via pyBigWig (<https://github.com/deeptools/pyBigWig>).

RNA-seq data analysis.

RNA-seq data for the mESC F123 cell line was pseudo-aligned and quantified directly using kallisto using the following parameters: -b 100 -l 180 -s 20-single. To determine gene expression abundance, protein-coding genes were ranked by the transcripts per million metric.

CAP-C data processing.

All CAP-C libraries were sequenced on the Illumina NextSeq500 or HiSeq2500. Paired-end FASTQ formatted files were aligned to the mm10, hg19 or dm3 reference genome using BWA-MEM⁴¹. Each read end was aligned independently before combining them into a single BAM file. Singletons and chimeric reads were filtered off. PCR duplicates were removed using Picard Tools. MboI restriction sites BED file was prepared containing the positions of all possible GATC cut sites using HiCPro⁴⁴. We performed an extra filtering step to remove all contact distances less than 1 kb in length to ensure that (1) all strand orientations were equally represented as a function of genomic distance and (2) because they were highly inconsistent across all replicates sequenced, and severely skewed the relative contact frequencies of long-range (≥ 20 kb) interactions. Only valid and filtered contact pairs were used for all downstream analysis (Supplementary Table 2). A preformatted text file was prepared before converting it into a .hic formatted file with the following resolutions (0.5, 1, 2, 5, 10, 25, 40, 50, 100, 250 and 500 kb and 1 and 2.5 megabase (mb)). Finally, a MAPQ ≥ 1 and MAPQ ≥ 30 .hic file was generated for each sample. To generate a CAP-C merge file, all G3, G5 and G7 primary and replicated libraries in the preformatted text file

were concatenated and merged-sorted. Juicer was then used to convert the merged-sorted file into a .hic formatted file. All contact matrices visually represented as contact maps were VC-normalized with Juicer⁴⁵.

DNase I fragmentation-based CAP-C data processing.

For DNase I fragmentation-based libraries, we have additional preprocessing steps to be performed before the instructions listed in the CAP-C data processing section. We first used SeqPrep to merge overlapping read pairs (2×150 bp). For the resulting merged and unmerged read pairs, cutadapt was used to trim the bridge linker adaptor (ACGCGATATCTTATCTGACT, both forward and reverse complementary) off the read from the 5' (-g) and 3' (-a) direction. Up to one adaptor count per read (-n) was used, and a minimum overlap length of ten threshold (-overlap) between the linker and read was used. Only read pairs where the length of at least one read was too short (<15 bp) or merged read pairs that failed to possess a linker were removed (Supplementary Table 2). Because DNase I was used to fragment BL-CAP-C datasets, no restriction cut site annotation files was done to annotate the paired reads.

Reproducibility of contact matrices.

The stratum-adjusted correlation coefficient using HiCRep⁴⁶ was calculated for 10, 25, 50, 100 and 1-mb resolution contact matrices for up to 5 mb in distance, with a smoothing factor of 20, 10, 5, 2 and 0. Comparisons were done between the primary and replicate libraries (MboI fragmented CAP-C) or among G3, G5 and G7 libraries (DNase I fragmented CAP-C).

In situ Hi-C data processing.

All in situ Hi-C libraries were sequenced on the Illumina NextSeq500 or HiSeq2500. Paired-end FASTQ formatted files were aligned to the mm10 reference genome using BWA-MEM⁴¹. Each read end was aligned independently before combining them into a single BAM file. Singletons and chimeric reads were filtered off. PCR duplicates were removed using Picard Tools. An MboI/DpnII restriction sites BED file was prepared containing the positions of all possible GATC cut sites using HiCPro⁴⁴. To be consistent with the CAP-C data, we performed an extra filtering step to remove all contact distances less than 1 kb in length. Only valid, and filtered contact pairs were used for all downstream analysis (Supplementary Table 2). A preformatted text file was prepared before converting it into a .hic formatted file with the following resolutions (0.5, 1, 2, 5, 10, 25, 40, 50, 100, 250 and 500 kb and 1 and 2.5 mb). Finally, a MAPQ \geq 1 and MAPQ \geq 30.hic file was generated for each sample.

PLAC-seq data processing.

We used Fit-HiC⁴⁷ to call high-confidence interactions in PLAC-seq for 1, 2, 5, 10 and 25-kb-resolution contact matrices (MAPQ > 30). Interactions with a *q* value of <0.05 and spanned distance greater than resolution used were kept.

Euclidean distance criteria.

Determining the overlap or concordance of 2D features is based on the criteria proposed by¹¹

$$\min(0.2 \times \text{length}, 50\text{kb}) < = \sqrt{(x_1 - x_2)^2 + (y_1 + y_2)^2}$$

Compartments, domains, loops calling and external validation.

Compartment eigenvectors were called at 500, 100, 50 and 25 kb resolution using the eigendecomposition of the correlation matrix from the O/E matrix⁷. Contact domains were called using the Arrowhead module of Juicer⁴⁵. To use the increase in short-range contacts in CAP-C, we sensitively call more short-range contact domains by making changes to the Blockbuster.java file to toggle the minimum width parameter. We set this to zero. A customized Juicer jar file was recompiled based on these changes. We called contact domains at 500 bp and 1, 2, 5, 10, 25 and 50 kb resolutions, and merged all nested and nonnested contact domains into a unique merged set. If two or more domains were considered to be similar (based on the Euclidean distance criteria), the contact domain with the highest corner score was retained. Loops or peaks in the deeply sequenced CAP-C and in situ Hi-C libraries were called using HiCCUPS¹¹ with the following high-resolution parameters (hiccup -m 1024 -r 5000,10000 -k VC -f .1,.1 -p 4,2 -i 7,5 -t 0.02,1.5,1.75,2 -d 20000,20000).

Detection of stripes.

The method used to detect all 5' and 3' stripe was a variation of the implementation described in Vian et al.³. For the detection of stripes, matrix-balanced contact frequencies at a 10-kb resolution were used. From the beginning of a chromosome, we query a 5-mb matrix and identify stripes within this window. We first iteratively test whether an element in the matrix is a member of a stripe. For each element (*i*, *j*) in the matrix, we test a predefined observed region against a predefined background region listed next. Using Poisson statistics, we determine the mean along the background region as the mu parameter and perform a one-sided test to determine whether the mean of the observed region is significantly greater (significance 0.1) than the background. We then repeat this procedure across the chromosome using overlapping windows with a step size of 2.5 mb to ensure stripes are not missed.

The predefined observed and background regions are listed as such:

Observed region:

Vertical : ContactMap[(*i* - *m*) : *i*, *j*] and ContactMap[*i* : (*i* + *m*), *j*]

Horizontal : ContactMap[*i*, (*j* - *m*) : *j*] and ContactMap[*i*, *j* : (*j* + *m*)]

where $m = 2$ and ContactMap is the square matrix containing matrix-balanced contact frequencies.

Background region:

$$\text{Top : ContactMap}[(i - \text{flank} + m) : (i - \text{flank})j]$$

$$\text{Bottom : ContactMap}[(i + (\text{flank} - 4) + m) : (i + (\text{flank} - 4), j)]$$

$$\text{Left : ContactMap}[i, (j - \text{flank} + m) : (j - \text{flank})]$$

$$\text{Right : ContactMap}[i, (j + (\text{flank} - 4) + m) : (j + (\text{flank} - 4))]$$

where flank = 6 and $m = 4$.

