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

Authors: Riccardo Calandrelli [1], Xingzhao Wen [2], Tri C. Nguyen [1], Chien-Ju Chen [2], Zhijie Qi [1], Weizhong Chen [1], Zhangming Yan [1], Weixin Wu [1], Kathia Zaleta-Rivera [1], Rong Hu [3,4], Miao Yu [4], Yuchuan Wang [5], 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.

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. We find that the relative abundance of trans-acting caRNA reflects the functional nuclear compartmentalization of the 3D genome, and the caRNA's sequence features are predictive of its spatial localization. We observe extensive caRNA-genome interactions that span several hundred kilobases to several megabases. These caRNA domains correlate with chromatin loops and enhancer-promoter interactions. We report a tradeoff of caRNA's promotive and suppressive roles to chromatin interactions. As a result, caRNA promotes chromatin interactions outside the anchors of chromatin loops, whereas caRNA exhibits an overall suppressive impact to chromatin interactions between loop anchors. Furthermore, caRNA suppresses chromatin loops' number and strengths genome-wide, and this suppression depends on electric charge-mediated RNA interactions. These results reveal caRNA's multifaced role in regulating 3D chromatin organization and highlight caRNA's ability to modulate chromatin looping.

Introduction

After initial debates [1], chromatin-associated RNA (caRNA) has been recognized as a "widespread component of interphase chromosomes" rather than artificial degradation products [1-6]. 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]. However, the global spatial organization of caRNAs has not been systematically characterized in the context of multiscale 3D genome organization. Here, we generate and integrate genome-wide caRNA-DNA [20,23-26] and DNA-DNA [27,28] interaction maps to comprehensively describe the spatial distribution of caRNAs in the nucleus and their role in 3D genome organization.

To map the associated genomic sequence of each caRNA, we used *in situ* Mapping of RNA-Genome Interaction (iMARGI) [17,24] to generate genome-wide RNA-DNA interaction maps. iMARGI captures RNA-genome interactions by proximity ligation followed by reverse transcription of RNA and high-throughput paired-end sequencing [24]. To describe the 3D genome organization, we used *in situ* Hi-C (Hi-C) [27,28] to detect long-range chromatin interactions genome-wide [27,28] and PLAC-seq to reveal long-range chromatin interactions at selected genomic regions [29]. We also used the genome's nuclear compartmentalization states ("SPIN states" [30]) relative to nuclear speckles, nuclear lamina, and putative nucleoli, derived from integrating TSA-seq [31], DamID [32], and Hi-C [30]. Combining these data, we investigate the spatial organization of caRNA and their relationship to multiscale 3D genome features, including A/B compartments [33], topologically associating domains (TADs) [34,35], and chromatin loops [27].

There is yet no consensus on how caRNA is expected to impact long-range chromatin interactions. On the one hand, a high-level of caRNA is thought to suppress chromatin looping [36]. This suppressive impact depends on caRNA's *electrostatic property* (ionic charge), where a high-level of caRNA promotes dissolution of nuclear condensates or liquid droplets, thus releases interacting cis-regulatory sequences [36-40]. On the other hand, caRNA is believed to promote chromatin interactions by retaining chromatin binding proteins including those that promote DNA looping in the physical vicinity of cis-regulatory sequences [22,36,41-43]. This promotive role depends on caRNA's base-pairing and protein-binding abilities, which we refer to as caRNA's *architectural property*. In this work, we report that a genome-wide reduction of caRNA promotes chromatin looping. Further, we reconcile caRNA's seemingly opposite impacts to chromatin interactions by accounting for the relative impacts of caRNA's *electrostatic property* (suppressive impact) and *architectural property* (promotive impact) on different genomic regions. Outside the loop anchors,

caRNA's architectural property outcompetes its *electrostatic property*, making caRNA's overall impact to chromatin interactions promotive (Supplementary Figure S1). On the loop anchors, caRNA's electrostatic property [36] exhibits a greater impact, where a high-level of caRNA suppresses loop formation.

Results

RNA attachment level (RAL) reflects 3D genome spatial compartmentalization

We generated iMARGI data from human embryonic stem (H1), foreskin fibroblast (HFF), and chronic myelogenous leukemia (K562) cells in duplicates and compared them with Hi-C data from the same cell types (Table S1). We derived the A/B compartment score based on the Hi-C contact matrix's first eigenvector [44] (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 [23], purple curve, Figure 1a, Supplementary Figure S2). Overall, Hi-C contact matrix's first eigenvector exhibits a genome-wide correlation with RAL (p-value < 2e-16, one way ANOVA) (Figure 1b, Supplementary Figure S2), suggesting an association of 3D genome spatial compartmentalization and caRNA distribution.

To test if different species of caRNAs have the same distribution on the genome, we separately analyzed the Alu-containing caRNA (Alu-caRNA) and LINE1-containing caRNA (L1-caRNA). Both Alu-caRNA and L1-caRNA's RALs exhibit similar genome-wide distributions as the RAL of all the caRNAs (Supplementary Figure S3), suggesting both trans-interacting Alu-caRNA and L1-caRNA are enriched in the A compartment. The enrichment of trans-interacting L1-caRNA in the A compartment does not correlate with the relative depletion of LINE1 elements in the A compartment [45,46]. However, the ratio of Alu-caRNA's RAL vs. L1-caRNA's RAL exhibits a genome-wide correlation with the A/B compartment score (Supplementary Figure S3). 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.

We next compared caRNA-DNA interactions with the more refined genome compartmentalization relative to various nuclear bodies. We identified 9 nuclear compartmentalization states (SPIN states) [30] (Figure 1a, Supplementary Figure S4). We removed any iMARGI read pair where the RNA and DNA ends were mapped to within 10 Mb on the same chromosome to filter out 99.99%

of the read pairs mapped within any contiguous genomic region of the same SPIN state (Supplementary Figure S4). The RNAs transcribed from 6,623 genes exhibit specificity in targeting the genomic sequences of some SPIN states (BH adjusted p-value < 0.01, Chi-square test, degrees of freedom = 8, Supplementary Figure S5a). To facilitate global comparisons, we projected the RNA-SPIN state matrix (Supplementary Figure S5a) to a two-dimensional space by UMAP [47], where each dot corresponds to an RNA. We found that the UMAP preserves RNA-to-RNA similarities of the nuclear localizations of each RNA's target genomic regions (Figure 1c, d).

Next, we asked if any sequence feature of the caRNA correlates with 3D nuclear localization of their targets. We considered 15 features, including lincRNA (whether this RNA is a lincRNA), gene length, 5' UTR length, 3' UTR length, exon/intron length ratio, the proportion (by length) of LINE/L1, LINE/L2, SINE/Alu, SINE/MIR, LTR/ERVL, LTR/ERVL-MaLR, LTR/ERV1, DNA/hAT, DNA/TcMar, and simple repeats. Both LASSO [48] and Generalized Linear Model (GLM) identified the same four features as the most predictive (Supplementary Figure S3b), namely, exon/intron length ratio (Figure 1c), gene type/lincRNA (Figure 1d), SINE/Alu, and 3' UTR length as associated with the target genomic sequences' SPIN states. Using these four features, GLM is able to predict the target genomic sequences' SPIN states with an average area under the ROC curve (AUC) of 0.854 (5-fold cross validations; Supplementary Figure S5c). The Area under precision-recall curve (AUPR) was 0.588, as compared to a prediction baseline AUPR of 0.216 (Supplementary Figure S5d). We further tested the model trained with H1's data on HFF's and K562's iMARGI data and SPIN states, which yielded an AUC of 0.795 for HFF (Figure 1e) (AUPR = 0.623, baseline AUPR = 0.327, Supplementary Figure S5d) and 0.777 for K562 (Figure 1e) (AUPR = 0.525, baseline AUPR = 0.258, Supplementary Figure S5d). These data suggest that specific sequence features are predictive of the RNA's target genomic sequences' 3D localization in the nucleus. Together, these results reveal genome-wide connections between caRNAs and 3D genome compartmentalization.

TADs demarcate RNA-DNA interaction densities

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 [34,35]. We sought to reveal TADs' role in RNA-DNA interactions genome-wide. We found that most TADs are either enriched or depleted with iMARGI read pairs (Figure 1f, Supplementary Figure S6a). Throughout the genome, the RNAs transcribed from within each TAD are more frequently attached to the genomic regions in the same TAD than to the flanking genomic regions of the TAD (p-value < 2.2e-16, t test) (Figure 1g). Conversely, the RNAs transcribed from the

genomic sequences outside a TAD are more frequently associated with the flanking sequences (p-value = 3e-4, t test) (Supplementary Figure S6b). These data suggest that TADs are self-interacting genomic regions of caRNA-DNA interactions.

caRNA-DNA contacts are stronger inside chromatin loops

Next, we asked whether RNA-DNA contacts more often appear within loops [27] than other parts of the genome. We used the Hi-C density ratio (HDR) to quantify the enrichment of Hi-C read pairs inside each loop as compared to this loop's nearby genomic regions (Methods, Supplementary Figure S7a) and iMARGI density ratio (MDR) to quantify the enrichment of iMARGI read pairs inside each loop (Methods, Supplementary Figure S7b). A total of 2,473 chromatin loops were identified based on H1's Hi-C data (a total of 616 million read pairs in the untreated H1 cells, Table S1b). MDRs were larger in the loops than in the randomly shuffled genomic regions that matched the loop number and loop size distribution (p-value < 1e-5, Wilcoxon test), suggesting a genome-wide tendency of more caRNA-DNA interactions within loops than other genomic regions. For example, the two anchors of a loop near the MEGF6 gene demarcated extensive caRNA-DNA interactions within this loop (Figure 2a). We repeated this analysis with a micro-C dataset generated from the H1 cells that has a deeper sequencing depth (3.22 billion read pairs in total, Accession ID: 4DNES21D8SP8) [49] than our Hi-C dataset, which yielded 30,220 chromatin loops, and the MDRs remained larger in the loops than randomly shuffled genomic regions with matched the loop number and loop sizes (p-value < 2.2e-16, Wilcoxon test).

Among the loops, the high-HDR loops were associated with the high-MDR loops (442 high-HDR loops vs. 239 high-MDR loops, p-value = 3e-11, Chi-square test) (Supplementary Figure S7c). As an alternative approach to the HDR analysis, we called loop domains (the loops with strong within-loop Hi-C contacts) as previously described [50]. Consistently, the identified loop domains were also enriched with high-MDR loops (620 loop domains vs. 442 high-MDR loops, p-value = 5.5e-4, Chi-square test) (Supplementary Figure S7c). Taken together, these results demonstrate that a chromatin loop with strong within-loop DNA-DNA interactions tend to exhibit strong within-loop RNA-DNA contacts.

caRNA domains: chromatin domains demarcated by caRNAs

Inspired by the discovery of TADs from Hi-C's contact matrix [34], we asked whether there is any recurring pattern in iMARGI's contact matrix, where the rows represent iMARGI's read pairs' RNA ends and the columns represent the corresponding DNA ends [24] (Figure 2a, b, Supplementary Figure S8a). A notable difference to Hi-C's symmetric contact matrix is that iMARGI's contact

matrix is asymmetric. This is because a caRNA transcribed from location A interacting with the DNA in location B, denoted in the entry of row A and column B in iMARGI's contact matrix, does not always have a reciprocal interaction, denoted in the entry of row B and column A.

Rectangular blocks of high-value entries emerged as a recurring pattern from iMARGI's contact matrix (Figure 2a, b). We identified the rectangular blocks by using HOMER to call peaks on the rows of the contact matrix (row peaks), and each 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, 2019, and 2,468 rectangular blocks from H1, HFF, and K562's iMARGI data (Supplementary Figure S8b). 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 this caRNA's gene. These rectangular blocks' heights (tens to hundreds kilobases) that correspond to the caRNAs are usually smaller than their widths (several hundred kilobases to several megabases) that correspond to the caRNA's target genomic regions (Supplementary Figure S8a, c), suggesting a caRNA usually interacts with its own gene and its own gene's flanking genomic regions. 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 rectangles' heights are similar to the lengths of the longest genes overlapping with each rectangle's y coordinates (Supplementary Figure S8e, f, g), suggesting that most caRNA domains are decorated by the caRNAs of single genes. Together, we identify caRNA domains as a unique characteristic of genome-wide caRNA-DNA associations.

Co-localization of caRNA domains and DNA loops

We then asked whether caRNA domains overlap with chromatin loops. We call a caRNA domain colocalized with a DNA loop if this caRNA domain overlaps with at least one anchor of this loop. Of the H1 cells' 2,473 DNA loops, 1,067 (43.1%) overlapped with caRNA domains (3,217 caRNA domains in H1 cells) (p-value < 9.3e-12, Chi-square test), suggesting a genome-wide colocalization of caRNA domains and chromatin loops.

Among the 1,067 loops that colocalized with caRNA domains, 600 (56.2%) involved a promoter as a loop anchor (p-value=7.2e-4, degree of freedom=1, Chi-square test), suggesting a genome-wide association of promoter-participating loops with caRNA domains. For example, the BLCAP's promoter formed loops with 3 downstream anchors, and the transcripts originated from the

genomic sequence near BLCAP's promoter spread across approximately 300 kb genomic regions to reach the three downstream anchors (Figure 2b).

