GRID-seq reveals the global RNA–chromatin interactome

Xiao Li1,3, Bing Zhou1,3, Liang Chen1, Lan-Tao Gou1, Hairi Li1 & Xiang-Dong Fu1,2

Higher eukaryotic genomes are bound by a large number of coding and non-coding RNAs, but approaches to comprehensively map the identity and binding sites of these RNAs are lacking. Here we report a method to capture in situ global RNA interactions with DNA by deep sequencing (GRID-seq), which enables the comprehensive identification of the entire repertoire of chromatin-interacting RNAs and their respective binding sites. In human, mouse, and Drosophila cells, we detected a large set of tissue-specific coding and non-coding RNAs that are bound to active promoters and enhancers, especially super-enhancers. Assuming that most mRNA–chromatin interactions indicate the physical proximity of a promoter and an enhancer, we constructed a three-dimensional global connectivity map of promoters and enhancers, revealing transcription-activity-linked genomic interactions in the nucleus.

Recent genomic research has revealed that mammalian genomes are more prevalently transcribed than previously thought1. Mammalian genomes express not only protein-coding mRNAs but also a large repertoire of non-coding RNAs (ncRNAs) that have regulatory functions in different layers of gene expression. Many ncRNAs appear to act directly on chromatin, as exemplified by various characterized long non-coding RNAs (lncRNAs)2,3. Some ncRNAs may mediate genomic interactions predominantly in cis, whereas others, such as MALAT1 and NEAT1, are capable of extensively acting in trans4. These findings suggest a role of specific RNA–chromatin interactions in regulating gene expression.

Various techniques have been developed to localize specific RNAs on chromatin. These include Chromatin Isolation by RNA Purification (ChiRP)5, Capture Hybridization Analysis of RNA Targets (CHART)6, and RNA Affinity Purification (RAP-DNA)7, which all rely on using complementary sequences to capture a specific RNA followed by deep sequencing to identify chromatin targets. However, these methods allow analysis of only one known RNA at a time, and consequently, a global view is lacking on all potential RNA–chromatin interactions, which is critical for addressing a wide range of functional genomics questions.

RNAs might also play a role in coordinating functional DNA elements in regulated gene expression. The chromatin structure has been analyzed with Hi-C, which detects all possible DNA–DNA interactions8,9, and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) which enriches specific factor-mediated interactions10–12. However, as these techniques detect both regulatory and static physical interactions that are largely confined within cell-type-independent topologically associating domains (TADs)13,14, chromatin-associated RNAs may help define chromatin interactions that are directly linked to transcriptional activities, and given the recently introduced concept of super-enhancers, chromatin-linked RNAs may also help differentiate super-enhancers from typical enhancers15–17.

To address these questions, we sought to develop a general approach for comprehensively localizing all potential chromatin-interacting RNAs in an unbiased fashion. Here we report a strategy for mapping global RNA interactions with DNA by deep sequencing (GRID-seq), which uses a bivalent linker to ligate RNA to DNA in situ on fixed nuclei. Application of GRID-seq to two human cell lines (MDA-MB-231 and MM.1S), one mouse embryonic stem cell line (mESC), and one Drosophila cell line (S2), exposed distinct classes of cis- and trans-chromosomal interacting RNAs that were linked to cell-type-specific gene expression programs. We discovered a large set of both coding mRNAs and ncRNAs that bind to active promoters and enhancers, especially super-enhancers. Assuming that most interactions represent a physical proximity between the site of transcription and the distal binding site, this comprehensive RNA–chromatin interactome permitted the identification of transcription-activity-associated promoter–enhancer interactions both within and beyond TADs.

RESULTS

Ligating RNA to proximal DNA in situ

We first chose a triple-negative breast cancer MDA-MB-231 cell line to develop an unbiased strategy to map RNA–chromatin interactions genome-wide. To this end, we stabilized RNAs on chromatin by double fixing cells with disuccinimidyl glutarate (DSG) and formaldehyde, isolated nuclei, and performed in situ DNA digestion with a frequent 4-base cutter AluI. We designed a biotin-labeled bivalent linker consisting of a single-stranded RNA (ssRNA) portion for ligation to RNA and a double-stranded DNA (dsDNA) portion for ligation to DNA (Supplementary Fig. 1a). The linker was pre-adenylated at the 5′ end of the RNA and characterized in vitro and in the cell (Supplementary Fig. 1b,c). We first performed in situ RNA ligation and then extended the DNA primer in the linker into ligated RNA with reverse transcriptase (Fig. 1a). After removing free linker, we performed in situ DNA ligation to AluI-digested genomic DNA followed by affinity purification on streptavidin beads. Next, we released...
ssDNA from the beads, generated dsDNA, and used a type II restriction enzyme MmeI to cleave DNA ~20 nt upstream and downstream from the two built-in recognition sites in the linker.

We resolved two defined DNA fragments in native gel, one (85 bp) corresponding to linker ligation to both RNA and DNA, and the other (65 bp) to linker ligation to either RNA or DNA (Fig. 1a and Supplementary Fig. 1c). We isolated the 85-bp band for adaptor ligation and PCR amplification followed by deep sequencing, typically generating ~200 million 100-nt raw reads (~40 million uniquely mapped RNA or DNA read mates) per library (Supplementary Fig. 2a).

Specific linker ligation to RNA and DNA was validated based on sequenced libraries by the lack of nucleotide preference at the RNA end, but with the expected nucleotide preference (AluI site) at the DNA end. Protein-DNA crosslinking was performed in solution. The two major bands resolved by native polyacrylamide gel correspond to the products of the linker ligated to both DNA and RNA (~85 bp corresponding to linker ligation to both RNA and DNA, and 65 bp to linker ligation to either RNA or DNA (Fig. 1a and Supplementary Fig. 2b). We isolated the 85-bp band for adaptor ligation and PCR amplification followed by deep sequencing, typically generating ~200 million 100-nt raw reads (~40 million uniquely mapped RNA or DNA read mates) per library (Supplementary Fig. 2a). Specific linker ligation to RNA and DNA was validated based on sequenced libraries by the lack of nucleotide preference at the RNA end, but with the expected nucleotide preference (AluI site) at the DNA end (Supplementary Fig. 2b). The RNA reads showed the same strand selection pattern as expected for RNA, with the expected nucleotide preference (AluI site) at the DNA end.
orientation as the original transcripts, but the DNA reads lacked any strand specificity (Supplementary Fig. 2c,d). Independent libraries showed a high concordance ($R^2 > 0.95$) (Supplementary Fig. 2e,f).

The RNA reads were primarily from genic regions (both intronic and exonic), indicating their origins from various partially spliced RNAs, whereas the DNA reads were predominantly from promoters and intergenic regions (Fig. 1b). In MDA-MB-231 cells, the chromatin-interacting RNAs were better correlated with nascent RNAs detected by global nuclear run-on (GRO-seq) than RNAs at the steady state measured by polyA+ RNA–seq (Fig. 1c,d). Positive correlations were also evident when compared to both rRNA-depleted RNAs and nascent RNAs in Drosophila S2 cells (Supplementary Fig. 2g,h). These data suggest that GRID-seq preferentially detects nascent RNA on chromatin in both human and Drosophila genomes.

We also detected various mature lncRNAs and small ncRNAs, likely due to RNA fragmentation that occurred either in intact cells, as reported earlier $^{18}$, or during the experimental procedure.

Validating GRID-seq and deducing background

Two well-characterized mammalian lncRNAs MALAT1 and NEAT1 were among the most significantly enriched RNAs on chromatin identified by GRID-seq in MDA-MB-231 cells (Fig. 1c,d). To enable direct comparison with the existing data as the first pass of validation for GRID-seq, we also performed GRID-seq on an mESC line where the high-quality Malat1 RNA capture data based on RAP-DNA are available $^{19}$ (Neat1 is not expressed in embryonic stem cells $^{20}$). We found that the data of the two assays were highly comparable across the whole genome (Fig. 1e), which we further highlighted in a mouse Chr. 17 region (Fig. 1f). As previously reported $^{4,19}$, MALAT1 interacted with active genes proportional to gene expression levels in both human and mouse cells (Supplementary Fig. 3a). However, MALAT1 appears to prefer transcription start sites (TSS) (Supplementary Fig. 3b), which is distinct from the pattern observed with RAP-DNA on mESCs or CHART on MCF-7 cells, where both showed MALAT1 preference for the gene body $^{4,19}$. Because RAP-DNA and GRID-seq detected Malat1 interactions with the same set of genes in mESCs (Supplementary Fig. 3c), the different patterns are likely due to local RNA–DNA contacts detected by GRID-seq versus total Malat1-associated DNA pulled down by the capture methods. As expected from the previous studies, MALAT1-decorated genes were distinct between MDA-MB-231 and MCF-7 cells (Supplementary Fig. 3d).

