Three-dimensional organization of chromatin associated RNAs and their role in chromatin architecture in human cells

Authors: Riccardo Calandrelli [1,*], Xingzhao Wen [2,*], John L Charles Richard [1,#], Zhifei Luo [1,#], Tri C. Nguyen [1], Chien-Ju Chen [2], Zhijie Qi [1], Shuanghong Xue [1], Weizhong Chen [1], Zhangming Yan [1], Weixin Wu [1], Kathia Zaleta-Rivera [1], Rong Hu [3,4], Miao Yu [4], Yuchuan Wang [5], Wenbo Li [6], Jian Ma [5], Bing Ren [3,4], Sheng Zhong [1]

[1] Department of Bioengineering, [2] Bioinformatics and Systems Biology Program, [3] Department of Cellular and Molecular Medicine, Center for Epigenomics, University of California San Diego, La Jolla, CA, USA

[4] Ludwig Institute for Cancer Research, La Jolla, CA, USA

[5] Computational Biology Department, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA

[6] Department of Biochemistry and Molecular Biology, McGovern Medical School, University of Texas Health Science Center, Houston, TX, USA

[*, #] Equal contribution

Correspondence: S.Z., email: szhong@ucsd.edu

Abstract

Chromatin-associated RNA (caRNA) is a vital component of the interphase nucleus; yet its distribution and role in the three-dimensional (3D) genome organization and function remain poorly understood. Here, we map caRNA's spatial distribution on the 3D genome in human embryonic stem cells, fibroblasts, and myelogenous leukemia cells. These maps reveal several characteristics of caRNA distribution. caRNA is preferentially associated with its transcription site and up to several megabases of flanking genomic sequence. Furthermore, topologically associating domain (TAD) boundaries insulate RNA-DNA contacts. Abolishment or creation of a TAD boundary by genomic editing changed the boundary's ability to insulate RNA-DNA contacts, demonstrating an impact of the 3D genome structure on the distribution of caRNA. Conversely, depletion of caRNA weakens TAD boundaries. We characterize caRNA based on their target genomic sequence into loop-anchor-associated RNA and those associated with any genomic sequence between loop anchors (between-anchor caRNA). Between-anchor caRNA suppresses chromatin loops, especially those with convergent CTCF binding sites in their loop anchors. These data suggest the indispensable roles of caRNA in the 3D genome organization.

Introduction

After initial debates[1], chromatin-associated RNA (caRNA) has been recognized as a "widespread component of interphase chromosomes" rather than artificial degradation products[2-6]. Growing evidence confirms that caRNA regulates gene transcription[7-16] and post-transcriptional RNA processing and localization[17,18]. These regulatory roles often depend on caRNA's spatial localization within the nucleus[12,19-22].

Select protein species, RNA species, and DNA sequences can co-cluster in 3D to form nuclear bodies and spatial compartments[21,22]. For example, nuclear speckle-associated proteins and RNAs including snRNAs and the MALAT1 lincRNA co-cluster with pre-mRNAs and their genomic sequences[21,22]. Thus, the spatial distribution of caRNA is expected to correlate with the 3D genome at a coarse scale (a citation to the 4DN Consortium's flagship paper (under review)). However, the spatial distribution of caRNA has not been systematically characterized in the context of multiscale organizational features of the 3D genome, including A/B compartments[23], topologically associating domains (TADs)[24,25], and chromatin loops[26] (a citation to the 4DN Consortium's flagship paper (under review)). It remains unclear whether and how the multiscale organization of the 3D genome can impact caRNA, and vice versa.

Here, we generate high-resolution genome-wide caRNA-DNA contact maps[7,27-30] in human cells using *in situ* Mapping of RNA-Genome Interaction (iMARGI)[27,28]. iMARGI captures RNA-genome interactions by proximity ligation followed by reverse transcription of RNA and high-throughput paired-end sequencing[27]. iMARGI can differentiate the sequencing reads originating from RNA (iMARGI RNA-end reads) or genomic DNA (iMARGI DNA-end reads). To describe the 3D genome organization, we use *in situ* Hi-C (Hi-C)[26,31] to detect long-range chromatin interactions genome-wide and PLAC-seq to reveal long-range chromatin interactions at selected genomic regions[32]. Combining these data, we investigate the spatial organization of caRNA and their relationship to multiscale 3D genome features. Further combining several strategies to modulate caRNA, we report caRNA's roles in stabilizing TAD boundaries and modulating chromatin looping.

Results

Nascent transcripts are associated with their transcription site and up to megabases of flanking sequence

We generated iMARGI data from human embryonic stem (H1), foreskin fibroblast (HFF), and chronic myelogenous leukemia (K562) cells in duplicates (Table S1). We used an unsupervised approach to discover any prominent feature in RNA-DNA contacts genome-wide. We represented iMARGI data as a contact matrix, where the rows represent the RNA ends of iMARGI read pairs, and the columns represent the corresponding DNA ends[28] (Figure 1a). A notable difference to Hi-C's symmetric contact matrix is that iMARGI's contact matrix is asymmetric. This is because RNA-DNA contacts are not necessarily reciprocal.

Rectangular blocks of high-value entries emerged as a recurring pattern from iMARGI's contact matrix (Figure 1a). We identified the rectangular blocks using HOMER to call peaks on the rows of the contact matrix (row peaks), and in each row peak using HOMER to call one strongest peak in the columns (column peak). A pair of row peak and column peak defines a rectangular block. We identified 3,217, 2,019, and 2,468 rectangular blocks from H1, HFF, and K562 iMARGI data (Figure 1b). All the identified rectangular blocks overlap with the diagonal entries of iMARGI's contact matrix, suggesting that they represent localized caRNA-genome interactions where a caRNA's target genomic sequences are near the transcription site of this caRNA. Each rectangular block corresponds to a unique chromatin domain, characterized by extensive genomic association of the caRNAs transcribed from within this domain. Hereafter we term such domains "caRNA domains". The size of a caRNA domain, represented by the width of a rectangular block, can reach tens of megabases (Figure 1c). At this point, we identified caRNA domains as a main feature of genome-wide caRNA-DNA associations. This feature is consistent with the notion that nascent transcripts are the major constituent of caRNA[33,34].

To clarify if iMARGI-captured caRNA is enriched with nascent transcripts, we compared iMARGI's RNA-end reads with several methods specialized in sequencing nascent

transcripts, including GRO-seq[35], PRO-seq[36], mNET-seq[37], and TT-seq[38]. iMARGI's cis-caRNA reads (those attached to any genomic regions on the same chromosome of their transcription loci) exhibited a comparable performance to TT-seq and outperformed GRO-seq, PRO-seq, mNET-seq in all three evaluations (Extended Text 1: Nascent transcription). Briefly, iMARGI exhibited high intron coverage (Figure 1d), slow decay of read coverage after TSS (arrowhead, Supplementary Figure S1a), and high correlation with TT-seq-derived nascent transcript levels (Supplementary Figure S1b), suggesting iMARGI's ability to measure the nascent transcripts is comparable to those specialized in nascent transcript sequencing. Notably, iMARGI data revealed that the further the genomic target sequence is from the transcription site of this caRNA, the fewer intronic sequences are detected on this caRNA (Figure 1d), confirming that not all caRNA are nascent transcripts (Extended Text 1: Nascent transcription). These data are consistent with microscopically observed clustering of nascent transcripts near their transcription locus[39-41] (Extended Text 1: Nascent transcription). Taken together, iMARGI-captured caRNA is enriched with nascent transcripts, and the caRNA located near their transcription sites tend to be un-spliced.

TAD boundaries insulate RNA-DNA contacts

The 3D genome is organized on different scales, including compartments, TADs, and chromatin loops[42]. We generated Hi-C data in H1, HFF, and K562 cells in duplicates and compared them with our iMARGI data (Table S1). We calculated the relative amount of caRNA attached to every genomic region, hereafter called RNA attachment level (RAL, defined as the number of iMARGI read pairs with the DNA ends mapped to this genomic segment[27] , purple curve, Figure 1e). Hi-C contact matrix's first eigenvector (PC1)[43] exhibits a genome-wide correlation with RAL of all caRNA (p-value < 2e-16, one way ANOVA), and with the RAL of Alu-containing caRNA (Alu-caRNA) or LINE1-containing caRNA (L1-caRNA) (Figure 1e, Supplementary Figure S2) (Extended text 2: caRNA in genome compartments). These data reveal a correlation between 3D genome compartmentalization and caRNA distribution.

TADs, where DNA sequences interact with each other more frequently than with the sequences outside, are important 3D genome features that are strongly correlated with transcriptional regulation[24,25]. We separately analyzed the caRNA transcribed from within a TAD or the other regions of the same chromosome outside of this TAD. The attachment level of any caRNA transcribed from within a TAD sharply decreases at the two boundaries of this TAD (p-value = 6.5e-16, Wilcoxon rank-sum test) (Figure 1f). Conversely, the attachment level of any caRNA transcribed from outside of a TAD exhibits an opposite change at the TAD boundaries (p-value = 2.6e-12, Wilcoxon rank-sum test) (Supplementary Figure S3b). These changes at TAD boundaries cannot be completely explained by the 1-dimensional genomic distance to the caRNA's transcription site. They suggest a possibility that a TAD boundary can insulate RNA-DNA contacts from the two sides of this boundary (cross-over RNA-DNA contacts).

We asked if altering the genomic sequence within a TAD boundary can affect the cross-over RNA-DNA contacts. First, we leveraged our previous finding that a CRISPR-mediated deletion of a HERV-H element (Chr13:55,578,227-55,584,087) (KO) within a TAD boundary from H9 human ES cells (WT) abolishes this TAD boundary[44]. We carried out iMARGI experiments on the KO and WT cells. We counted the numbers of cross-over and non-cross-over iMARGI read pairs in WT and KO. Compared to WT, KO exhibited an increased proportion in the cross-over read pairs (OR = 1.3, p-value = 0.013, Chi-square test) (Figure 1g). Thus, deleting a fraction of a TAD boundary reduced its insulation to cross-over RNA-DNA contacts.

Second, we previously created an insertion cell line (KI) using piggyBac transposon-mediated genomic insertion of this HERV-H sequence[44]. In this KI cell line, 3 insertion sites exhibited large increases in insulation, as measured by the changes in directionality index (delta_DI > 30)[44]. We carried out iMARGI in KI and WT cells. Compared to WT, all three insertion sites exhibited reduced proportions of cross-over read pairs in KI (OR = 0.74, 0.89, and 0.66), of which two were statistically significant (p-value = 3.7e-5, 3.1e-7, Chi-square test, corresponding to OR = 0.74, 0.66). In particular, the insertion site with the largest increase in insulation (delta_DI = 66.3) is the only insertion site that is directly detected as a *de novo* TAD boundary, i.e., a boundary called in the KI Hi-C but not called

in the WT Hi-C (de_novo-boundary-site). This de_novo-boundary-site site is one of the two insertion sites with a significant loss of cross-over RNA-DNA contacts (Figure 1h). Thus, the de novo creation of a TAD boundary led to increased insulation of cross-over RNA-DNA contacts.

In contrast, we previously also identified 4 other insertion sites with smaller increases in insulation (delta_DI < 30), which did not result in *de novo* TAD boundaries. None of these 4 insertion sites led to detectable changes in the proportion of RNA-DNA contacts (the smallest p-value > 0.63, Chi-square test). Thus, inserting the same DNA sequence without sufficient subsequent changes in 3D genome structure cannot lead to a change in cross-over RNA-DNA contacts. Taken together, these data show an impact of the 3D genome structure to the distribution of caRNA.

caRNA on CTCF binding sites exhibit resistance to ribonuclease activity

The CCCTC-binding factor (CTCF), a key regulator of chromosomal structure, is an RNA binding protein[45,46]. We asked whether CTCF binding regions are enriched with caRNA. To this end, we obtained a total of 76,230 peaks from CTCF ChIP-sequencing data in H1 ES cells[47] and plotted caRNA levels (RAL) in every peak and its 10 kb flanking regions. RAL also peaked at the CTCF ChIP-seq peaks and sharply decreased in the flanking regions (Figure 2a), suggesting that CTCF-bound genomic regions are hotspots of caRNA attachment.

Hereafter, we call the CTCF ChIP-seq peak colocalized caRNA "CTCF co-bound caRNA". We asked whether CTCF's DNA recognition sequence can affect CTCF co-bound caRNA. To this end, we leveraged the allelic sequence difference in the phased genome of H1 human ES cells (https://data.4dnucleome.org/joint-analysis#phased-genomes). Scanning the CTCF ChIP-seq peaks with CTCF DNA recognition motif (JASPAR CTCF_MA0139.1.pfm), we identified 59,438 ChIP-seq peaks that contain CTCF recognition sites (CRS). Among them, 1,094 peaks harbor between-allele single-nucleotide variation (SNV) inside the CRS (a.k.a. CRS_SNV). On the other hand, 378 CRS-containing peaks exhibited between-allele differences in ChIP-seq signals (FDR < 0.05) and were identified as allele-specific peaks (asPeaks). Intersecting CRS_SNVs with

the asPeaks yielded 206 CRS_SNV asPeaks, which are CTCF peaks harboring CRS_SNV and exhibiting between-allele differences in ChIP-seq signals. We quantified the allelic difference in the CRS motif score by the ratio of the matching scores to the CTCF DNA recognition motif (Motif score ratio). A motif score ratio greater than 1 means that the reference allele has a stronger match to the CTCF motif (JASPAR CTCF_MA0139.1.pfm) than the alternative allele. We also quantified the allele-difference of CTCF co-bound caRNA by the ratio of the RALs of the two alleles (RAL allele ratio) and the allele-difference in the CTCF binding by the ratio of the ChIP-seq signals of the two alleles (CTCF-binding allele ratio). Motif score ratio exhibits a positive correlation to RAL allele ratio (p-value = 7e-6, Spearman correlation) (Figure 2b), suggesting a reduction of caRNA attachment level on the allele with a weaker match to CTCF's recognition motif. As a sanity check, the motif score ratio is predictive of the CTCF-binding allele ratio (p-value < 2.2e-16, paired t test), confirming that CTCF's DNA recognition sequence drives the allele-difference of CTCF binding. For example, a CTCF asPeak on the intron of TBC1D9B gene exhibited 9.85 folds of ChIP-seq reads on the reference allele as compared to the alternative allele, and consistently, the caRNA attachment level is 2.38 folds on the reference allele as compared to the alternative allele (Figure 2c, d). Taken together, not only caRNA is enriched on CTCF ChIP-seq peaks, but the allelic sequence difference in CTCF recognition sites can further explain the allelic difference in RAL on these CTCF peaks.

CTCF has an RNA binding domain and in the presence of RNA, CTCF can self-associate[45,46]. The self-associated CTCFs may create a steric hindrance to ribonucleases (RNase)[48] and RNA decay machinery[49]. With this idea, we treated H1 cells with RNase A[50,51] and generated iMARGI data in the RNase-treated cells (Table S1, Supplementary Figure S5). We asked if RNase removes CTCF co-bound caRNA and other caRNA to the same extent. RNase reduced RAL both inside and outside CTCF peaks, suggesting that RNase can remove caRNA regardless of whether it is CTCF co-bound caRNA or not (Average RAL track, Figure 2a, Control). However, the degree of reduction is stronger outside than inside of the CTCF peaks (p-value < 2.2e-16, paired t test) (Figure 2a, RNase). As a result, the contrast of RAL inside CTCF peaks to the flanking sequences

became more pronounced in RNase than in control (Figure 2e, f). These data suggest that CTCF co-bound caRNA is protected from ribonuclease activity.

caRNA stabilizes TAD boundaries

RNase treatment resulted in an approximately 20% reduction of the number of TAD boundaries (p-value = 3.217e-8, paired t test) (Supplementary Figure S5a), indicating a negative impact of RNA on TAD boundaries. Furthermore, we compared the change of caRNA on each TAD boundary to the change in this boundary's strength. We quantified the degree of caRNA removal on a boundary by the difference in this boundary's RAL between RNase and control (Delta RAL). The greater extent of caRNA removal on a boundary is represented by a smaller (more negative) Delta RAL. We quantified the change of a boundary's strength by the difference in insulation scores[52] between RNase and control (Delta insulation). By definition, a stronger boundary has a smaller insulation score[52]. The TAD boundaries with large reductions of caRNA exhibited increased Delta insulation, i.e., weakened TAD strengths (p-value = 2e-16, ANOVA test) (Supplementary Figure S5b). Furthermore, we calculated the proportions of boundaries that disappeared (Diminished boundaries) or shared (Persistent boundaries) in RNase. A higher degree of caRNA reduction (Delta RAL is more negative) resulted in a larger fraction of Diminished boundaries (p-value = 2.2e-16, Chi-square test) (Figure 2g). Specifically, all the top 30 boundaries with the greatest reduction of RAL were diminished, and only two out of the top 100 boundaries with the greatest reduction of RAL were persistent (odds ratio = 5.58, Chi-square p-value = 0.011). These data suggest that the removal of TAD boundary caRNA weakens the TAD boundary. In line with these data, targeted suppression of a transcription locus located within a TAD boundary led to a loss of insulation[44].