Detection of a 3' stripe (horizontal on the contact map) and a 5' stripe (vertical on the contact map) is dealt with differently. For 3' stripes, the horizontal region is tested twice against both the top and bottom regions. For 5' stripes, the vertical region is tested twice against both the left and right regions. The resulting matrix is a binarized matrix with 0 representing a failure to reject the background parameterized null hypotheses. We postprocess the binary matrix by reverting values to 0 if there is not more than 20 one elements per row as well as rows that do not have more than 15 consecutive one elements. Finally, the genomic position of the row possessing resulting one elements are tabulated as stripe occurring loci.

Enrichment of histone modifications and transcription factor ChIP-seq at domain boundaries and loops.

Enrichment was calculated based on the ChIP-seq signals at the domain boundary or loop locus pair against a random domain or loop set. Random domain or loop genomic intervals were drawn using BEDTools shuffle based on the size distribution of either domains or loops.

CAP-C eigenvectors versus compartmental eigenvectors.

Given a distribution of physical distances for each locus pair along the formaldehyde-crosslinked genome, each dendrimer is expected to independently sample and UV-crosslink locus pairs within the dendrimer's effective distance (or capture) range. We sought to describe patterns of contacts yielded from different dendrimers for each locus (bin) along the 1D genome or locus pair (bin pair) from the 2D contact map. To explore how contact maps generated from varying dendrimer sizes relate with compartments, we performed principal component analysis jointly on all three dendrimer experiments. The relative contact frequency data matrix, M , is described to have m dendrimers at the rows and n locus bins (1D) as features on the columns. The data matrix is mean-centered and decomposed into

their eigenvectors and eigenvalues by SVD. The top-ranking eigenvectors (either first or second) were used to correlate with known histone modification marks (H3K36me3 or H3K9me2) that were informative of compartments. Data matrix was also transformed by projecting it onto the selected eigenvectors to identify the dendrimer sample's attribution (positive or negative). To standardize the direction of the eigenvector between principal component analysis done on different chromosomes, we arbitrarily swapped signs of the selected eigenvector if transformed value of the G3 sample is positive (that is, negative if G3). We term this the CAP-C eigenvectors (or CAP-C eigen in short) to differentiate between compartmental eigenvectors: the traditional analysis for compartment detection. Unlike CAP-C eigenvector, compartment eigenvectors⁷ are characterized from the eigendecomposition of the correlation matrix initially calculated using the symmetrical matrix-balanced observed over expected contact matrix (O/E).

CAP-C eigenvector (2D).

For 2D CAP-C eigenvector analysis, the relative contact frequency data matrix, M , is described to have m dendrimer experiments at the rows and $({}^n C_2 + n)$ locus bins (2D) as features on the columns. After performing SVD, we reorder the values of the selected eigenvector back into the original (i, j) locus pair position on the contact map. We term these dendrimer maps.

Advantages of CAP-C eigenvector (1D) or (2D).

Because compartmental eigenvectors are computed based on the co-variability of locus pairs while CAP-C eigenvectors are computed based on the differences in contact frequencies obtained from each dendrimer, d , for each locus pair (2D) to compute the covariance matrix, dendrimer maps can be computed and visualized just for a selected region (or subset) of the data. We could also compute 1D CAP-C eigenvector values quickly and easily for up to the maximum map resolution for each dendrimer's contact map. From our results, we reason that smaller sized dendrimers (G3) are able to access regions of compacted and closed chromatin while larger ones (G5, G7) are a better fit to the open chromatin configurations.

ChromHMM.

For the F123 mESC cell line, we retrieved ChromHMM states and performed liftOver from the mm9 to mm10 genomic coordinates. For the HepG2 cell line, 15 ChromHMM states were called using ChromHMM from H3K4me1, H3K4me3, H3K9me3, H3K27ac, H3K27me3, H3K36me3 and Pol II, and subsequently annotated using terms described previously.

Definition of loop versus nonloop domains.

To investigate the boundary properties of loop versus nonloop domains, we first built a nonoverlapping and nonnested contact domain set. Contact domains called from 0.5, 1, 2, 5, 10, 25 and 50 kb resolution were agglomeratively (starting from the highest resolution contact domains; that is, 0.5 kb resolution) filled into any available genomic regions not already placed by a contact domain. From this resulting contact domain set, domains were classified as loop domains if peaks overlapped the domain corners within the Euclidean

distance criteria. This definition of loop domains was previously established⁴⁸. All remaining contact domains were classified as nonloop domains.

Characterizing boundaries loop and nonloop domains.

For all ChIP-seq datasets of histone modification marks and transcription factor peaks, we extracted only CPM values ± 5 kb around the boundary, and CPM values at 5 to 95% around the domain body using pyBigWig.

Analyses of perturbation experiments.

To investigate the role of CTCF or transcription in mediating intradomain interactions, we visualize the loss of intradomain interactions across the stratum of the domain. We use loop and nonloop domains annotated from the deep sequenced CAP-C. Boundaries of domains were first extended by 30% in both directions, and then rescaled domains to fit $(16 \times 17)/2$ bins (length of each bin represents 10% of the domain length). Contact frequencies for each bin in each domain was rescaled to the total intradomain counts per domain (total to 100%). Once rescaled, we take the paired difference between the condition and WT, and visualize the average differences for all rescaled loop, nonloop and size-matched nonloop domains.

Meta-analysis of contact domains with varying sizes.

To build mean contact maps of contact domains with varying sizes, boundaries of each contact domain were first extended both upstream and downstream by 30%, and contact frequencies recomputed into 32 bins (six upstream and 20 for the body plus six downstream). The rescaled contact probability was then computed for each contact map, and a mean contact map visualized based on the array of rescaled contact matrices. We then performed meta-analysis on loop domains, nonloop domains, and domains whose boundaries overlaps a promoter and enhancer state. At the same time, we also visualized cohesin ChIA-PET interactions⁴⁹ overlapping promoter and enhancers. Promoter-enhancer interactions are unlikely to have the same characteristics as CTCF-cohesin loops we see in loop domains.

Size-matching nonloop domains.

Using weighted random sampling, we generated a set of nonloop domains ($n = 1,016$) that was domain-sized matched with the distribution of domain size for loop domains ($n = 2,595$).

Analyses involving distribution of contact domain sizes and number of genes per domains.

Contact domains overlapping at least 50% of the length of the gene is considered to overlap a single gene.

Directionality index and TADs.

Directionality indices (DI) were calculated based on the following according to Dixon et al.⁸ for quantifying strength of domain boundaries (that is, the upstream versus downstream bias) in a one-dimensional linear fashion.

$$DI = \left(\frac{B - A}{|B - A|} \right) \left(\frac{(A - E)^2}{E} + \frac{(B - E)^2}{E} \right)$$

where

A is the total number of contacts pairs for all bins between the locus to the locus $R \times p$ kb upstream.

B is the total number of contact pairs for all bins between the locus to the locus $R \times p$ kb downstream.

E is the expected number of contact pairs for each locus.

R is the resolution used to bin the contact matrix. Generally, we use $R = 5$ kb resolution.

p is the number of bins. Generally, we use $p = 50$ bins.

Analyses involving domain boundary formation based on the orientation of gene pairs.

Ensembl gene annotations (v.87) of the mm10 reference genome was used to determine the longest transcript of a gene. We classified the orientation of each consecutive gene pair as divergent (reverse-forward strand), convergent (forward-reverse strand), and tandem (forward-forward and reverse-reverse strands). We plot average contact maps of VC-normalized counts and O/E values by centering the TSS or transcription end site of the second gene in the gene pair.