Other than lncRNAs, most RNA reads were from protein-coding genes, which might reflect both specific and non-specific RNA–chromatin interactions, thus requiring a background model to identify specific interactions. To assess the background, we mixed isolated nuclei from MDA-MB-231 and S2 cells in equal genome content (Fig. 1g). By using uniquely and unambiguously mapped RNA–DNA read maps to the human or Drosophila genome, we detected 6.8% human RNA linked to Drosophila DNA and 8.4% Drosophila RNA linked to human DNA (Fig. 1g). Using these cross-species reads, we took advantage of the small Drosophila genome (thus having sufficient read density from human RNAs) to construct a true background for non-specific RNA–chromatin interactions (Fig. 1h, top panel). We next used mRNA–chromatin interactions in the same cells to develop an endogenous background model, inspired by the strategy developed for processing Hi-C data $^{21}$. As illustrated in a stepwise fashion on MDA-MB-231 cells (Supplementary Fig. 4a), we deduced the background based on endogenous mRNAs engaged in trans-chromosomal interactions, and after normalization to equal density in comparison with specific RNA, true trans-chromosomal interactions were still preserved. This endogenous background was highly concordant with the exogenous background (Fig. 1h and Supplementary Fig. 4b), and further quantitative analysis showed <1% discrepancy in identifying specific RNA–chromatin interactions by using either background model in both MDA-MB-231 and S2 cells (Supplementary Fig. 4c).

Differentiating specific and non-specific RNA–chromatin interactions

Using the background models we developed, we estimated the false-positive rate of trans-chromosomal RNA–chromatin interactions at 3.3% in MDA-MB-231 cells, 6.9% in S2 cells, and 4.7% in mESCs (Supplementary Fig. 4d). After peak calling, we found that 70.5% chromatin-enriched RNAs showed at least one significant trans peak in MDA-MB-231 cells, 77.1% in S2 cells, and 87.8% in mESCs. Notably, in MDA-MB-231 cells, only 3.6% of 71.4% total trans reads were in peaks, compared to 14.7% of 28.6% total cis reads in peaks, and similar results were also obtained on S2 cells and mESCs (Supplementary Fig. 4d). These data suggest that the majority of trans reads resulted from RNAs released from their sites of transcription that non-specifically interacted with different chromosomes, whereas about half of cis reads were engaged in specific interactions with chromatin.

We also noted that non-specific trans interactions tended to occur on open chromatin regions when compared to RNA Pol II binding as well as H3K4me1 and H3K27ac marks (Supplementary Fig. 4e–g). This further confirmed the importance of developing a reliable background model in order to identify specific RNA–chromatin interactions. After background correction and peak filtering, all true trans-chromosomal interactions stood out, as demonstrated with Malat1 whose signals closely tracked nascent RNA production detected by GRO-seq, indicating the association of Malat1 with actively transcribed genes in mESCs (Fig. 1i).

Global view of RNA–chromatin interactions

After removing non-specific signals, we detected 868 (88.75%) mRNAs and 72 (7.36%) ncRNAs highly enriched on chromatin in MDA-MB-231 cells at the current sequencing depth. We obtained comparable data on mESCs (Supplementary Table 1). Displaying all specific chromatin-enriched RNAs on chromosomes, we observed that only a limited number of RNAs were extensively engaged in trans interactions across the genome (Fig. 2a). In MDA-MB-231 cells, for example, MALAT1 and NEAT1, as well as U2 snRNA and two variant U2 snRNAs, interacted with numerous loci on all chromosomes (Fig. 2b). By contrast, the majority of RNAs, whether protein-coding (pc) or non-coding (nc), interacted with chromatin near their sites of transcription. These extensive RNA–chromatin interactions were highly reproducible based on duplicated GRID-seq experiments, even with increasing resolution in all cell types we examined (Supplementary Fig. 5a,b).

In S2 cells, we also detected a large number of chromatin-enriched RNAs (Supplementary Table 1). For example, an enlarged chromosomal view showed that roX2, a known lncRNA involved in dosage compensation in Drosophila$^{22}$, exclusively decorated Chr. X (Fig. 2c). Comparing this profile with the published roX2 ChIRP and CHART data $^{5,6}$, as well as the ChIP-seq data on MSL3, a known roX2-interacting factor $^{23}$, we observed high concordance among all data sets, as indicated by examples on an expanded view of Chr. X (Fig. 2d) and by the overlaps in both peak number (Fig. 2e) and position (Fig. 2f). Even at the raw data levels, the concordance was strong among the data generated by different methods (Supplementary Fig. 5c). In fact, GRID-seq showed the highest specificity for Chr. X and was more concordant with MSL3 ChIP-seq signals on Chr. X than
other RNA capture results (Supplementary Fig. 5d–f). Moreover, roX2 GRID-seq peaks recovered >96% of previously defined chromosomal entry sites23 or high-affinity sites24 for the roX-MSL complex (Supplementary Fig. 5g). Together, these data suggest that our unbiased GRID-seq approach is able to recapitulate known specific RNA–chromatin interactions with high specificity and sensitivity. However,

Figure 2 Global view of RNA–chromatin interactions. (a) Heatmap showing chromatin-enriched RNAs across the whole human genome in MDA-MB-231 cells. Row: chromatin-enriched RNAs. Column: human genome in 1-Mb resolution. Major trans-chromosomal interacting RNAs are labeled on the right. U2-36P and U2-2P are transcripts from two variant U2 snRNA genes. (b) Two enlarged representative regions boxed in a, showing detailed RNA–chromatin interaction patterns on Chr. 11 (left) and Chr. 17 (right) at 100 Kb resolution. pc: protein-coding RNAs, nc: non-coding RNAs. (c) Heatmap showing background-corrected RNA–chromatin interactions on Chr. X across the Drosophila genome. Right: an enlarged view of the boxed region on the left, showing decoration of roX2 RNA on Chr. X in Drosophila S2 cells. (d) A region of Drosophila Chr. X, illustrating roX2 RNA on chromatin detected by ChIRP and CHART in comparison with GRID-seq signals as well as ChIP-seq signals for the roX2 binding protein MSL3. (e) Overlaps of peaks identified by different methods. (f) Meta-analysis of roX2-chromatin interactions detected by ChIRP, CHART, and GRID-seq relative to MSL3 ChIP-seq peaks. (g–i) Ternary plots of chromatin-enriched RNAs in called peaks in human MDA-MB-231 cells (g), mESCs (h), and Drosophila S2 cells (i), showing percentages of individual chromatin-interacting RNAs engaged in local (±10 Kb around their genes), cis (in the same chromosome except local), and trans (in other chromosomes) interactions. The three axes were shown as dashed lines perpendicular to the edges of the triangle. The arrowhead on the tip of each axis indicates the direction of increase from 0 (at the edge) to 100% (at the vertex). Three ticks on each axis mark 25%, 50%, and 75% positions. Colors of dots represent different types of RNAs and sizes represent chromatin interaction levels. Labeled are representative coding mRNAs and lncRNAs in each genome.
given the all-to-all nature of GRID-seq, each chromatin-enriched RNA is expected to have fewer reads compared to the capture technologies that focus on a single target at comparable sequencing depths. For example, at the current sequencing depth, the GRID-seq data on roX2 gave rise to a medium peak width of 83 Kb from a total of 42K reads, whereas ChIRP roX2 exhibited a medium peak width of 4.5 Kb from a total of 40M reads, indicating a lower resolution of GRID-seq compared to ChIRP on this particular RNA.

To further characterize newly identified chromatin-enriched RNAs, we classified their chromatin-interactions in local (±10 Kb from their encoding genes), cis (beyond local regions, but in the same chromosomes), and trans (across different chromosomes) modes. Notably, with a few exceptions of specific IncRNAs and small ncRNAs, the majority of RNAs exhibited predominant local and cis-interactions in all cell types (Fig. 2g–i). Compared to human MDA-MB-231 cells, we noted a much lower degree of trans-interactions in mESCs (Fig. 2g,h), and relative to mammalian cells, we saw more restricted local interactions in Drosophila S2 cells (Fig. 2i).

Individual RNAs each showed specific preference for different modes of interactions, as illustrated by Circos plots25 of representative coding mRNAs and IncRNAs in each cell type. Some RNAs exhibited rather local and cis-interactions, whereas others engaged in more extensive trans-interactions (Supplementary Fig. 5h–j). These data provide rich resources for future investigation of individual chromatin-interacting RNAs.

Cell-type-specific interactions

We next determined whether specific RNA–chromatin interactions reflected cell-type-specific activities and analyzed another well-characterized human multiple myeloma cell line MM.1S, which enabled us to take advantage of previously generated functional data on this cell type26. Similar to MDA-MB-231 cells, we detected MALAT1 and NEAT1 (Supplementary Fig. 6a) and numerous other mRNAs and ncRNAs on chromatin (Supplementary Table 1). We also detected exclusive decoration of the IncRNA XIST on Chr. X in this human cell type (Supplementary Fig. 6b).

Cross-analysis between MDA-MB-231 and MM.1S cells revealed that the repertoire of chromatin-enriched RNAs were largely cell-type-specific (Fig. 3a,b), as illustrated on a representative region of Chr. 4 (Fig. 3c) and on the whole genome (Supplementary Fig. 7a), whereas background RNA–chromatin interactions were relatively similar (Supplementary Fig. 7b). Even a common set of RNAs showed distinct chromatin-interaction patterns in the two cell types, as exemplified on Chr. 6 (Fig. 3d), indicating that chromatin–RNA interactions likely reflected cell-type-specific gene regulation programs. Consistently, we also observed a genome-wide trend of chromatin-enriched RNAs that specifically bound DNA elements marked with H3K4me1, H3K4me3, and H3K27ac, as well as RNAPII (Supplementary Fig. 7c,d).