Depletion of CTCF resulted in the loss of a subset of TAD boundaries, although it remains unclear why some boundaries are more sensitive to CTCF depletion than others[53]. Considering caRNA's role in TAD boundary's stability, we predicted that the TAD boundaries with little caRNA tend to be lost when CTCF is depleted. To test this prediction, we compared iMARGI data with a published CTCF ChIP-seq dataset of auxin-induced degron (AID)-mediated CTCF degradation (CTCF AID) and no-auxin control[54].

We identified 3,323 TAD boundaries in control, among which 1,340 persisted in CTCF depleted cells (persistent boundaries) while the rest were diminished in CTCF AID (diminished boundaries). The diminished boundaries exhibited lower RAL than the persistent boundaries (p-value < 2.2e-16, t test) (Figure 2h), confirming our prediction. In summary, a TAD boundary's caRNA level is predictive of its stability when CTCF's protein level drastically decreases.

Removal of caRNA increases chromatin loops' number and strengths

We treated H1 cells with ammonium acetate (NH4OAc) to disrupt RNA's electrostatic interactions (the interactions due to electric charges)[55-57], Flavopiridol (FL) to suppress transcription elongation[58-60], and RNase A to reduce caRNA[50,51] based on established protocols (NH4OAc[56], FL[58], RNase A[50]). NH4OAc disrupts RNA's electrostatic interactions in living cells by providing monovalent cations without perturbing intracellular pH[56]. To check the expected effects of the three treatments, we immunostained nuclear speckle-associated proteins SON[61] and SC35[62] in control and each treatment. NH4OAc reduced the numbers of SON and SC35's foci (p-value = 0.001 for SON, 0.009 for SC35, Wilcoxon test) (Figure 3a, b, Supplementary Figure S6e, h), consistent with the role of RNA's electrostatic interactions in maintaining nuclear speckles[63]. Conversely, FL made SON and SC35 foci larger and more distinct[64] (Figure 3c, Supplementary Figure S6f, i). RNase A increased the numbers of SON and SC35's foci (p-value = 0.034 for SON, 0.010 for SC35, Wilcoxon test) (Figure 3d, Supplementary Figure S6g, j), consistent with the observations that "low RNA/protein ratios promote phase separation into liquid droplets"[65] and condensate formation[66].

We generated Hi-C and iMARGI data after each treatment in duplicates (Table S1) and analyzed these data together with those of the unperturbed H1 cells (control). As expected, FL exhibited the largest reduction of the heights of the rectangular blocks in iMARGI's contact matrix (p-value < 3e-104, Wilcoxon rank-sum test) (Supplementary Figure S4e), consistent with FL's inhibitory effect on transcription elongation[58]. RNase exhibited the largest reduction of caRNA domains' number (3,217 in control and 357 in RNase, p-value < 3e-9, paired t test) (Supplementary Figure S4d) and sizes (widths of

the rectangular blocks) (p-value < 5e-210, Wilcoxon rank-sum test) (Supplementary Figure S4f).

We sequenced the Hi-C libraries of each condition (control, NH4OAc, FL, RNase) to similar sequencing depths (~600-650 million read pairs per condition, Table S1). Compared to control, NH4OAc had little impact on the proportions of intra- and inter-chromosomal interactions (Figure 3e) or intra-chromosomal contact frequency as a function of the genomic distance between the contacting loci[23] (Supplementary Figure S7). FL increased the proportion of intra-chromosomal interactions (Figure 3e) and the proportion of sub-megabase-distance intra-chromosomal contacts (p-value < 2e-16, Wilcoxon test) (Supplementary Figure S7b), consistent with the idea that transcription weakens chromatin contacts within gene bodies[50]. On the other hand, RNase decreased the proportion of intra-chromosomal interactions (Figure 3e) and the proportion of sub-megabase-distance intra-chromosomal contacts (p-value < 2e-16, Wilcoxon test) (Supplementary Figure S7b). These results indicate both the transcription process and this process' RNA products can impact chromatin contacts, but their impacts are not identical.

We called chromatin loops from our Hi-C data in each of the four conditions that have comparable sequencing depths (Table S1) using HiCCUPS[67]. The loop numbers were similar in control (2,473 loops) and NH4OAc (2,437 loops) (p-value = 0.55, paired t test) and were increased in FL (5,039 loops) (p-value < 1.1e-8, paired t test) and RNase (4,963 loops) (p-value < 2.3e-9, paired t test) (Figure 3f). These loop number differences cannot be attributed to different sequencing depths or batch effects because the samples were prepared in the same batch and sequenced to comparable depths (600 – 650 million read pairs per condition, Table S1b). Most of the emerged loops in FL colocalized with the emerged loops in RNase (first column, Figure 3f). For example, a loop linking ATF7 and KRT18 genes that was absent in control and NH4OAc emerged in both FL and RNase (arrows, Figure 3g, Supplementary Figure S8a).

The overall loop strength was similar in control and NH4OAc, but stronger in FL and RNase, as reflected by both Peak to Lower Left (P2LL) (Figure 3h) and Z-score Lower

Left (ZscoreLL) scores[26] (Supplementary Figure S8b). We repeated these analyses based on the union of the loops in the four conditions and quantified every loop's strength by Peak to Mean (P2M) in each condition. P2Ms were greater in FL and RNase than in control (p-value < 2.2e-16, Wilcoxon test), whereas NH4OAc's P2Ms were not different from the control's (p-value = 0.41, Wilcoxon test) (Figure 3i). The consistent increases of loop strengths in FL and RNase as compared to control support our detected increases of loop numbers in FL and RNase and suggest that our conclusion of loop number increases does not depend on the threshold of loop calls. Complementary to FL-based transcription inhibition, our experiments on depletion of RPB1, the largest subunit of RNA Polymerase II also led to increases in loop number and strengths as measured by Hi-C (Extended Text 3: RPB1 depletion). Similarly, the depletion of RPB1 in another cell line (DLD-1) led to the emergence of chromatin loops[68]. Taken together, our genome-wide perturbations reveal a negative correlation between caRNA attachment and chromatin loops.

We tested if we could re-create one of the loop changes as detected from our genome-wide perturbations by targeted suppression of one caRNA-producing gene. The ZMYND8 gene is approximately 150 kb in size. Transcription inhibition (FL) depleted caRNA from an approximately 90 kb genomic region including ZMYND8's promoter sequence and approximately the first half of ZMYND8's gene body and induced a ~110 kb chromatin loop straddling across this caRNA-depleted region (Figure 3j). We suppressed ZMYND8 by CRISPR interference (CRISPRi) in an H1 ES cell line with doxycycline-inducible dCas9-KRAB[69,70]. Compared to scrambled gRNA control, our gRNA targeting ZMYND8's promoter reduced ZMYND8's transcription level to approximately 25% (Figure 3k). We designed chromosome conformation capture (3C) primers[71] for (1) a negative control "loop" (Negative Ctrl) that is located 200 kb upstream of the emerged loop and has approximately the same size as the emerged loop, which is not detected as a loop in any Hi-C experiment, (2) a positive control loop (Positive Ctrl) detected by Hi-C in both control and FL, which is not on the same chromosome as ZMYND8, and (3) the emerged loop (also termed the "to-be-tested loop"). We carried out 3C after treating the cells with doxycycline without supplying gRNA (gRNA:None Ctrl), supplying with a scrambled gRNA (gRNA:Scramble Ctrl), and with gRNA targeting ZMYND8's promoter (gRNA:ZMYND8).

The Negative Ctrl primers did not yield any product in any experiment (the first 3 lanes), and the Positive Ctrl primers yielded products at the expected sizes in all three experiments (the last 3 lanes, Figure 3l). In contrast, the primers for the to-be-tested loop yielded a unique product with ZMYND8 gRNA (arrow, Figure 3l), which is absent from the gRNA:None and gRNA:Scramble controls. Thus, we reproduced the change of the chromatin loop at the ZMYND8 locus by targeted suppression of the ZMYND8 gene.

Between-anchor caRNA suppresses chromatin loops

We separated the caRNA associated with chromatin loops into two groups, namely those associated with loop anchors (anchor caRNA), and those between loop anchors (between-anchor caRNA). We asked if the changes in the level of between-anchor caRNA correlates with the changes of chromatin loops. To answer this question, we analyzed the union of the loops (Union loops) detected in every condition (Control, NH4OAc, FL, RNase). These Union loops represent all possible loop locations, including those detected as loops in the Control (control loop) or in an RNA perturbation experiment (emergent loop). We used the ratio of between-anchor caRNA and anchor caRNA levels (Inside-loop To Anchor (ITA) ratio) to represent the relative level of between-anchor caRNA for any Union loop.

First, we tested if the relative level of between-anchor caRNA in the untreated cells (Control) correlates with the control loops by the Gene Set Enrichment Analysis (GSEA)[72]. We first sorted the Union Loops by increasing levels of between-anchor caRNA, i.e., increasing ITA ratios, creating a ranked list (Figure 4a). We then plotted the corresponding GSEA score at every rank (Figure 4b), where a positive/negative GSEA score indicates an enrichment/depletion of the control loops in the subset of Union loops from the top to the currently ranked loop. The GSEA scores stayed positive in the top portion (~30%) of this rank list (Figure 4b), suggesting that the control loops are enriched in the ~30% of the Union loops with the lowest between-anchor caRNA. Thus, among all the possible loop locations, the control loops are enriched in those locations that in the Control exhibited low levels of between-anchor caRNA.

Second, RNase reduced caRNA levels in all Union loops (Figure 4c) and nearly doubled the number of detected loops as compared to Control (Figure 3f). We tested whether the loops in RNase (RNase loops) appeared at the locations where the between-anchor caRNA levels are low in RNase. To this end, we re-ordered the Union loops by increasing levels of between-anchor caRNA, i.e., the ITA ratio in RNase (Figure 4c). As expected, all the GSEA scores are positive in this analysis (Figure 4d), which is because the RNase loops comprise the majority (~75%) of the Union loops, and therefore a majority in any top ranked subset. The GSEA scores in this rank list of Union loops first increased and then decreased, which means that the RNase loops are enriched the higher ranked subset, which are the Union loops exhibiting lower levels of between-anchor caRNA in RNase. These data suggest that the RNase emergent loops are created at the genomic locations where the between-anchor caRNA is more rigorously removed than the anchor caRNA by RNase.

Convergent CSB in the loop anchors is a characteristic of the loops created by loop extrusion[66]. We categorized all the possible locations of loops, i.e., the Union loops, into three groups based on the orientations of the CTCF binding sites at their anchors, namely the loops with convergent CBS, non-convergent CBS, or no CBS. We used Peak to Lower Left (P2LL) to quantify the strength of each loop[26]. Compared to control, RNase treatment increased P2LL in the Union loops with convergent CBS (p-value < 1.6E-9, Wilcoxon test, Figure 4e). In comparison, RNase did not increase P2LL in the Union loops with non-convergent CBS (p-value = 0.4663, Wilcoxon test, Figure 4e) or in the loops without CBS (p-value = 0.6277, Wilcoxon test, Figure 4e). Thus, among all the possible locations of chromatin loops, RNase increased loop strengths at those locations with convergent CBS in their loop anchors.

We asked if the results from RNase treatment can be reproduced by selective removal of a subset of caRNA. To prioritize a subset of caRNA, we reasoned that if a type of caRNA is most rigorously removed by RNase, removal of this type of RNA is more likely to reproduce the results in RNase. We analyzed 50 types of caRNAs, including mRNA, lincRNA, and the RNA of 42 types of repeat families (Repeatmasker, hg38). We quantified the extent of removal of each RNA type by the odds ratio between RNase (as opposed

to control) and this RNA type (as opposed to the other RNA types) based on iMARGI's RNA-end read counts. A smaller odds ratio indicates a stronger depletion of RNase to this type of RNA from chromatin than to the other RNA types. The RNA of the HERV-H repeat family[73] exhibited the smallest odds ratio (odds ratio = 0.350, p-value < 2.2e-16, Chi-square test, Supplementary Figure S9), and thus were selected for further analyses.

We identified the HERV-H caRNA-associated genomic sequences (HERV-RNA target regions) in Control and compared them with the locations of RNase emergent loops. The RNase emergent loops are enriched at the locations that exhibit between-anchor HERV-H caRNA in Control (odds ratio = 1.38, p-value = 5.621e-5, Chi-square test, Figure 4f), suggesting that those loops that stride across between-anchor HERV-H caRNA are suppressed in Control. Next, we analyzed the subset of RNase emergent loops that stride across HERV-H caRNA-attached genomic sequences in Control. Hereafter, we call this subset of RNase emergent loops as "candidate HERV-H caRNA insulated loops" (CHRI-loops). CHRI-loops are enriched with convergent CBS in their loop anchors as compared to (1) control loops striding across HERV-H caRNA-attached genomic sequences (OR = 1.34, p = 0.0068, Chi-square test), or (2) the other control loops not striding across HERV-H caRNA-attached genomic sequences (OR = 1.44, p = 5.8e-6, Chi-square test), or (3) the RNase emergent loops not striding across any HERV-H caRNA-attached genomic sequences in Control (OR = 1.60, p = 4.1e-9, Chi-square test, Figure 4g). Thus, convergent CBS are enriched in the loop anchors of CHRI-loops.

Finally, we tested whether deleting a copy of the HERV-H repeats from the genome can lead to increase the loop strengths of any CHRI-loops. We recently used CRISPR to knock out a specific HERV-H element at Chr13:55,578,227-55,584,087 (Chr13:55.5MB_HERV) in human ES cells (Chr13:55.5MB_HERV KO), however, we did not analyze its impact on chromatin loops in our previous work[44]. Here, we identified the caRNA transcribed from Chr13:55.5MB_HERV and its target genomic sequences in Control (Chr13:55.5MB_HERV targets). We call the loops that stride across any Chr13:55.5MB_HERV targets as "target-crossing loops". We compared Hi-C data in Chr13:55.5MB_HERV KO and Control. We identified two target-crossing loops with increased loop strengths in Chr13:55.5MB_HERV KO (Figure 4h). Neither loop locates

on Chromosome 13, where the HERV-H element is deleted from. In contrast, we did not detect any target-crossing loops with decreased loop strengths in Chr13:55.5MB_HERV KO. Additionally, both Chr13:55.5MB_HERV KO-induced target-crossing loops had convergent CBS in their loop anchors (Figure 4h). Thus, deleting a caRNA-producing genomic sequence led to strengthening of the loops that stride across this caRNA. These data suggest a suppressive role of between-anchor caRNA to chromatin loops.

Based on these results, we predicted that we could suppress a chromatin loop by inducing the transcription of this loop's between-anchor sequence. To test this prediction, we leveraged that there is a ~55 kb loop straddling the AAVS1 locus (AAVS1 loop), and nearby, there is a non-overlapping loop with a similar size to the AAVS1 loop[74] (Nearby Ctrl loop, Figure 4i). We induced transcription at the AAVS1 locus by addition of doxycycline (Dox+) to the H1 ES cells with a doxycycline-inducible dCas9-KRAB gene[69] and subsequently tested the loops with chromosome conformation capture (3C)[71]. Compared to without doxycycline (Dox-), Dox+ weakened the AAVS1 loop (star, Figure 4j) but had little impact on the Nearby Ctrl loop (Figure 4j). Thus, as predicted, we suppressed a chromatin loop by inducing the transcription of this loop's between-anchor sequence.

Discussion

The initial challenges to caRNA as a distinct class of RNA were focused on whether these RNAs are exclusively nascent transcripts[1]. Such a concern was alleviated by the discoveries of long-range RNA-chromatin interactions[2-6], suggesting that caRNA does not completely overlap with nascent transcripts. Our genome-wide analyses reveal several features of caRNA distribution. First, caRNA is preferentially associated with its transcription site and up to several megabases of flanking genomic sequence. Second, TAD boundaries insulate RNA-DNA contacts, evidencing the impact of 3D genome structure on caRNA distribution. Third, CTCF binding sites are hotspots of caRNA. We speculated that this enrichment is in part due to the slower decay[75] of the CTCF co-bound caRNA. In line with this idea, CTCF co-bound caRNA is protected from ribonuclease activity.

It remains unresolved whether caRNA affects the genome's spatial organization. On the one hand, because several 3D features of the genome can be reproduced by computational models without considering caRNA[64,65] and *in vitro* experiments to recapitulate loop extrusion without RNA[66], caRNA was not expected to affect the genome's 3D organization. On the other hand, caRNA demarcates active and repressive chromosomal domains[20,21,76-78], modulates nuclear condensates and spatial compartments[22,66], and promotes transcription factor-chromatin interaction[45,65,79] and folding[45], indicating the involvement of caRNA in the spatial organization of the genome.

At the TAD scale, our data reveal the role of caRNA in maintaining TAD boundaries' stability, suggesting a critical role of caRNA in maintaining the genome's 3D structure. Consistent with this result, depletion of nuclear RNA weakened CTCF's binding to the genome[45]. At the loop scale, our data suggest the relative level of anchor caRNA and between-anchor caRNA can modulate chromatin loops. Removal of the between-anchor caRNA creates chromatin loops or intensifies existing chromatin loops. The created or intensified loops are enriched with convergent CTCF binding sites in their loop anchors, consistent with those loops created by loop extrusion. These results nominate RNA as an indispensable component in shaping the genome's 3D structure.