Alternative promoter analyses.

RefSeq annotated transcripts were used to discover alternative promoters by selecting genes with different TSSs (1,046 genes or 4.2% of all RefSeq genes had more than one TSS). Because certain genes have more than two known and unique TSSs, we simplified the analysis by selecting genes with only two unique TSSs that are at least 5 kb apart. Genes that were less than 10 kb at their longest were filtered away (689 genes remained). We classified these genes with multiple alternative promoters into four different classes based on the arrangement in which their TSS overlapped an active promoter ChromHMM state and noted that 128 (18.6%) did not have either TSS overlap any active promoter states; 99 (14.4%) had their upstream TSS, and not their downstream TSS, overlap an active promoter region; 300 (43.5%) had their downstream TSS, but not their upstream TSS, overlap an active promoter region and 162 (23.5%) had both TSS overlapped active promoters. Based on these classifications, we plotted VC-normalized mean contact maps and O/E maps to determine the strength of their domain boundaries. We also overlapped these regions with RNA Pol II ChIP-seq, PRO-seq (determine presence and strength of divergent transcription) and directionality index (strength of domain boundary) values. Contact maps and mean signal values of these genes were all adjusted to the forward orientation.

Autocorrelation analysis.

To quantify whether there was increased compartmentalization (fragmentation into more compartments) or decreased compartmentalization (grouping or spreading into less compartments) between WT and induction samples, we calculated the autocorrelation of the compartment eigenvectors as described previously⁵⁰.

Detection of TICs.

To detect regions with higher levels of interactions in flavopiridol-treated samples, we first binarized a 2×2 mb² matrix, M (using a 5 kb resolution) by using a $\log_2(\text{treated}/\text{WT}) > 0.5$ threshold. Any element, M_{ij} greater than the threshold is assigned a one, while the rest remained zero. To detect consecutive regions of ones, we applied the following rule for each bin pair to filter out ones that are unlikely to belong to any interacting region:

$$D_{ij} = \begin{cases} 1 & \text{if } \sum_{j=-5}^5 M_{i,j} \geq 5 \text{ and } \sum_{i=-5}^5 M_{i,j} \geq 5 \\ 0 & \text{otherwise} \end{cases}$$

The matrix D , holds the resulting filtered matrix. This criterion is applied iteratively over all bin pairs in the matrix, M , to decrease noise and increase the chance for detection. Next, we used an image detection framework, OpenCV, to find contours and apply a rectangle boundary over D . Last, because we were looking for larger regions instead of single pixel loops, we apply an area threshold of six to obtain these detected regions. We perform this analysis genome-wide while using a sliding-window step size of 1 mb. To further reduce the false-positive rate of our set of detected regions, we manually inspected 5,211 regions and pruned away those that did not show a substantial increase in interactions in the flavopiridol-treated samples. We obtained a final set of 491 detected regions. These regions largely overlapped protein-coding genes. Hence, we assigned the two intervals of the region to its closest respective gene. Finally, using the 365 gene pairs, we construct a graph, and detect all subgraphs with connected components. Each subgraph is a cluster we subsequently term TICs. In total, 228 TICs were discovered genome-wide.

Randomization of sequential order distance gene pairs.

To randomly generate a set of gene pairs that matches the distance between pairs as closely as possible, we randomly chose gene pairs that were d th order away to match the d th order distance between the TIC gene pair. For example, if the gene's pair is found two genes away downstream, we match the randomly selected gene with another gene that is also two genes away in the downstream direction.

Statistics and reproducibility.

The DNA–FISH results on loop validation in Supplementary Figs. 12g, and 13c,d were conducted at least three times with similar results. The immunofluorescence combined with DNA–FISH experiments on TIC validation in Supplementary Figs. 22c,d, 24a,b, 25b,c and 27c were conducted two times with similar results. The results in Supplementary Fig. 26 were generated from two independent experiments. The gel shift results confirming

dendrimer–DNA crosslinking in Supplementary Figs. 3a and 30a,b were performed a number of times with similar results before the final data were presented for a typical single experiment.

Reporting Summary.

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

Data availability

CAP-C, in situ Hi-C, ChIP–seq and PLAC–seq raw sequencing data are available on Gene Expression Omnibus accession: [GSE110061](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE110061). ChIP–seq data can also be found in the following links on UCSD genome browser: CAP-C–mESC: <https://genome.ucsc.edu/s/anton386/CAPC%2DmESC>; CAP-C–CTCF-AID: <https://genome.ucsc.edu/s/anton386/CAPC%2DCTCF%2DAID>; CAP-C–induction: <https://genome.ucsc.edu/s/anton386/CAPC%2DInduction>. Source data are provided with this paper.

Code availability

Code for CAP-C and ChIP–seq analysis is available on GitHub: <http://github.com/ouyang-lab/CAPC>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank A. Andersen (Life Science Editors) for editing the manuscript, P.W. Faber for helping with high-throughput sequencing, and J. Fei and J. Zhang for helping with DNA–FISH experiments. This work was supported by grant nos. NIH F32CA221007 (B.T.H.), NIH RM1HG008935 (C.H.), NIH U54CA193419 (C.H.), the Ludwig Institute for Cancer Research (B.R. and C.H.) and NIH/NIGMS grant no. R35 GM124998 (Z.O.). C.H. is an investigator of the Howard Hughes Medical Institute.

References

1. Sexton T & Cavalli G The role of chromosome domains in shaping the functional genome. *Cell*. 160, 1049–1059 (2015). [PubMed: 25768903]
2. Pombo A & Dillon N Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol* 16, 245–257 (2015). [PubMed: 25757416]
3. Vian L et al. The energetics and physiological impact of cohesin extrusion. *Cell* 173, 1165–1178. e1120 (2018). [PubMed: 29706548]
4. Dekker J, Rippe K, Dekker M & Kleckner N Capturing chromosome conformation. *Science* 295, 1306–1311 (2002). [PubMed: 11847345]
5. Zhao Z et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nat. Genet* 38, 1341–1347 (2006). [PubMed: 17033624]
6. Dostie J et al. Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res.* 16, 1299–1309 (2006). [PubMed: 16954542]