Then, we identified each promoter's interacting sequences (PINS) from PLAC-seq (Table S1) and categorized promoters into two groups, namely those with and without PINS. The caRNA-DNA interactions surrounding the promoters with PINS were stronger than those surrounding the promoters without PINS (Figure 2c). We also compared PINS' positions (upstream or downstream to TSS) with the caRNA domains' positions. The ratio of the numbers of PINSs that are upstream and downstream to a TSS (PINS asymmetric ratio) correlated with the ratio of the upstream and the downstream length (with respect to the TSS) of the caRNA domain overlapping with this promoter (RNA asymmetric ratio) (SCC=0.85, p-value < 1e-32, permutation test) (Figure 2d). Thus, if a promoter has more upstream PINSs than downstream PINSs, the caRNA domain overlapping with this promoter tends to extend further to the upstream direction, and vice versa. Together, caRNA domains tend to overlap with chromatin loops including regulatory loops.

Transcription arrest and RNase A treatment impact intra-chromosomal contact frequency

We treated H1 cells with ammonium acetate (NH4OAc) to disrupt RNA's electrostatic interactions (the interactions due to electric charges) [51-53], Flavopiridol (FL) to suppress transcription elongation [54-56], and RNase A to reduce caRNA [39,57] based on established protocols (NH4OAc [52], FL [54], RNase A [39]) (Supplementary Figure S9). NH4OAc disrupts RNA's electrostatic interactions in living cells by providing monovalent cations without perturbing intracellular pH [52]. To check the expected effects of the three treatments, we immunostained nuclear speckle-associated proteins SON [58] and SC35 [59] 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 S10), consistent with the role of RNA's electrostatic interactions in maintaining nuclear speckles [52]. Conversely, FL made SON and SC35 foci larger and more distinct [60] (Figure 3c, Supplementary Figure S10). 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 S10), consistent with the observations that "low RNA/protein ratios promote phase separation into liquid droplets" [40] and condensate formation [36].

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 S8g), consistent with FL's inhibitory effect on transcription elongation [54]. 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 S8h) and sizes (widths of the rectangular blocks) (p-value < 5e-210, Wilcoxon rank sum test) (Supplementary Figure S8i).

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 impacts to 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 [33] (Figure 3f, g, h). 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) (Figure 3f, g, h), consistent with the idea that transcription weakens chromatin contacts within gene bodies [38]. 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) (Figure 3f, g, h). These results indicate both the transcription process and this process' RNA products can impact chromatin contacts, but their impacts are not identical.

TADs and A/B compartments remain mostly unchanged after perturbations of caRNA.

Based on the "Measure of Concordance (MoC)", a summary statistic on the genome-wide differences of TAD numbers and positions between experimental groups [61], the genome-wide TAD numbers and positions of all four conditions were categorized as "highly concordant" (pairwise MoCs=0.93, 0.90, 0.82, which were all above the threshold 0.75 for being "highly concordant" [61]). Slightly more TADs were detected in FL and less TADs were detected in RNase (Supplementary Figure S6c), consistent with increase and decrease of the proportions of intra-chromosomal and sub-megabase-distance genomic interactions in FL and RNase (Figure 3e, g).

Hi-C contact matrix's first eigenvector in each treatment was strongly correlated with that in control (Figure 3i, j, k, l), suggesting small impacts of FL, RNase, and NH4OAc to the genome's A/B compartments, if any. FL and RNase attenuated the resemblance of RAL to A/B compartments (SCC between RAL and Hi-C's first eigenvector=0.63 in control, 0.55 in FL, 0.52 in RNase) (Figure 1b, Figure 3n, o). NH4OAc further reduced this resemblance (SCC=0.28) (Figure 3i, m). For example, Malat1's RAL (calculated from the subset of iMARGI's read pairs where the RNA reads were mapped to Malat1) correlated with Hi-C's first eigenvector in control (SCC = 0.61,

Supplementary Figure S11a), however this correlation was diminished in NH4OAc (SCC = -0.01, Supplementary Figure S11b). These data suggest that electrostatic interactions are required for RAL to reflect the genome's A/B compartments.

FL and RNase treatment increases chromatin loop numbers and loop strengths.

We called chromatin loops from our Hi-C data in each of the four conditions that have comparable sequencing depths (Table S1) using HiCCUPS (Methods). 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 4a). These loop number differences cannot be attributed to different sequencing depths or batch effect 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 4a). For example, a loop linking ATF7 and KRT18 genes that was absent in control and NH4OAc emerged in both FL and RNase (arrows, Figure 4b, c). For another example, a loop linking the two ends of ZMYND8 gene in control was enriched with caRNA-DNA interactions (Arrow, Figure 4g). NH4OAc rewired this loop into a larger loop by shifting one of the two anchors (Arrow, Figure 4h). FL retained the loop in control (Arrow) and induced two additional loops (Figure 4i). RNase also induced two loops (Figure 4j), including a loop coinciding with a FL-induced loop (Arrowhead, Figure 4j). The other two loops in FL shrank into smaller sizes in RNase (Arrows 1 and 2, Figure 4i, j).

The loop size (distance between the two loop anchors) distributions were similar in control and NH4OAc (p-value = 0.93) and were increased in FL (p-value = 0.005) and RNase (p-value = 3.3e-5, Wilcoxon test) (Figure 4d). 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) and Z-score Lower Left (ZscoreLL) scores [27] (Supplementary Figure S12a, b). 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 4e, f, Supplementary Figure S12c). 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 calling (Supplemental Text: Accounting for experimental method

and sequencing depth variations). Taken together, these results suggest caRNA's indispensable role in regulating the genome's 3D structure, especially chromatin looping.

Discussion

Our data revealed pronounced correlations of caRNA-DNA interactions and the multiscale 3D genome organization [62,63]. "caRNA domains" emerged from the caRNA-DNA contact map as the most prominent pattern. Each caRNA domain represents a caRNA-enriched chromosomal domain, where the enriched caRNAs are transcribed from one or several genes inside this domain. The presence and the relative position of caRNA domains correlate with the existence of promoter-enhancer loops and the enhancer's relative positions.

These correlations begged the question of whether there is any dependence between the genome's 3D structure and the caRNA? 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 (RNA-independence model). On the other hand, caRNA demarcates active and repressive chromosomal domains [19,21,45,67,68], and specific lncRNAs or chromatin-bound RNA binding proteins [15,69,70] can assist chromatin folding, indicating an involvement of caRNA in spatial organization the genome (RNA-dependence model).

In line with the RNA-dependence model, both FL and RNase treatments increased the number of chromatin loops and the strengths of pre-existing loops (figure 4a, e). Most FL- and RNase-induced loops overlapped (Figure 4a), suggesting FL and RNase invoked a shared loop formation mechanism. This shared mechanism likely involves caRNA rather than the transcriptional process' polymerase reaction because both FL and RNase reduces caRNA and only FL stops the polymerase reaction.

Previous data revealed two seemingly opposite impacts of caRNA to chromatin looping. Removal [15] or mutation of the RNA-binding domain [43,71] of chromatin-bound RNA binding proteins (RBP) suppressed chromatin interactions, suggesting caRNAs promote chromatin contacts. However, high levels of caRNAs dissolve chromatin-participating condensates and thus release the chromatin loops [36], suggesting a negative impact of high-level caRNA to chromatin looping. We speculated that these two seemingly opposite impacts can be explained by differentiating RNA's architectural and electrostatic properties. The RBP knockdown and mutation experiments [15,43,71]

disrupted protein-RNA binding, which we loosely categorize as a part of caRNA's *architectural impact*. On the other hand, the dissolution of chromatin-participating condensates by high-level of caRNA [36] depends on a principle of RNA-mediated phase separation, that is high and low RNA/protein ratios lead towards dissolution and formation of phase separated droplets or condensates, respectively [40]. We call these effects caRNA's *electrostatic impact*.

Our questions are, first, which impact dominates the genome-wide distance distribution of Hi-C contacts, and second, which impact dominates the Hi-C detected chromatin loops? NH4OAc, which disrupts RNA's electrostatic interactions without perturbing intracellular pH [51-53], did not change the distance distribution of Hi-C contacts (Figure 3f, h), suggesting negligible *electrostatic impact* on the distance distribution. Thus, if caRNA can affect Hi-C contacts' distance distribution, we expect the *architectural impact* to dominate. Therefore, the predicted direction of change would be a reduction of the proportion of sub-megabase-distance Hi-C contacts in RNase, which is indeed the case (Figure 3g). As a control, transcription arrest (FL) increased the proportion of sub-megabase-distance Hi-C contacts as expected [38] (Figure 3g).

Both FL and RNase reduced caRNA as expected (RAL tracks, Figure 3i) and increased Hi-C detected chromatin loops (Figure 4a). This direction of change suggests that caRNA's *electrostatic impact* dominates the loops detected by Hi-C. If the emerged chromatin loops in FL and RNase were indeed mediated by caRNA's electrostatic interactions, we would expect that NH4OAc, which also reduces caRNA genome-wide (RAL tracks, Figure 3i), did not increase Hi-C detected chromatin loops, because NH4OAc disrupts electrostatic interactions [53]. This is indeed the case (Figure 4a). Furthermore, NH4OAc did not increase the strengths of Hi-C detected loops either (Figure 4e; Supplemental Text: Accounting for experimental method and sequencing depth variations).

Taken together, our data reveal the relative importance of the transcription process, caRNA's *architectural impact* and *electrostatic impact* to chromatin interactions. Hi-C contacts' distance distribution is modulated by a balance of the transcription process that reduces and caRNA's *architectural impact* that enhances sub-megabase-distance Hi-C interactions. caRNA's *electrostatic impact* has the largest effect on Hi-C derived loops. Reducing caRNA without depleting them or disrupting their electrostatic interactions lead to more Hi-C loops. Finally, the union of chromatin loops in control, FL, and RNase provide a genome-wide "backbone" map of chromatin loops, including those loops in the current cell type and the locations where chromatin loops can emerge.

Acknowledgements

This work is funded by NIH grants DP1DK126138, R01GM138852, DP1HD087990, and NIH Common Fund 4D Nucleome grants UM1HG011593, UM1HG011585, U54DK107977, and U01CA200147.

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.

References

1 Holmes, D. S., Mayfield, J. E., Sander, G. & Bonner, J. Chromosomal RNA: its properties. *Science* **177**, 72-74, doi:10.1126/science.177.4043.72 (1972).

2 Rodriguez-Campos, A. & Azorin, F. RNA is an integral component of chromatin that contributes to its structural organization. *PLoS One* **2**, e1182, doi:10.1371/journal.pone.0001182 (2007).

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, doi:10.1016/j.gde.2016.04.005 (2016).

4 Nozawa, R. S. & Gilbert, N. RNA: Nuclear Glue for Folding the Genome. *Trends Cell Biol* **29**, 201-211, doi:10.1016/j.tcb.2018.12.003 (2019).

5 Nozawa, R. S. *et al.* SAF-A Regulates Interphase Chromosome Structure through Oligomerization with Chromatin-Associated RNAs. *Cell* **169**, 1214-1227 e1218, doi:10.1016/j.cell.2017.05.029 (2017).

6 Pederson, T. Half a century of locating DNA and RNA in cells. *FASEB J* **33**, 8693-8694, doi:10.1096/fj.190801ufm (2019).

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

8 Place, R. F., Li, L. C., Pookot, D., Noonan, E. J. & Dahiya, R. MicroRNA-373 induces expression of genes with complementary promoter sequences. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 1608-1613, doi:10.1073/pnas.0707594105 (2008).

9 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, doi:10.1126/science.1101372 (2004).
10 Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-137, doi:10.1038/379131a0 (1996).
11 Yang, F. *et al.* The lncRNA Firre anchors the inactive X chromosome to the nucleolus by binding CTCF and maintains H3K27me3 methylation. *Genome biology* **16**, 52, doi:10.1186/s13059-015-0618-0 (2015).
12 Calandrelli, R. *et al.* Stress-induced RNA-chromatin interactions promote endothelial dysfunction. *Nat Commun* **11**, 5211, doi:10.1038/s41467-020-18957-w (2020).
13 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, doi:10.1126/science.1203919 (2011).
14 Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323, doi:10.1016/j.cell.2007.05.022 (2007).
15 Xiao, R. *et al.* Pervasive Chromatin-RNA Binding Protein Interactions Enable RNA-Based Regulation of Transcription. *Cell* **178**, 107-121 e118, doi:10.1016/j.cell.2019.06.001 (2019).
16 Dumelie, J. G. & Jaffrey, S. R. Defining the location of promoter-associated R-loops at near-nucleotide resolution using bisDRIP-seq. *Elife* **6**, doi:10.7554/eLife.28306 (2017).
17 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, doi:10.1073/pnas.1819788116 (2019).
18 Yin, Y. *et al.* U1 snRNP regulates chromatin retention of noncoding RNAs. *Nature* **580**, 147-150, doi:10.1038/s41586-020-2105-3 (2020).
19 Percharde, M. *et al.* A LINE1-Nucleolin Partnership Regulates Early Development and ESC Identity. *Cell* **174**, 391-405 e319, doi:10.1016/j.cell.2018.05.043 (2018).
20 Li, X. *et al.* GRID-seq reveals the global RNA-chromatin interactome. *Nat Biotechnol* **35**, 940-950, doi:10.1038/nbt.3968 (2017).
21 Chen, W. *et al.* RNAs as Proximity-Labeling Media for Identifying Nuclear Speckle Positions Relative to the Genome. *iScience* **4**, 204-215, doi:10.1016/j.isci.2018.06.005 (2018).
22 Quinodoz, S. A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *bioRxiv*, 2020.2008.2025.267435, doi:10.1101/2020.08.25.267435 (2020).
23 Sridhar, B. *et al.* Systematic Mapping of RNA-Chromatin Interactions In Vivo. *Curr Biol* **27**, 610-612, doi:10.1016/j.cub.2017.01.068 (2017).
24 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).
25 Bell, J. C. *et al.* Chromatin-associated RNA sequencing (ChAR-seq) maps genome-wide RNA-to-DNA contacts. *Elife* **7**, doi:10.7554/eLife.27024 (2018).
26 Bonetti, A. *et al.* RADICL-seq identifies general and cell type-specific principles of genome-wide RNA-chromatin interactions. *Nat Commun* **11**, 1018, doi:10.1038/s41467-020-14337-6 (2020).
27 Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-1680, doi:10.1016/j.cell.2014.11.021 (2014).
28 Oksuz, B. A. *et al.* Systematic evaluation of chromosome conformation capture assays. *bioRxiv*, 2020.2012.2026.424448, doi:10.1101/2020.12.26.424448 (2020).
29 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).
30 Wang, Y. *et al.* SPIN reveals genome-wide landscape of nuclear compartmentalization. *Genome Biol* **22**, 36, doi:10.1186/s13059-020-02253-3 (2021).
31 Chen, Y. *et al.* Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler. *J Cell Biol* **217**, 4025-4048, doi:10.1083/jcb.201807108 (2018).