Acknowledgments

Data availability

All high-throughput data supporting the current study have been deposited on the 4D Nucleome Data Portal (https://data.4dnucleome.org) with the following IDs. iMARGI datasets: H1 control, 4DNESNOJ7HY7; H1 NH4OAc, 4DNESGRI8A8N; H1 FL, 4DNES8B3R3P8; H1 RNase, 4DNESOBRUQ12; HFF, 4DNES9Y1GHK4; K562, 4DNESIKCVASO. Hi-C datasets: H1 control, 4DNESFSCP5L8; H1 NH4OAc, 4DNES2253IBO; H1 FL, 4DNES65I3RQG; H1 RNase, 4DNES4AABNEZ; HFF,

4DNESNMAAN97; K562, 4DNESI7DEJTM. PLAC-seq datasets: H1, 4DNESQMO66LZ; HFF, 4DNESIF5UIQE; K562, 4DNESWX1J3QU.

Code availability

The codes used for the analysis have been deposited and made publicly available on GitHub at https://github.com/Zhong-Lab-UCSD/RNA3Dgenome-code-repository.

Extended Text 1: Nascent transcription

All data used in this comparison are based on K562 cells. We reused the evaluations methods proposed by TT-seq's authors, which are based on (1) intron/exon ratio, (2) uniformity of read coverage downstream to the transcription start site (TSS), and (3) nascent transcription levels.

First, the intron/exon ratio is the ratio of length-normalized read coverage in intron and exons (intronic read counts per base / exonic read counts per base). The regular RNA-seq data exhibited small intron/exon ratios (mean = 0.277, median = 0.134), consistent with the notion that RNA-seq primarily sequences spliced RNA (RNA-seq column, Figure 1c). In comparison, PRO-seq, mNET-seq, GRO-seq, and TT-seq exhibited larger intron/exon ratios (largest p-value < 2.2e-16, Kolmogorov test) with TT-seq's intron/exon ratio being the largest (mean = 0.816, median = 0.618), consistent with these methods' design goal to target nascent RNA and TT-seq being the latest method (Columns 2-4, Figure 1c). Next, we categorized iMARGI's RNA-end reads by their target genomic regions into cis- and trans- caRNA reads, where a cis-caRNA read is associated with the same chromosome as its transcription locus and a trans-caRNA read is associated with another chromosome. iMARGI's cis-caRNA reads exhibited larger intron/exon ratios (mean = 0.837, median = 0.755) than iMARGI's trans-caRNA reads (p-value < 2.2e-16, Kolmogorov test), consistent with the idea that nascent transcripts are enriched at the neighboring genomic regions of their transcription loci[39-41]. Reinforcing this idea, the subset of iMARGI's cis-caRNA reads where the caRNA read's within 1 Mb in genomic distance to its paired DNA read (proximal cis-caRNA) exhibited slightly larger intron/exon ratios (mean = 0.868, median = 0.778) than the entire collection of cis-caRNA reads (p-value = 1.269e-09, Kolmogorov test). Importantly, iMARGI's cis-caRNA reads exhibited larger intron/exon ratios than TT-seq (p-value < 2.2e-16, Kolmogorov test), suggesting iMARGI's cis-caRNA reads efficiently captured intronic RNA.

iMARGI's trans-caRNA's intron/exon ratios (mean = 0.578, median = 0.402) are significantly larger than RNA-seq's ratios (p-value < 2.2e-16, Kolmogorov test) and are close to PRO-seq's ratios (mean = 0.567, median = 0.404, p-value = 0.233, Kolmogorov test), suggesting that intronic transcripts remain a major constituent of trans-caRNA.

These data are consistent with the speedy diffusion of newly transcribed RNA observed by single-molecule tracking in live cells[80]. Thus, we used all iMARGI's caRNA reads (iMARGI in short) that include trans- and cis-caRNA reads in our subsequent comparisons. As expected, iMARGI's overall intron/exon ratios are ranked in the middle of PRO-seq, mNET-seq, GRO-seq, and TT-seq's ratios (The column of "All iMARGI RNA-end reads", Figure 1c),

Second, the main rationale for developing TT-seq was to overcome a limitation of the previous nascent RNA sequencing methods, that is their read coverage quickly decays after the TSS (arrowheads in PRO-seq, GRO-seq, mNET-seq rows, Figure 1d) and thus cannot efficiently capture the RNA transcribed from genomic regions beyond several hundred bases downstream from TSS[34]. Our reanalysis of TT-seq data reproduced TT-seq's improved uniformity of read coverage downstream to the TSS (arrowhead in TT-seq row, Figure 1d). Like TT-seq, iMARGI also improved on the uniformity of read coverage (arrowhead in iMARGI row, Figure 1d) as compared to GRO-seq, PRO-seq, and mNET-seq. Thus, iMARGI also overcomes the limitation of a fast decay of read coverage after TSS. Interestingly, our reanalysis of TT-seq data reproduced an increasing slope of read coverage downstream to TSS (arrowhead in TT-seq row, Figure 1d), which likely reflects a technical caveat due to TT-seq's overcompensation in sequencing transcripts' 3' ends. In contrast, iMARGI does not exhibit TT-seq's increasing slope of read coverage downstream to TSS, in line with iMARGI's experimental strategy that does not have include steps to enrich for transcripts' 5' (GRO-seq, PRO-seq, mNET-seq) or 3' ends (like TT-seq).

Third, we compared nascent transcription levels measured by these technologies. We used TT-seq as the reference technology and compared every other technology to TT-seq. We quantified nascent transcription levels by coverage per kilobase (CPK) and separately analyzed four RNA types, namely coding transcripts (mRNA), lincRNA (lincRNA), antisense RNA (asRNA). As expected, every other technology exhibited a positive correlation with TT-seq in every analyzed RNA type (the largest p-value < 2.2e-16, spearman correlation test, Figure Nascent-CPK). Importantly, iMARGI exhibited the highest correlation with TT-seq in every RNA type Figure Nascent-CPK), suggesting that

iMARGI-based quantification of nascent transcription levels is the closer to TT-seq's quantification than PRO-seq, GRO-seq, and mNET-seq. Taken together, iMARGI's ability to measure the nascent transcripts is comparable to the other methods that are specialized in nascent transcript sequencing. This could be useful because iMARGI is applicable to study untreated cells and tissues, i.e. without requiring incubation of cells with metabolically labeled nucleotides.

Extended Text 2: caRNA in 3D genome compartments

We derived the A/B compartment score based on the Hi-C contact matrix's first eigenvector[43] (PC1 track, Figure 1a), and the relative amount of caRNA attached to every genomic region based on iMARGI's RNA attachment level (RAL, defined as the number of iMARGI read pairs with the DNA ends mapped to this genomic segment[27], purple curve, Figure 1a). Overall, Hi-C contact matrix's first eigenvector exhibits a genome-wide correlation with RAL (p-value < 2e-16, one way ANOVA) (Figure 1a, Supplementary Figure S2), suggesting an association between 3D genome spatial compartmentalization and caRNA distribution.

To test if different species of caRNA have the same distribution on the genome, we separately analyzed the Alu-containing caRNA (Alu-caRNA) and LINE1-containing caRNA (L1-caRNA)[81,82]. Both Alu-caRNA and L1-caRNA's RALs exhibit similar genome-wide distributions as the RAL of all the caRNAs (Figure 1a), suggesting both Alu-caRNA and L1-caRNA are enriched in the A compartment. The enrichment of L1-caRNA in the A compartment does not correlate with the relative depletion of LINE1 genomic sequence in the A compartment[76,83]. However, the ratio of Alu-caRNA's RAL vs. L1-caRNA's RAL exhibits a genome-wide correlation with the A/B compartment score (log2(Alu/L1) track, Figure 1a). These data suggest that not all caRNAs have the same distributions on the genome; moreover, the difference between Alu-caRNA's and L1-caRNA's distributions corresponds to A/B compartments.

Extended Text 3: RBP1 depletion

RBP1, encoded by the POLR2A gene, is the largest subunit of RNA polymerase II. We used the second generation of the auxin-inducible degron (AID2) technology to deplete RBP1[84]. In HCT116 RPB1-Dox-OsTIR1-mClover-mAID cells (RPB1-AID2 cells), where RBP1 is tagged for acute depletion upon addition of doxycycline and 5-Ph-IAA[84]. As the control, we treated RPB1-AID2 cells by doxycycline for 24 hours without 5-Ph-IAA and followed with Hi-C (No-depletion Ctrl), which yielded 632,233,849 read pairs. For the depletion experiment, we treated RPB1-AID2 cells by doxycycline for 24 hours and 5-Ph-IAA for 6 hours and followed with Hi-C (Depletion group), which yielded a comparable number (716,607,191) of read pairs to the IAA- Ctrl. We subjected these data loop calling with HiCCUPS[67]. The Depletion group yielded 3,307 loops, which is approximately 16% more than the detected loops in No-deletion Ctrl (2,619) (p-value = 7.9e-9, paired t-test, chromosome by chromosome). These data suggest depletion of RBP1 led to an increase in loop number.

Taking the union of the loops in the No-depletion Ctrl and the Depletion group, we obtained a total of 5,241 loops (Union loops). We compared the loop strengths (P2LL)[26] between the No-depletion Ctrl and the Depletion group using all the Union loops. The Depletion group exhibited higher P2LLs than the No-depletion Ctrl (fold change = 1.06, p-value < 2.07e-14, paired t-test), suggesting an increase in loop strengths. Taken together, acute depletion of RBP1 resulted in more and stronger chromatin looping.

References

1   Holmes, D. S., Mayfield, J. E., Sander, G. & Bonner, J. Chromosomal RNA: its properties. *Science* **177**, 72-74 (1972). https://doi.org:10.1126/science.177.4043.72
2   Rodriguez-Campos, A. & Azorin, F. RNA is an integral component of chromatin that contributes to its structural organization. *PLoS One* **2**, e1182 (2007). https://doi.org:10.1371/journal.pone.0001182
3   Hall, L. L. & Lawrence, J. B. RNA as a fundamental component of interphase chromosomes: could repeats prove key? *Curr Opin Genet Dev* **37**, 137-147 (2016). https://doi.org:10.1016/j.gde.2016.04.005
4   Nozawa, R. S. & Gilbert, N. RNA: Nuclear Glue for Folding the Genome. *Trends Cell Biol* **29**, 201-211 (2019). https://doi.org:10.1016/j.tcb.2018.12.003
5   Nozawa, R. S. *et al.* SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs. *Cell* **169**, 1214-1227 e1218 (2017). https://doi.org:10.1016/j.cell.2017.05.029
6   Pederson, T. Half a century of locating DNA and RNA in cells. *FASEB J* **33**, 8693-8694 (2019). https://doi.org:10.1096/fj.190801ufm

7   Li, X. *et al.* GRID-seq reveals the global RNA-chromatin interactome. *Nat Biotechnol* **35**, 940-950 (2017). https://doi.org:10.1038/nbt.3968

8   Miao, Y. *et al.* Enhancer-associated long non-coding RNA LEENE regulates endothelial nitric oxide synthase and endothelial function. *Nat Commun* **9**, 292 (2018). https://doi.org:10.1038/s41467-017-02113-y

9   Place, R. F., Li, L. C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proc Natl Acad Sci U S A* **105**, 1608-1613 (2008). https://doi.org:10.1073/pnas.0707594105

10   Morris, K. V., Chan, S. W., Jacobsen, S. E. & Looney, D. J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289-1292 (2004). https://doi.org:10.1126/science.1101372

11   Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-137 (1996). https://doi.org:10.1038/379131a0

12   Yang, F. *et al.* The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome Biol* **16**, 52 (2015). https://doi.org:10.1186/s13059-015-0618-0

13   Calandrelli, R. *et al.* Stress-induced RNA-chromatin interactions promote endothelial dysfunction. *Nat Commun* **11**, 5211 (2020). https://doi.org:10.1038/s41467-020-18957-w

14   Watanabe, T. *et al.* Role for piRNAs and noncoding RNA in de novo DNA methylation of the imprinted mouse Rasgrf1 locus. *Science* **332**, 848-852 (2011). https://doi.org:10.1126/science.1203919

15   Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323 (2007). https://doi.org:10.1016/j.cell.2007.05.022

16   Xiao, R. *et al.* Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell* **178**, 107-121 e118 (2019). https://doi.org:10.1016/j.cell.2019.06.001

17   Dumelie, J. G. & Jaffrey, S. R. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *Elife* **6** (2017). https://doi.org:10.7554/eLife.28306

18   Yan, Z. *et al.* Genome-wide colocalization of RNA-DNA interactions and fusion RNA pairs. *Proc Natl Acad Sci U S A* **116**, 3328-3337 (2019). https://doi.org:10.1073/pnas.1819788116

19   Yin, Y. *et al.* U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature* **580**, 147-150 (2020). https://doi.org:10.1038/s41586-020-2105-3

20   Percharde, M. *et al.* A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174**, 391-405 e319 (2018). https://doi.org:10.1016/j.cell.2018.05.043

21   Chen, W. *et al.* RNAs as Proximity-Labeling Media for Identifying Nuclear Speckle Positions Relative to the Genome. *iScience* **4**, 204-215 (2018). https://doi.org:10.1016/j.isci.2018.06.005

22   Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *Cell* **184**, 5775-5790 e5730 (2021). https://doi.org:10.1016/j.cell.2021.10.014

23   Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293 (2009). https://doi.org:10.1126/science.1181369

24   Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380 (2012). https://doi.org:10.1038/nature11082

25   Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385 (2012). https://doi.org:10.1038/nature11049

26   Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680 (2014). https://doi.org:10.1016/j.cell.2014.11.021

27   Sridhar, B. *et al.* Systematic Mapping of RNA-Chromatin Interactions In Vivo. *Curr Biol* **27**, 602-609 (2017). https://doi.org:10.1016/j.cub.2017.01.011

28   Wu, W. *et al.* Mapping RNA-chromatin interactions by sequencing with iMARGI. *Nat Protoc* **14**, 3243-3272 (2019). https://doi.org:10.1038/s41596-019-0229-4

29   Bell, J. C. *et al.* Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *Elife* **7** (2018). https://doi.org:10.7554/eLife.27024

30   Bonetti, A. *et al.* RADICL-seq identifies general and cell type-specific principles of genome-wide RNA-chromatin interactions. *Nat Commun* **11**, 1018 (2020). https://doi.org:10.1038/s41467-020-14337-6

31   Akgol Oksuz, B. *et al.* Systematic evaluation of chromosome conformation capture assays. *Nat Methods* **18**, 1046-1055 (2021). https://doi.org:10.1038/s41592-021-01248-7

32   Fang, R. *et al.* Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res* **26**, 1345-1348 (2016). https://doi.org:10.1038/cr.2016.137

33   Li, X. & Fu, X. D. Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat Rev Genet* **20**, 503-519 (2019). https://doi.org:10.1038/s41576-019-0135-1

34   Wissink, E. M., Vihervaara, A., Tippens, N. D. & Lis, J. T. Nascent RNA analyses: tracking transcription and its regulation. *Nat Rev Genet* **20**, 705-723 (2019). https://doi.org:10.1038/s41576-019-0159-6

35   Core, L. J., Waterfall, J. J. & Lis, J. T. Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845-1848 (2008). https://doi.org:10.1126/science.1162228

36   Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**, 950-953 (2013). https://doi.org:10.1126/science.1229386

37   Nojima, T., Gomes, T., Carmo-Fonseca, M. & Proudfoot, N. J. Mammalian NET-seq analysis defines nascent RNA profiles and associated RNA processing genome-wide. *Nat Protoc* **11**, 413-428 (2016). https://doi.org:10.1038/nprot.2016.012

38   Schwalb, B. *et al.* TT-seq maps the human transient transcriptome. *Science* **352**, 1225-1228 (2016). https://doi.org:10.1126/science.aad9841

39   Femino, A. M., Fay, F. S., Fogarty, K. & Singer, R. H. Visualization of single RNA transcripts in situ. *Science* **280**, 585-590 (1998). https://doi.org:10.1126/science.280.5363.585

40   Shah, S. *et al.* Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* **174**, 363-376 e316 (2018). https://doi.org:10.1016/j.cell.2018.05.035

41   Su, J. H., Zheng, P., Kinrot, S. S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641-1659 e1626 (2020). https://doi.org:10.1016/j.cell.2020.07.032

42   Dekker, J. *et al.* The 4D nucleome project. *Nature* **549**, 219-226 (2017). https://doi.org:10.1038/nature23884

43   Mourad, R. & Cuvier, O. Predicting the spatial organization of chromosomes using epigenetic data. *Genome Biol* **16**, 182 (2015). https://doi.org:10.1186/s13059-015-0752-8

44   Zhang, Y. *et al.* Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet* **51**, 1380-1388 (2019). https://doi.org:10.1038/s41588-019-0479-7

45   Saldana-Meyer, R. *et al.* RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol Cell* **76**, 412-422 e415 (2019). https://doi.org:10.1016/j.molcel.2019.08.015