7. Lieberman-Aiden E et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* 326, 289–293 (2009). [PubMed: 19815776]
8. Dixon JR et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380 (2012). [PubMed: 22495300]
9. Phillips-Cremins JE et al. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell* 153, 1281–1295 (2013). [PubMed: 23706625]
10. Downen JM et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. *Cell* 159, 374–387 (2014). [PubMed: 25303531]
11. Rao SS et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680 (2014). [PubMed: 25497547]
12. Ramani V et al. Mapping 3D genome architecture through in situ DNase Hi-C. *Nat. Protoc* 11, 2104–2121 (2016). [PubMed: 27685100]
13. Maiti PK, Çağın T, Wang G & Goddard WA Structure of PAMAM dendrimers::generations 1 through 11. *Macromol.* 37, 6236–6254 (2004).
14. Astruc D, Boisselier E & Ornelas C Dendrimers designed for functions: from physical, photophysical, and supramolecular properties to applications in sensing, catalysis, molecular electronics, photonics, and nanomedicine. *Chem. Rev* 110, 1857–1959 (2010). [PubMed: 20356105]
15. Eichman BF et al. The crystal structures of psoralen cross-linked DNAs: drug-dependent formation of Holliday junctions. *J. Mol. Biol* 308, 15–26 (2001). [PubMed: 11302703]
16. Liang Z et al. BL-Hi-C is an efficient and sensitive approach for capturing structural and regulatory chromatin interactions. *Nat. Commun* 8, 1622 (2017). [PubMed: 29158486]
17. Kolb HC, Finn M & Sharpless KB Click chemistry: diverse chemical function from a few good reactions. *Angew. Chem. Int. Ed* 40, 2004–2021 (2001).
18. Hnisz D, Day DS & Young RA Insulated neighborhoods: structural and functional units of mammalian gene control. *Cell* 167, 1188–1200 (2016). [PubMed: 27863240]
19. Nora EP et al. Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* 169, 930–944 e922 (2017). [PubMed: 28525758]
20. Davuluri RV et al. The functional consequences of alternative promoter use in mammalian genomes. *Trends Genet.* 24, 167–177 (2008). [PubMed: 18329129]
21. Rowley MJ et al. Evolutionarily conserved principles predict 3D chromatin organization. *Mol. Cell* 67, 837–852 e837 (2017). [PubMed: 28826674]
22. Rowley MJ et al. Condensin II counteracts cohesin and RNA polymerase II in the establishment of 3D chromatin organization. *Cell Rep.* 26, 2890–2903.e2893 (2019). [PubMed: 30865881]
23. Mitchell JA & Fraser P Transcription factories are nuclear subcompartments that remain in the absence of transcription. *Genes Dev.* 22, 20–25 (2008). [PubMed: 18172162]
24. Mifsud B et al. Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet* 47, 598 (2015). [PubMed: 25938943]
25. Cubeñas-Potts C et al. Different enhancer classes in *Drosophila* bind distinct architectural proteins and mediate unique chromatin interactions and 3D architecture. *Nucleic Acids Res.* 45, 1714–1730 (2016).
26. Quinodoz SA et al. Higher-order inter-chromosomal hubs shape 3D genome organization in the nucleus. *Cell.* 174, 744–757.e724 (2018). [PubMed: 29887377]
27. Fang R et al. Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res.* 26, 1345 (2016). [PubMed: 27886167]
28. Hsieh T-HS et al. Resolving the 3D landscape of transcription-linked mammalian chromatin folding. *Mol. Cell* 78, 539–553.e538 (2020). [PubMed: 32213323]
29. Du Z et al. Allelic reprogramming of 3D chromatin architecture during early mammalian development. *Nature.* 547, 232–235 (2017). [PubMed: 28703188]
30. Hug CB, Grimaldi AG, Kruse K & Vaquerizas JM Chromatin architecture emerges during zygotic genome activation independent of transcription. *Cell.* 169, 216–228 e219 (2017). [PubMed: 28388407]

31. Bonev B et al. Multiscale 3D genome rewiring during mouse neural development. *Cell*. 171, 557–572 e524 (2017). [PubMed: 29053968]
32. Lu H et al. Phase-separation mechanism for C-terminal hyperphosphorylation of RNA polymerase II. *Nature* 58, 318–323 (2018). 1.
33. Schoenfelder S et al. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nat. Genet* 42, 53 (2010). [PubMed: 20010836]
34. Busslinger GA et al. Cohesin is positioned in mammalian genomes by transcription, CTCF and Wapl. *Nature*. 544, 503 (2017). [PubMed: 28424523]
35. Larson AG et al. Liquid droplet formation by HP1 α suggests a role for phase separation in heterochromatin. *Nature*. 547, 236–240 (2017). [PubMed: 28636604]
36. Sabari BR et al. Coactivator condensation at super-enhancers links phase separation and gene control. *Science*. 361, eaar3958 (2018). [PubMed: 29930091]
37. Guo YE et al. Pol II phosphorylation regulates a switch between transcriptional and splicing condensates. *Nature*. 572, 543–548 (2019). [PubMed: 31391587]
38. Kubo N et al. Preservation of chromatin organization after acute loss of CTCF in mouse embryonic stem cells. Preprint at bioRxiv 10.1101/118737 (2017).
39. Landt SG et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res*. 22, 1813–1831 (2012). [PubMed: 22955991]
40. Law C et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Res*. 5, 10.12688/f1000research.9005.3 (2018).
41. Li H Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at <https://arxiv.org/abs/1303.3997v2> (2013).
42. Zhang Y et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol*. 9, R137 (2008). [PubMed: 18798982]
43. Quinlan AR & Hall IM BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 26, 841–842 (2010). [PubMed: 20110278]
44. Servant N et al. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol*. 16, 259 (2015). [PubMed: 26619908]
45. Durand NC et al. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst*. 3, 95–98 (2016). [PubMed: 27467249]
46. Yang T et al. HiCRep: assessing the reproducibility of Hi-C data using a stratum-adjusted correlation coefficient. *Genome Res*. 27, 1939–1949 (2017). [PubMed: 28855260]
47. Ay F, Bailey TL & Noble WS Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts. *Genome Res*. 24, 999–1011 (2014). [PubMed: 24501021]
48. Rao SSP et al. Cohesin loss eliminates all loop domains. *Cell*. 171, 305–320 e324 (2017). [PubMed: 28985562]
49. DeMare LE et al. The genomic landscape of cohesin-associated chromatin interactions. *Genome Res*. 23, 1224–1234 (2013). [PubMed: 23704192]
50. Schwarzer W et al. Two independent modes of chromatin organization revealed by cohesin removal. *Nature*. 551, 51–56 (2017). [PubMed: 29094699]