32 Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protoc* **2**, 1467-1478, doi:10.1038/nprot.2007.148 (2007).
33 Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293, doi:10.1126/science.1181369 (2009).
34 Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-380, doi:10.1038/nature11082 (2012).
35 Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-385, doi:10.1038/nature11049 (2012).
36 Henninger, J. E. *et al.* RNA-Mediated Feedback Control of Transcriptional Condensates. *Cell* **184**, 207-225 e224, doi:10.1016/j.cell.2020.11.030 (2021).
37 Rhine, K., Vidaurre, V. & Myong, S. RNA Droplets. *Annu Rev Biophys* **49**, 247-265, doi:10.1146/annurev-biophys-052118-115508 (2020).
38 Heinz, S. *et al.* Transcription Elongation Can Affect Genome 3D Structure. *Cell* **174**, 1522-1536 e1522, doi:10.1016/j.cell.2018.07.047 (2018).
39 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).
40 Maharana, S. *et al.* RNA buffers the phase separation behavior of prion-like RNA binding proteins. *Science* **360**, 918-921, doi:10.1126/science.aar7366 (2018).
41 Sigova, A. A. *et al.* Transcription factor trapping by RNA in gene regulatory elements. *Science* **350**, 978-981, doi:10.1126/science.aad3346 (2015).
42 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, doi:10.1038/s41589-019-0422-3 (2020).
43 Saldana-Meyer, R. *et al.* RNA Interactions Are Essential for CTCF-Mediated Genome Organization. *Mol Cell* **76**, 412-422 e415, doi:10.1016/j.molcel.2019.08.015 (2019).
44 Mourad, R. & Cuvier, O. Predicting the spatial organization of chromosomes using epigenetic data. *Genome Biol* **16**, 182, doi:10.1186/s13059-015-0752-8 (2015).
45 Lu, J. Y. *et al.* Homotypic clustering of L1 and B1/Alu repeats compartmentalizes the 3D genome. *Cell Res*, doi:10.1038/s41422-020-00466-6 (2021).
46 Lu, J. Y. *et al.* Genomic Repeats Categorize Genes with Distinct Functions for Orchestrated Regulation. *Cell Rep* **30**, 3296-3311 e3295, doi:10.1016/j.celrep.2020.02.048 (2020).
47 Vermeulen, M., Smith, K., Eremin, K., Rayner, G. & Walton, M. Application of Uniform Manifold Approximation and Projection (UMAP) in spectral imaging of artworks. *Spectrochim Acta A Mol Biomol Spectrosc* **252**, 119547, doi:10.1016/j.saa.2021.119547 (2021).
48 Tibshirani, R. Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society. Series B (Methodological)* **58**, 267-288 (1996).
49 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).
50 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324, doi:10.1016/j.cell.2017.09.026 (2017).
51 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, doi:10.1038/nchem.2414 (2016).
52 Saha, S. & Hyman, A. A. RNA gets in phase. *J Cell Biol* **216**, 2235-2237, doi:10.1083/jcb.201706034 (2017).
53 Jain, A. & Vale, R. D. RNA phase transitions in repeat expansion disorders. *Nature* **546**, 243-247, doi:10.1038/nature22386 (2017).

54 Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432-445, doi:10.1016/j.cell.2010.03.030 (2010).
55 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, doi:10.1074/jbc.M102306200 (2001).
56 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, doi:10.1016/j.cell.2014.08.018 (2014).
57 Thakur, J., Fang, H., Llagas, T., Disteche, C. M. & Henikoff, S. Architectural RNA is required for heterochromatin organization. *bioRxiv*, 784835, doi:10.1101/784835 (2019).
58 Ilik, I. A. *et al.* SON and SRRM2 are essential for nuclear speckle formation. *Elife* **9**, doi:10.7554/eLife.60579 (2020).
59 Fu, X. D. & Maniatis, T. Factor required for mammalian spliceosome assembly is localized to discrete regions in the nucleus. *Nature* **343**, 437-441, doi:10.1038/343437a0 (1990).
60 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**, doi:10.1242/jcs.226563 (2019).
61 Zufferey, M., Tavernari, D., Oricchio, E. & Ciriello, G. Comparison of computational methods for the identification of topologically associating domains. *Genome Biol* **19**, 217, doi:10.1186/s13059-018-1596-9 (2018).
62 Pederson, T. Genome architecture and expression 2019-2020: the transition phase. *Curr Opin Genet Dev* **67**, 1-4, doi:10.1016/j.gde.2020.09.003 (2021).
63 Dekker, J. *et al.* The 4D nucleome project. *Nature* **549**, 219-226, doi:10.1038/nature23884 (2017).
64 Bianco, S. *et al.* Polymer physics predicts the effects of structural variants on chromatin architecture. *Nat Genet* **50**, 662-667, doi:10.1038/s41588-018-0098-8 (2018).
65 Fudenberg, G., Kelley, D. R. & Pollard, K. S. Predicting 3D genome folding from DNA sequence with Akita. *Nat Methods* **17**, 1111-1117, doi:10.1038/s41592-020-0958-x (2020).
66 Kim, E., Kerssemakers, J., Shaltiel, I. A., Haering, C. H. & Dekker, C. DNA-loop extruding condensin complexes can traverse one another. *Nature* **579**, 438-442, doi:10.1038/s41586-020-2067-5 (2020).
67 Nozawa, R. S. & Gilbert, N. Interphase chromatin LINEd with RNA. *Cell* **156**, 864-865, doi:10.1016/j.cell.2014.02.005 (2014).
68 Hall, L. L. *et al.* Stable C0T-1 repeat RNA is abundant and is associated with euchromatic interphase chromosomes. *Cell* **156**, 907-919, doi:10.1016/j.cell.2014.01.042 (2014).
69 Lee, J. T. Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nat Rev Mol Cell Biol* **12**, 815-826, doi:10.1038/nrm3231 (2011).
70 Li, X. & Fu, X. D. Chromatin-associated RNAs as facilitators of functional genomic interactions. *Nat Rev Genet* **20**, 503-519, doi:10.1038/s41576-019-0135-1 (2019).
71 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, doi:10.1016/j.molcel.2019.07.039 (2019).

Figure legends

Figure 1. Comparison of RNA-DNA interactions and genomic interactions. All plots are based on H1 cells unless otherwise specified. (a) 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, juxtaposed with SPIN states (SPIN, color bars) derived from Hi-C, TSA-seq, and DamID data. (b) Scatterplot of Hi-C's first eigenvector (y axis) and RAL (x axis) on every 500 kb genomic bin (dot) of the entire genome. (c, d) UMAP projection of the RNA-SPIN state contact matrix, where each dot corresponds to an RNA and a row in panel c. Each RNA gene's exon/intron ratio is marked from high (dark) to low (c), or whether an RNA is a lincRNA (black) or not (grey) (d). (e) ROC curves of a cross-validation in H1 (red) and external validations in HFF (blue) and K562 (green). Values: AUCs. (f) iMARGI's contact matrix from RNA (rows) to DNA (columns) of the corresponding genomic region. Boxes: TADs called from Hi-C, including caRNA-enriched (blue) and caRNA-depleted TADs (green boxes). (g) The RAL calculated from only the RNAs transcribed within each TAD (row), plotted across this TAD (center block) and its equal-length flanking regions on both sides. Blue curve at the bottom: the average RAL of all TADs (rows).

Figure 2. RNA stripe. (a) A loop (blue dots) superimposed on iMARGI's RNA(row)-DNA(column) contact matrix. Bottom: arc plots of iMARGI read pairs and loop. (b) Three loops (colored dots) superimposed on RNA-DNA contact matrix. Middle: arc plots of iMARGI read pairs, loops, H3K4me3 PLAC-seq read pairs. Bottom: H3K27ac, H3K4me3, H3K4me1, CTCF, RAD21, and YY1 ChIP-seq tracks. (c) RNA-chromatin interaction intensities (y axis) surrounding the promoters with (blue) and without PINS (yellow). (d) Scatterplot the relative positions of PINS (log PINS asymmetric ratio, x axis) and RNA stripes (log RNA asymmetric ratio, y axis) with respect to the TSS for every gene (dot).

Figure 3. RNA perturbations. (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-h) Distance distribution of Hi-C's intra-chromosomal read pairs in each condition. (g,h) Expanded views of the distance distribution (f) in the range of 0-2 Mb (g) and ≥ 2 Mb (h). (i) A/B compartment scores (red/green) based on Hi-C's first eigenvector (A/B) and RAL (purple curve) in each condition. (j-l) Scatterplots of A/B compartment scores based on Hi-C's first eigenvector (A/B) in NH4OAc (y axis) (j), FL (k), and RNase (l) vs. in control (x axis). Each dot is a 500 kb bin. (m-o) Scatterplots of A/B compartment scores (A/B) vs. RAL in NH4OAc (m), FL (n), and RNase (o).

Figure 4. Loop comparisons. (a) Upset plot of loop numbers of the four conditions (rows). (b) Hi-C contact matrix of every replicate (row). (c) iMARGI contract matrix of each condition for the corresponding genomic region of (b). Dots: loops. Arrows in panels b & c: a shared loop in FL and RNase that is absent in control and NH4OAc. (D) Loop size distribution in each condition (column). **: p-value < 5.0e-3, ****: p-value < 3.3e-5. (e) Loop strength (P2M) distribution in each condition (column). ****: p-value < 2.2e-16. (f) A ternary plot for each loop (dot)'s P2M change in each treatment as compared to control. A loop with equal P2M in a treatment and in the control corresponds to a point at the center (50% mark) of this treatment's axis. (g-j) iMARGI contract matrix and loops (blue dots) in each condition.