In addition to various unannotated DNA elements, RNA interactions were enriched on active promoters and enhancers in a cell-type-specific manner (Fig. 3e and Supplementary Fig. 7e,f), positively correlated with gene expression levels (Supplementary Fig. 7g). For example, we observed cell-type-specific chromatin-enriched RNAs that were able to interact with enhancers several hundred Kb away from their promoters (Fig. 3f,g). Representative of many commonly captured RNAs, FAM49B RNA showed similar chromatin interaction density in both MDA-MB-231 and MM.1S, but reached out to distinct enhancers (arrows in Fig. 3h). Although we do not have sufficient read density to detect enhancer-produced RNAs (eRNAs), which are believed to link enhancers to promoters27,28, these data indicate that RNAs from actively transcribing genes are also associated with their enhancers, perhaps reflecting spatial proximity between specific promoters and enhancers in the nucleus.

Prevalent RNAs on super-enhancers

Recent studies suggest that enhancers may be segregated into typical and super-enhancers, the latter being defined by a much higher density of enhancer marks, such as Mediator and BRD4 ChIP-seq signals that generally track H3K27ac signals, and super-enhancers also appear to be more potent than typical enhancers in activating nearby genes16,26. Because of the RNA decoration on active enhancers, we sought to determine whether GRID-seq signals might also reflect relative strengths of typical versus super-enhancers. We found that enhancers highly associated with RNAs predominantly corresponded to super-enhancers in both MDA-MB-231 and MM.1S cells (Fig. 4a and Supplementary Fig. 8a), which was also evident from quantitative analysis (Fig. 4b and Supplementary Fig. 8b). Therefore, the levels of chromatin-interacting RNAs may provide an independent criterion to differentiate typical from super-enhancers.

We next sorted enhancers based on their levels of bound RNA and compared the expression of neighboring genes from flanking enhancers using the same RNA expression data and analysis strategy as previously reported on MM.1S cells26. We found that genes adjacent (±50 Kb) to the top 10% RNA-decorated enhancers were more active than those adjacent to the bottom 10% (Fig. 4c,d). Consistently, the genes associated with the top 10% RNA-decorated enhancers were more responsive to functional perturbation with the BRD4 inhibitor JQ1 than those associated with the bottom 10% (Fig. 4e). We performed a similar analysis on MDA-MB-231 cells by using GRO-seq to score nascent RNA production and transcriptional response to JQ1 treatment and reached the same conclusion (Supplementary Fig. 8c–e). Combined, these data demonstrated that the levels of chromatin-enriched RNAs reflected enhancer activities in activating gene expression.

To further establish specific RNA–chromatin interactions on enhancers, we took advantage of the REDfly database, which listed a large number of genomic fragments tested for enhancer activities using a reporter-based assay in Drosophila embryos29. To enable comparison with our GRID-seq data, we first identified active enhancers based on the published H3K27ac ChIP-chip data on S2 cells30. By examining S2-cell-specific RNA interaction levels on different classes of distal regulatory elements, we found that active enhancers marked by H3K27ac were indeed preferentially linked to chromatin-enriched RNAs, compared to a similar number of randomly selected genomic regions (Fig. 4f). These data provided further support of the significance of chromatin-enriched RNAs on active enhancers.

RNA–chromatin interactions relative to TADs

A fundamental genomics question regards how various DNA elements interact with one another in the three-dimensional (3D) space of the nucleus. Hi-C experiments revealed predominant DNA–DNA interactions within TADs with a median interval of ~800 Kb33, but RNAPII ChIA-PET studies showed RNAPII-tethered genomic interactions both within and beyond TADs10,11. However, it has been challenging to use either Hi-C or RNAPII ChIA-PET data to differentiate actively transcribing genes from inactive or transcriptionally poised genes31. As GRID-seq has the ability to detect chromatin sites associated with RNA production, we observed in both human cell types that RNAs often covered chromatin ~1 Mb away from their transcription sites (Fig. 5a). mESCs showed the same trend as human cells (Fig. 5b, upper panel). By contrast, the cis-interaction range of RNAs was on average about tenfold shorter in Drosophila S2 cells (Fig. 5b, lower...
**Figure 3** Cell-type-specific RNA–chromatin interactions in human cells. (a) Overlap of chromatin-enriched RNAs between MDA-MB-231 and MM.1S cells, illustrating the cell-type specificity. (b) Comparison of individual chromatin-enriched RNA levels in the two human cell types. Colored dots represent enriched RNAs on chromatin, a few of which are highlighted. (c) Illustration of both common and cell-type-specific RNA–chromatin interactions in MDA-MB-231 and MM.1S cells. Two representative cell-type-specific regions are shaded (see further details in panel f and g). (d) Heatmap illustrating differential RNA–chromatin interactions on Chr. 6 from common chromatin-enriched RNAs in the two human cell types. Red and blue, respectively, indicate lower and higher levels of chromatin interactions in MDA-MB-231 cells relative to MM.1S cells. (e) Meta-analysis of RNA interactions at gene promoters (top panel) or enhancers (middle panel) in MDA-MB-231 and MM.1S cells. Bottom panel: heatmaps of enhancers ranked by normalized RNA signals in the two cell types, showing both cell-type-specific and common RNA–chromatin interactions at enhancers. (f–h) Examples showing broad chromatin interactions by LEF1 (MM.1S cell-specific) (f), VEGFC (MDA-MB-231 cell-specific) (g), and FAM49B (common) (h), in comparison with key chromatin marks (in autoscale). RPM, RNA reads per million.
data on mESCs\textsuperscript{22} and S2 cells\textsuperscript{23}, we compared RNA–DNA interactions scored by GRID-seq and DNA–DNA contacts established by Hi-C. Because Hi-C detects all types of DNA–DNA interactions, whereas GRID-seq captures only the interactions of RNA-producing genes with DNA elements, we extracted Hi-C contacts from individual gene bodies for direct comparison with RNA–chromatin interactions detected by GRID-seq (Fig. 5c). Pearson’s Correlation Coefficient (PCC) between GRID-seq and Hi-C signals for each gene quantitatively demonstrated the high global concordance between the two data sets within ±1 Mb in mESCs and ±200 Kb in S2 cells (Fig. 5d). This was further illustrated by representative examples of mESCs (Fig. 5e) and S2 cells (Supplementary Fig. 9a).

We further marked GRID-seq-detected RNA–chromatin interactions relative to previously assigned TAD boundaries, observing that GRID-seq signals were predominantly confined within TADs in both mouse and fly cells (Supplementary Fig. 9b,c, left panels). However, a small fraction of RNAs were clearly capable of interacting with chromatin across TAD boundaries (red line), spreading >50% of their chromatin interaction signals into neighboring TADs on both mESCs and S2 cells (Supplementary Fig. 9b,c, right panels). These data suggest that chromatin-interacting RNAs were largely embedded in the high-order organization of nuclear territories. Such similarity demonstrates that GRID-seq signals could be applied to infer genomic interactions that are linked to RNA production, providing yet another transcription-focused approach to complement the existing 3D genomic technologies.

**Global connectivity of promoters and enhancers**

To further use GRID-seq to infer transcription-linked genomic interactions, we turned to a long-standing problem of how enhancers and active gene promoters contact one another in the 3D genome. Although GRID-seq \textit{per se} does not distinguish between \textit{cis}-interactions by RNAs with DNA elements in the proximity of their sites of transcription and \textit{trans}-interactions due to traveling RNAs after they are released from chromatin, we took advantage of their distinct features to construct a statistical model to differentiate \textit{cis}- versus \textit{trans}-interactions. We reasoned that the collective \textit{trans}-chromosomal signals from mRNAs were statistically unlikely to reflect proximal interactions, which we defined as the null distribution. Thus, any \textit{cis}-chromosomal signal that rejects the null distribution at a stringent significance level would most likely indicate chromatin proximity between active genes and their underlying DNA elements (Supplementary Fig. 10a). Based on this model and the requirement for a significance level of \(Z \geq 3\), we identified 10,933 significant promoter–enhancer and 8,142 promoter–promoter interactions in MM.1S cells (Supplementary Table 2). We visualized the resultant promoter–promoter and promoter–enhancer networks with Cytoscape by using a self-organized layout\textsuperscript{14}, as illustrated on Chr. 1 from MM.1S cells (Fig. 6a).

Based on this network, we observed that typical enhancers appeared to have slightly longer interaction ranges than super-enhancers (Supplementary Fig. 10b,c). We next calculated the frequencies of promoter–promoter and promoter–enhancer interactions, finding that each promoter attracted RNAs from up to four other genes in most cases (Fig. 6b), suggesting that one gene promoter may serve as an enhancer for other genes, as previously proposed\textsuperscript{11}. However, in contrast to an earlier report\textsuperscript{12}, we rarely detected promoter–promoter interactions between chromosomes. We also found that each chromatin-enriched RNA was able to interact with multiple typical enhancers, but only one or two super-enhancers (Fig. 6c). By contrast, each enhancer, whether typical or super, mainly interacted with RNAs from one or two

Panel. Such prevalent local interactions are likely due to closely spaced genes in the fly genome compared to mammalian genomes. These data indicate that the genomic organization dictates the range of specific RNA–chromatin interactions.