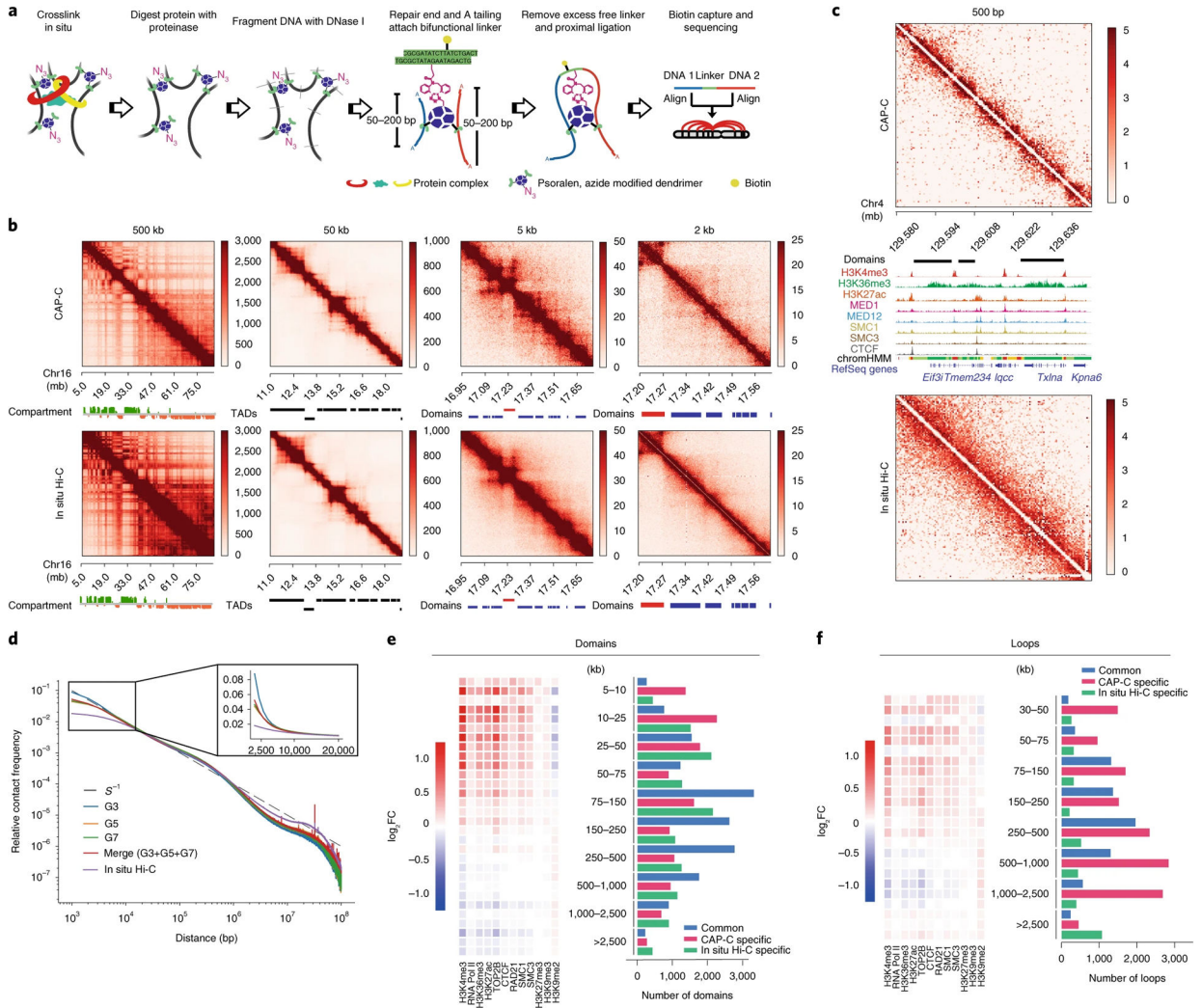


Fig. 1 | CAP-C resolves high-resolution local chromatin structure.

a, Scheme of CAP-C. **b**, Contact maps at various resolutions (upper panel, CAP-C; bottom panel, in situ Hi-C). HiCRep analysis on contact maps obtained by the two methods at various resolutions (500 kb, HiCRep = 0.966; 50 kb, HiCRep = 0.931; 5 kb, HiCRep = 0.977 and 2 kb, HiCRep = 0.857). Chromatin features resolved at each resolution are shown as a track aligned below each contact map. Compartment A is labeled as green box while compartment B is labeled in orange. TADs are shown as black bars. Contact domains are labeled as blue bars. Loop domains are shown as red bars. **c**, CAP-C (top) and in situ Hi-C (bottom) contact matrices at 500 bp resolution for Chr4, 129.58–129.65 mb. Histone modifications and ChIP profiles are aligned in the middle. Three domains are detected (black bar) in CAP-C, enveloping *Eif3i*, *Tmem234* and *Txlna*, respectively. No domains are detected for in situ Hi-C at this resolution. **d**, Relative contact frequency versus genomic distance curves of CAP-C and in situ Hi-C. The enlarged inset window shows approximately six- to eightfold contact enrichment for CAP-C over in situ Hi-C. G3, G5, G7 represent CAP-C performed with psoralen-functionalized PAMAM dendrimer generation 3, 5 and 7 (with diameters of 3.6, 5.4 and 8.1 nm), respectively. CAP-C merge represents merging of

data generated from G3, G5 and G7. **e**, CAP-C identifies more small-sized domains than in situ Hi-C and the identified domain boundaries correlate well with active histone marks, CTCF and cohesin binding sites. Domains were called by Arrowhead algorithm¹¹ (see Methods) and classified into ten groups based on their sizes. **f**, CAP-C identifies more loops, and the identified loops correlate well with active histone marks, CTCF and cohesin, genome-wide. Loops were called by HiCCUPS¹¹ and classified into ten groups based on their genomic spans.