46   Hansen, A. S. *et al.* Distinct Classes of Chromatin Loops Revealed by Deletion of an RNA-Binding Region in CTCF. *Mol Cell* **76**, 395-411 e313 (2019). https://doi.org:10.1016/j.molcel.2019.07.039

47   Mei, S. *et al.* Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res* **45**, D658-D662 (2017). https://doi.org:10.1093/nar/gkw983

48   Liang, X. H., Sun, H., Nichols, J. G. & Crooke, S. T. RNase H1-Dependent Antisense Oligonucleotides Are Robustly Active in Directing RNA Cleavage in Both the Cytoplasm and the Nucleus. *Mol Ther* **25**, 2075-2092 (2017). https://doi.org:10.1016/j.ymthe.2017.06.002

49   Moore, M. Nuclear RNA turnover. *Cell* **108**, 431-434 (2002). https://doi.org:10.1016/s0092-8674(02)00645-1

50   Barutcu, A. R., Blencowe, B. J. & Rinn, J. L. Differential contribution of steady-state RNA and active transcription in chromatin organization. *EMBO Rep* **20**, e48068 (2019). https://doi.org:10.15252/embr.201948068

51   Thakur, J. & Henikoff, S. Architectural RNA in chromatin organization. *Biochem Soc Trans* **48**, 1967-1978 (2020). https://doi.org:10.1042/BST20191226

52   Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244 (2015). https://doi.org:10.1038/nature14450

53   Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944 e922 (2017). https://doi.org:10.1016/j.cell.2017.05.004

54   Kubo, N. *et al.* Promoter-proximal CTCF binding promotes distal enhancer-dependent gene activation. *Nat Struct Mol Biol* **28**, 152-161 (2021). https://doi.org:10.1038/s41594-020-00539-5

55   Aumiller, W. M., Jr. & Keating, C. D. Phosphorylation-mediated RNA/peptide complex coacervation as a model for intracellular liquid organelles. *Nat Chem* **8**, 129-137 (2016). https://doi.org:10.1038/nchem.2414

56   Saha, S. & Hyman, A. A. RNA gets in phase. *J Cell Biol* **216**, 2235-2237 (2017). https://doi.org:10.1083/jcb.201706034

57   Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243-247 (2017). https://doi.org:10.1038/nature22386

58   Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432-445 (2010). https://doi.org:10.1016/j.cell.2010.03.030

59   Chao, S. H. & Price, D. H. Flavopiridol inactivates P-TEFb and blocks most RNA polymerase II transcription in vivo. *J Biol Chem* **276**, 31793-31799 (2001). https://doi.org:10.1074/jbc.M102306200

60   Engreitz, J. M. *et al.* RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. *Cell* **159**, 188-199 (2014). https://doi.org:10.1016/j.cell.2014.08.018

61   Ilik, I. A. *et al.* SON and SRRM2 are essential for nuclear speckle formation. *Elife* **9** (2020). https://doi.org:10.7554/eLife.60579

62   Fu, X. D. & Maniatis, T. Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**, 437-441 (1990). https://doi.org:10.1038/343437a0

63   Smith, K. P., Hall, L. L. & Lawrence, J. B. Nuclear hubs built on RNAs and clustered organization of the genome. *Curr Opin Cell Biol* **64**, 67-76 (2020). https://doi.org:10.1016/j.ceb.2020.02.015

64   Kim, J., Han, K. Y., Khanna, N., Ha, T. & Belmont, A. S. Nuclear speckle fusion via long-range directional motion regulates speckle morphology after transcriptional inhibition. *J Cell Sci* **132** (2019). https://doi.org:10.1242/jcs.226563

65   Sigova, A. A. *et al.* Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978-981 (2015). https://doi.org:10.1126/science.aad3346

66   Henninger, J. E. *et al.* RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**, 207-225 e224 (2021). https://doi.org:10.1016/j.cell.2020.11.030

67   Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98 (2016). https://doi.org:10.1016/j.cels.2016.07.002

68   Zhang, S. *et al.* RNA polymerase II is required for spatial chromatin reorganization following exit from mitosis. *Sci Adv* **7**, eabg8205 (2021). https://doi.org:10.1126/sciadv.abg8205

69   Genga, R. M. J. *et al.* Single-Cell RNA-Sequencing-Based CRISPRi Screening Resolves Molecular Drivers of Early Human Endoderm Development. *Cell Rep* **27**, 708-718 e710 (2019). https://doi.org:10.1016/j.celrep.2019.03.076

70   Gonzalez, F. *et al.* An iCRISPR platform for rapid, multiplexable, and inducible genome editing in human pluripotent stem cells. *Cell Stem Cell* **15**, 215-226 (2014). https://doi.org:10.1016/j.stem.2014.05.018

71   Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-1311 (2002). https://doi.org:10.1126/science.1067799

72   Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005). https://doi.org:10.1073/pnas.0506580102

73   Smit, A. F. Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* **9**, 657-663 (1999). https://doi.org:10.1016/s0959-437x(99)00031-3

74   Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell* **78**, 554-565 e557 (2020). https://doi.org:10.1016/j.molcel.2020.03.003

75   Zhou, H. *et al.* Rixosomal RNA degradation contributes to silencing of Polycomb target genes. *Nature* **604**, 167-174 (2022). https://doi.org:10.1038/s41586-022-04598-0

76   Lu, J. Y. *et al.* Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res* **31**, 613-630 (2021). https://doi.org:10.1038/s41422-020-00466-6

77   Nozawa, R. S. & Gilbert, N. Interphase chromatin LINEd with RNA. *Cell* **156**, 864-865 (2014). https://doi.org:10.1016/j.cell.2014.02.005

78   Hall, L. L. *et al.* Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* **156**, 907-919 (2014). https://doi.org:10.1016/j.cell.2014.01.042

79   Hansen, A. S., Amitai, A., Cattoglio, C., Tjian, R. & Darzacq, X. Guided nuclear exploration increases CTCF target search efficiency. *Nat Chem Biol* **16**, 257-266 (2020). https://doi.org:10.1038/s41589-019-0422-3

80   Shav-Tal, Y. *et al.* Dynamics of single mRNPs in nuclei of living cells. *Science* **304**, 1797-1800 (2004). https://doi.org:10.1126/science.1099754

81   Sharp, J. A., Perea-Resa, C., Wang, W. & Blower, M. D. Cell division requires RNA eviction from condensing chromosomes. *J Cell Biol* **219** (2020). https://doi.org:10.1083/jcb.201910148

82   Xiong, F. *et al.* RNA m(6)A modification orchestrates a LINE-1-host interaction that facilitates retrotransposition and contributes to long gene vulnerability. *Cell Res* **31**, 861-885 (2021). https://doi.org:10.1038/s41422-021-00515-8

83   Lu, J. Y. *et al.* Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Rep* **30**, 3296-3311 e3295 (2020). https://doi.org:10.1016/j.celrep.2020.02.048

84   Yesbolatova, A. *et al.* The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat Commun* **11**, 5701 (2020). https://doi.org:10.1038/s41467-020-19532-z

Figure legends

Figure 1. caRNA and 3D genome organization. (a) An RNA-DNA contact matrix in a 2M bp sequence on Chromosome 6. The blue box marks an identified caRNA domain, shown as a rectangular block in iMARGI's contact matrix. (b) Upset plot of the numbers of the detected caRNA domains in H1, HFF, and K562. (c) Box plots of caRNA domains' sizes (blue) and of the lengths of caRNA-producing genomic sequences (red), corresponding to the widths (blue) and heights (red) of the rectangular blocks in iMARGI's contact matrix. (d) Comparison of iMARGI's RNA-end sequences with nascent-transcript-sequencing methods. Y axis: Intron/exon ratio, that is the number of reads mapped to intron per base divided by the number of reads mapped to exon per base. Columns: sequencing methods. iMARGI's RNA-end reads (last four columns) are categorized by their target genomic regions into cis- and trans- caRNA reads, where a cis-caRNA read is associated with the same chromosome as its transcription locus ("Cis only" track) and a trans-caRNA read is associated with another chromosome ("Trans only" track). All: iMARGI's trans-caRNA and cis-caRNA reads combined. "Cis < 1 Mb" column: a cis-caRNA read targeting a genomic sequence that is within 1 Mb of its transcription locus. (e) Comparison of iMARGI-derived RNA attachment level (RAL, purple curve) and A/B compartments inferred by Hi-C's contact matrix's first eigenvector (PC1 track) on Chromosome 11. RAL: the RAL based on all caRNA, RAL w/o Alu and L1: the RAL based on all caRNA except Alu- and L1-caRNA. Alu RAL: RAL calculated with Alu-caRNA. L1 RAL: RAL calculated with L1-caRNA. Log2(Alu/L1): log ratio of the Alu RAL and L1 RAL. (f) The RAL based only on the caRNA transcribed within each TAD (row), plotted across this TAD (center block) and its equal-length flanking regions on both sides. Curve at the bottom: the average RAL of all TADs (rows). (g) Comparison of normalized RNA-DNA contact matrices in WT and KO cell lines. The arrowhead points to the HERV-H element in WT that is deleted in KO. The increased RNA-DNA contacts in KO are mostly cross-over contacts (in the box at the upper right corner). The lost RNA-DNA contacts in KO are mostly non-cross-over (lower right corner). (h) Comparison of normalized RNA-DNA contact matrices in WT and KI cell lines. The arrowhead points to the insertion site. KI exhibited increased contacts (+) on either side of the insertion site (upper left and lower right corners) and decreased (-) cross-over contacts (upper right and lower left corners).

Figure 2. caRNA on CTCF binding sites. (a) RNA attachment level (RAL) on CTCF ChIP-seq peaks and their flanking regions (rows) in Control (left) and RNase (right). Top: the average RAL of all the rows. (b) Scatterplot of the allele-difference in sequence similarities to CTCF's DNA recognition motif (motif score ratio, x axis) against the allele-difference of the attached caRNA, as the ratio of the RALs of the two alleles (RAL allele ratio, y axis). Each dot represents a CRS_SNV asPeak. CRS_SNV: A CTCF recognition site (CRS) harboring a between-allele single-nucleotide variation (SNV). asPeak: allele-specific peak. A positive correlation means the allele with the higher sequence similarity to the CTCF DNA recognition motif is attached with more caRNA. (c-d) A CRS_SNV where the reference allele exhibits a higher similarity to the CTCF DNA recognition motif (c), which also exhibits more attached caRNA (RAL), as compared to the alternative allele (d). (e) Comparison of RAL ratios in Control and RNase (columns). A RAL ratio is the RAL on a peak (defined as a 1 kb bin centered at the peak center) divided by the RAL on the flanking sequence (defined as 3 kb flanking sequence). ***: p-value < 1e-15, paired t test. (f) RAL on a CTCF binding site derived from two Control (blue tracks) and two RNase (red tracks) iMARGI experiments. (g) Changes in TAD boundary's caRNA (Delta RAL, x axis) vs. the proportions of diminished (red bar) and persistent boundaries (blue bar, y axis) with the corresponding Delta RAL. A smaller (more negative) Delta RAL represents a greater depletion of caRNA from the TAD boundary (x axis). (h) Distributions of RAL (x axis) of the TAD boundaries in control that diminished (red) or persisted (blue) in CTCF depletion.

Figure 3. RNA perturbation-induced changes. (a-d) Immunostaining of SON in Control (a), NH4OAc (b), FL (c), and RNase treated H1 cells (d). Scale bar = 6 μm. (e) Proportions of Hi-C's intra- (yellow) and inter-chromosomal (blue) read pairs in each condition. (f) Upset plot of loop numbers of the four conditions (rows). (g) An example of loop change. Hi-C contact matrix of every replicate (row). Arrows: a shared loop in FL and RNase that is absent in control and NH4OAc. (h-i) FL and RNase increase loop strengths. (h) Aggregate loop strength (P2LL,

y axis) in each condition (column). Color bars: the loops detected in each condition (red) or their union (blue). (i) Loop strength (P2M) distribution in each condition (column). ****: p-value < 2.2e-16. (j-l) An induced chromatin loop in transcriptional suppression of the ZMYND8. (j) Changes in RNA-DNA contact maps in Control (upper panel) and FL (lower panel). FL reduced the caRNA in the upstream region of the ZMYND8 gene (blue box) and induced a chromatin loop near the caRNA-depleted region (curve at the bottom). CBS: CTCF binding site. Arrowheads point to CBSs' directions. Arrow: the emergent loop in FL. (k) Comparison of normalized ZMYND8's expression levels (y axis) in CRISPRi experiments with the scrambled gRNA (Scramble Ctrl) and ZMYND8-targeting gRNA (ZMYND8 gRNA). (l) 3C products from the Negative Ctrl primers (the first 3 lanes), the primers for the To-be-tested loop (3 middle 3 lanes), and the Positive Ctrl primers (the last 3 lanes) in CRISPRi experiments without a gRNA (gRNA: None), with a scrambled gRNA control (gRNA: Scramble), or with the ZMYND8-targeting gRNA (gRNA: ZMYND8). The Negative Ctrl primers did not yield any product in any experiment (Lanes 3-5). The Positive Ctrl primers yielded products of the same sizes in all three experiments (Lanes 9-11). The primers for the to-be-tested loop yielded a product with ZMYND8-targeting gRNA (arrow) but not with a scrambled gRNA or without gRNA (Lanes 6-8), confirming that a loop is created by ZMYND8 CRISPRi. Lane 1: E-Gel™ 1 Kb DNA Ladder. Lanes 2: E-Gel™ 50bp DNA Ladder.