We next investigated how RNA–chromatin interactions were related to the 3D chromatin structure. Taking advantage of the high-quality Hi-C

![Figure 4](https://example.com/image4.png)  
**Figure 4** Preferential RNA decoration on super-enhancers in MM.1S cell. (a) Super-enhancers in relationship to RNA–chromatin interaction levels. 10,567 active enhancers marked by H3K27ac in MM.1S cells were ranked by RNA interaction levels. Each red bar on top represents a super-enhancer. Red curve: the accumulative curve of rank-ordered RNA–chromatin interactions; gray dashed line: random distribution. (b) Density distribution of chromatin-enriched RNAs at super-enhancers (SE, red) and typical enhancers (TE, blue). (c) Left: Rank-ordered RNA–chromatin interaction levels on all active enhancers. Right: the top 10% and bottom 10% enhancers with the most and least RNA interactions, respectively, selected for functional analysis. (d) Gene expression associated with top 10% enhancers (orange box) and bottom 10% enhancers (green box). (e) Accumulative curves of expression changes of genes associated with top 10% and bottom 10% enhancers in response to JQ1 treatment in MM.1S cell. Statistical significance of comparison was determined by Student’s \(t\)-test in panels b, d, and e with degrees of freedom of 631.5, 1965.8, and 1926.4, respectively. Lines and whiskers in the box plot in d depict the median and 1.5 IQR of values. (f) Meta-analysis of chromatin-enriched RNAs on H3K27ac-positive enhancers detected in Drosophila S2 cells (red) versus all enhancers scored positively in Drosophila embryos from the REDfly database (blue) or random genomic fragments (black).
genes (Fig. 6d). These findings suggest that, while each gene is controlled by a large number of enhancers, each enhancer, regardless of typical or super status, is dedicated to regulate a highly selective set of target genes.

We next sought functional evidence for these deduced global promoter–enhancer interactions. We chose specific examples in MM.1S cells: RNAs from two transcribing genes SNX5 and RPBP1 were interacting with one super-enhancer and six typical enhancers (Supplementary Fig. 10d). In response to the BRD4 inhibitor JQ1, both genes were downregulated (Supplementary Fig. 10e), and the super-enhancer showed more reduction in BRD4 binding than typical enhancers (Supplementary Fig. 10f). We then extended the analysis...
to all promoters and enhancers connected by chromatin-enriched RNAs in MM.1S cells by asking whether genes associated with at least one super-enhancer (plus typical enhancers) might be more sensitive to perturbation by JQ1 than those linked only to typical enhancers. We found that this was indeed the case (Fig. 6e). We performed parallel analysis on mESCs based on super-enhancers previously defined by Mediator binding16 and reached the same conclusion from the transcriptional response to Mediator depletion (Fig. 6f).

In addition to network analysis on individual chromosomes, we displayed the whole genome network detected by GRID-seq with Cytoscape (Fig. 6g). Notably, the resulting global network revealed the organization of individual chromosomes that resembled nuclear territories detected by chromosome painting35, which is similar to those reconstructed with Hi-C data in budding yeast36 and mammalian cells37. Despite the fact that we rarely detected promoter–promoter proximity between chromosomes, we did observe various specific interchromosomal interactions, suggesting potential neighboring relationships between different chromosomes in a given cell type. Although verifying such putative interchromosomal interactions clearly requires future work, especially at the single-cell level, the elucidated global interaction network establishes a foundation to understand genomic organization in the 3D space of the nucleus.

**DISCUSSION**

We present here a technology for global detection and analysis of RNA–chromatin interactions. One of the major findings of applying GRID-seq to mammalian and fly cells is that few RNAs are capable of engaging in broad trans-chromosomal interactions, with the exceptions of the major lncRNAs MALAT1 and NEAT1 in mammals, as reported in the literature4,19. However, we cannot rule out the possibility that many less abundant lncRNAs may escape detection at our...
current sequencing depth. Interestingly, we detected a large number of small nucleolar RNAs (snoRNAs) on chromatin interactions in fly cells, raising the possibility that various snoRNAs may have important roles at the chromatin level. We also detected many unannotated chromatin-enriched transcripts in all cell types we examined, thus providing rich resources for future studies of their functions. Numerous RNAs were able to reach out to chromatin regions that are megabases away in linear DNA distance, and in some extreme cases, some specific RNAs can decorate an entire chromosome arm, posing the question whether some of those RNAs might broadly modulate gene expression on various autosomes.

Although our GRID-seq data provide rich resources to study individual chromatin-interacting RNAs, we have used this information to elucidate general patterns of RNA–chromatin interactions. Here, it is important to emphasize that, although GRID-seq detects genomic interactions that are directly linked to transcription, it does not necessarily establish the functionality of detected distal DNA elements for regulating their contact genes, which requires functional perturbation in their native genomic contexts. We also used the GRID-seq data to infer contacts between the sites of RNA transcription and distal DNA elements. In contrast to the RNAPiP ChiP-PET-deduced promoter–promoter interactions, which suggested one promoter interacts with an average of about eight other promoters, and about half of these interactions occur between chromosomes, we mainly detected intrachromosomal interactions based on our statistical model. Although resolving this discrepancy will require future studies, we argue that intrachromosomal interactions are expected to dominate interchromosomal interactions because of the confinement imposed by nuclear territories.

The majority of chromatin-enriched RNAs are pre-mRNAs, implying that many pre-mRNAs may function in the regulation of gene expression as IncRNAs before being processed into mRNAs in the nucleus. This is in line with increasing evidence for a role of both IncRNAs and nascent RNAs in mediating a range of regulatory activities on chromatin, as exemplified by the RNA-dependent recruitment of a de novo DNA methyltransferase or transcriptional activators or repressors. Thus, the GRID-seq technology is expected to expedite the discovery of a variety of RNA-mediated regulatory activities on chromatin.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We wish to express our gratitude to X. Ji and R. Young (Massachusetts Institute of Technology) for sending us MM.15 cells; B. Zhou and S. Wasserman (University of California San Diego) for sending us S2 cells; G. Li and B. Ren (University of California San Diego) for sharing mESC with us; C.-M. Chiang (University of Texas Southwestern) for sharing the IQ1 inhibitor; C. Class, S. Dowdy, and N. Chi for critical comments; and members of the Fu laboratory for stimulating discussion and advice during this investigation. This work was supported by NIH grants (HG004659, HG007005, GM049369 and DK098808) to X.-D.F.