Figure 1

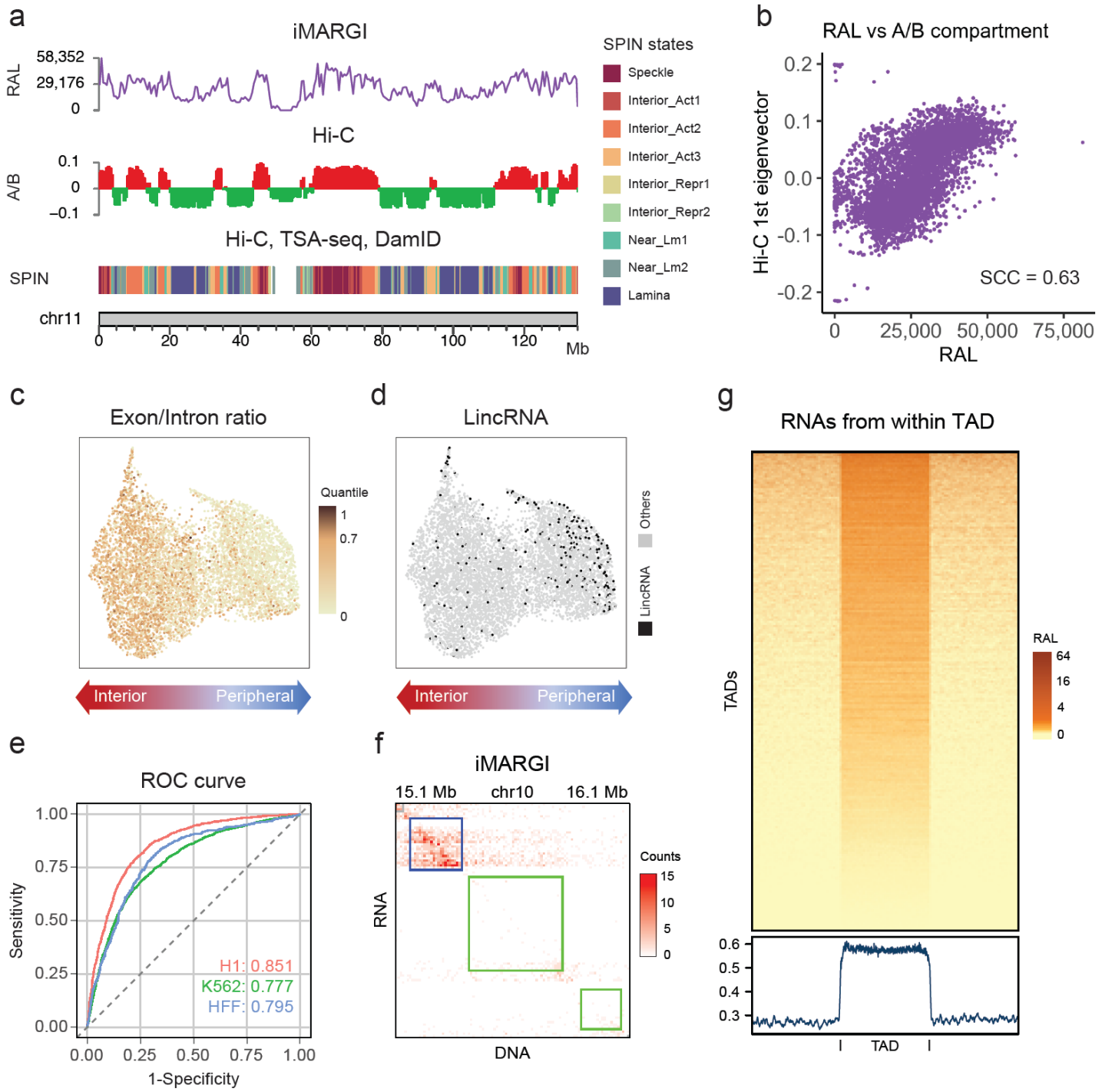

a
iMARGI
RAL
58,352
29,176
0
Hi-C
A/B
0.1
0
−0.1
Hi-C, TSA-seq, DamID
SPIN
chr11
0
20
40
60
80
100
120
Mb
SPIN states
Speckle
Interior_Act1
Interior_Act2
Interior_Act3
Interior_Repr1
Interior_Repr2
Near_Lm1
Near_Lm2
Lamina
b
RAL vs A/B compartment
Hi-C 1st eigenvector
0.2
0.1
0.0
-0.1
-0.2
SCC = 0.63
0
25,000
50,000
75,000
RAL
c
Exon/Intron ratio
Quantile
1
0.7
0
Interior
Peripheral
d
LincRNA
Others
LincRNA
Interior
Peripheral
e
ROC curve
Sensitivity
1.00
0.75
0.50
0.25
0.00
H1: 0.851
K562: 0.777
HFF: 0.795
0.00
0.25
0.50
0.75
1.00
1-Specificity
f
iMARGI
15.1 Mb
chr10
16.1 Mb
RNA
DNA
Counts
15
10
5
0
g
RNAs from within TAD
TADs
RAL
64
16
4
0
0.6
0.5
0.4
0.3
I
TAD
I

Figure 2

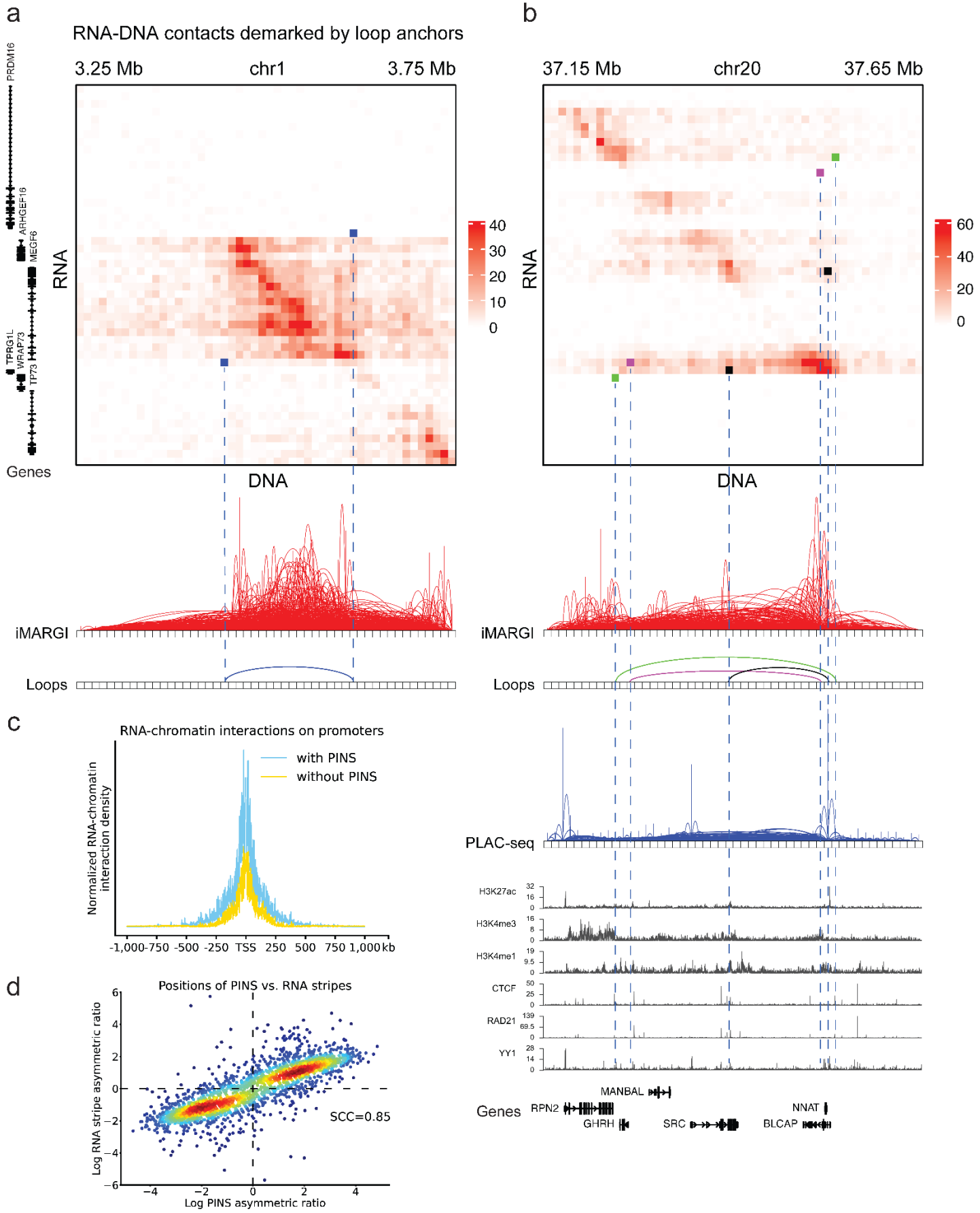

a
RNA-DNA contacts demarked by loop anchors
3.25 Mb
chr1
3.75 Mb
PRDM16
ARHGEF16
MEGF6
TPRG1L
WRAP73
TP73
Genes
RNA
DNA
40
30
20
10
0
iMARGI
Loops
b
37.15 Mb
chr20
37.65 Mb
RNA
DNA
60
40
20
0
iMARGI
Loops
c
RNA-chromatin interactions on promoters
with PINS
without PINS
Normalized RNA-chromatin interaction density
-1,000 -750 -500 -250 TSS 250 500 750 1,000 kb
PLAC-seq
H3K27ac
H3K4me3
H3K4me1
CTCF
RAD21
YY1
MANBAL
Genes
RPN2
GHRH
SRC
NNAT
BLCAP
d
Positions of PINS vs. RNA stripes
Log RNA stripe asymmetric ratio
Log PINS asymmetric ratio
SCC=0.85

Figure 3

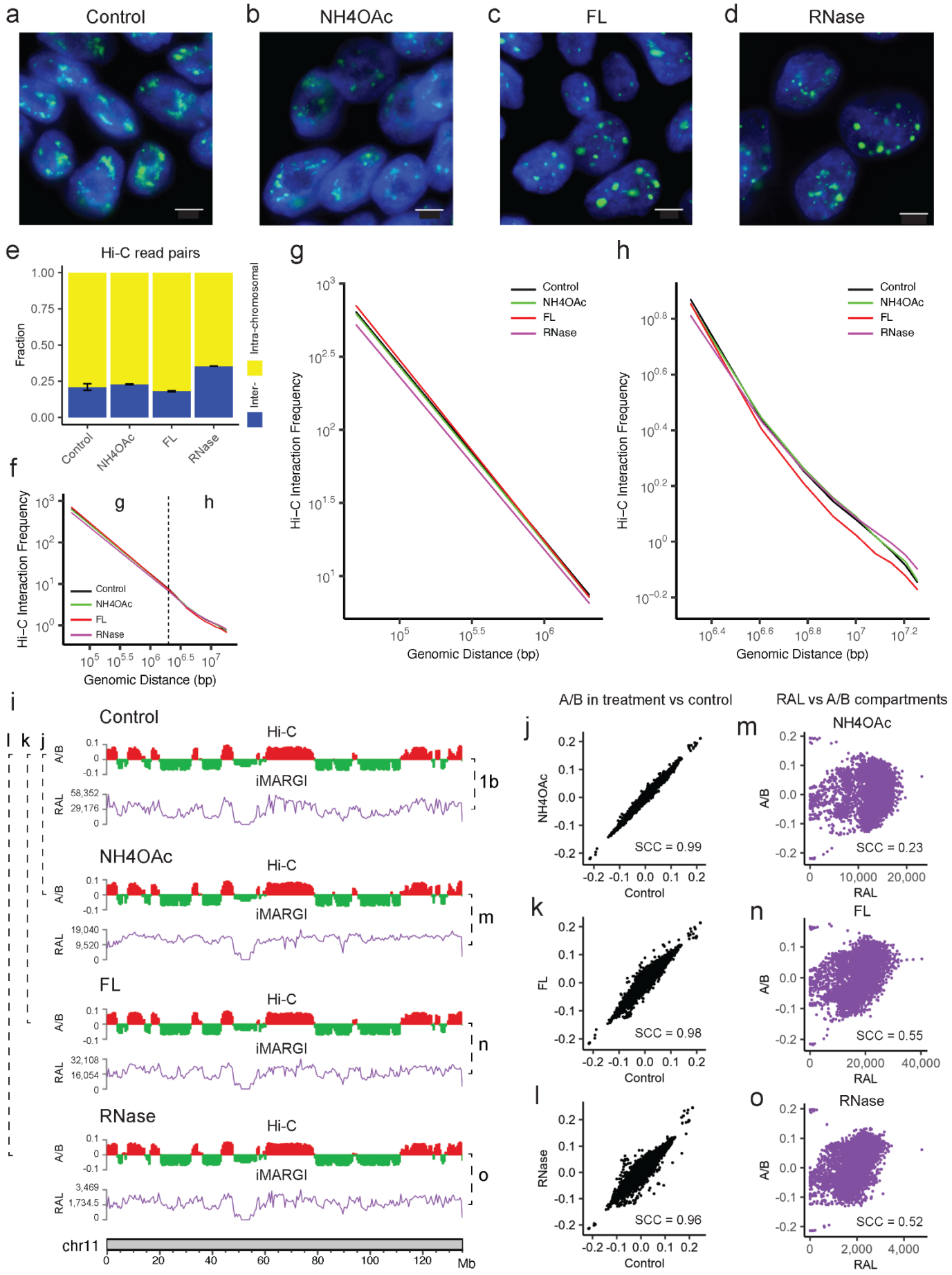
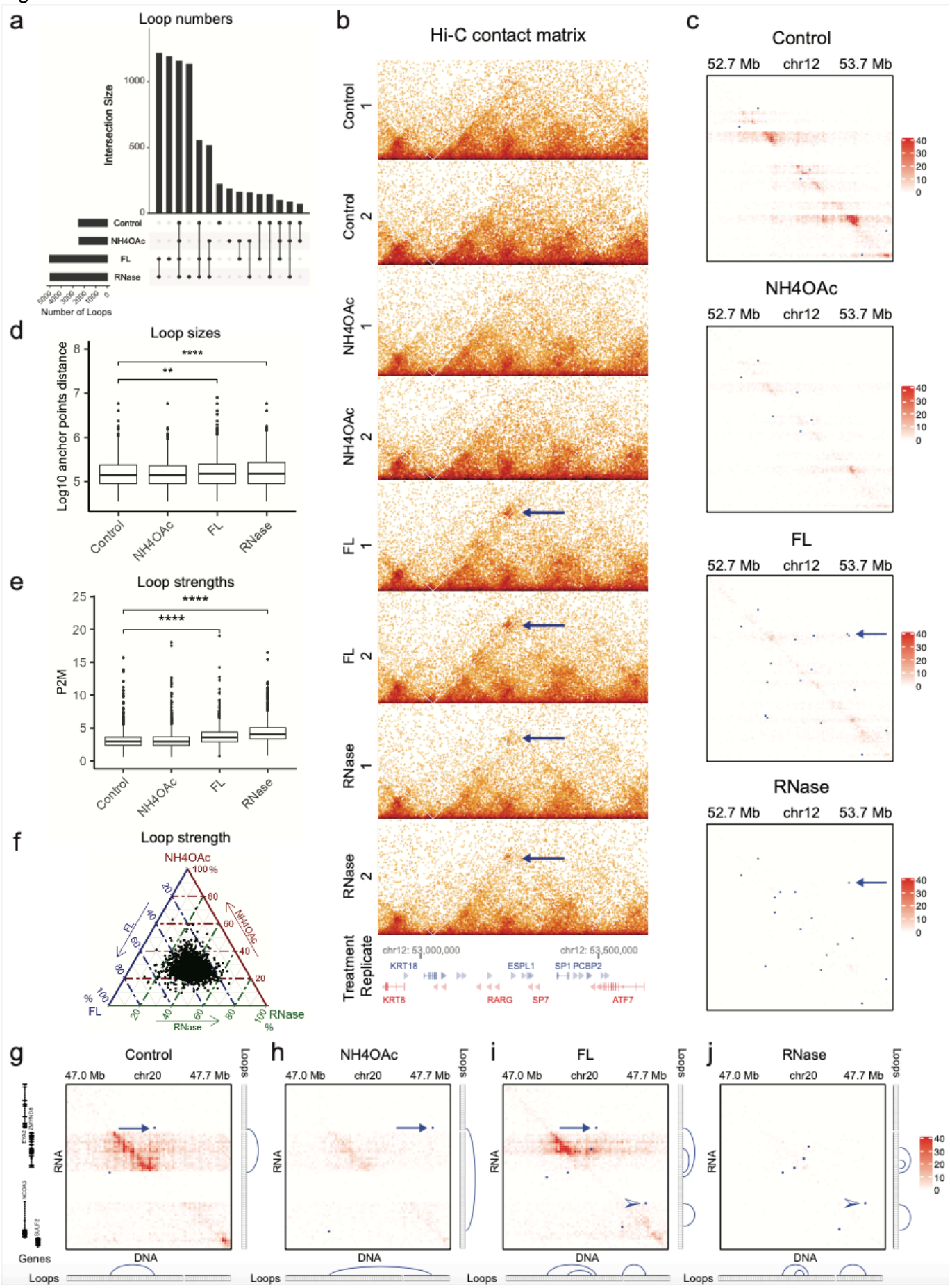