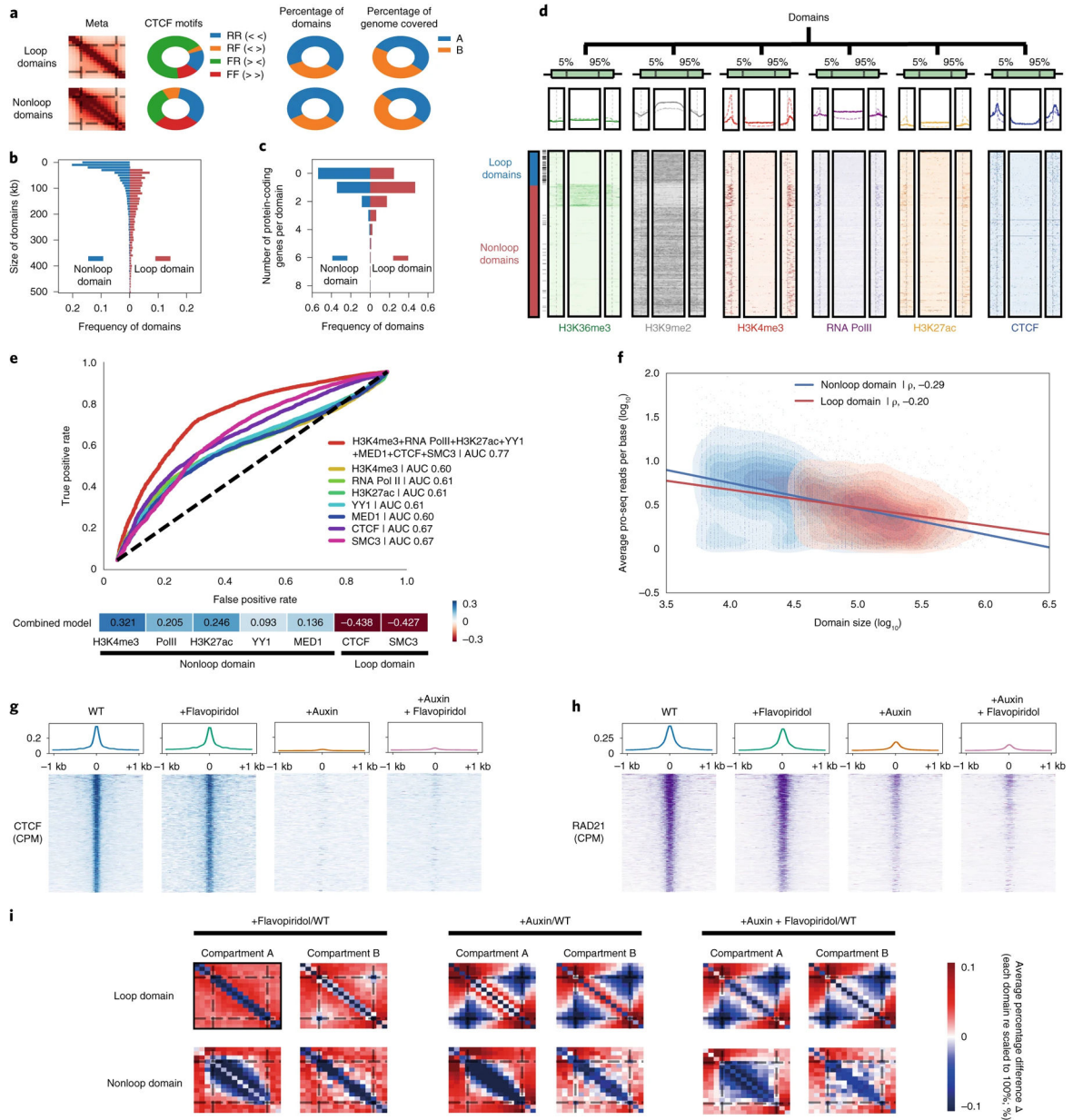


Fig. 2 | CAP-C identifies loop and nonloop domains as two distinct types of chromatin domains. **a**, Meta-contact map showing intradomain interactions, pie chart showing the frequency of CTCF motif orientations and pie charts showing the fraction of domains and percentage of genome covered by compartments A and B within the loop and nonloop domains. **b**, Plot of the distribution of domain sizes for nonloop and loop domains. **c**, Plot of the distribution of the number of protein-coding genes found in nonloop versus loop domains. **d**, Domains were segregated into left boundaries, domain bodies and right boundaries. The distribution of indicated ChIP-seq signals around left boundaries, domain bodies and right boundaries of nonloop domains (shown as dotted lines) and loop domains (shown as solid lines) are aligned below. Heatmap displaying the reads density distribution of indicated ChIP-seq signals in nonloop and loop domains. Each row represents a domain. **e**, A logistic regression

model was trained using either individual features or all seven features (combined model) to discriminate between loop domains (negative set) and nonloop domains (positive set). The receiver operating characteristic curve is plotted for each regression model. **f**, A linear regression was performed between domain sizes and transcription rates. Spearman's correlation coefficient = -0.29 for nonloop domains; -0.20 for loop domains. **g,h**, Metagene plots of ChIP-seq data for CTCF (**g**) and RAD21 (**h**) were centered around a ± 1 kb region for mESCs with different perturbations: WT, inhibition of transcription elongation (+Flavopiridol), acute depletion of CTCF (+Auxin), combined acute depletion of CTCF and inhibition of transcription elongation (+Flavopiridol +Auxin). **i**, Meta-domain analyses showing differential domain contacts. Domains were classified into two main types: loop domain (LD) and nonloop domain (nLD). Each type was further divided into two groups based on their presence in A or B compartments.

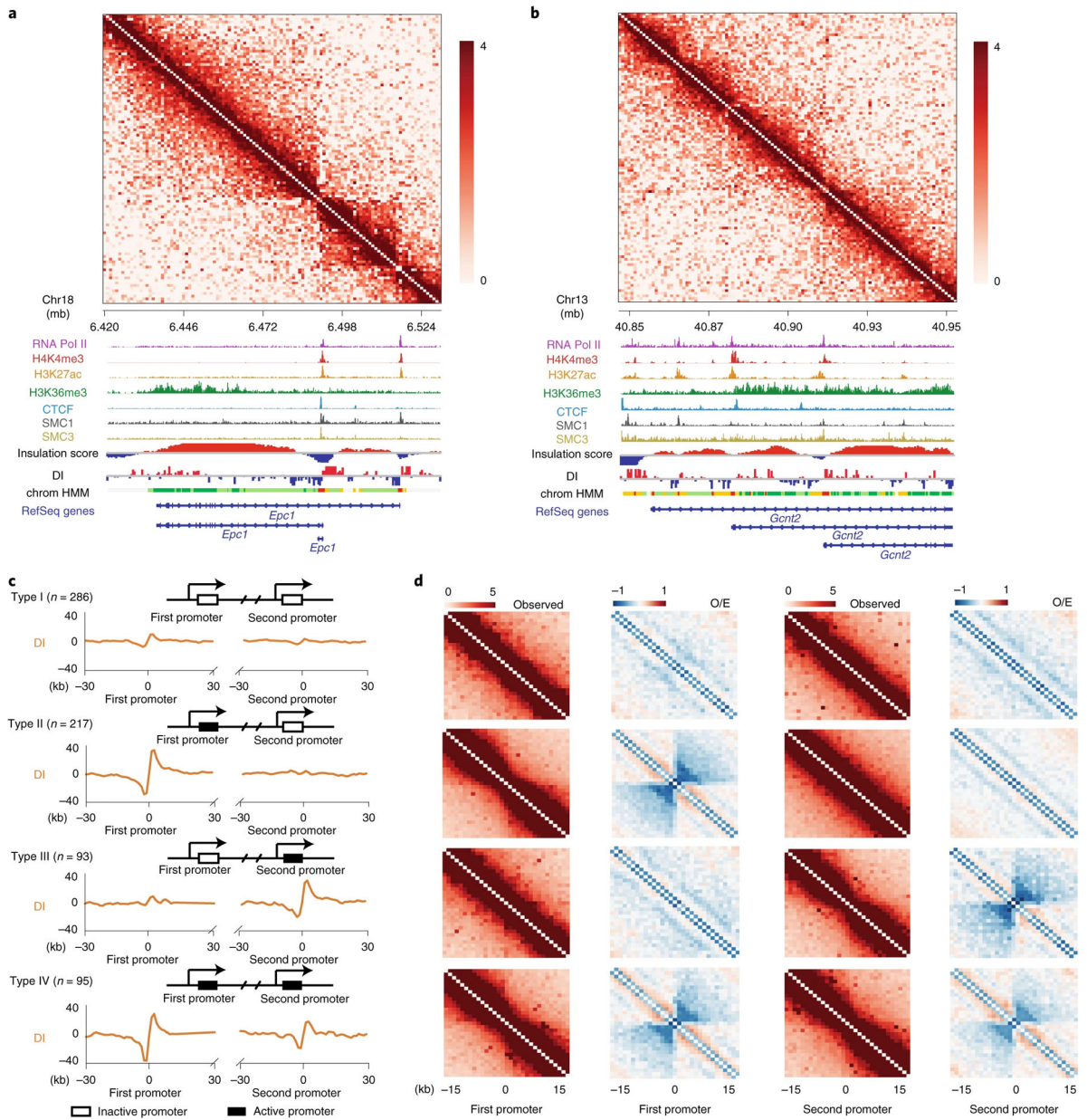


Fig. 3 | Active promoters are involved in chromatin boundary formation.

a,b, Genes were segregated into different contact domains with active promoters on the boundary. Two examples of CAP-C contact matrices (at 1 kb resolution) are shown corresponding to Chr18, 6.42–6.53 mb (**a**) and Chr13, 40.85–40.96 mb (**b**). The black line depicts the domains called by Arrowhead. Directionality index (DI), histone modifications and ChIP–seq peaks are shown below each matrix. *Epc1* and *Gcnt2* are insulated by their active promoters, respectively. **c,** Genes with alternative promoters were selected and classified into four different types based on the transcription state of their first and second promoters. Numbers of each type are shown inside the corresponding brackets. Directionality index values around each promoter are shown below. **d,** Interaction counts and

the \log_2 ratio of observed interactions divided by expected interactions for a given genomic distance are shown side by side for each type.