AUTHOR CONTRIBUTIONS

X.L. and X.-D.F. designed GRID-seq; X.L. performed most experiments; B.Z. and X.L. analyzed the data; L.C. performed GRO-q; L.-T.G. contributed to characterization of the global gene network; H.L. sequenced all GRO-q and GRID-seq libraries; X.L., B.Z., and X.-D.F. wrote the paper.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/ reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. Djebali, S. et al. Landscape of transcription in human cells. Nature 489, 101–108 (2012).
2. Fu, X.D. Non-coding RNA: a new frontier in regulatory biology. Nat. Sci. Rev. 1, 191–204 (2014).
3. Rinn, J.L. & Chang, H.Y. Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. 81, 145–166 (2012).
4. West, J.A. et al. Super-enhancers in the control of cell identity and disease. Nature 514, 462–467 (2014).
5. Giorgetti, L. et al. The Xist IncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 341, 1237973 (2013).
6. Engele, J.F. et al. The genomic binding sites of a noncoding RNA. Proc. Natl. Acad. Sci. USA 108, 20497–20502 (2011).
7. Engreitz, J.M. et al. The Xist IncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science 341, 1237973 (2013).
8. Lieberman-Aiden, E. et al. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293 (2009).
9. Hassan, D. et al. Super-enhancers in the control of cell identity and disease. Cell 155, 934–947 (2013).
10. Whyte, W.A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319 (2013).
11. Li, G. et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell 148, 84–92 (2012).
12. Zhang, X. et al. Chromatin connectivity maps reveal dynamic promoter–enhancer–long-range associations. Nature 504, 306–310 (2013).
13. Dixon, J.R. et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 485, 376–380 (2012).
14. Jin, F. et al. A high-resolution map of the three-dimensional chromatin interactome in human cells. Nature 503, 290–294 (2013).
15. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. Cell 155, 934–947 (2013).
16. Whyte, W.A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319 (2013).
17. Pott, S. & Lieb, J.D. What are super-enhancers? Nat. Genet. 47, 8–12 (2015).
18. Grosswendt, S. et al. Unambiguous identification of miRNA-target site interactions by different types of ligation reactions. Mol. Cell 54, 1042–1054 (2014).
19. Engele, J.F. et al. RNA-RNA interactions enable specific targeting of noncoding RNAs to nascent Pre-mRNAs and chromatin sites. Cell 159, 188–199 (2014).
20. Chen, L.L. & Carmichael, G.G. Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. Mol. Cell 43, 467–478 (2009).
21. Yaffe, E. & Tanay, A. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. Nat. Genet. 43, 1059–1065 (2011).
22. Gelbart, M.E. & Kuroda, M.I. Drosohila dosage compensation: a complex voyage to the X chromosome. Development 136, 1399–1410 (2009).
23. Alekseyenko, A.A. et al. A sequence motif within chromatin entry sites directs MSL establishment on the Drosohila X chromosome. Cell 134, 599–609 (2008).
24. Straub, T., Grimaud, C., Gilliland, G.D., Mitterweger, A. & Becker, P.B. The chromosomal high-affinity binding sites for the Drosohila dosage compensation complex. PLoS Genet. 4, e1000302 (2008).
25. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645 (2009).
26. Lovén, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. Cell 153, 320–334 (2013).
27. Wang, D. et al. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature 474, 390–394 (2011).
28. Li, W. et al. Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature 498, 516–520 (2013).
29. Gallo, S.M. et al. REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosohila. Nucleic Acids Res. 39, D118–D123 (2011).
30. Nègre, N. et al. A cis-regulatory map of the Drosohila genome. Nature 471, 527–531 (2011).
31. Adamian, K. & Liu, J.T. Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. Nucleic Acids Res. 42, 720–731 (2012).
32. Giorgetti, L. et al. Structural organization of the inactive X chromosome in the mouse. Nature 535, 575–579 (2016).
33. Ulanov, S.V. et al. Active chromatin and transcription play a key role in chromosome partitioning into topologically associating domains. Genome Res. 26, 70–84 (2016).
34. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003).
35. Cremer, T. & Cremer, C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat. Rev. Genet. 2, 292–301 (2001).
36. Duan, Z. et al. A three-dimensional model of the yeast genome. Nature 465, 363–367 (2010).
37. Di Stefano, M., Paulsen, J., Lien, T.G., Hovig, E. & Micheletti, C. Hi-C-constrained physical models of human chromosomes recover functionally-related properties of genome organization. Sci. Rep. 6, 35985 (2016).
38. Schmitz, K.M., Mayer, C., Postepska, A. & Grummt, I. Interaction of noncoding RNA with the rDNA promoter mediates recruitment of DNMT3b and silencing of rRNA genes. Genes Dev. 24, 2264–2269 (2010).
39. Sigova, A.A. et al. Transcription factor trapping by RNA in gene regulatory elements. Science 350, 978–981 (2015).
40. Lai, F. et al. Activating RNAs associate with Mediator to enhance chromatin architecture and transcription. Nature 494, 497–501 (2013).
41. Wei, C. et al. RBFox2 binds nascent RNA to globally regulate polycomb complex 2 targeting in mammalian genomes. Mol. Cell 62, 875–889 (2016).
42. Yin, Y. et al. Opposing roles for the IncRNA haunt and its genomic locus in regulating HOXA gene activation during embryonic stem cell differentiation. Cell Stem Cell 16, 504–516 (2015).
43. Lee, N., Moss, W.N., Yario, T.A. & Steitz, J.A. EBV noncoding RNA binds nascent RNA to drive host PAX5 to viral DNA. Cell 160, 607–618 (2015).
44. Yu, Y. et al. Panoramix enforces piRNA-dependent cotranscriptional silencing. Science 350, 339–342 (2015).
ONLINE METHODS

Cell culture MDA-MB-231 breast cancer cells (HTB-26, ATCC) were cultured at 37 °C and 5% CO₂ in Dulbecco's Modified Eagle Medium (Thermo Fisher), supplemented with 10% FBS. MM.15 cells were cultured at 37 °C and 5% CO₂ in RPMI-1640 supplemented with 1% GlutaMAX (Thermo Fisher) and 10% FBS. For JQ1 treatment, MDA-MB-231 cells were resuspended in fresh media containing 500 mM Q1 (a gift from C.-M. Chiang, UT Southwestern) or 0.05% DMSO as vehicle for a duration of 6 h. mESC (C57BL/6) were cultured in Knockout DMEM (Thermo Fisher) supplemented with 15% Knockout Serum Replacement (Thermo Fisher), 2 mM L-glutamine (Thermo Fisher), 1 x non-essential amino acids (Thermo Fisher), 0.1 mM 2-mercaptoethanol (Thermo Fisher), 1,000 U/ml LIF (Millipore), 3 μM CHIR99021 (Stemgent) and 1 μM PD0325901 (Stemgent). Drosophila S2 cells were cultured at room temperature and ambient CO₂ in Schneider's DSG solution. Cells were washed and further crosslinked for 10 min at room temperature with 3% PBS-diluted formaldehyde followed by quenching formaldehyde with 350 mM glycine. Cells were washed twice with PBS and incubated in 500 μl of Buffer A (10 mM Tris-Cl pH 7.5, 10 mM NaCl, 0.2% IGPAL, 1 U/μl RiboLock (Thermo Fisher), 1 x protease inhibitor (Sigma-Aldrich)) for 15 min on ice. To prepare nuclei, fixed cells were washed in 200 μl of 1 x Tango Buffer (Thermo Fisher) and then incubated in 320 μl Buffer B (1 x Tango Buffer, 0.2% SDS) for 10 min at 62 °C. Nuclei were immediately quenched with 50 μl of 10% Triton X-100, and the integrity of nuclei was examined under microscope. Nuclei were collected, washed twice with 1 x Tango Buffer, resuspended in 500 μl of AluI solution (1 x Tango Buffer, 1 U/μl RiboLock, 1 x protease inhibitor, 1% Triton X-100, 0.5 U/μl AluI) (Thermo Fisher), and incubated at 37 °C for 2 h with agitation. Nuclei were collected, resuspended in 400 μl of PKN solution (1 x Tango Buffer, 1 U/μl RiboLock, 1 x protease inhibitor, 1 mM ATP, 0.35 U/μl T4 PNK (Thermo Fisher)), and incubated at 37 °C for 1.5 h with agitation.

For in situ linker ligation to RNA, prepared nuclei were washed twice with 200 μl of 1 x RNA Ligase Buffer (NEB), resuspended in 500 μl of RNA ligation solution (1 x RNA Ligase Buffer, 1 U/μl RiboLock, 0.4 pmol/μl pre-adenylated linker, 4 U/μl T4 RNA Ligase 2-truncated KQ (NEB), 15% PEG-8000), and incubated at 25 °C for 2 h. For primer extension, 10 μl of H₂O₂, 36 μl of 1 M KCl, 32 μl of 10 mM dNTP mix, 28 μl of 5 x RT First Strand Buffer (Thermo Fisher), 28 μl of 100 mM DTT, and 5 μl of SuperScript III Reverse Transcriptase were added directly into the suspension, and the reaction was incubated at 50 °C for 45 min. For in situ linker ligation to Alu-cut genomic DNA, nuclei were collected, washed twice with 200 μl of 1 x DNA Ligase Buffer (NEB) to remove free linker, resuspended in 1.2 ml of DNA Ligation Solution (0.2 U/μl RiboLock, 1 x DNA Ligase Buffer, 1 mg/ml BSA, 1% Triton X-100, 1 U/μl T4 DNA Ligase (Thermo Fisher)) and incubated overnight at 16 °C with rotation.

Nuclei were collected, washed with PBS, resuspended in 266 μl of Proteinase K solution (50 mM Tris-Cl pH 7.5, 100 mM NaCl, 1 mM EDTA, 1% SDS, 1 mg/ml Proteinase K (Thermo Fisher)), and incubated at 65 °C for 30 min. After adding 20 μl of 5 M NaCl, protease-treated nuclei were incubated for another 1.5 h. Total DNA was extracted and dissolved in 200 μl of B&W Buffer (5 mM Tris-Cl pH 7.5, 1 M NaCl, 0.5 mM EDTA, 0.02% Tween-20). Isolated DNA was mixed with 300 μg of streptavidin-conjugated magnetic beads that had been washed with B&W Buffer for biotin-affinity purification. After incubation at 37 °C for 30 min, the beads were extensively washed five times with B&W Buffer, and incubated in 100 μl of 150 mM NaOH at room temperature for 10 min. Cleared supernatant was collected, neutralized with 6.5 μl of 1.25 M acetic acid, and diluted with 11 μl of 10x TE Buffer (100 mM Tris-Cl pH 7.5, 10 mM EDTA). Released single-stranded DNA (ssDNA) was precipitated with isopropanol and dissolved in 30 μl H₂O. Second strand synthesis was performed by mixing ssDNA with 250 ng Random Hexamer Primers and 5 μl of 10x NEB Buffer CutSmart. The reaction was incubated at 98 °C for 5 min, chilled on ice, added with 8.5 μl H₂O₂, 5 pmol dNTP and 5 U Klenow Fragment (3′ to 5′ exo-) enzyme (NEB), and further incubated at 37 °C for 1 h. After heat inactivation at 70 °C for 10 min, 5 pmol S-adenosylmethionine (NEB), and 1 U Mmelt enzyme (NEB) were added to the reaction followed by incubation at 37 °C for 30 min. Another 3 U Mmelt was added and the reaction was incubated for another 30 min. The reaction was terminated by adding 40 μg Proteinase K at 65 °C for 20 min. Digested DNA was extracted and purified before loading to 12% native polyacrylamide TBE gel for size-selection. This gel electrophoresis is a critical quality control for the first half of the protocol.