a
Control
b
NH4OAc
c
FL
d
RNase
e
Hi-C read pairs
Fraction
Intra-chromosomal
Inter-
Control
NH4OAc
FL
RNase
f
Hi-C Interaction Frequency
Genomic Distance (bp)
g
h
Control
NH4OAc
FL
RNase
i
Control
NH4OAc
FL
RNase
Hi-C
iMARGI
A/B
RAL
1b
m
n
o
chr11
Mb
A/B in treatment vs control
RAL vs A/B compartments
j
k
l
m
n
o
SCC = 0.99
SCC = 0.98
SCC = 0.96
SCC = 0.23
SCC = 0.55
SCC = 0.52
Control
RAL

Figure 4


a
Loop numbers
Intersection Size
Control
NH4OAc
FL
RNase
Number of Loops
b
Hi-C contact matrix
Control 1
Control 2
NH4OAc 1
NH4OAc 2
FL 1
FL 2
RNase 1
RNase 2
Treatment Replicate
chr12: 53,000,000
chr12: 53,500,000
KRT18
ESPL1
SP1 PCBP2
KRT8
RARG
SP7
ATF7
c
Control
NH4OAc
FL
RNase
52.7 Mb
chr12
53.7 Mb
d
Loop sizes
Log10 anchor points distance
Control
NH4OAc
FL
RNase
e
Loop strengths
P2M
f
Loop strength
NH4OAc
FL
RNase
g
Control
h
NH4OAc
i
FL
j
RNase
47.0 Mb
chr20
47.7 Mb
RNA
DNA
Genes
Loops

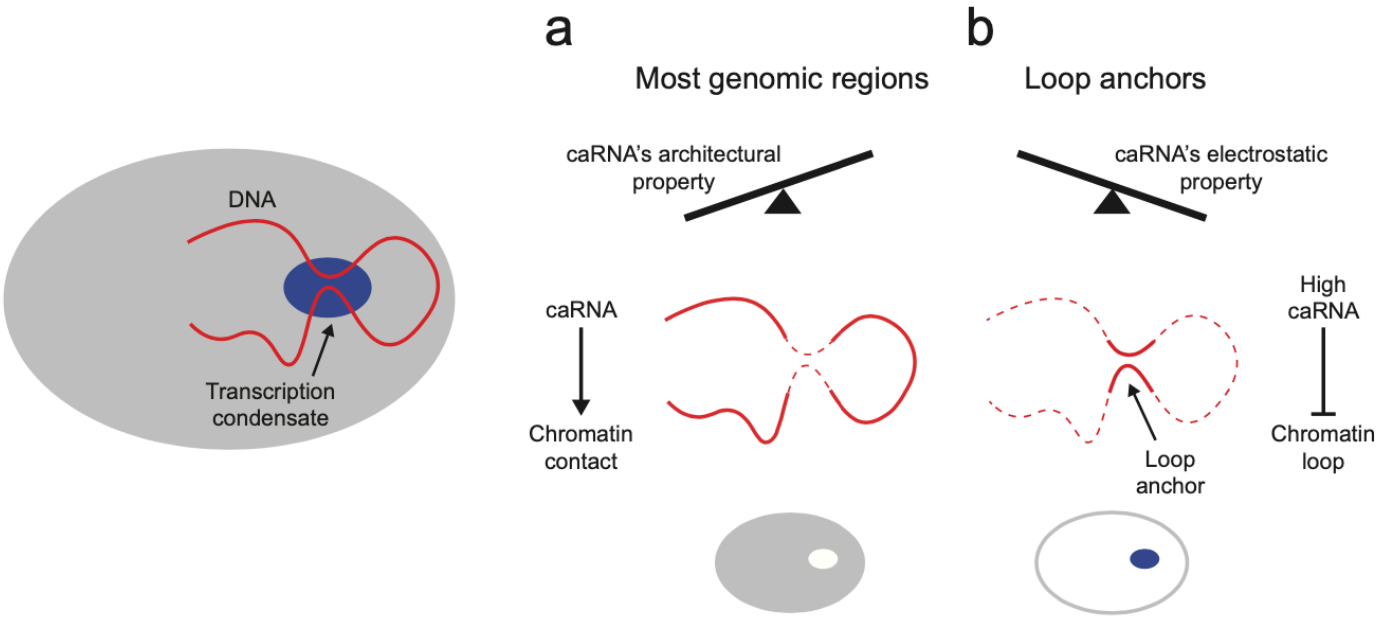
Supplementary Text

Accounting for experimental method and sequencing depth variations

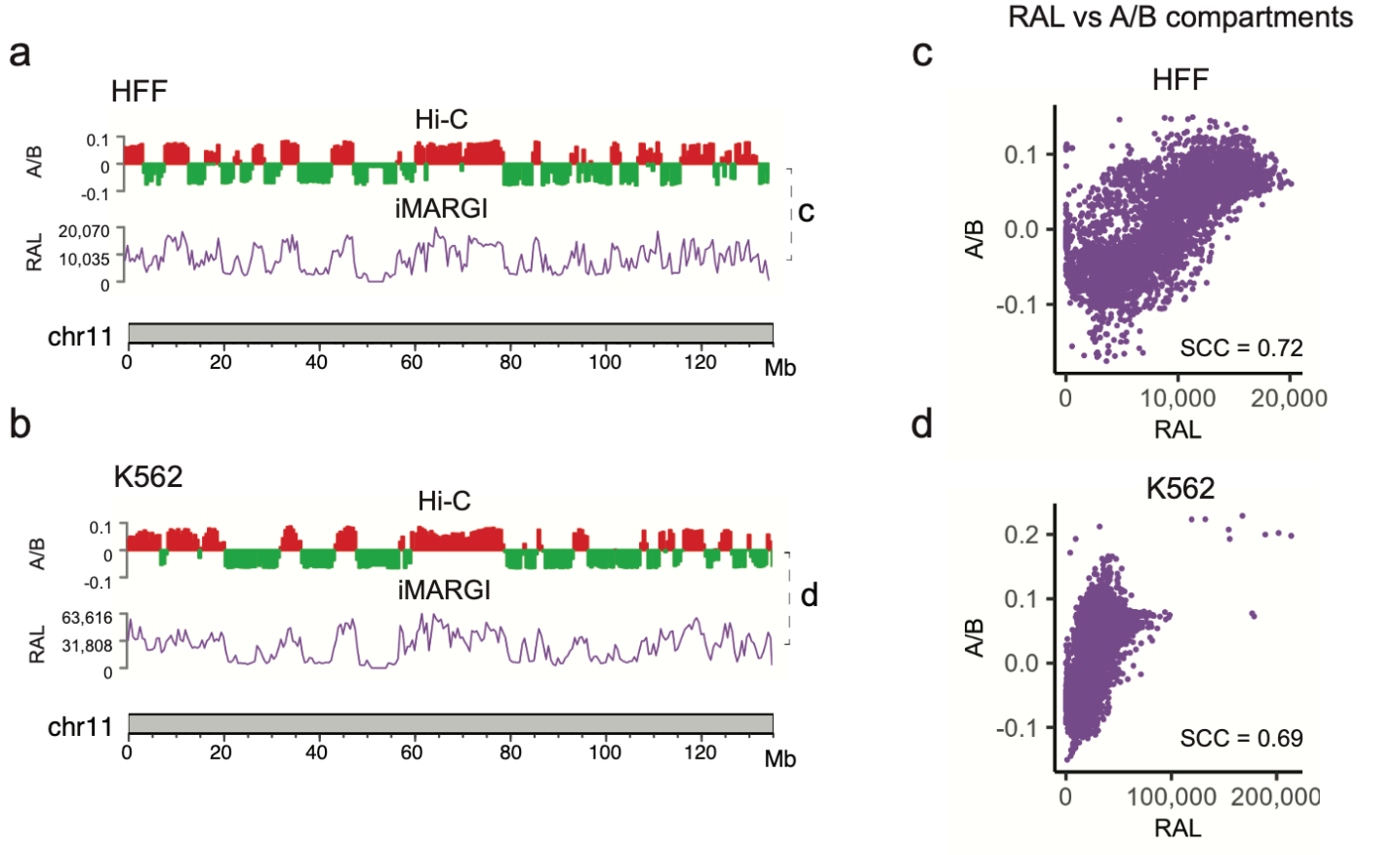
The number of detected chromatin loops depends on the experimental method, e.g. Hi-C or micro-C and the sequencing depth, making it challenging to compare the numbers of chromatin loops across different experimental conditions (control, NH4OAc, FL, RNase). We used two approaches to make our comparisons fair. First, we used the same experimental method and comparable sequencing depths amongst the experimental conditions. Second, we took the union of the loops detected in every experimental condition and compared every loop's loop strength across the experimental conditions. By taking the union of loops across all experimental conditions, even if a loop is not detected in an experimental condition (does not reach the threshold to be called as a loop), this "candidate loop's" position is still considered in the loop-strength comparison. Because this approach considers all the candidate loop positions, it ameliorates the concern that some loops are missed by a particular experimental technique or sequencing depth. Because most loops' loop-strength changes between any treatment and the control are in the consistent direction as the change of the loop numbers, we expect that our conclusion on the loop number increase or decrease would not change even if we were to change our experimental technique or sequencing depth.

Supplementary Figures

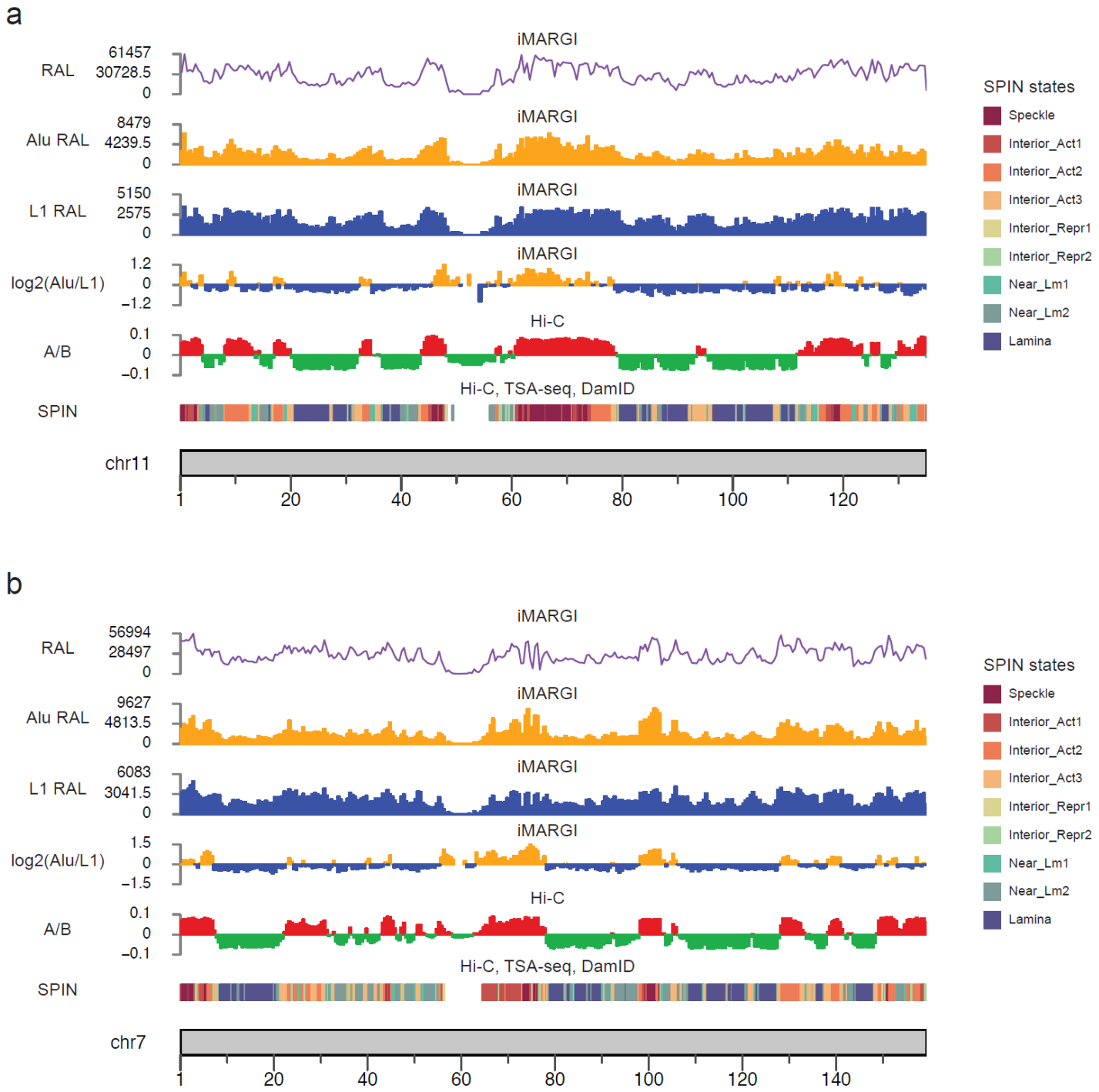
Supplementary Figure S1. A tradeoff between caRNA's roles in promoting and suppressing chromatin interactions. caRNA's architectural property promotes chromatin contacts (a), whereas caRNA's electrostatic property regulates chromatin loop formation and dissolution in a dosage-dependent manner (b). caRNA's architectural property exhibits a greater impact than caRNA's electrostatic property to chromatin contacts in most of the genomic regions (a), whereas caRNA's electrostatic property exhibit a greater impact on loop anchors (b).