Figure 4. Between-anchor caRNA suppresses chromatin loops. (a-b) The loops in the Control (control loops) are depleted with between-anchor caRNA. (a) The caRNA levels in the control (Control RAL) on loop anchors (two sides) and between the anchors (middle) is color-coded (blue: low, yellow: high) for every loop detected in any condition (Union loops, rows). Loops are ranked by the relative level of their between-anchor caRNA (Inside-loop To Anchor (ITA) ratio) from low (top) to high (bottom). (b) The enrichment/depletion level (GSEA score, x axis) of the control loops in the subset of loops from the top-ranked loop (first row) and the currently ranked loop (current row, y axis). A positive/negative GSEA score indicates an enrichment/depletion of the control loops in this subset of loops. The control loops are enriched in the top ranked loops, i.e., those with low levels of between-anchor caRNA (blue bar on the right), and are depleted in the bottom-ranked loops, i.e., those with high levels of between-anchor caRNA (yellow bar). (c-d) RNase emergent loops are those with low levels of between-anchor caRNA. (c) The caRNA levels in the RNase (RNase RAL) on loop anchors (two sides) and between the anchors (middle) is color-coded (blue: low, yellow: high) for every loop detected in any condition (Union loops, rows). The union loops (rows) are ordered by the relative level of their between-anchor caRNA (the ITA ratio calculated in RNase) from low (top) to high (bottom). (d) The enrichment level (GSEA score, x axis) of the RNase loops (x axis) in the subset of loops from the top-ranked loop (first row) and the currently ranked loop (current row, y axis). The RNase-specific loops are enriched in the top-ranked loops, i.e., the loops with low levels of between-anchor caRNA in RNase, as indicated by the increasing GSEA scores (blue bar on the right). In contrast, the loops detected in other conditions are enriched in the bottom-ranked loops, i.e., the loops with high levels of between-anchor caRNA in RNase, as indicated by the decreasing GSEA scores (yellow bar). (e) Comparison of loop strengths (P2LL, y axis) in the loops with convergent CBS, non-convergent CBS, and without CBS in the control and RNase (columns). ****: p-value < 1.6e-9. ns: not significant. (f) Enrichment of RNase emergent loops with between-anchor HERV-H caRNA in control (odds ratio, x axis). **: p-value < 5.6e-5, Chi-square test. (g) Enrichment of "candidate HERV-H caRNA insulated loops" (CHRI-loops) with convergent CBS in their loop anchors as compared to control loops striding across HERV-H caRNA-attached genomic sequences (first row), the other control loops not striding across HERV-H caRNA-attached genomic sequences (second row), and the RNase emergent loops not striding across any HERV-H caRNA-attached genomic sequences (third row). *: p < 6.8e-3, **: p < 5.8e-6, ***: p < 4.1e-9. (h) The two target-crossing loops (rows) with increased Hi-C contacts in HERV-H KO (KO column) as compared to control (Control column). Arrow: direction of CTCF binding site in the loop anchor. (i) Transcription induction of a gene suppresses a loop straddling this gene. Genomic coordinates of the AAVS1 locus, the loop straddling the AAVS1 locus (AAVS1 loop), and a nearby loop with a similar size (Nearby Ctrl loop). (j) 3C products without doxycycline (Dox: -) and with transcription induction by doxycycline (Dox: +), based on primers against the AAVS1 loop (Lanes 3, 4), the Nearby Ctrl loop (Lanes 5, 6), and a size-matched control region without any Hi-C detected loop (Negative Ctrl, Lanes 7, 8). *: Difference in 3C products between Dox- and Dox+. Lane 1: E-Gel™ 1 Kb DNA Ladder. Lanes 2 and 9: E-Gel™ 50bp DNA Ladder.
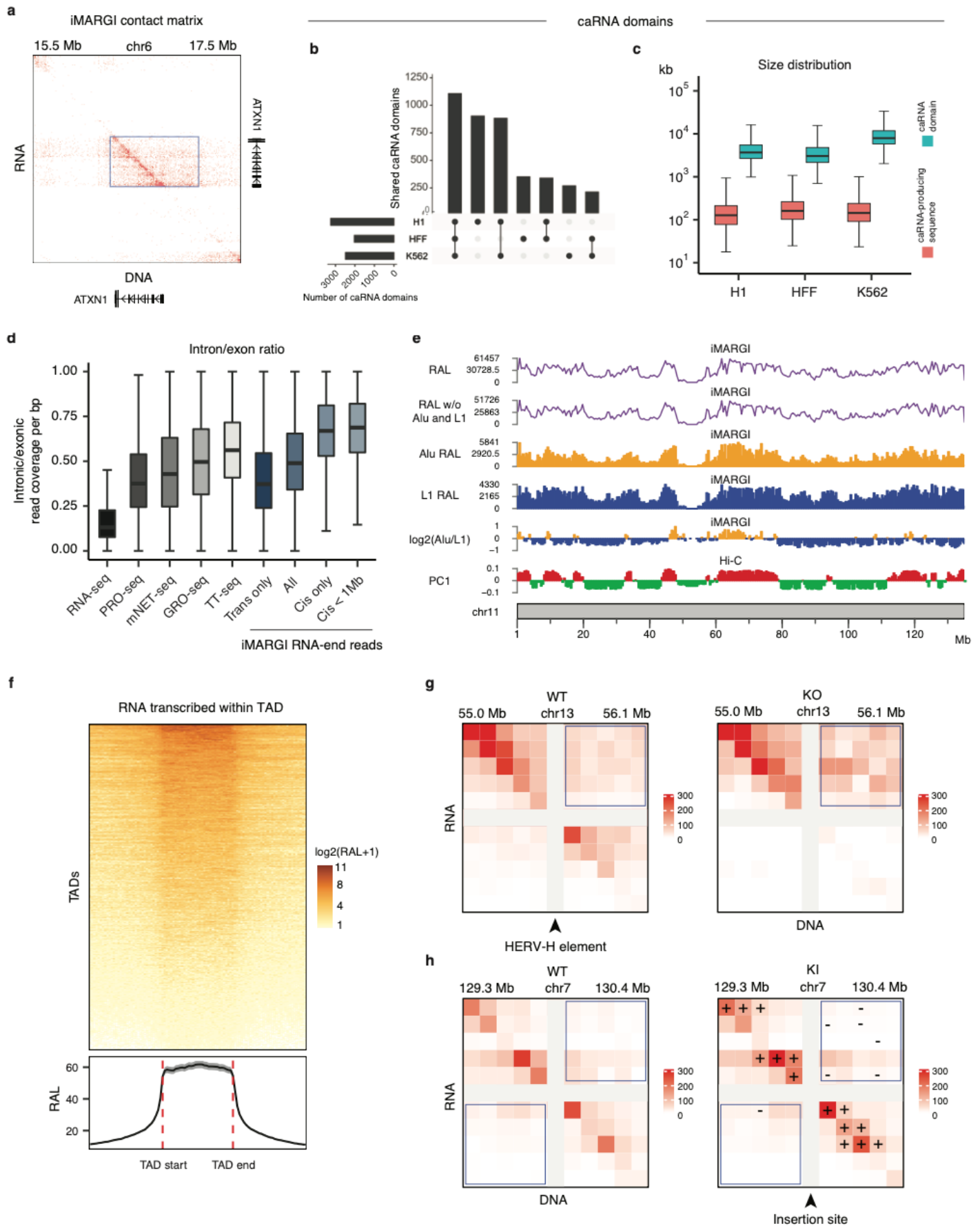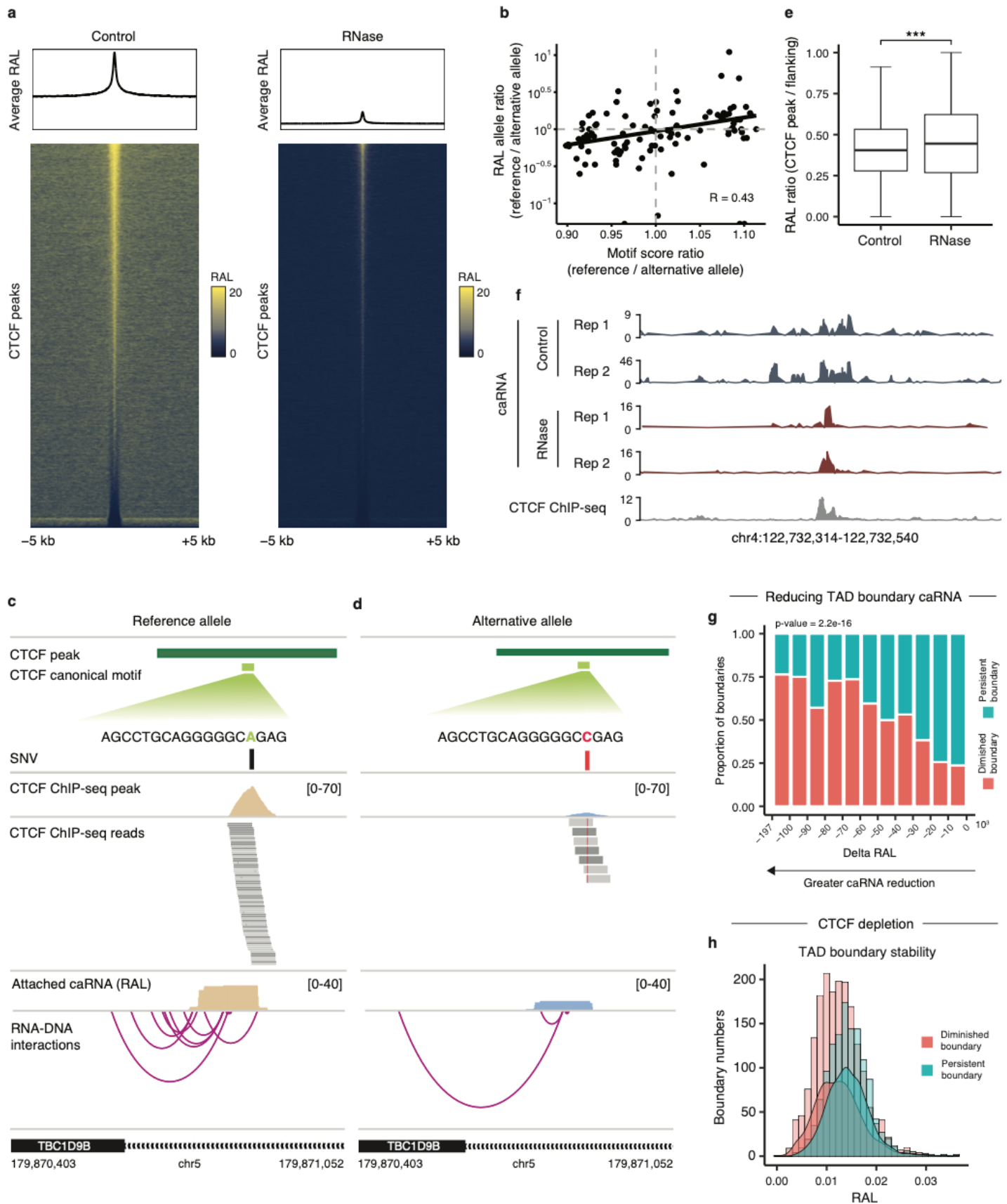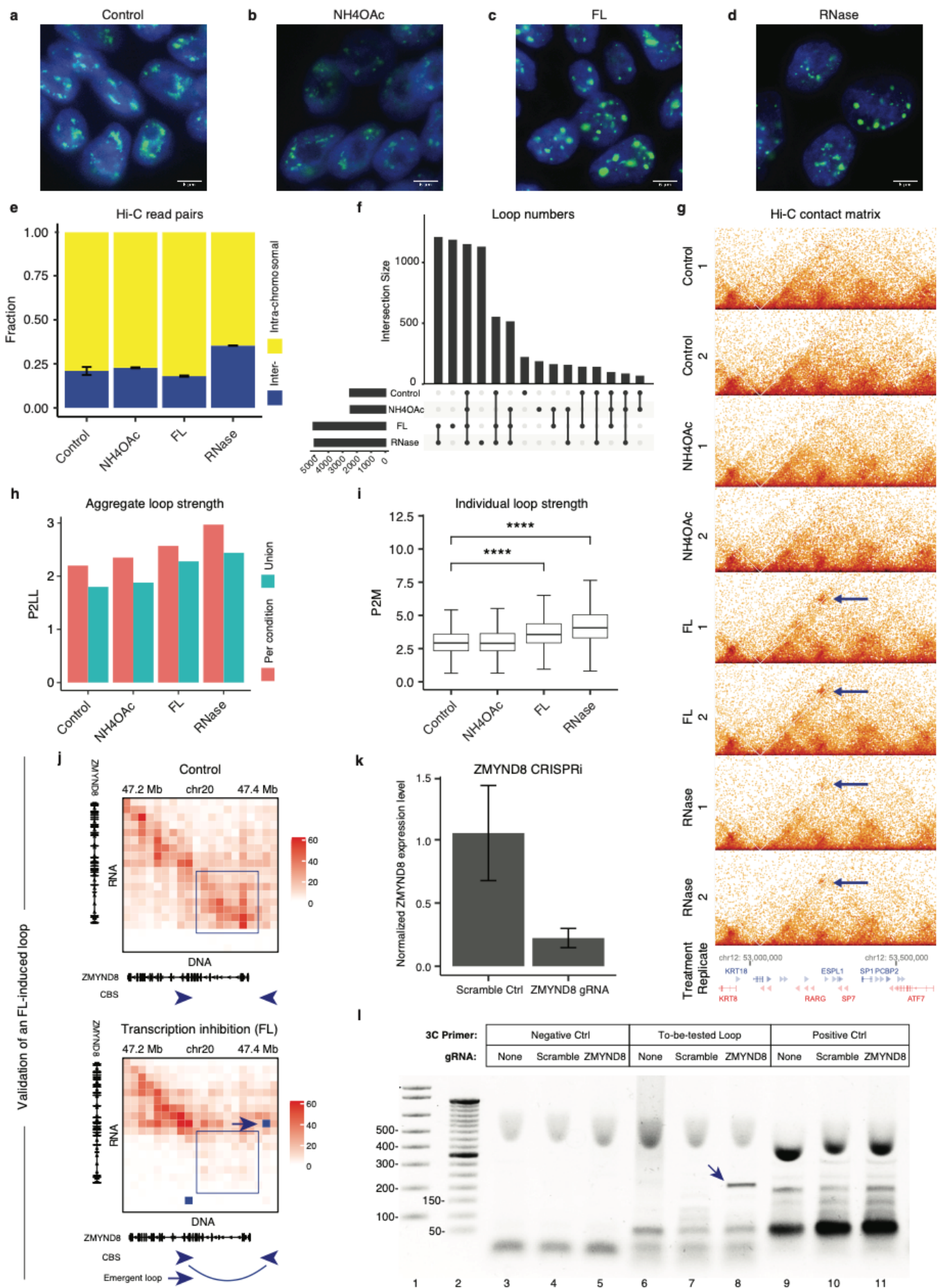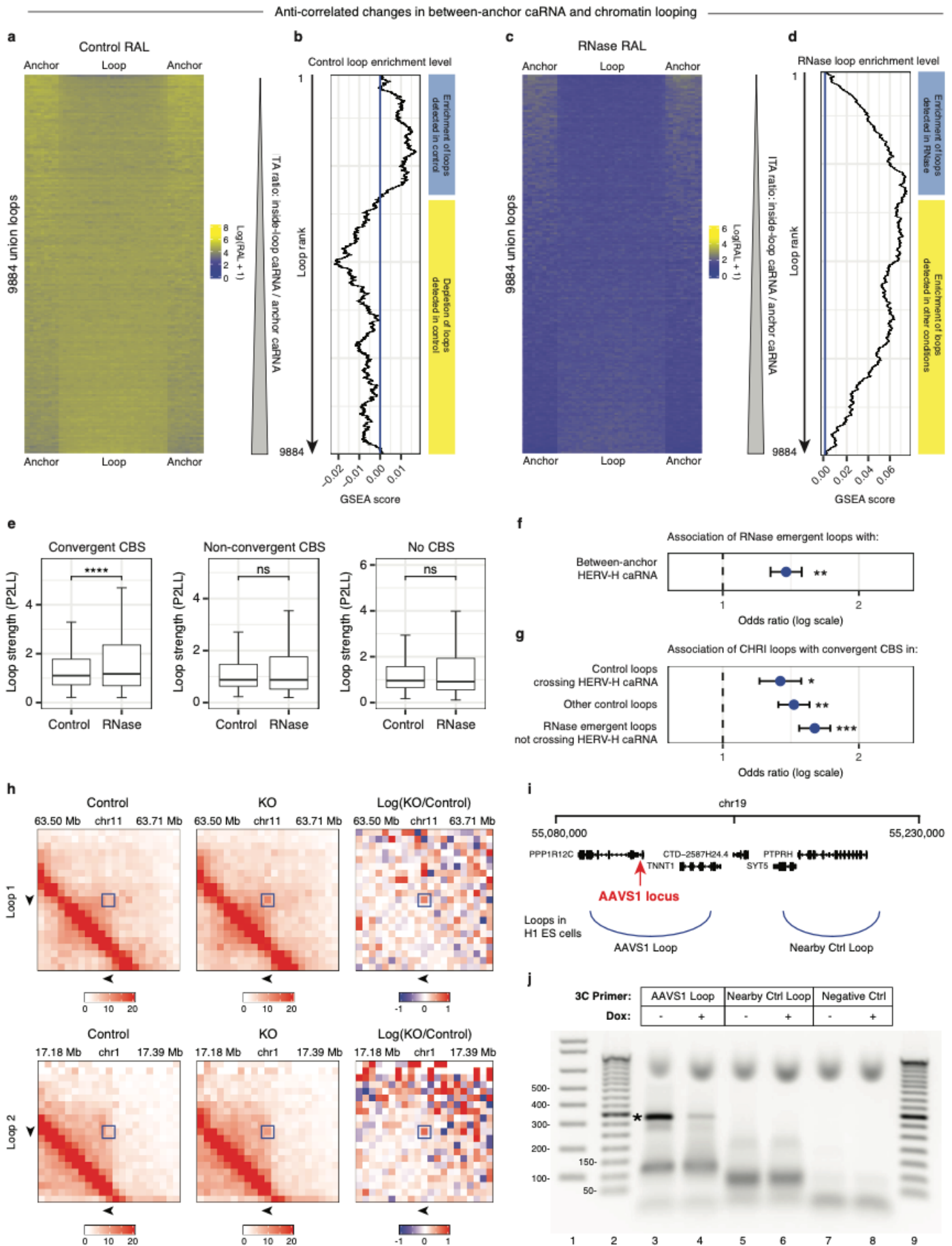
Figure 1