Construction of GRID-seq library. A bivalent linker was chemically synthesized (IDT), as illustrated in Supplementary Figure 1a,b. The DNA strand consists of 5′-5′Phos/GTGGAGGACCGGAGGTGTCGACATGCT-3′, and the DNA/RNA hybrid strand consists of 5′-5′Phos/GrUrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUrGrUr
The 85-bp band should be clearly visible by the naked eye on top of the background DNA smear (Fig. 1a, part 2, Supplementary Fig. 1c, left panel). The presence of the 65-bp band was diagnostic of whether inefficient ligation occurred at RNA ligation or DNA ligation step. When a linker was not ligated at RNA, it would eventually produce single-stranded products that had MmeI motif at the 3′ end before random priming. The absence of the 65-bp band was due to the extremely low probability of random priming from the very 3′ end, which is required to produce a double-stranded MmeI motif. The desired band at 85 bp was excised and purified for adaptor ligation. Moreover, a negative control sample should also be harvested from the gel (e.g., from 95-bp smear region above the 85-bp band) and processed in parallel to ensure the lack of products in subsequent procedures.

Adapters were prepared by annealing the following two oligonucleotides (IDT) in 1× NEB Buffer 2 to a final concentration of 25 mM: 5′-5Phos/AGATCGCAAGACGTGTGCTCTTCCGATCT-3′ and 5′-ACACTCTTTCCCTACACGACGCTATCT-3′, where N represents random nucleotide. Purified DNA was dissolved in 10 µl of 1× NEB Buffer CutSmart and 0.5 µl U Shrimp Alkaline Phosphatase (NEB), incubated at 37 °C for 30 min, and heat inactivated at 65 °C for 5 min. The reaction was diluted with 56 µl H2O, mixed with 10 µl 10× T4 DNA Ligase Buffer (NEB), 32 µl PEG-6000, 200 pmol Adapters and 1,600 U T4 DNA Ligase (NEB), and incubated at room temperature for 1 h. Unligated nick was phosphorylated by 20 U T4 PNK (NEB) supplemented with 100 pmol ATP at 37 °C for 30 min. Single-strand nicks were then sealed by addition of 1 µl 10× T4 DNA Ligase Buffer, 100 pmol ATP, and 1,600 U T4 DNA ligase (NEB) at room temperature for 30 min. DNA along with excessive Adapters were extracted and purified before loading into 10% native polyacrylamide TBE gel for size selection. The desired band appeared in a compact single band that could be empirically determined to be 165 to 185 bp in size. This variability might be caused by the “Y”-shaped Adaptor that migrate differently under different gel electrophoresis conditions. It is important to ensure the absence of the desired band in the negative control sample when isolating the desired band for subsequent steps (Supplementary Fig. 1c, middle panel). DNA was extracted and dissolved in 20 µl H2O, in parallel of a new negative control sample. To amplify each library, 20 µl of PCR amplification mix (9.4 µl of H2O, 5 µl of DNA sample, 4 µl of 5× Phusion HF Buffer, 40 pmol dNTP, 5 pmol Primer#1, 5 pmol Primer#2, 0.4 U Phusion High-Fidelity DNA Polymerase (Thermo Fisher)) was prepared. PCR primers consist of Primer#1 (5′-AATGATACGGCGACCACCGAGATCTACACBBBBBACA CTCTTCTCCCTACAGAAGCTCTCCGATCT-3′ (BBBBB: 5 nt barcode for multiplexing libraries)) and Primer#2 (5′-CAAGCAGAAGACGGCATATA CGAGAATCTCGTGTCTCTCAGATCT-3′). PCR was done initially for 30 s at 98 °C, and then 16 cycles of 10 s denaturation at 98 °C, 30 s annealing at 65 °C, and 15 s extension at 72 °C. The PCR product was resolved by native 10% polyacrylamide gel and a band 194 bp in size was recovered (Supplementary Fig. 1c, right panel). The negative control sample should not yield any visible band in a similar range. DNA was subsequently subjected to single-end 100-bp sequencing on Illumina HiSeq 2500 with the sequencing primer (5′-ACACTCTTTCTGCTCTTCACACGAGTCTTCTATGATCT-3′).

To set up a human–Drosophila mix, MDA-MB-231 and S2 cells were independently double-crosslinked and collected, from which nuclei were isolated and counted (related to Fig. 1g,h). Pilot experiments indicated that human MDA-MB-231 nuclei and Drosophila S2 nuclei at a 1:5 ratio contain roughly equal amounts of total nucleic acid, and accordingly, 1 million MDA-MB-231 nuclei and 5 million S2 nuclei were mixed (Fig. 1g). The construction of the mix library was performed in parallel on 2 million MDA-MB-231 cell nuclei and 10 million S2 cell nuclei.

GRID-seq raw data processing. Upon sequencing, reads from individual libraries were segregated according to multiplexing barcodes and then both barcode and residual adaptor sequences were removed, producing trimmed reads that predominantly ranged from 84 bp to 87 bp (read count of each library is shown in the third column in Supplementary Fig. 2a). To precisely remove the linker sequence from each read, MmeI motifs were used for defining linker boundaries. Linker orientation also dictated whether a given read at each end was originated from RNA or genomic DNA. In most sequenced GRID-seq libraries, >70% of reads unambiguously contain the complete linker sequence at the expected positions. To minimize the loss of reads that do not contain the exact full linker sequence due to errors introduced in sequencing and/or PCR, we first filtered reads based on the presence of two opposite-oriented MmeI motifs, then the read segment in-between the two MmeI motifs were aligned to the linker sequence from both directions to determine its orientation. With this strategy, ~85% of raw reads could be further clipped at MmeI motifs to produce paired DNA and RNA read mates (fourth column in Supplementary Fig. 2a). Paired reads ranging from 18 bp to 23 bp in size were assigned uniquely paired IDs and deposited at Gene Expression Omnibus.

All processed read mates were separately aligned to their indicated genome builds using Bowtie2 with parameter of -local46. Human samples were aligned to genome build hg38, mouse samples to mm9 or Drosophila samples to dm3. Read pairs containing ambiguously mapped DNA or RNA read were filtered out by SAMtools47 with the parameter of -q2 (fifth column in Supplementary Fig. 2a). To estimate the numbers of cross-species RNA–DNA read mates in the mix of MDA-MB-231 and S2 nuclei, RNA reads were first aligned independently to the transcriptome builds of hg38 and dm3, and only uniquely mapped reads were identified by SAMtools47 with the most stringent parameter of -q44. DNA reads with their RNA read mates uniquely aligned to the human transcriptome were then aligned to the human genome and filtered with the parameter of -q2. Those DNA reads that failed to align to the human genome were then aligned to the Drosophila genome, with the parameter of -q2. Similarly, DNA reads with their RNA read mates uniquely aligned to the Drosophila transcriptome were first aligned to the Drosophila genome, and those unaligned DNA reads were then aligned to the human genome (related to Fig. 1g).

Identifying chromatin-enriched RNAs. Genomic regions with enriched GRID-seq RNA reads were detected by MACS2 using the broad-peak detection model48. Mapped regions with significant enrichment (P < 0.0001) and overlapping with known-gene annotation (Ensemble genes GRCh38.83 for human, NCBI37 for mouse and BDGP5.78 for Drosophila) were assigned to their respective largest annotated transcripts. Enriched regions without any known annotation were assigned as “unannotated transcripts.” To ensure a high specificity, we filtered all detected RNAs and unannotated transcripts with stringent cutoffs based on the enrichment of their RNA and DNA read mates. First, transcripts with detectable coverage of RNA reads on their genes above the sliding-window threshold were considered as “abundant chromatin-interacting RNAs.” The sliding-window threshold was determined by the following requirements: (N_i – N_{i+1}) ≥ n, where i is the rank of given RNA; N is the read counts of this RNA; and n is the 1/100 of the total number of ranked RNAs. Second, we evaluated the read densities of mate RNA and DNA reads for each of these abundant chromatin-interacting RNAs by BEDtools49 and SAMTools52. A subset of abundant chromatin-interacting RNAs (1) with sufficient RNA read density on gene body (RPK (reads per Kb) ≥ 200) or (2) with significant RNA read density (RPK ≥ 100) associated with any given genomic region was identified as chromatin-enriched RNAs. By applying this strategy, we took into consideration both the abundance of transcript and targeting chromatin site.

Comparison of Malat1 GRID-seq raw signal with RAP-DNA and CHART. RAP-DNA-detected chromatin-interaction of mouse Malat1 were obtained from public data set (Supplementary Table 3). Genome-wide and local comparison of interaction density was calculated as RPK (reads per Kb genome). Malat1 coverage values were averaged into 100-Kb intervals when displayed in the genome and specific chromosomes (Fig. 1e). Pairwise Pearson’s Correlation Coefficient (PCC) for the whole genome was calculated at 10-Kb resolution. Human MALAT1 and mouse Malat1 targeting genes were assigned by the overlap of active gene-bodies with significant peaks of DNA reads linked to MALAT1/Malat1 RNA, called by MACS2 with FDR < 0.05.