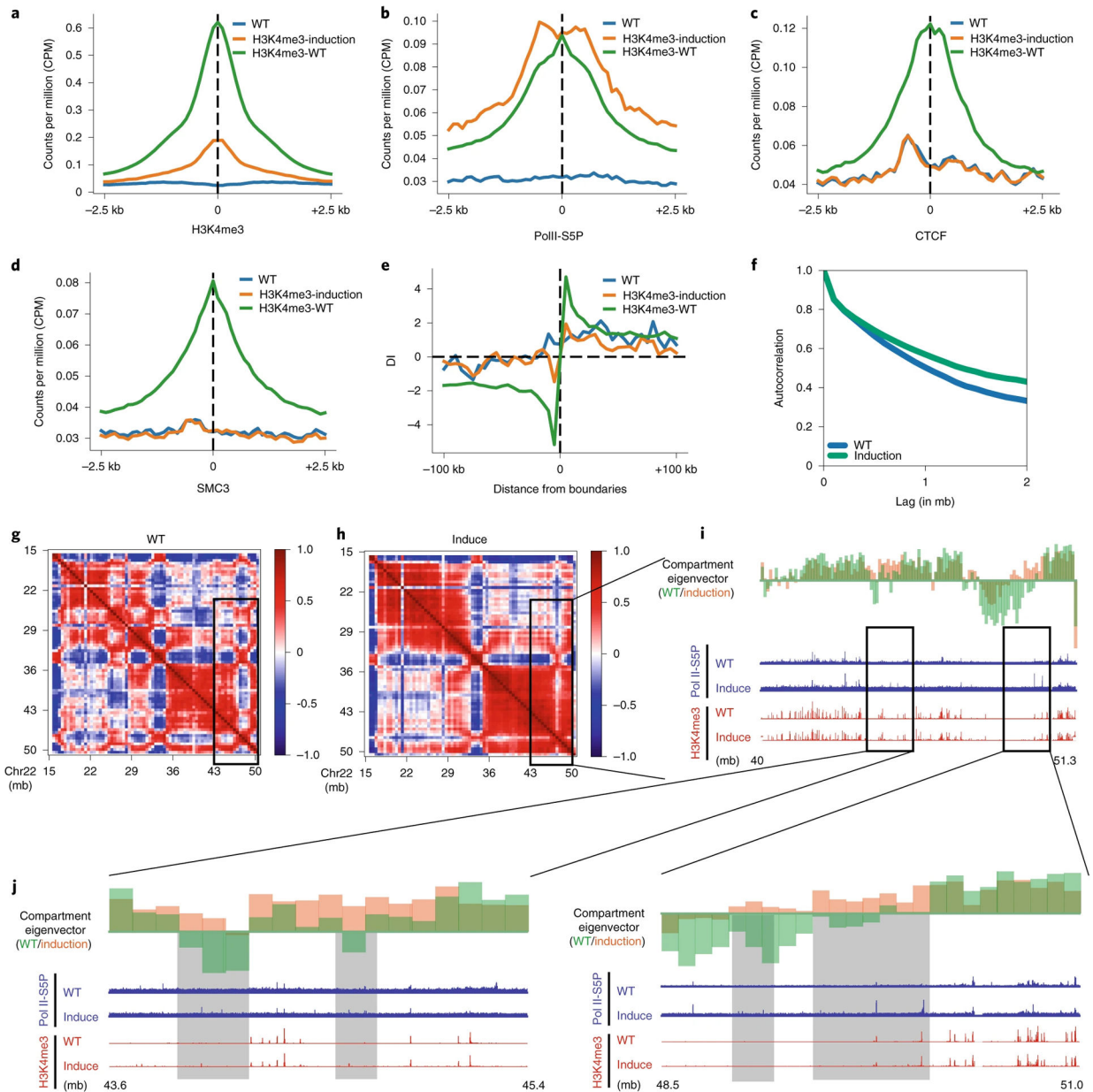


Fig. 4 | Induction of transcription on nonannotated TSS by DNMTi results in the formation of weak chromatin boundaries and compartment changes.

a–d, DNMTi treatment caused formation of weak chromatin boundaries on newly activated TSS loci in HCT116 cells. Induced loci (H3K4me3-induce) were identified based on \log_2FC of H3K4me3 greater than 1.5-fold than WT. The originally active loci, marked with H3K4me3 ChIP-seq peaks in WT (H3K4me3-WT), were selected as positive controls. Metagene plots of ChIP-seq data for Pol II-S5P (**a**); H3K4me3 (**b**); CTCF (**c**) and SMC3 (**d**) were centered ± 2.5 kb around these loci. **e**, Directionality index (DI) values were calculated ± 100 kb around these loci. **f**, Autocorrelation between compartment changes and distance was plotted. **g,h**, DNMTi caused compartment changes in HCT116 cells. The Pearson correlation was plotted for chr22, 16.0–51.3 mb in WT (**g**) or induced HCT116 cells (**h**).

Regions with compartment changes are highlighted with a black box. **i**, Compartment changes caused from the B to A transition after DNMTi treatment. Compartment eigenvectors were computed for both WT (marked in green) and induction (marked in orange). Regions with positive eigenvalues represent compartment A, and negative eigenvalues represent compartment B. Pol II-S5P and H3K4me3 ChIP-seq peaks are aligned at bottom for the WT and induction samples. Compartment transition regions are highlighted with a black box. **j**, Enlarged views of two black boxes highlighting regions with compartment transitions usually coupled with Pol II-S5P ChIP-seq peaks.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

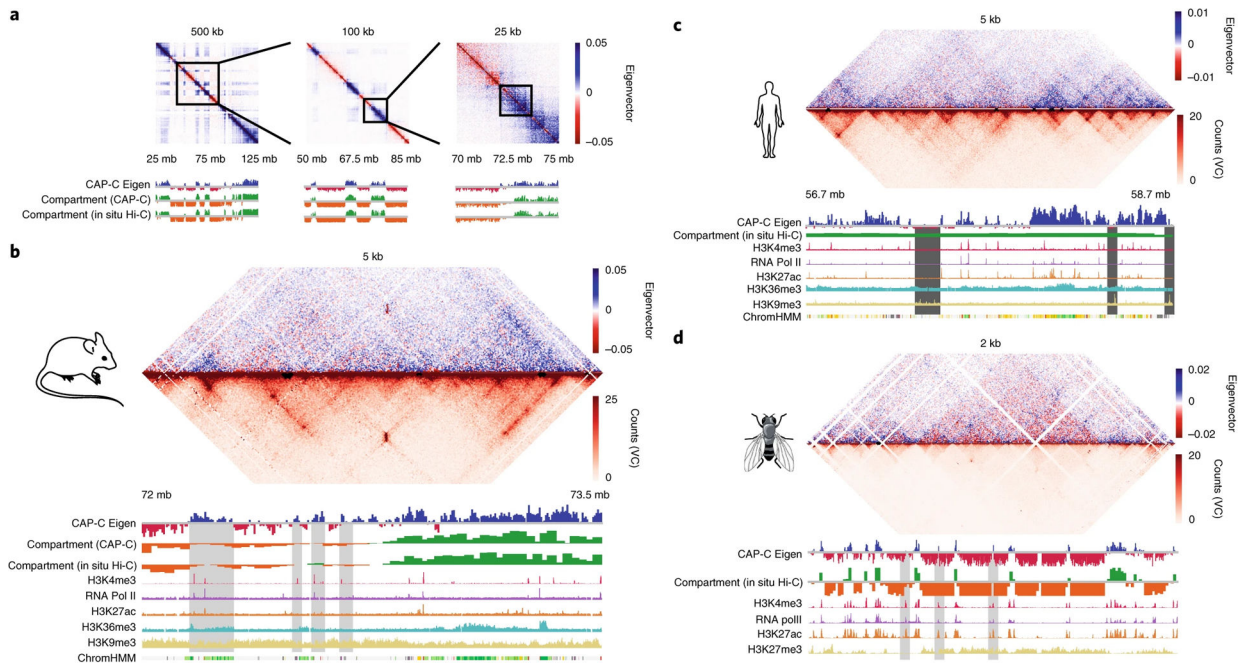


Fig. 5 | CAP-C identifies conserved small-scale chromatin compartmentalization shared among species.