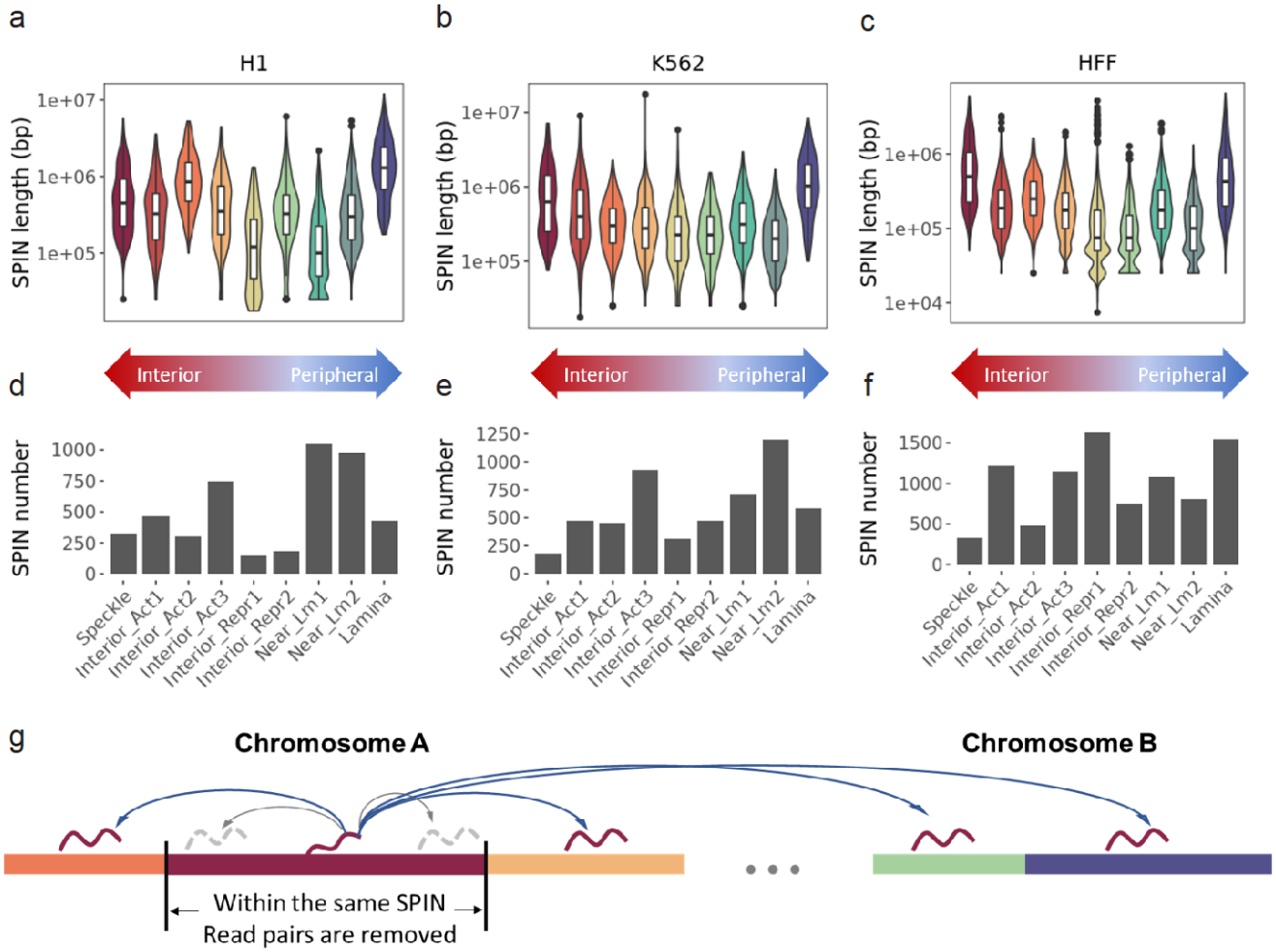
Supplementary Figure S2. Comparison of RNA-chromatin and genomic interactions. (a, b) iMARGI-derived RNA attachment level (RAL, purple curve) and A/B compartments represented by Hi-C's contact matrix's first eigenvector (A/B, red and green bars) of each 500 kb bin on Chromosome 11 in HFF (a) and K562 (b). (c, d) 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 HFF (c) and K562 (d).



Supplementary Figure S3. Comparison of Alu-caRNA, L1-caRNA, and genomic compartments. All plots are based on H1 cells unless otherwise specified, plotted for chromosome 11 (a) and 7 (b). RAL track: RAL distribution across the chromosome. All caRNAs are included in calculating the RAL. Alu RAL track: Alu-caRNA's RAL distribution across the chromosome. L1 RAL track: L1-caRNA's RAL distribution across the chromosome. Log2(Alu/L1): log ratio of the Alu-caRNA's RAL and L1-caRNA's RAL. A/B: A/B compartment score, that is the Hi-C's contact matrix's first eigenvector. SPIN: SPIN states derived from Hi-C, TSA-seq, and DamID data.



Supplementary Figure S4. SPIN states and RNA's target sequences in every SPIN state. (a, b, c) Number of genomic regions in each SPIN state (column) in H1 (a), HFF (b), and K562 cells (c). (d, e, f) the length distribution of the genomic regions in each SPIN state. (g) A cartoon showing the iMARGI read pairs mapped to a contiguous genomic region annotated by the same SPIN state are removed (grey dotted curve) from further analysis. (h) The contingency table used for testing the association of a gene (gene A)'s RNA and the SPIN states of the RNA's target genomic regions.