Construction of non-specific background. To determine specific RNA–chromatin interactions, we developed both experimental and computational approaches to evaluating the genome-wide background of non-specific RNA interactions on chromatin. The experimental approach was to utilize mixed nuclei from human MDA-MB-231 and Drosophila S2 cells for library construction and parallel data analyses in the Drosophila genome, these alien RNAs.
from the mixed library represent true non-specific RNAs, and their chromatin interactions were therefore considered as background for non-specific RNA–chromatin interactions (Fig. 1h, top track). Similarly, in the human genome, a background was built of non-specific interactions based on alien RNAs from Drosophila nuclei. Because of the dramatic difference in genome sizes, the DNA read density on the human genome mated with Drosophila RNAs was often too scattered to provide a reliable background. The performance evaluation led to the conclusion that we could not achieve a satisfactory density by simple increase of sequencing depth (data not shown). Therefore, we sought to computationally construct the background based on endogenous RNA reads. Analogous to the strategy proposed by algorithms for background correction of Hi-C data30, in which signals of interchromosomal DNA–DNA interactions were combined to deduce noise distribution, we used trans-chromosomal RNA–chromatin interactions to deduce the background. Considering potential bias that might be introduced by certain non-coding RNAs, such as MALAT1 and NEAT1, known to have significant trans-chromosomal interactivity, we excluded all non-coding RNAs during background construction. DNA reads mated with all detected protein-coding RNAs that engaged in trans-chromosomal interactions were thus combined, normalized, and used to calculate the coverage in each 1-Kb bin of the genome (Supplementary Fig. 4a, step 1). Such read density at each bin was then smoothed by a moving window of flanking 10 bins and normalized by the total read number and chromosome size (Supplementary Fig. 4a, step 2). The final value $B_i$ at each bin $i$ is:

$$B_i = \frac{1}{m} \sum_{k=1}^{i} \sum_{j} \frac{Read_{d_{i,j}}}{\sum_{k} Read_{d_{i,k}} \times L_i / 1000}$$

where $m$ is the number of RNAs mated with specific DNA reads; for each RNA $k$, $Read_{d_{i,j}}$ is the read counts in bin $i$ from RNA $k$; $C_i$ is the total number of DNA reads mated with RNA $k$ in the chromosome of bin $i$; and $L_i$ is the length of the chromosome where bin $i$ is located (Supplementary Fig. 4a, step 3). As shown in the main text, the resulting background in the Drosophila genome was highly correlated with the cross-species background (Fig. 1h, bottom track and scatter plot). This strategy enabled us to deduce the background on any cell type by using endogenous RNA reads.

**Identification of specific RNA–chromatin interaction.** To evaluate specific RNA–chromatin interactions for each RNA at each genomic bin, we first summarized the coverage of DNA reads mated with each RNA in the 1-Kb binned genome, and then normalized by the total number of mapped reads and the length of chromosome where the bin was located. This part was similar to the formula described for background construction. We next calculated the fold enrichment by dividing the normalized DNA read density with background read density, giving rise to value $V_i$ in bin $i$:

$$V_i = \frac{\text{Read}_d_{i} \times L_i / 1000}{B_i}$$

where the ratio $V_i$ represents the fold enrichment of this RNA on the chromosome $C_i$ at bin $i$ location, compared to background. For each chromatin-enriched RNA, bins without sufficient enrichment of DNA reads ($V < 2$) were filtered out as false positives. To construct a robust genome-wide pattern of specific RNA binding, fold enrichment of RNA at genomic bins with robust levels (at least three bins with fold enrichment $\geq 2$ in every 10 bin-window) were preserved and smoothed by a moving-window of 10 bins (Fig. 1i and Supplementary Fig. 4a step 4). In this way, we identified enriched peaks for individual RNAs on the genome and considered them as specific RNA–chromatin interactions in downstream analyses.

We further combined interaction patterns of all chromatin-enriched RNAs into a 2D matrix (each gene as one row and each genomic bin as one column), based on which we performed subsequent analyses on genomic features. By partitioning the linear genome into bins of fixed size (e.g., 1 Mb or 1 Kb) on one dimension, and partitioning into gene annotations on the other dimension, the map can be represented as an interaction matrix $M$, where the entry $M_{i,j}$ is the background-corrected interaction density (fold-enrichment) observed when the RNA of gene $i$, interacted with genomic bin $j$. Such matrices were used to generate the global GRID-seq interaction maps.

GRID-seq interaction heatmap, ternary plot, and Circos plot. GRID-seq interaction maps is a list of RNA–DNA interactions produced by the background-corrected interaction matrix (see previous section). An interval on the $x$-axis refers to a set of consecutive genomic bins; while the interval on the $y$-axis refers to a gene body where individual chromatin-enriched RNAs were derived. We defined the resolution of a GRID-seq interaction map as the genomic bin size used to construct a particular matrix. Such interaction maps could be directly plotted into heatmaps, each row in the interaction heatmap representing one chromatin-enriched RNA, which was coordinate-ranked based on the gene location in the genome; and each column represented one genomic bin at a given resolution. Thus, the color at each position in this matrix represented the level of this RNA (row) interacting with this binned interval of genome (column) (Fig. 2a–c, Supplementary Fig. 5a and Supplementary Fig. 6a,b). In the ternary plots, each point corresponded to one chromatin-enriched RNA. The size of each data point was proportional to the level of background-corrected interaction with chromatin in log scales. The position of each data point in the triangular coordinates reflected the relative percentages of interaction levels in local, cis and trans modes, as determined by the interaction matrix (Fig. 2g–i). The Circos plots23 exemplified representative chromatin-enriched RNAs. The links in the center of the plot were drawn based on the original RNA–DNA read mates. Links were bundled into 1-Mb resolution for simplicity of the plots. The outer circle of histograms was plotted based on background-corrected interaction levels of the RNAs. $Y$-axis of histograms was auto-scaled to the highest peak on the genome (Supplementary Fig. 5h–j).

Comparison of roX2 GRID-seq with ChIRP, CHART, and MLS3 ChIP-seq. Peaks of roX2 ChIRP and CHART were extracted from original published data sets without modification (Supplementary Table 3). GRID-seq peaks were filtered based on the distribution of peak density ($Z > 1.7$), resulting in 108 significant peaks. MLS3 ChIP-seq reads were mapped to dm3 genome build, and peaks were called by MACS2 with default narrow peak parameters (FDR < 0.05), stitched within 5-Kb range and filtered by $Z$ score (>0.8), resulting in 285 top peaks that agree with the original report25, which used dm1 genome build. We first counted the overlapping peaks between each pair of GRID-seq, ChIRP, CHART, and among all three, then merged them into sets of uniformed peaks (Fig. 2e, intersections of Venn diagram). RNA interaction signals on chromatin (RPK) detected by ChIRP and CHART, as well as background-corrected signals by GRID-seq (fold-enrichment) were piled on the composite MLS3 ChIP-seq peaks flanked by 10 Kb (Fig. 2f). Mean levels of signals were normalized to the highest value, generating curves of relative occupancy.

Assigning active promoters and enhancers. In MDA-MB-231 and MM.1S cells, active promoters and enhancers were identified based on genomic regions enriched with histone marks of H3K4me3 and H3K27ac, extracted from published data (Supplementary Table 3). Briefly, enriched peaks of H3K4me3 and H3K27ac were detected by MACS2 in narrow-peak mode with default parameters. H3K4me3-enriched peaks in regions $\pm 5$ Kb from TSS were filtered out as active promoters by BEDTools49, and enriched H3K27ac peaks in regions $\pm 2.5$ Kb from known promoters were removed. The remaining peaks were then stitched together if they were clustered within a 12.5-Kb region. These stitched H3K27ac peaks were defined as active enhancers. The coverage of H3K4me3 in active promoters and the coverage of H3K27ac in active enhancers were then calculated by BEDTools49. Super-enhancers were defined based on the H3K27ac coverage on active enhancers using the algorithm as previously described26. On mESCs, active promoters were defined by H3K4me3 and H3K27ac marked peaks with same criteria as on human cells. Enhancer annotation in mESCs was according to Whyte et al.16, by exploiting the co-occupancy of Oct4, Sox2, and Nanog. Super-enhancers were defined by the high Mediator binding on the transcription factor defined enhancers, instead of H3K27ac, as described by Whyte et al.16. On Drosophila cells, enhancers were annotated by the REDfly database29, of which a subset of active enhancers was defined by H3K27ac ChIP-chip data set on Drosophila S2 cells according to published modENCODE data (Supplementary Table 3). The total RNA interaction density on the gene body, enhancer or promoter was calculated as the sum of the fold enrichment of RNA-chromatin interactions.
Hi-C data processing. In *Drosophila* S2 cells, raw Hi-C reads from published data sets (Supplementary Table 3) were first separated in paired fragment mates, and independently aligned on the *Drosophila* genome (dm3 build) by Bowtie2 in end-to-end mapping mode. Reads that were aligned but unpaired were discarded, and paired read mates were converted into a paired-end BAM file. Aligned read mates were further filtered by the assignment of HindIII sites in the genome by HIC-PRO. To construct a high-resolution contact map, raw contact densities were further allocated and smoothed into 10 Kb binned genome by 10-step-overlapping using HiTC R package, resulting in a new contact map at 1 Kb resolution. Next, intra- and interchromosomal interactions in the contact map were normalized by the ICE algorithm. Topeologically associating domains in S2 cells were directly sourced from published data at the 10 Kb resolution (Supplementary Table 3).