Figure 2

Figure 3

Figure 4

Anti-correlated changes in between-anchor caRNA and chromatin looping

Supplementary material

Supplementary Figure S1. Comparison of iMARGI's RNA-end sequences with nascent transcription sequencing methods. (a) Comparison based on the read coverage on TSS' flanking sequences. Arrowheads point to the fast decay of read coverage on the sense strand (red) downstream to TSS in PRO-seq, GRO-seq, and mNET-seq (last three rows), that is eliminated in TT-seq and iMARGI (first rows). (b) Correlation of each technology (color bar)'s measured transcription level to TT-seq's (y axis), among the coding genes (mRNA), lincRNAs, and antisense RNAs (asRNA) (columns). To illustrate the data behind the first orange bar, the expanded view provides the scatterplot of measured transcription level (CPK: coverage per kilobase) of every mRNA (dot) by TT-seq (y axis) and iMARGI (x axis).
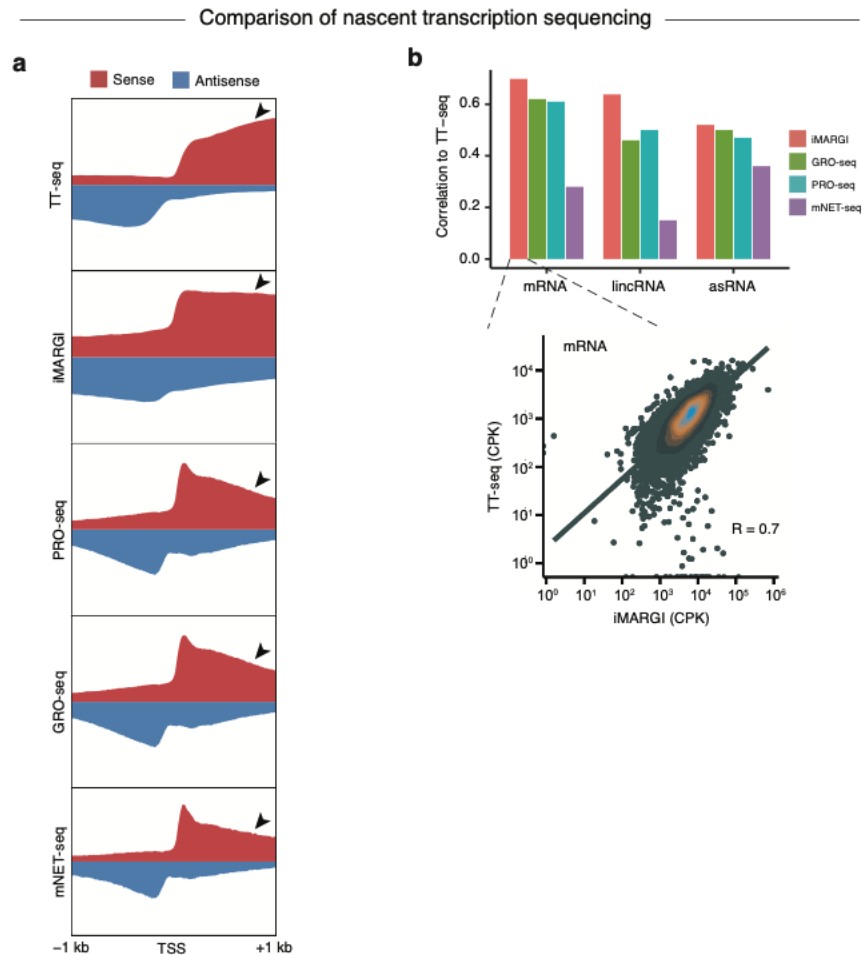
Supplementary Figure S2. Comparison of RNA-chromatin and genomic interactions. (a, b, c) iMARGI-derived RNA attachment level (RAL, purple curve) and A/B compartments represented by Hi-C's contact matrix's first eigenvector (A/B, red and green bars) of each 500 kb bin on Chromosome 11 in H1 (a), HFF (b) and K562 (c). (d, e, f) Scatterplot of Hi-C's first eigenvector (y axis) and RAL (x axis) on every 500 kb genomic bin (dot) of the entire genome in H1 (d), HFF (e) and K562 (f). (g, h, i) Scatterplot of gene density (y axis) and RAL (x axis) on every 500 kb genomic bin (dot) of the entire genome in H1 (g), HFF (h) and K562 (i). SCC: Spearman correlation coefficient.
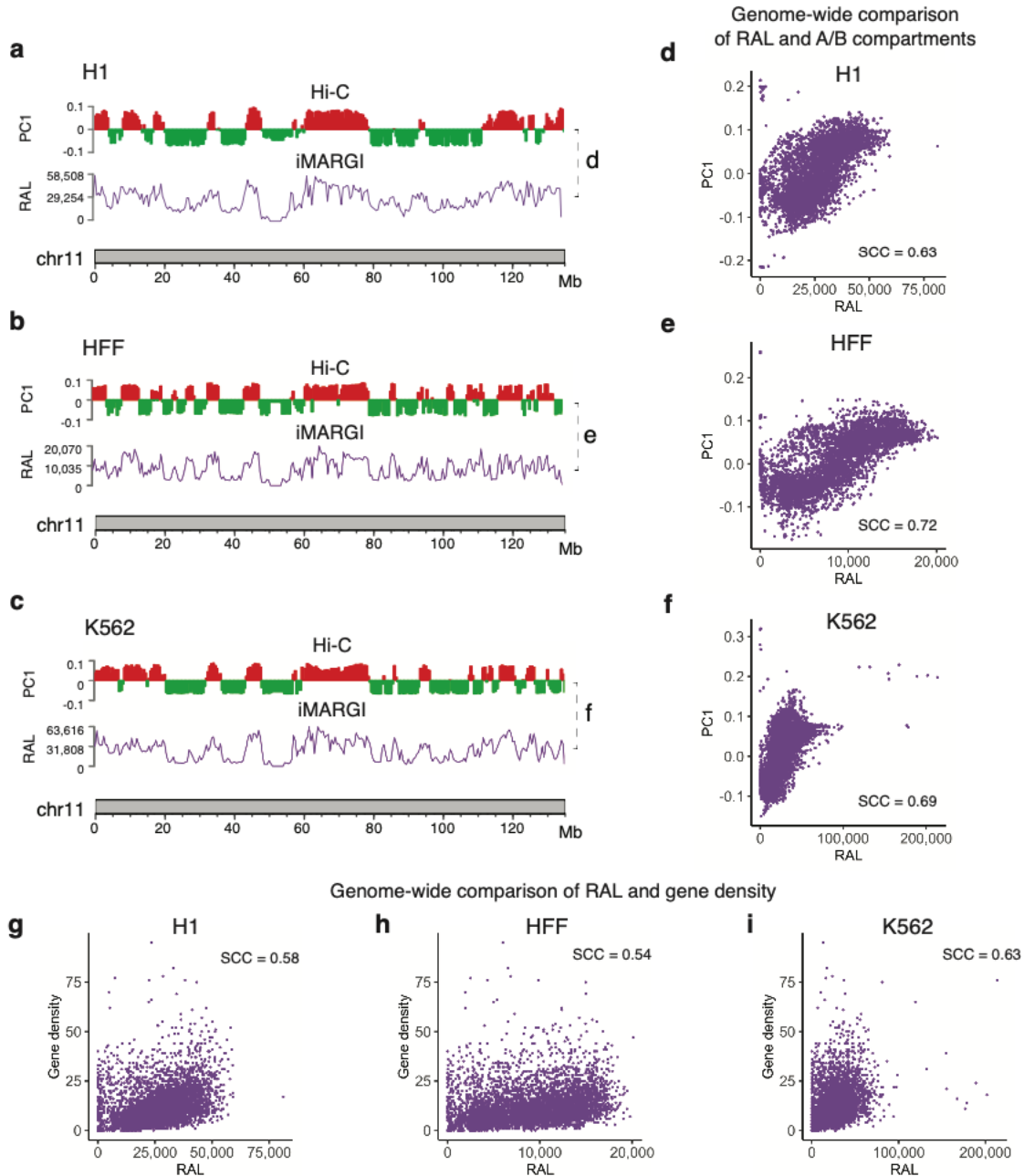
Supplementary Figure S3. TADs. (a) Proportions of caRNA-enriched (blue) -and caRNA-depleted (green) TADs in each condition (column). Error bar: SEM. (b) The RAL of every TAD (row) and its equal-length flanking regions, based on all the RNAs transcribed from any genomic sequences outside of this TAD (row). The TAD lengths are normalized (center, x axis). Blue curve at the bottom: average RALs of all TADs. (c) TAD numbers in each condition (column). (d) A reproduction of Figure 1b with Alu- and L1-caRNA removed from the analysis. (e) A reproduction of Figure S2b with Alu- and L1-caRNA removed from the analysis.
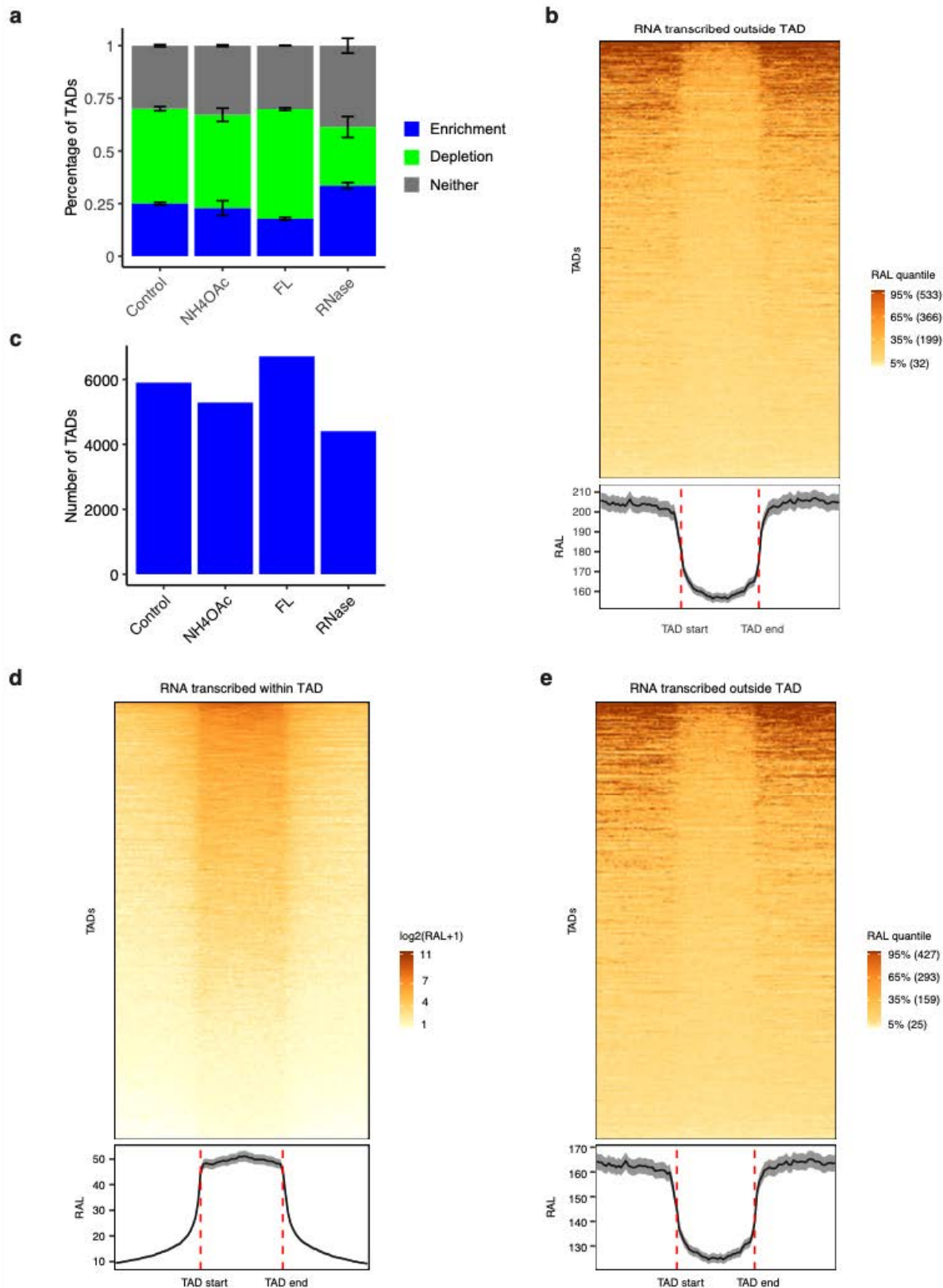
Supplementary Figure S4. caRNA domains. (a, b, c) Scatter plots of the rectangular blocks' widths that correspond to caRNA domain's sizes (x axis) and the lengths of the longest gene in each caRNA domain (y axis). The rectangles' heights are similar to the lengths of the longest genes overlapping with each rectangle's y coordinates, suggesting that most caRNA domains are decorated by the caRNAs of single genes. (d) Upset plot of the numbers of caRNA domains in untreated H1 (Control) and H1 treated with NH4OAc, FL, and RNase. (e) Distributions of the heights, *: p-value < 1e-25, ***: p-value < 1e-75, ****: p-value <1e-100. (f) Distributions of widths. *: p-value < 1e-50, ***: p-value < 1e-180.
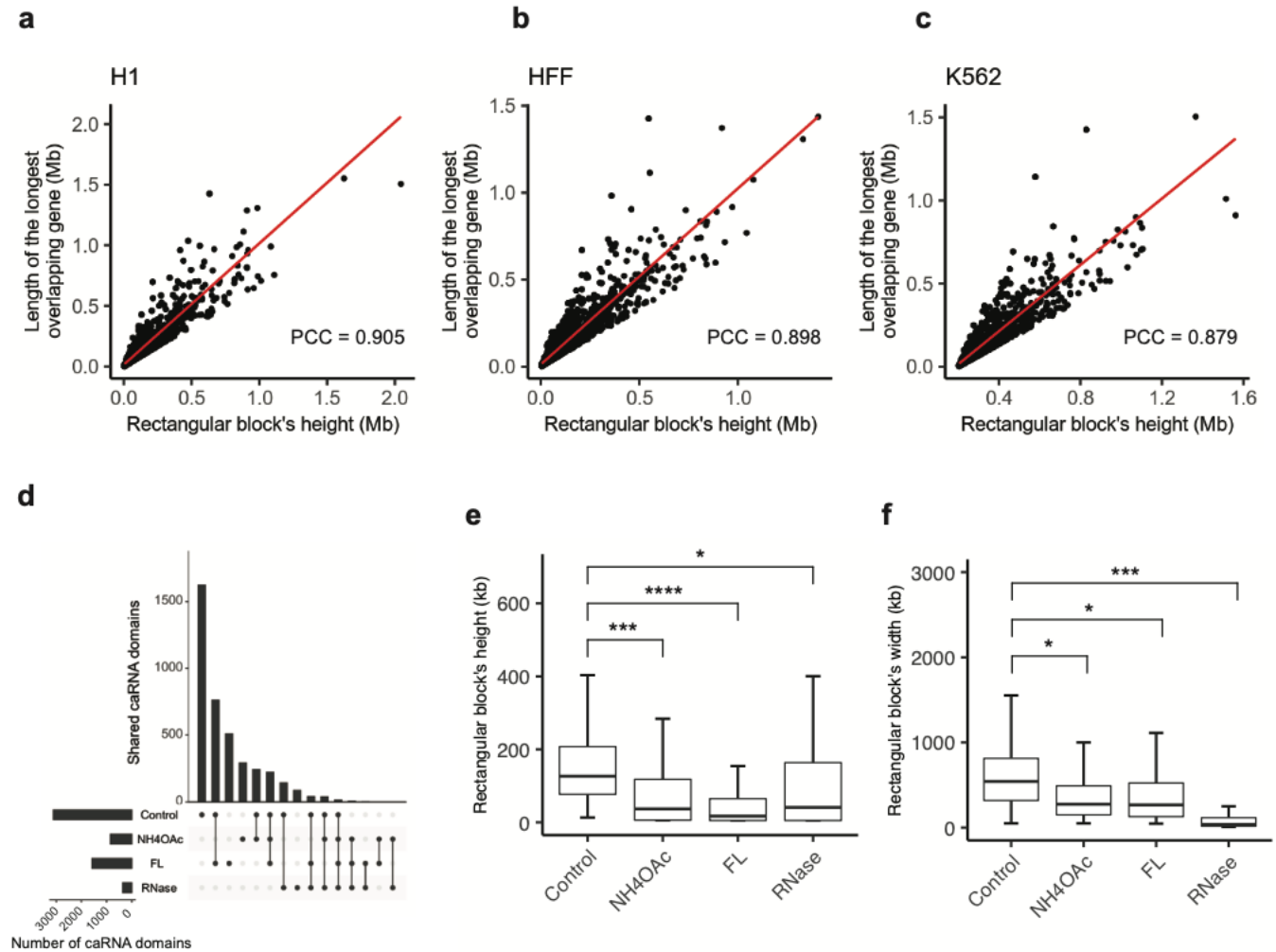
Supplementary Figure S5. RNase-induced TAD boundary changes. (a) The number of TAD boundaries (y axis) in control (red) and RNase (blue) with respect to the range of TAD boundaries' sizes (x axis). p-value: the p-value of a paired t test. (b) Changes in TAD boundary's caRNA (Delta RAL, x axis) vs. changes in boundary strength (Delta insulation, y axis). A smaller (more negative) Delta RAL represents a greater depletion of caRNA (x axis). A positive Delta insulation represents the weakening of a TAD boundary, and the larger Delta insulation is the greater the degree of weakening is (y axis). Error bar: standard error of Delta insulations in this group of TAD boundaries (column). p-value: p-value from ANOVA.
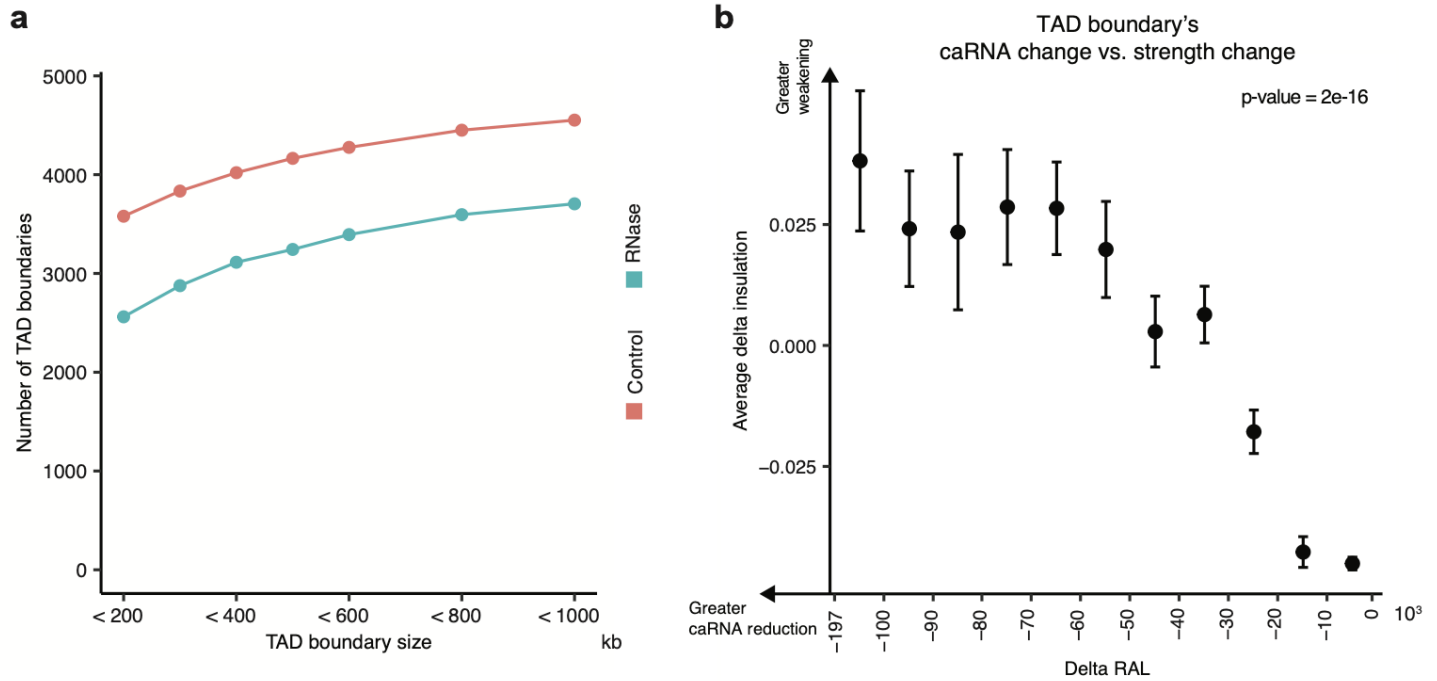
Supplementary Figure S6. Immunofluorescence analyses of SC35 and SON. (a-d) Immunostaining of SC35 in control (a), NH4OAc (b), FL (c), and RNase treated H1 cells (d). Scale bar = 6 µm. (e-g) Distribution of SON's average number of foci per nucleus in control and each treatment (first row). *: p-value < 0.05. **: p-value < 0.01. In comparison, SON's mean background fluorescence (last row) does not change between control (pink) and each treatment (green). ns: not significant. (h-j) Distribution of SC35's average number of foci per nucleus in control and each treatment (first row). **: p-value < 0.01. ***: p-value < 1.0e-3. In comparison, SC35's mean background fluorescence (last row) does not change between control (pink) and each treatment (green). ns: not significant.
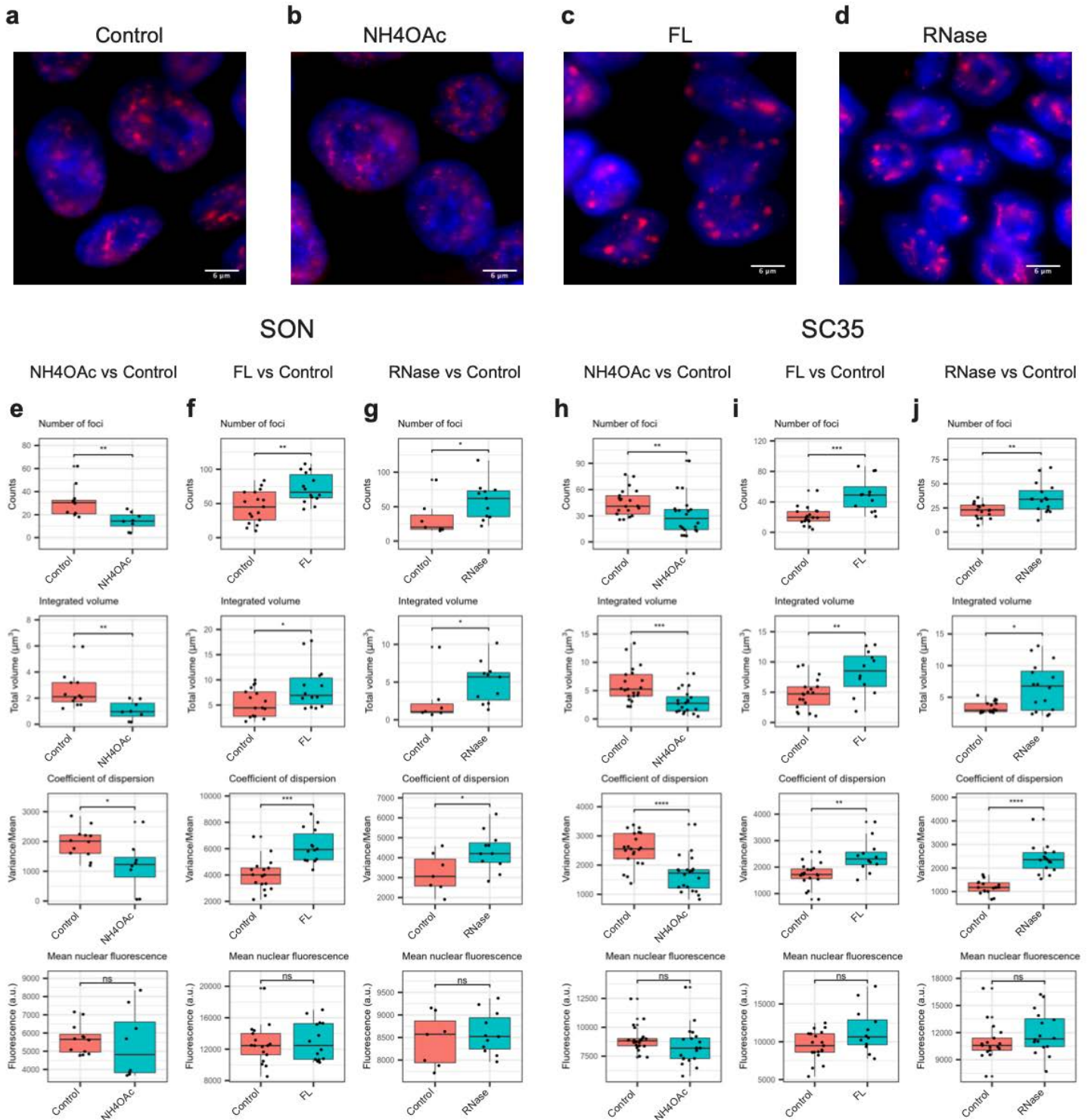
Supplementary Figure S7. (a) Distance distribution of Hi-C's intra-chromosomal read pairs in control (black), NH4OAc (green), FL (red), and RNase (purple). Expanded views of the distance distribution are shown in the ranges of 0-2 Mb (b) and ≥ 2 Mb (c).
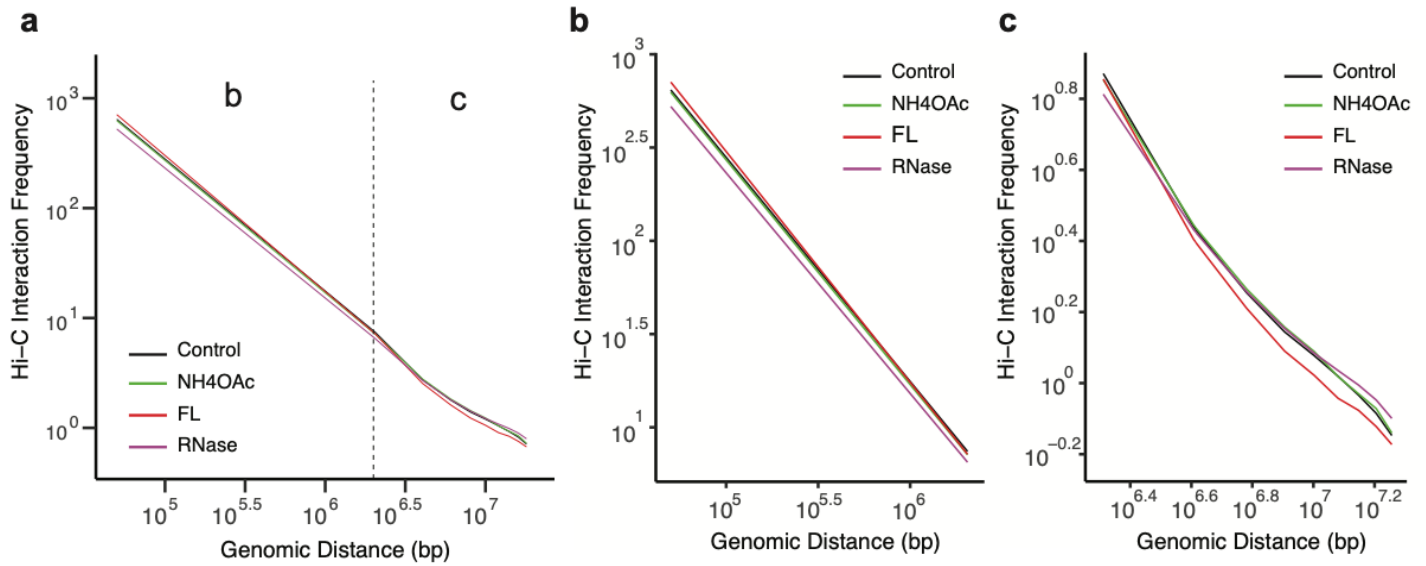
Supplementary Figure S8. (a) An example of loop change. iMARGI contract matrix of each condition for the corresponding genomic region of Figure 4g. Dots: loops. Arrows: a shared loop in FL and RNase that is absent in control and NH4OAc. (b) Loop strengths as measured by Z-score Lower Left (ZscoreLL) scores, based on the loops detected in each condition (red) and the union of loops of the four conditions (green). ZscoreLL is an overall score of the entire Hi-C dataset.
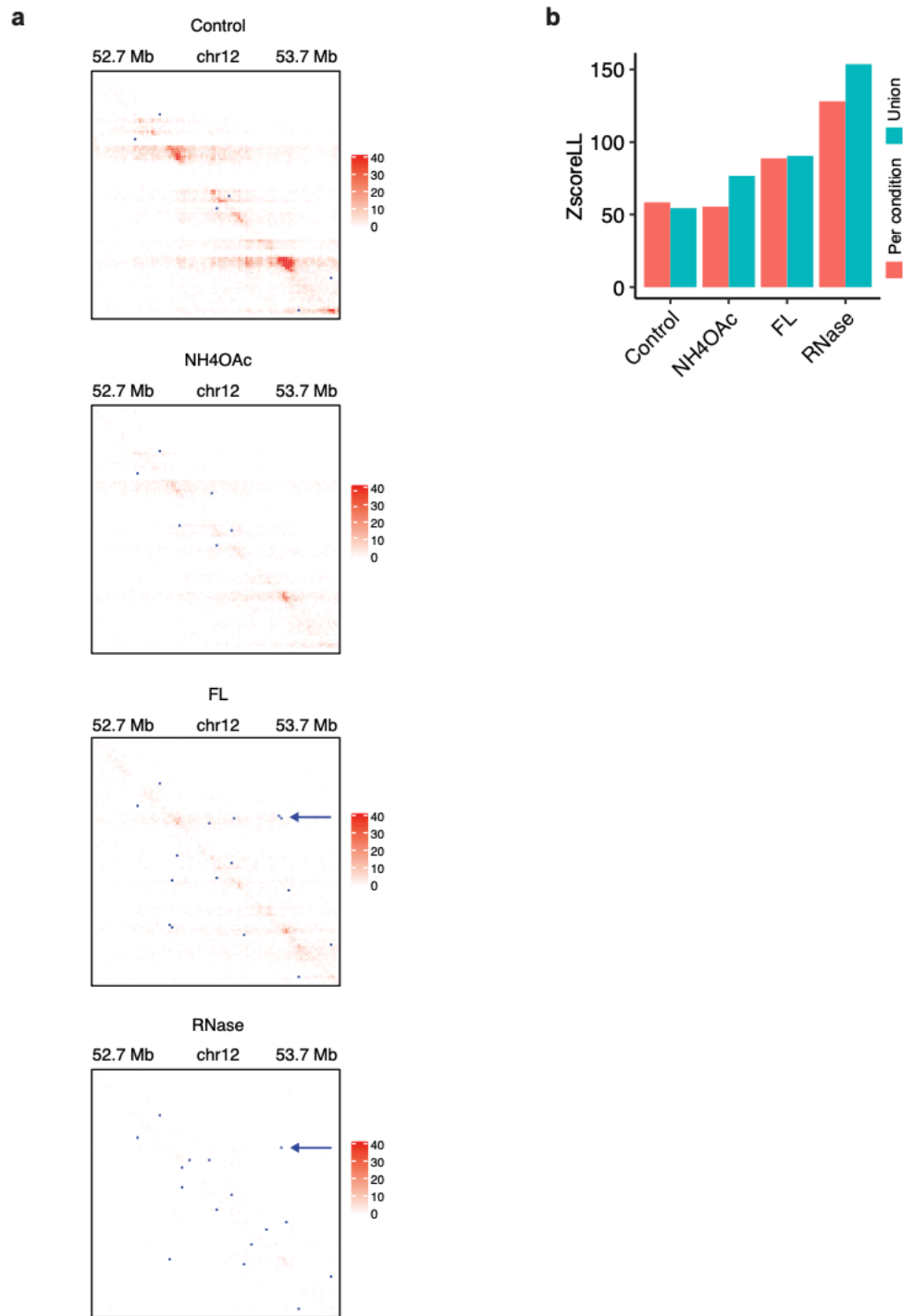
Supplementary Figure S9. Ranking of different types of caRNAs including mRNA, lincRNA, and genomic repeat-transcribed RNA (rows) by the relative extent of depletion in RNase to this type of RNA (log odds ratio, x axis) as compared to the other RNA types. The most depleted RNA type ranks at the top.
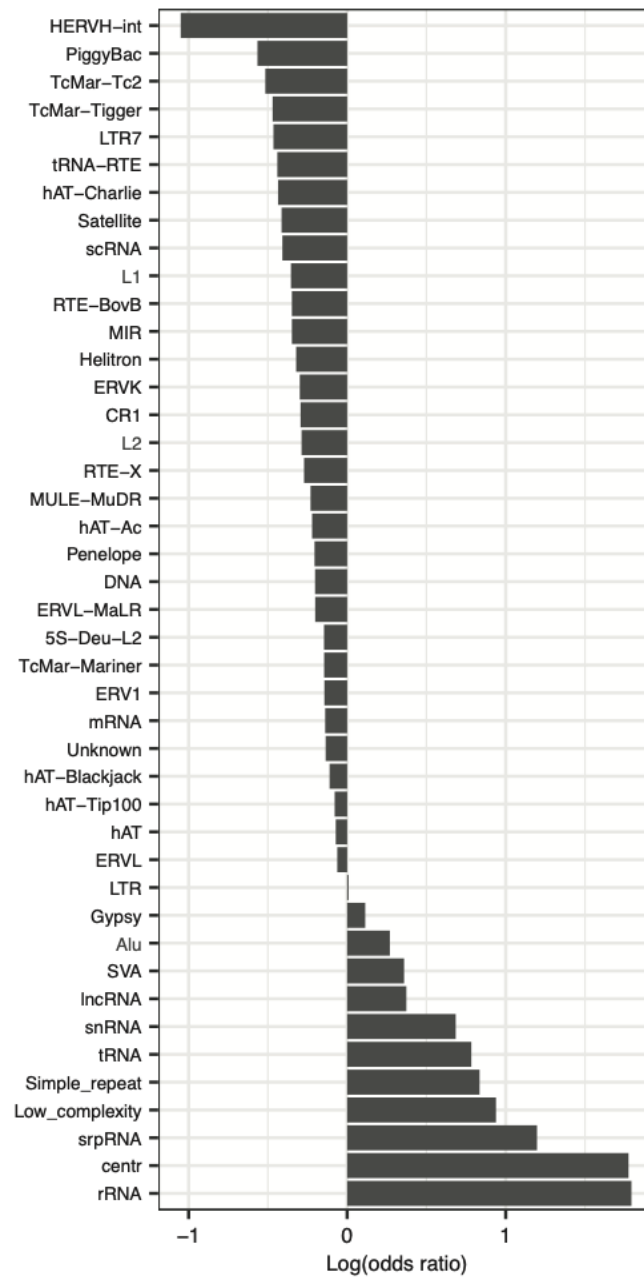
Table S1. Summary of iMARGI, Hi-C, and PLAC-seq datasets.