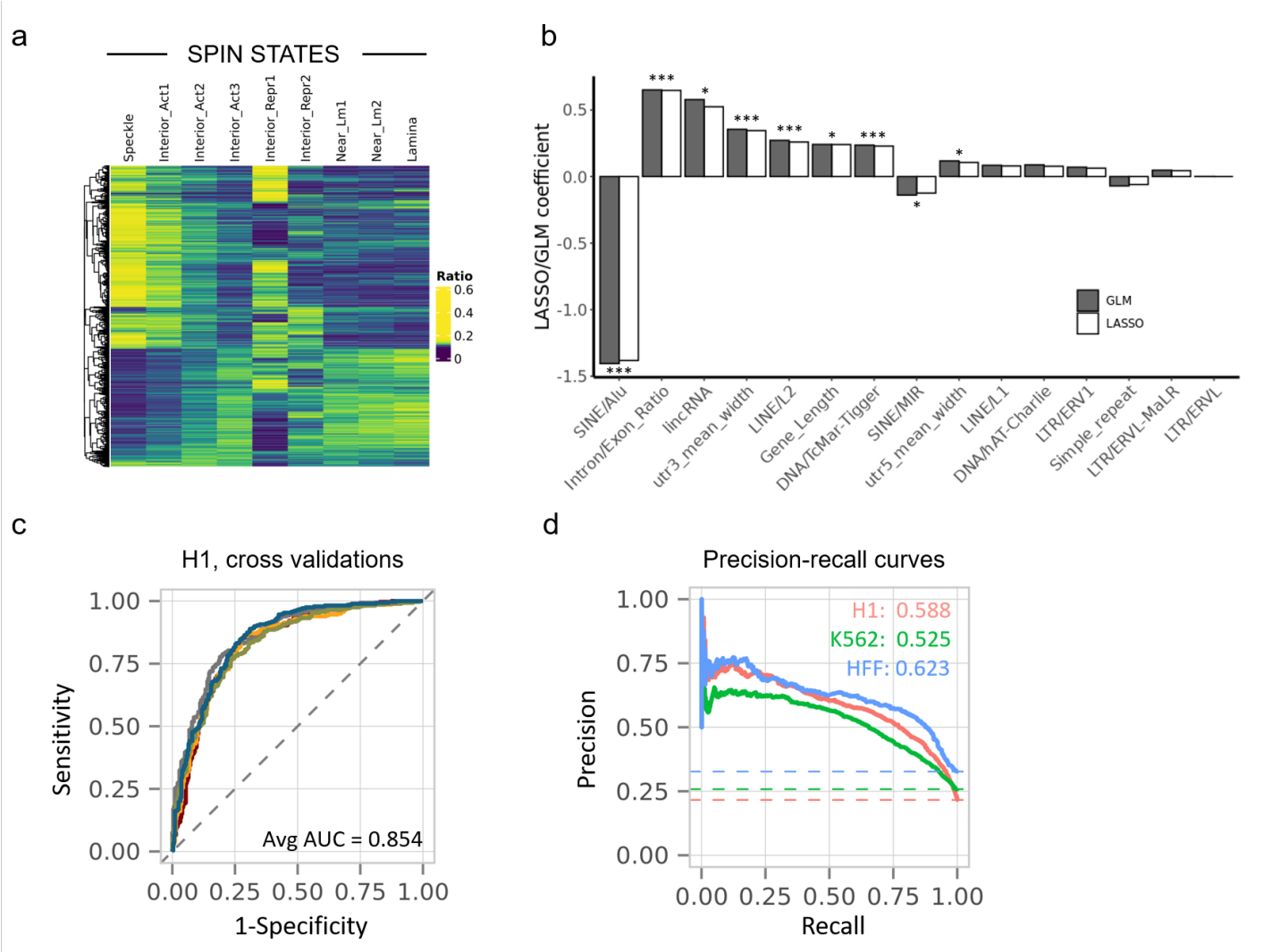


h

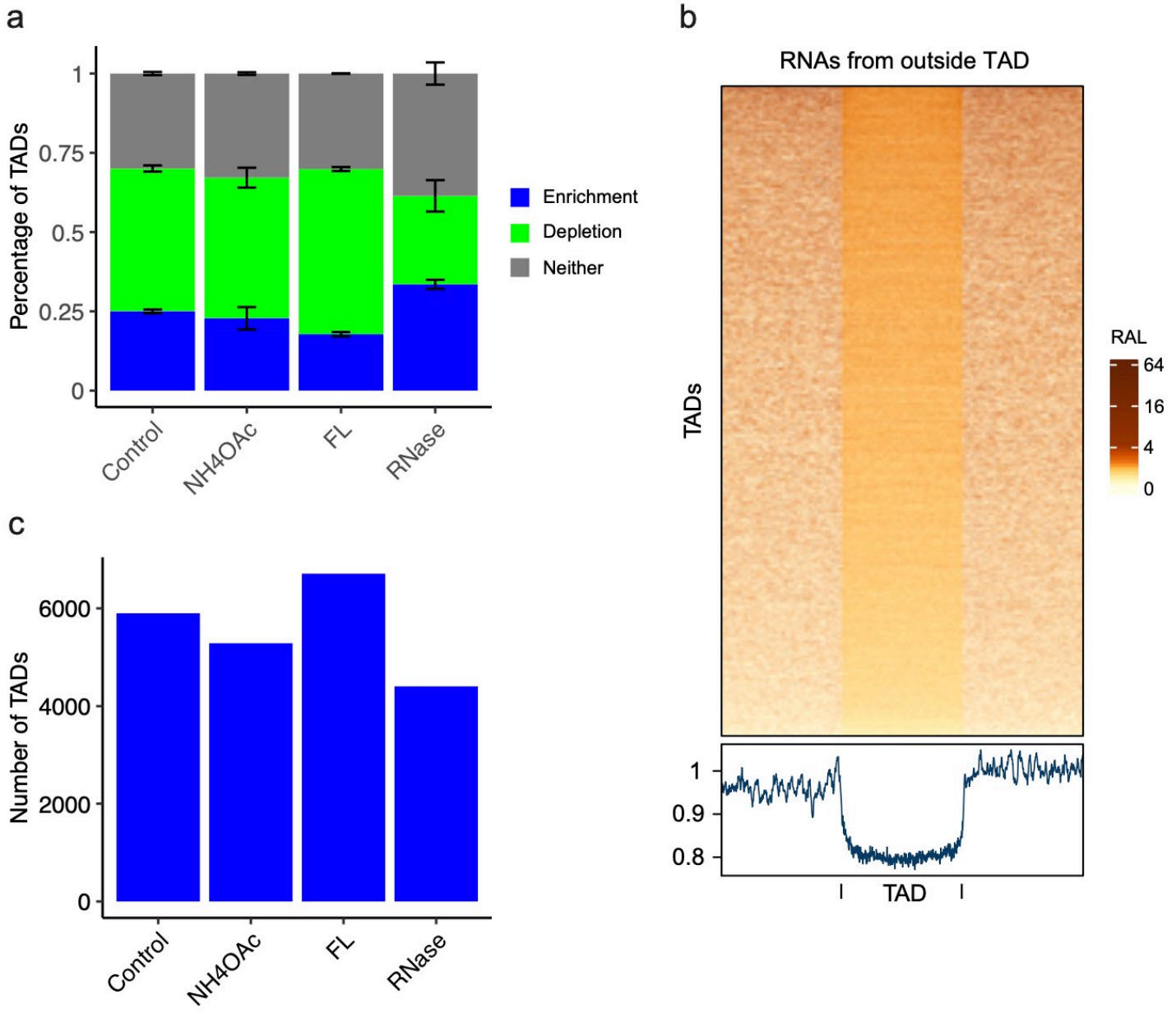
Contingency table

		The DNA-end of this read mapped to SPIN								
		Speckle	Interior Act1	Interior Act2	Interior Act3	Interior Repr1	Interior Repr2	Near Lm1	Near Lm2	Lamina
The RNA-end of a read pair mapped to gene A	Yes									
	No									

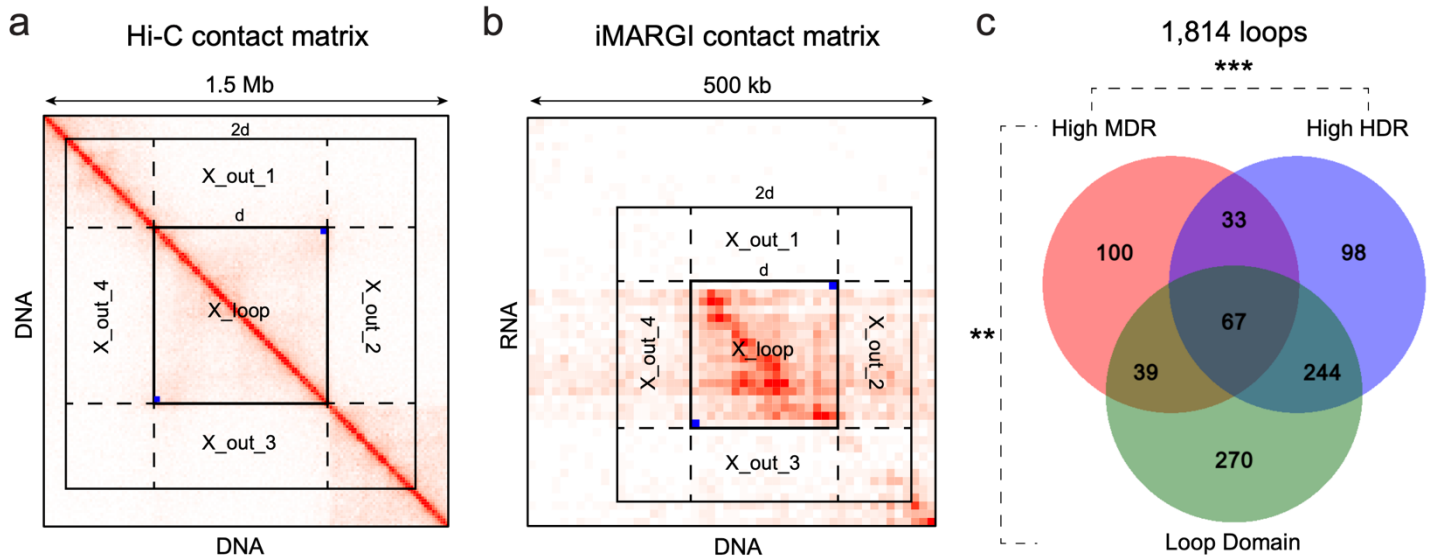
Supplementary Figure S5. RNA's target genomic sequences in SPIN states. (a) Heatmap of the RNA (row)-SPIN state (column) contact matrix. Each matrix entry is the normalized sum of iMARGI read pairs from this RNA (row) to all the genomic regions of this SPIN state (column). (b) Coefficients of 15 features using generalized linear model (GLM) and LASSO predicting RNA localization specificity as interior (Y=-1) or exterior (Y=+1). Significance of each feature considering feature mean and variance (z-score) were denoted as: 0<p(|z-score|)<0.001: ***, 0.001<p(|z-score|)<0.01: ** and 0.01 <p(|z-score|)<0.1 *. (c) ROC curves from 5-fold cross validations in H1. (d) Precision recall curves for a cross-validation test in H1 (red) and external validations with HFF (blue) and K562 (green). Dotted lines: prediction baselines. Values: the area under PR curve (AUPR) values.



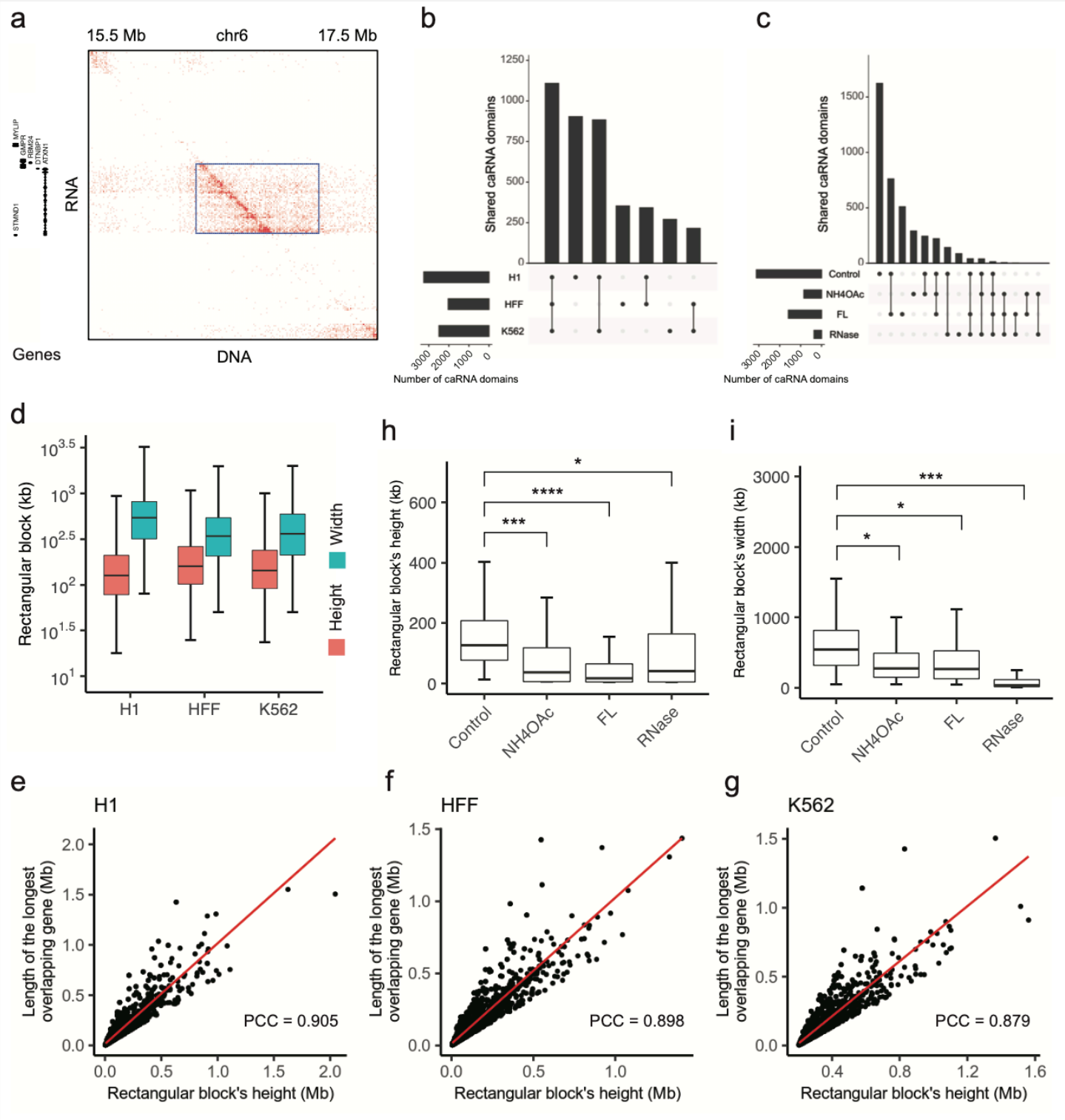
Supplementary Figure S6. TADs. (a) Proportions of caRNA-enriched (blue) -and caRNA-depleted (green) TADs in each condition (column). Error bar: SEM. (b) The RAL of every TAD (row) and its equal-length flanking regions, based on all the RNAs transcribed from any genomic sequences outside of this TAD (row). 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).



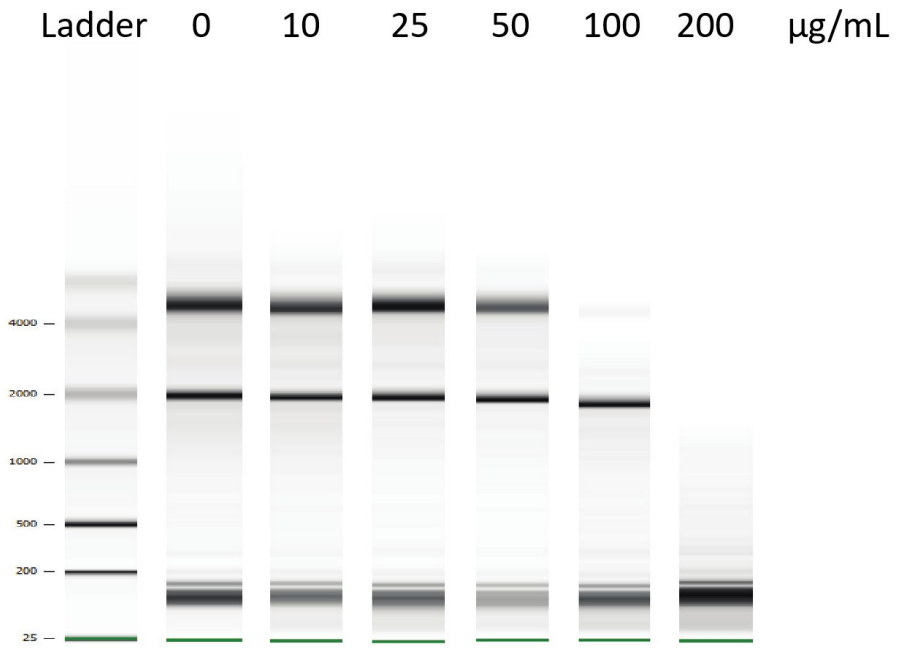
Supplementary Figure S7. Calculation of Hi-C density ratio (HDR) (a) and iMARGI density ratio (b). X_loop: the square sub-matrix as demarked by a given loop. X_out_1, 2, 3, 4: the rectangular sub-matrices, spanning d/2 to the outside of the loop, where d is the loop size. HDR/MDR is defined as the ratio of the sum of all the cells in X_loop versus the sum of the cells in X_out_1, 2, 3, 4 in the Hi-C/iMARGI contact matrix at 10 kb resolution. (c) Venn diagram of high MDR regions, high HDR regions, and loop domains. **: p-value < 1.0e-3, ***: p-value < 1.0e-10.