a, SVD was performed on an m rows by n columns data matrix (where m is the number of dendrimer experiments, and n is the number of loci bins across a specified region and resolution) of relative contact frequencies. The eigenvector with the highest eigenvalue yields a ‘dendrimer map’ that shows bifurcated separation (see Methods). A separate SVD analysis on the row sums of a contact matrix yields CAP-C eigenvector (see Methods). Compartment eigenvectors derived from CAP-C and in situ Hi-C are aligned below. Projection onto the selected eigenvector showing G5 and G7 dendrimer experiments enriched for open configurations, while the G3 dendrimer experiment enriched for closed configurations. **b–d**, Close-up of the ‘dendrimer map’ showing fine level of compartment detail for mESC at chr5, 72.0–73.5 mb (**b**); HepG2 at chr3, 56.7–58.7 mb (**c**) and *Drosophila* S2 at chr2L, 21.55–22.25 mb (**d**). CAP-C eigenvector, compartment eigenvector and indicated ChIP-seq profiles are aligned below each dendrimer map. The species icons were produced using Servier Medical Art by Servier (<https://smart.servier.com>) and modified under a Creative Commons Attribution 3.0 Unported License (<https://creativecommons.org/licenses/by/3.0/>).

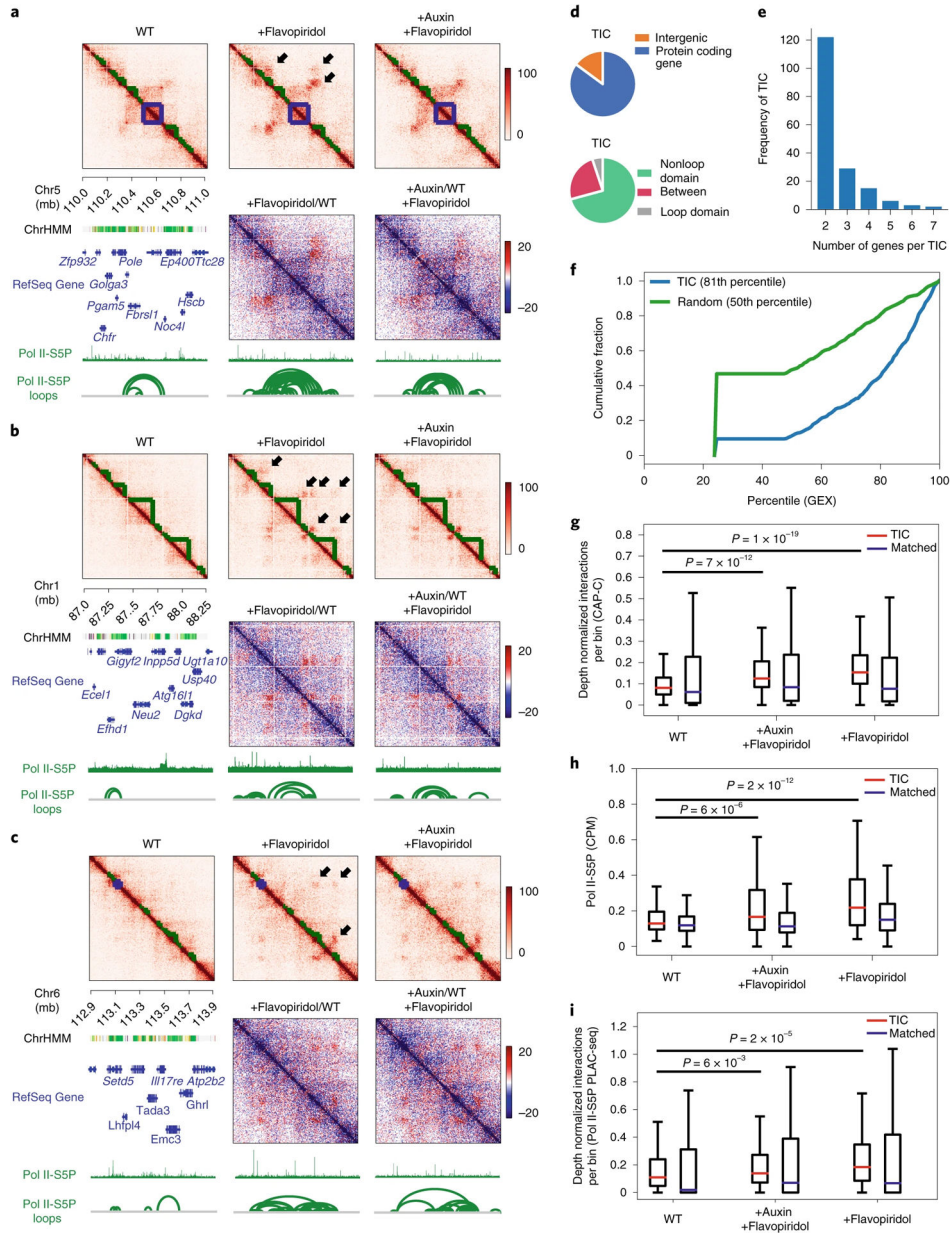


Fig. 6 | TICs as a type of nuclear subcompartment.

a–c, TICs displayed enhanced interactions after transcription elongation inhibition; each TIC brings two or more active genes together in close proximity. Three of the largest TICs are represented along with their characteristic features. Connected components (genes) are listed for each TIC: Cluster 1 (*Pole*, *Ep400*, *Golga3*, *Pgam5*, *Noc4l*, *Pxmp2*, *P2rx2*, *Mir7026*) (**a**); Cluster 2 (*Gigyf2*, *Sag*, *Kcnj13*, *Inpp5f*, *Atg16l1*, *Usp40*, *3110079015Rik*, *Dgkd*) (**b**) and Cluster 3 (*Setd5*, *Fancd2*, *Tatdn2*, *Emc3*, *Thumpd3*, *Fancd2os*, *Sec13*) (**c**). Each example is visually represented by a depth-normalized contact map annotated with detected regions (black arrowhead) showing increased interactions between flavopiridol-treated and WT, nonloop domains (marked as green) and loop domains (marked as blue). Differences between flavopiridol-treated and WT heatmaps highlight increased interactions (in red)

within detected regions. Pol II-S5P ChIP-seq peaks and Pol II-S5P mediated interactions (loops) are aligned below. **d**, Upper panel, pie chart showing the proportion of TICs that span protein-coding regions. Bottom panel, pie chart showing the proportion of TICs that span nonloop domains. **e**, Distribution of the number of gene members per TIC (median, 2 and maximum 7). **f**, Cumulative density function of genes ranked by reads per million of bulk RNA-seq (TIC median, 81th percentile; random median, 50th percentile). **g**, Normalized CAP-C interactions per bin at TIC under each treatment (WT, +Auxin/+Flavopiridol, +Flavopiridol only). **h**, Pol II-S5P ChIP-seq levels at TICs under each treatment. **i**, Pol II-S5P-mediated loops (PLAC-seq) under each treatment. Boxes in **g–i** indicate the median and interquartile ranges, with whiskers indicating 1.5× the interquartile range. *P* values were calculated using two-sided Kolmogorov–Smirnov test (no adjustments made for multiple testing).