a. iMARGI datasets				
Cell line	Treatment	Number of replicates	Total # of read pairs	Source
H1	None	4	2,642,778,166	This work
	NH4OAc	2	1,247,204,613	
	FL	2	1,438,312,761	
	RNase	2	1,670,230,715	
HFF	None	2	2,755,576,893	
K562	None	2	1,293,950,206	

b. Hi-C datasets				
Cell line	Treatment	Number of replicates	Total # of read pairs	Source
H1	None	2	616,625,628	This work
	NH4OAc	2	613,098,350	
	FL	2	604,503,572	
	RNase	2	654,798,738	
HFF	None	2	2,764,855,452	4DNESNMAAN97
K562	None	6	907,136,828	4DNESI7DEJTM

c. PLAC-seq datasets				
Cell line	Treatment	Number of replicates	Total # of read pairs	Source
H1	None	2	237,966,651	This work
HFF	None	2	134,809,941	
K562	None	2	314,344,830	

d. iMARGI datasets in engineered cells				
Cell line	Treatment	Number of replicates	Total # of read pairs	Source
H9 MLC2v:H2B	None	1	692,140,673	This work
H9 MLC2v:H2B HERV2-KO	None	1	566,783,631	
H9 MLC2v:H2B HERV2-ins-clone2	None	1	967,136,454	

e. Hi-C datasets in engineered cells				
Cell line	Treatment	Number of replicates	Total # of read pairs	Source
HCT116 RPB1-Dox-OsTIR1-mClover-mAID	Doxycycline (control)	1	632,233,849	This work
	Doxycycline and 5-Ph-IAA	1	716,607,191	
H9 MLC2v:H2B	None	2	552,532,207	GSE116862
H9 MLC2v:H2B HERV2-KO	None	2	635,909,624	

Online methods

Cell culture and treatments. Human embryonic stem cells (H1), hTert-immortalized human foreskin fibroblasts (HFF), and chronic myelogenous leukemia lymphoblasts (K562) were obtained from the 4D Nucleome (4DN) Cell Repository and cultured following the 4DN Consortium's approved culture protocol for each cell line (https://www.4dnucleome.org/cell-lines.html). The cell lines in the 4DN Cell Repository were established by the 4DN Consortium in collaboration with WiCell and ATCC for providing quality-controlled cells from the identical batch to minimize cell source and culture condition variations. The cell culture protocols were developed by the 4DN Cell Line Working Group and approved by the 4DN Steering Committee.