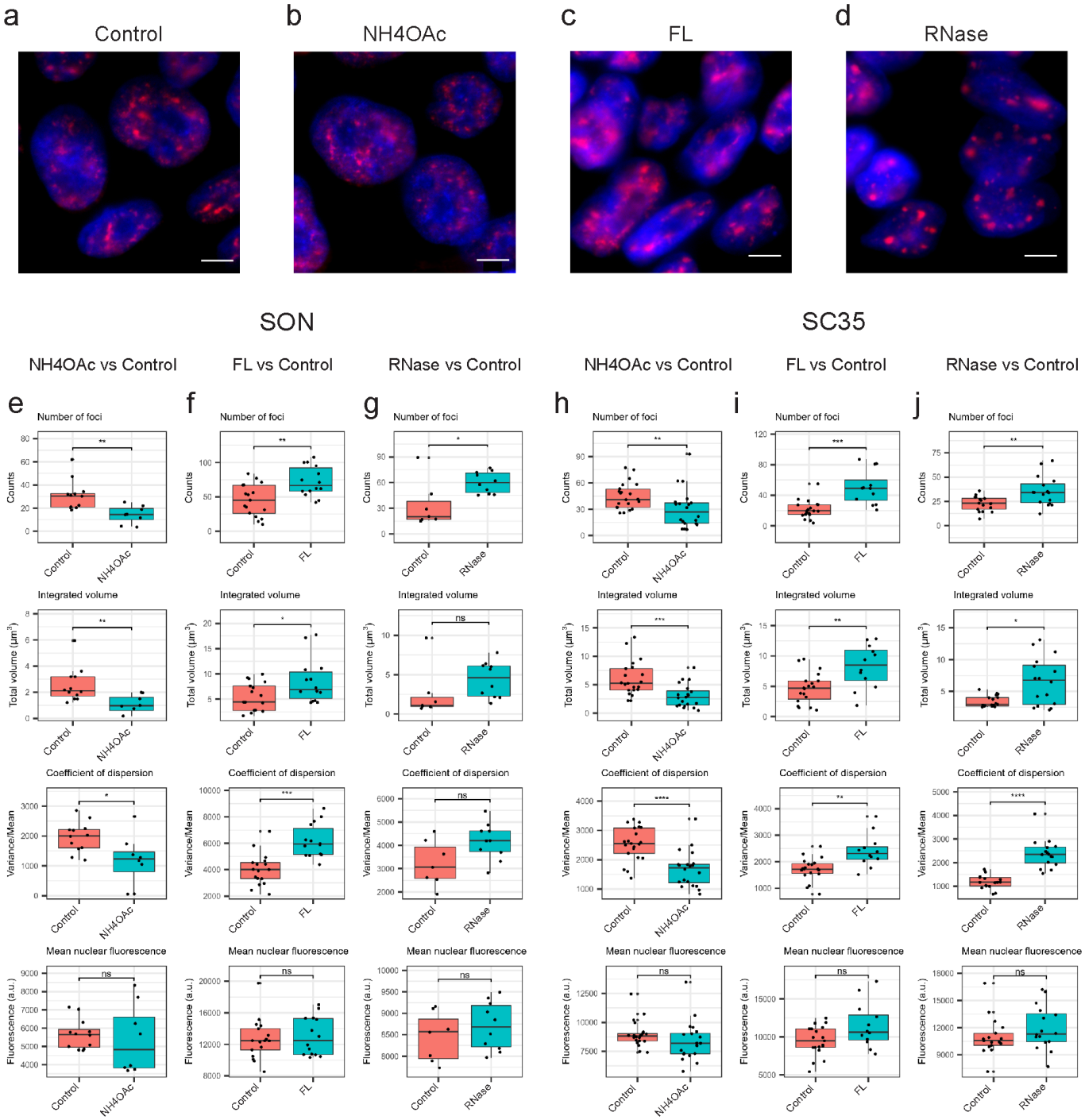
Supplementary Figure S8. caRNA domains. (a) an identified rectangular block (blue) from iMARGI's contact matrix. (b) Upset plot of the numbers of caRNA domains in H1, HFF, and K562. (c) Upset plot of the numbers of caRNA domains in untreated H1 (Control) and H1 treated with NH4OAc, FL, and RNase. (d) Box plots of the heights (blue) and widths (green) of the detected rectangular blocks. (e-g) 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). (h) Distributions of the heights, *: p value < 1e-25, ***: p value < 1e-75, ****: p value <1e-100. (i) Distributions of widths. *: pvalue < 1e-50 , ***: p value < 1e-180.



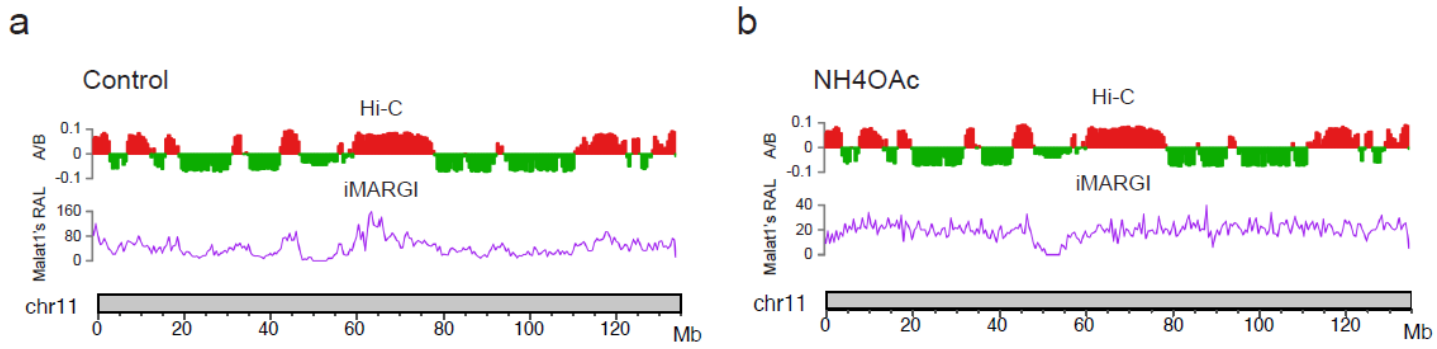
Supplementary Figure S9. RNA size distributions in several RNase A titrations. H1 cells were gently permeabilized with 0.01% Triton X-100 in PBS and then treated with 0, 10, 25, 50, 100, 200 μg/mL of RNase A in PBS for 10 min. The extracted RNAs was normalized to 0.1 μg/mL and the size distribution was measured by Bioanalyzer's Pico RNA chip. The concentration of 200 μg/mL was chosen for all the experiments presented in this paper.



Supplementary Figure S10. 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 S11. iMARGI-derived Malat1's RNA attachment level (Malat1's 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 control (a) and NH4OAc (b).



Supplementary Figure S12. Loop strengths. (a) Peak to Lower Left (P2LL) and (b) 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). Either P2LL or ZscoreLL is an overall score of the entire Hi-C dataset. (c) Distribution of every loop's Peak to the Mean (P2M) in each condition, based on the union of loops of the four conditions. ****: p-value < 2.2e-16.

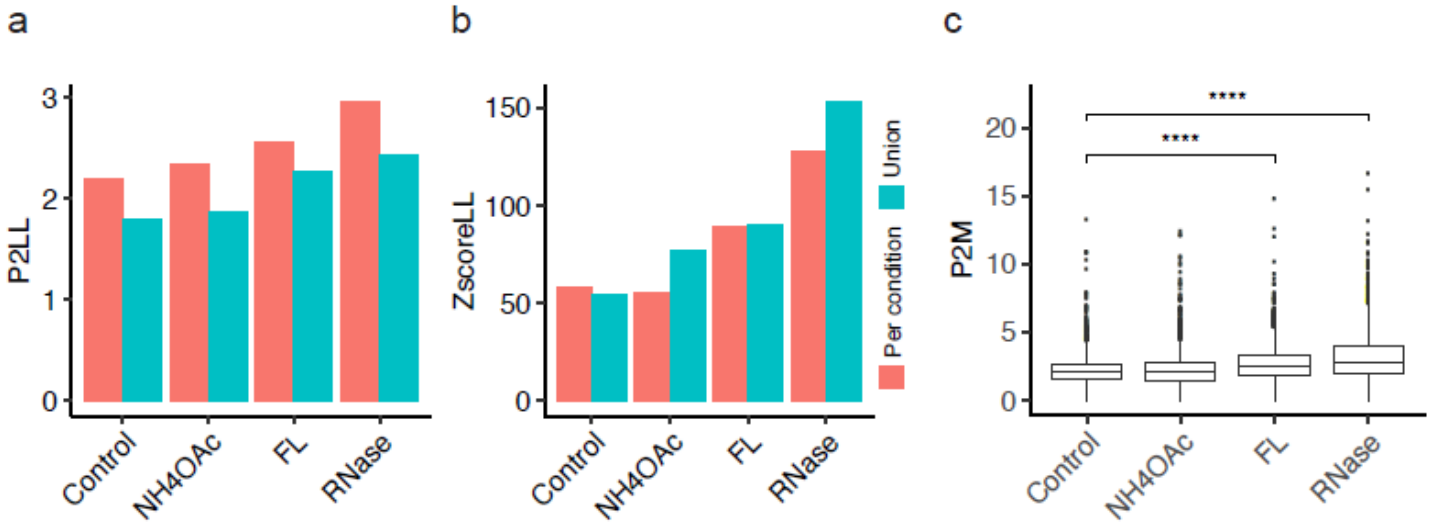


Table S1. Summary of iMARGI (a), Hi-C (b), PLAC-seq (c) datasets.

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	

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[2]. 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 incubator as described previously[3] . 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 equivalent amount of DMSO in medium in 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 (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.

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 file were provided to Cooler (v0.8.10) and Juicer Tools (v1.22.01) to generate .mcool and .hic files. The .mcool file was used in HiGlass 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 and 10 kb resolutions. Except for the resolution parameter, all the other parameters were left as the default.

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 (7,041 unique loops in 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) was re-assigned to a sample *j* if both anchors of loop *i* were within +/-10 kb flanking regions of a loop in sample *j* before the taking the union. Aggregate Peak Analysis was performed using the Juicer's "APA" tool with default parameters. Loop domains were called as previously described [1], by searching for loop-domain pairs where the peak pixel was within the smaller of 50 kb or 0.2 of the length of the domain at the corner of the domain.

iMARGI library generation and data processing. iMARGI libraries were generated and processed as previously described [2]. According to 4DN's approved iMARGI's data processing protocol [2], 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 [3]. 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 the correlation analyses.

PLAC-seq library generation and data processing. PLAC-seq libraries were generated and processed as previously described [4]. The promoter interacting sequences (PINS) were identified

by MAPS [5] 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).

Summary of RNA's target genomic regions by SPIN states. SPIN states were called following the approach as previously described [6] with additional steps to ensure that the SPIN states are comparable across multiple cell types (Wang et al., manuscript in preparation). We removed any iMARGI read pair of which the two reads are mapped to the same chromosome and within 10 Mb from each other. Only the genes with at least one iMARGI read pair's RNA-end mapped to this gene and the DNA-end mapped to a SPIN annotated genomic region are used. Three outlier genes, GRID2, WDR74 and SNHG14, which accounted for 25.58%, 6.78% and 4.69% of the variance of the first principal component of the RNA-SPIN state matrix, were removed.

The association of a RNA's target genomic sequences and the SPIN states was identified by a Chi-square test. A null hypothesis is that this RNA's target genomic sequences are not associated with any SPIN state. The continency table of this test is shown in Supplementary Figure S2h.

The Chi-square test identified RNAs were further tested with APEX-seq data. The processed APEX-seq data were downloaded from GEO (GSE116008). The processed data include the p-values and fold-changes for every APEX analyzed component/localization. Lamina-specific RNAs were identified with the Initial Criteria: lamina's p-value < 0.1, lamina's fold-change > 0, and nucleolus fold-change < 0.1, and the Second criterion: the BH adjusted p-value of satisfying the Initial Criteria < 0.01. Nuclear interior-specific RNAs were identified with the Initial Criteria: nuclear localization sequence (NLS)'s fold-change > 0, NLS's p-value < 0.1, lamina's fold-change < 0.1, nuclear pore's fold-change < 0.1, and nucleolus's fold-change < 0.1, and the Second Criterion: the BH adjusted p-value of satisfying the Initial Criteria < 0.01.

Analysis of RNA's sequence features. Gene features are extracted using R package GenomicRanges (version 1.38.0) from GENCODE hg38 gtf v24. Repeats of hg38 were downloaded from RepeatMasker (Smit, AFA, Hubley, R & Green, P. RepeatMasker Open-4.0). The sequence features were ranked by their Z-scores derived from fitting a generalized linear model (GLM). The same sequence features were ranked again by LASSO with L1-norm.

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 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 the 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.

Hi-C density ratio (HDR) and iMARGI density ratio (MDR). HDR and MDR were calculated based on the five sub-matrices of contact matrix, which are a square sub-matrix (X_loop) demarked by a given loop of size d and four rectangular sub-matrices spanning d/2 to the outside of the loop (sizes d x d/2) (X_out_1, X_out_2, X_out_3, X_out_4). HDR/MDR was calculated as the ratio of the sum of all the cells in X_loop versus the sum of the cells in X_out_1, 2, 3, 4 in the Hi-C/iMARGI contact matrix (Supplementary Figure S5).

Reference

1 Rao, S. S. P. *et al.* Cohesin Loss Eliminates All Loop Domains. *Cell* **171**, 305-320 e324, doi:10.1016/j.cell.2017.09.026 (2017).

2 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).

3 Sridhar, B. *et al.* Systematic Mapping of RNA-Chromatin Interactions In Vivo. *Curr Biol* **27**, 610-612, doi:10.1016/j.cub.2017.01.068 (2017).

4 Chen, P. B. *et al.* Discovery and Functional Characterization of Pro-growth Enhancers in Human Cancer Cells. *bioRxiv*, 2021.2002.2004.429675, doi:10.1101/2021.02.04.429675 (2021).

5 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).

6 Wang, Y. *et al.* SPIN reveals genome-wide landscape of nuclear compartmentalization. *Genome Biol* **22**, 36, doi:10.1186/s13059-020-02253-3 (2021).