Ammonium acetate treatment. H1 cells were treated with 0.1 M NH_4OAc in complete mTeSR medium for 10 min as described in a previous study[1]. Briefly, a crystalline NH_4OAc (Sigma-Aldrich, Cat# A1542-500G) was dissolved in nuclease-free water and further diluted in cell medium. Aspirate medium in each well and H1 cells were treated with 0.1 M NH_4OAc in medium for 10 min at RT.

Flavopiridol treatment. H1 cells were treated with 1µM flavopiridol in complete mTeSR medium for 1h in an incubator as described previously[2]. Specifically, a crystalline flavopiridol (hydrochloride) (Cayman Chemical, item# 10009197) was dissolved in DMSO to prepare 1mM flavopiridol (FL) stock solution. 1mM FL stock solution was further diluted with complete mTeSR medium. Aspirate cell medium in each well and H1 cells were either treated with 1µM FL in medium or an equivalent amount of DMSO in the medium in an incubator at 37°C for 1h.

RNase A treatment. H1 cells were harvested from cell culture plate and aliquoted cell suspension to 10 million H1 cells per 1.5 mL tube. Wash the cells with 1 mL 1X PBS and centrifuge at 500 X g for 3 min at RT. Then, cells were gently permeabilized by resuspending cell pallets with 0.01% PBST (TritonX-100 in PBS) and treated for 5 min at RT. After permeabilization, cells were treated with 200 µg/mL RNase A as described previously[3] (Thermo Fisher Scientific, Cat# EN0531) on rotator for 10 min at RT. The treated cells were fixed with 4% formaldehyde (Thermo Fisher Scientific, Cat# 28906) for immunofluorescence imaging. For Hi-C and iMARGI library generation, the treated cells were fixed with 1 mL 1% formaldehyde on rotator for 10 min at RT. Then, the reactions were terminated with 250 µL 1M glycine on rotator for 10 min at RT. The treated sample was centrifuged at 2000Xg for 5 min at 4°C and washed with 1 mL cold 1X PBS.

dCas9-KRAB inducible cells. The doxycycline-inducible dCas9-KRAB H1 ES cell line is generated and karyotyped by the 4D Nucleome Consortium (Danwei Huangfu Laboratory) (https://4dnucleome.org), with TRE-dCas9-KRAB and CAGGS-M2rtTA targeted into the AAVS1 locus.

HERV-H deletion and insertion cells. The control H9 human ES cells (H9 MLC2v:H2B), HERV-H deletion cell line (H9 MLC2v:H2B HERV2-KO), and HERV-H insertion cell line (H9 MLC2v:H2B HERV2-ins-clone2) were generated by Bing Ren lab and described in reference[4].

RPB1 The auxin-inducible degron 2 cells. The RPB1 auxin-inducible degron 2 cells (HCT116 RPB1-Dox-OsTIR1-mClover-mAID) were generated by Masato Kanemaki lab and described in reference[5].

Immunofluorescence imaging. The cells on coverslip (Fisher Scientific, Cat# 12-541A) were fixed with 4% formaldehyde at RT for 30 min. The fixed cells were washed with 1X PBS once and permeabilized with 0.1% TritonX-100 in PBS (PBST) at RT for 15 min on shaker. Afterwards, cells were blocked with 5% BSA (VWR, Cat# 97061-420) in PBST at RT for 30 min with gentle shaking. For SC35 staining, H1 cells were incubated with 1 mL diluted mouse monoclonal anti-SC35

primary antibody (1:250) (Abcam, Cat# ab11826) in 5% BSA at 37°C for 1h, and subsequently washed three times with PBST on shaker for 10 min. Cells were further incubated with 1 mL diluted goat anti-mouse secondary antibody with Alexa Fluor 568 (1:500 dilution) (Invitrogen, Cat# A-11004) in 5% BSA at 37°C for 30 min. For SON staining, the cells were incubated with 1 mL diluted rabbit anti-SON primary antibody (1:2000 dilution) (Atlas Antibodies, HPA023535) in 5% BSA at 37°C for 1h, and subsequently washed three times with PBST on shaker for 10 min. The cells were incubated with 1 mL diluted goat anti-rabbit secondary antibody with Alexa Fluor 488 (1:500 dilution) (Invitrogen, Cat# A-11008) in 5% BSA at 37 in 5% BSA at 37°C for 30 min. After staining, the cells were washed three times with PBST on shaker for 10 min. The cells on coverslips were mounted on slides (Fisher Scientific, Cat# 12-544-2) with 10 µL ProLong antifade glass mountant with NucBlue stain (Thermo Fisher Scientific, Cat# P36981), placed in dark room for air-dry overnight. Images in the size of 512×512 pixels were acquired on Applied Precision OMX Super Resolution Microscope using a 100X/1.518 oil objective (GE Healthcare Life Sciences) (pixel size = 0.079 µm). Z-stack images were acquired with thickness of 0.3 µm sample thickness.

Identification of nuclear speckle foci. Nuclear speckle foci were identified by a previously described method[2]. Briefly, the nuclei were manually segmented and the mean fluorescence intensity in nuclei were measured with FIJI. The nuclear speckle foci were identified by FIJI 3D Object Counter plugin, with an appropriate intensity threshold of the mean fluorescence intensity in the cell nuclei and a size cut-off of more than 50 adjoining pixels (pixel size, 79 nm X 79 nm).

In situ Hi-C library generation and data processing. The Hi-C libraries were generated with the Arima-HiC kit (Arima Genomics, material# A510008, Document# A160134 v00) following the manufacturer's instructions. Hi-C data was processed following 4DN consortium's Hi-C data processing protocol (https://www.4dnucleome.org/protocols.html). Briefly, the Hi-C data were processed using the 4D Nucleome (4DN)'s Hi-C Processing Pipeline (v0.2.5) (https://data.4dnucleome.org/resources/data-analysis/hi_c-processing-pipeline), with MAPQ > 30 to filter out multiple mappings.

The output .pairs files were provided to Cooler[6] (v0.8.10) and Juicer Tools[7] (v1.22.01) to generate .mcool and .hic files. The .mcool file was used in HiGlass[8] for visualization. The .hic files were inputted in Juicer Tools for A/B compartment, TAD, and loop analyses. A/B compartments were called by Juicer's "Eigenvector" tool based on KR normalized observed/expected (O/E) contacts at 500 kb resolution. TADs were called by Juicer's "Arrowhead" tool based on KR-normalized contacts at 10 kb resolution. Loops were called by Juicer's "CPU HiCCUPS" tool based on KR-normalized contacts simultaneously at 5 kb and 10 kb resolutions. Except for the resolution parameter, all the other parameters were left as the default.

TAD boundaries were extracted as the genomic regions between TADs in each sample. TAD boundary insulation score was calculated according to the definition in Crane et. al, 2015[9].

Unique loops and overlapping loops were determined as follows. First, the Juicer called loops from each condition were merged into "unique loops" by taking the union. Then the unique loops in the union were reassigned to each condition by the following rule: a unique loop *i* (in the union) with anchor size *s* (either 5 or 10 kb) was re-assigned to a sample *j* if both anchors of loop *i* were within +/-*s* flanking regions of a loop in sample *j*. Aggregate Peak Analysis was performed using the Juicer's "APA" tool with default parameters. Metrics to define the loop strength such as Peak to Lower Left (P2LL), Z-score Lower Left (ZscoreLL), and Peak to Mean (P2M) were calculated as defined in Juicer's APA[7]. The control loop straddling the AAVS1 locus was detected from H1-hESC Micro-C data[10]. To select RNase emergent loops that stride across HERV-H caRNA-attached genomic sequences in Control, i.e., the candidate HERV-H caRNA insulated loops (CHRI-loops), we used a threshold of at least 2 iMARGI read pairs with their RNA ends

overlapping with HERV-H and their DNA ends mapped to the between-loop-anchor sequence. To check if deleting a copy of the HERV-H repeats led to increase of loop strengths of CHRI-loops, we used a threshold of 0.05 for the delta peak (KO peak - control peak), where peak is the normalized Hi-C read count at the loop's pixel, normalized by the total number of read pairs in each sample.

iMARGI library generation and data processing. iMARGI libraries were generated and processed as previously described[11]. According to 4DN's approved iMARGI's data processing protocol[11], any iMARGI read pair in which the RNA end and the DNA end mapped to within 1,000 bp of each other on the genome are removed from the data analysis. The RNA attachment level (RAL) of each genomic segment is the count of the DNA-ends mapped to this genomic segment[12]. Only the inter-chromosomal and the intra-chromosomal iMARGI read pairs that are separated by at least 200 kb apart were used for calculating RAL in any of the correlation analyses. Repeats of hg38 were downloaded from RepeatMasker (Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0). RAL of Alu-containing caRNA (Alu-caRNA) and LINE1-containing caRNA (L1-caRNA) were calculated as the count of the DNA ends mapped to each genomic segment (500 kb size) whose RNA ends mapped to a repeat segment of the Alu or LINE1 family respectively.

PLAC-seq library generation and data processing. PLAC-seq libraries were generated and processed as previously described[13]. The promoter interacting sequences (PINS) were identified by MAPS[14] with default parameters and the input data of PLAC-seq and H3K4me3 ChIP-seq (doi:10.17989/ENCSR443YAS for H1, ENCSR813CFB for HFF, and ENCFF699EUP for K562).

caRNA domains. Each rectangular block on iMARGI's contact matrix was identified as a peak of the iMARGI's read pairs' RNA ends (the height of this RNA peak) and a corresponding DNA peak of the DNA ends (the width of this RNA peak). Homer's findPeaks function was applied to the RNA ends of iMARGI' read pairs (peak size = 5,000 bp, minimum peak interval = 12,000 bp) to identify the peaks on the RNA ends (RNA peak). For reach RNA peak, all the iMARGI's read pairs with their RNA ends inside this RNA peak were retrieved. The retrieved read pairs' DNA ends were subjected to Homer's findPeaks (peak size=25,000 bp, minimum peak interval=50,000 bp) to identify the peaks on the DNA ends (DNA peaks). If multiple DNA peaks were reported, the DNA peak with the highest read number was designated as the corresponding DNA peak.

Normalized RNA-chromatin interaction densities on TSS' flanking regions. Each TSS' flanking region (-1,000 bp to 1,000 bp) was segmented into 250 bp bins. For each caRNA domain overlapping this this TSS, a normalized interaction intensity of this caRNA domain on each bin is calculated by the ratio of the number of iMARGI read pairs' DNA ends in this caRNA domain in each bin vs. the total number of iMARGI read pairs' DNA ends in this caRNA domain in this 2,000 bp flanking region. When there is more than one caRNA domain overlapping with a TSS, we used the sum of every caRNA domain's normalized interaction intensities of each bin as the normalized interaction intensity of this bin.

TSS metagene pileup analysis. The TSS+/- 1kbp region was divided into equal length bins with each bin of 10bp. Intensity for each bin is defined as the average for all transcripts at each bin's log transformed total coverage. Log scale was applied for the raw coverage to minimize the effect of extreme values dominating the mean and magnifies the low valued signatures.

Intron per base coverage ratio in mRNA calculation. There are 19,844 protein coding genes annotated in GENECODE v24. For each gene, all exon regions were extracted and reduced to derive a non-overlapping exon annotation. We then calculated the total reads coverage for the whole gene, all non-overlapped exons, and their corresponding total length. The intron per base coverage for one gene is calculated by subtracting exon coverage and length from the full gene. Genes that have zero exon coverage were excluded.

Phasing ChIP-seq and RNA-DNA interaction. H1 cell line phased genome was obtained from the 4D Nucleome Consortium Data Portal (https://data.4dnucleome.org/joint-analysis#phased-genomes). CTCF canonical motifs on hg38 reference genome were identified using FIMO with parameters --max-stored-scores 1000000 --thresh 1e-2 with CTCF motif MA0139.1 downloaded from JASPAR. H1 cell line CTCF ChIP-seq bam files were downloaded from ENCODE (ENCFF237PQJ, ENCFF397FGK, ENCFF551FQW, ENCFF767YLK). CTCF binding peaks in the H1 cell line were obtained from the cistrome database. The CTCF CBS motifs that simultaneously overlap with one CTCF binding peak and one SNV were filtered using R package GenomicRanges. Allele specific binding at CTCF motifs were tested using chi-square test (since chi-square test applies to large count value, we require the sum of CTCF ChIP-seq reads at both alleles is larger than 5) under the null hypothesis that there is no difference of CTCF binding on the two alleles. We also pre-excluded DNA copy number variant regions based on ClinVar database before carrying out the allele specific binding test. P-values were corrected for multiple testing using FDR. CTCF ChIP-seq or iMARGI DNA ends are assigned to each allele using customized script based on sequence similarities (smith-waterman alignment) between the bam recorded read sequences, reference sequences or alternative sequences encompassing SNV. CBS motif scores for a given sequence with the same length of CTCF motif (MA0139.1), we calculated the sum of scores at each base according to the PWM.

References to online methods

1 Saha, S. & Hyman, A. A. RNA gets in phase. *J Cell Biol* **216**, 2235-2237, doi:10.1083/jcb.201706034 (2017).
2 Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432-445, doi:10.1016/j.cell.2010.03.030 (2010).
3 Barutcu, A. R., Blencowe, B. J. & Rinn, J. L. Differential contribution of steady-state RNA and active transcription in chromatin organization. *EMBO Rep* **20**, e48068, doi:10.15252/embr.201948068 (2019).
4 Zhang, Y. *et al.* Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nat Genet* **51**, 1380-1388, doi:10.1038/s41588-019-0479-7 (2019).
5 Yesbolatova, A. *et al.* The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat Commun* **11**, 5701, doi:10.1038/s41467-020-19532-z (2020).
6 Abdennur, N. & Mirny, L. A. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. *Bioinformatics* **36**, 311-316, doi:10.1093/bioinformatics/btz540 (2020).
7 Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95-98, doi:10.1016/j.cels.2016.07.002 (2016).
8 Kerpedjiev, P. *et al.* HiGlass: web-based visual exploration and analysis of genome interaction maps. *Genome Biol* **19**, 125, doi:10.1186/s13059-018-1486-1 (2018).
9 Crane, E. *et al.* Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature* **523**, 240-244, doi:10.1038/nature14450 (2015).
10 Krietenstein, N. *et al.* Ultrastructural Details of Mammalian Chromosome Architecture. *Mol Cell* **78**, 554-565 e557, doi:10.1016/j.molcel.2020.03.003 (2020).
11 Wu, W. *et al.* Mapping RNA-chromatin interactions by sequencing with iMARGI. *Nat Protoc* **14**, 3243-3272, doi:10.1038/s41596-019-0229-4 (2019).
12 Sridhar, B. *et al.* Systematic Mapping of RNA-Chromatin Interactions In Vivo. *Curr Biol* **27**, 602-609, doi:10.1016/j.cub.2017.01.011 (2017).
13 Fang, R. *et al.* Mapping of long-range chromatin interactions by proximity ligation-assisted ChIP-seq. *Cell Res* **26**, 1345-1348, doi:10.1038/cr.2016.137 (2016).

14   Juric, I. *et al.* MAPS: Model-based analysis of long-range chromatin interactions from PLAC-seq and HiChIP experiments. *PLoS Comput Biol* **15**, e1006982, doi:10.1371/journal.pcbi.1006982 (2019).