We adopted the ICE-normalized Hi-C contact matrix at 40 Kb resolution on mESCs, as originally reported by Giorgetti et al. and combined the diploid contact map into haploid contact matrix before further analysis. TADs were recalculated based on the merged Hi-C matrix by using the same scripts with the same parameters provided in the report. Visualization of Hi-C data was aligned by the center of the TAD. The RNA interaction density heatmaps were generated by transforming the RNA–chromatin interaction data sets (transformed and centered in the same way as RNA–chromatin interactions) to the binned genome by 10-step-overlapping using HiTC R package, resulting in a new contact matrix at 1 Kb resolution. Next, intra- and interchromosomal interactions in the contact map were normalized by the ICE algorithm. Topologically associating domains in S2 cells were directly sourced from published data at the 10 Kb resolution (Supplementary Table 3).

To compare with gene-oriented interactions deduced from the GRID-seq data, the normalized Hi-C contact map was transformed into a gene–chromatin matrix, very similar to the GRID-seq interaction matrix described in the previous sections. Specifically, all of the intra- and interchromosomal interactions that failed to connect with any known genic regions were discarded, while those interactions at genic regions were kept. Interactions of these genes were first summarized gene by gene and then transformed into a gene-centered contact matrix (Fig. 5c and Supplementary Fig. 9c).

**Comparison of RNA local and cis-interactions with Hi-C.** RNA interactions with flanking chromatin regions around its gene locus were displayed as heatmaps (Fig. 5a,b). The matrices underlying the heatmaps were generated by transforming the RNA–chromatin interaction matrices and aligning all RNAs’ gene-body to the center bin. The RNA interaction density (fold-enrichment) value at each genomic bin was the same as in the RNA–chromatin interaction matrix. The rows of heatmaps representing chromatin-enriched RNAs were sorted in decreasing order based on the total interaction density across the displayed genomic interval (Fig. 5a,b). The matrices were generated in 10-Kb resolution on both human cells, but in 40 Kb on mESCs, and 1 Kb on *Drosophila* S2 cells for the comparison with respective Hi-C matrices. Hi-C gene–chromatin contacts matrices (described in the previous section) were transformed and centered in the same way as RNA–chromatin interactions matrix (Fig. 5c). The sorting of Hi-C gene–chromatin contacts matrices were kept in the same order as in the RNA–chromatin heatmaps (sorting orders displayed in Fig. 5b). Global concordance of Hi-C and GRID-seq matrices were evaluated by Pairwise Pearson’s Correlation Coefficient of each gene (row of heatmap) at all genomic bins (columns of heatmap) within ±1 Mb on mESCs or ±200 Kb on *Drosophila* S2 cells (Fig. 5d). When displaying exemplary tracks, all GRID-seq RNA interactions were plotted in their original 1-Kb resolution (Fig. 5e and Supplementary Fig. 9a).

**Comparison of RNA local and cis-interactions with TADs.** RNA interactions with flanking genomic regions around its gene locus were displayed as heatmaps (Supplementary Fig. 9b,c, left panels). The matrices underlying the heatmaps were generated by transforming the RNA–chromatin interaction matrices. Each row of the new matrix represents one chromatin-interacting RNA. The genomic bins at TAD boundaries were labeled red. Each row was aligned by the center of the TAD. The RNA interaction density (fold-enrichment) value at each genomic bin was the same as in the RNA–chromatin interactions matrix. The rows of heatmaps were sorted in decreasing order based on the size of TADs. Note that the same TAD could appear multiple times in different rows as some chromatin-enriched RNAs share the same encompassing TADs. To quantitatively evaluate how much chromatin-enriched RNAs interacted with chromatin across TAD boundaries with significant levels (≥50%), we plotted the portion of each RNAs interaction signals that reached beyond TAD boundaries as a relative percentage of its total signals as a bar graph (Supplementary Fig. 9b,c, right panels). The sorting order of the bar graphs was kept the same as the heatmaps.

**Inference of GRID-seq detected networks.** Hi-C intrachromosomal contact networks, which is generally believed to represent spatial proximity, exponentially decline along the linear genomic distance following power-law distribution. We observed in GRID-seq that most cis RNA–chromatin interaction signals also exponentially decline along the linear distance from their sites of transcription, following similar power-law distribution. A few exceptions in GRID-seq represented by well-known trans-acting lncRNAs, such as MALAT1 and NEAT1, showed similar intensity at their sites of transcripts as well as on numerous loci in other chromosomes. However, mRNAs, as a collection, strictly follows such power-law distribution and thus signify chromatin proximity similar to that of Hi-C. Such concordance therefore enables our rationale to use mRNA signals to deduce genomic proximity. In contrast to cis interactions, most mRNA signals landed on all other chromosomes other than the ones they were transcribed from were likely due to trans interactions, and those trans signals on individual loci were generally orders-of-magnitude weaker compared to cis signals. Thus, we used these trans-chromosomal interactions of mRNAs as “true negatives” to build a statistical null model as non-proximal interactions. The principle of our null model is that, as mRNA signals decline along genomic distance from the site of transcription, the signal levels become indistinguishable from trans-interactions on other chromosomes. Therefore, any signal that rejects the null model is considered as cis-interaction that occurs in the spatial proximity of individual transcribed loci. According to this null model, for each given RNA–DNA interaction peak of mRNAs, we calculated a Z score to evaluate its significance of deviation from the trans null distribution. Specifically, for each RNA K interacting with enhancer t, the normalized interaction density $G_{kt}$ was calculated based on the GRID-seq value $V_{kt}$ as:

$$G_{kt} = \frac{\log_{10} \left( \sum_{i \in G_k} V_{kt} \right) - \log_{10} \left( \sum_{i \notin G_k} V_{kt} \right)}{\log_{10} \left( \sum_{i \in G_k} V_{kt} \right)}$$

The $G_{kt}$ value was observed to fit normal distribution in all human, mouse, and *Drosophila* data. We used trans-chromosomal distribution of $G$ value as statistical null model and these mRNA–chromatin interactions significantly greater than the null distribution were considered to reflect the spatial proximity of looped chromatin.

**Functional perturbation of enhancer activities.** MDA-MB-231 cells were treated with the BRD4 inhibitor JQ1 or DMSO for 6 h, and then immediately harvested for analysis by global nuclear run-on. To quantify transcription activities in an unbiased manner, GRO-seq read densities were first normalized by total uniquely mapped reads to remove variations between libraries. To minimize potential bias introduced by promoter passing or due to variations in gene length, only reads aligned in between 2 to 3 Kb downstream of TSS were selected to calculate the transcription activity. Genes smaller than 3 Kb were excluded from this analysis. For genes expressing multi-isofoms, the transcripts from the most active promoter were selected to represent the gene’s transcription activity. GRO-seq data on mESCs were processed with the same criteria as on human cells. GRO-seq data on *Drosophila* S2 cells were analyzed as previously described.

---

**Comparison of RNA local and cis-interactions with TADs.** RNA interactions with flanking genomic regions around its gene locus were displayed as heatmaps (Supplementary Fig. 9b,c, left panels). The matrices underlying the heatmaps were generated by transforming the RNA–chromatin interaction matrices. Each row of the new matrix represents one chromatin-interacting RNA. The genomic bins at TAD boundaries were labeled red. Each row was aligned by the center of the TAD. The RNA interaction density (fold-enrichment) value at each genomic bin was the same as in the RNA–chromatin interactions matrix. The rows of heatmaps were sorted in decreasing order based on the size of TADs. Note that the same TAD could appear multiple times in different rows as some chromatin-enriched RNAs share the same encompassing TADs. To quantitatively evaluate how much chromatin-enriched RNAs interacted with chromatin across TAD boundaries with significant levels (≥50%), we plotted the portion of each RNAs interaction signals that reached beyond TAD boundaries as a relative percentage of its total signals as a bar graph (Supplementary Fig. 9b,c, right panels). The sorting order of the bar graphs was kept the same as the heatmaps.

**Inference of GRID-seq detected networks.** Hi-C intrachromosomal contact networks, which is generally believed to represent spatial proximity, exponentially decline along the linear genomic distance following power-law distribution. We observed in GRID-seq that most cis RNA–chromatin interaction signals also exponentially decline along the linear distance from their sites of transcription, following similar power-law distribution. A few exceptions in GRID-seq represented by well-known trans-acting lncRNAs, such as MALAT1 and NEAT1, showed similar intensity at their sites of transcripts as well as on numerous loci in other chromosomes. However, mRNAs, as a collection, strictly follows such power-law distribution and thus signify chromatin proximity similar to that of Hi-C. Such concordance therefore enables our rationale to use mRNA signals to deduce genomic proximity. In contrast to cis interactions, most mRNA signals landed on all other chromosomes other than the ones they were transcribed from were likely due to trans interactions, and those trans signals on individual loci were generally orders-of-magnitude weaker compared to cis signals. Thus, we used these trans-chromosomal interactions of mRNAs as “true negatives” to build a statistical null model as non-proximal interactions. The principle of our null model is that, as mRNA signals decline along genomic distance from the site of transcription, the signal levels become indistinguishable from trans-interactions on other chromosomes. Therefore, any signal that rejects the null model is considered as cis-interaction that occurs in the spatial proximity of individual transcribed loci. According to this null model, for each given RNA–DNA interaction peak of mRNAs, we calculated a Z score to evaluate its significance of deviation from the trans null distribution. Specifically, for each RNA K interacting with enhancer t, the normalized interaction density $G_{kt}$ was calculated based on the GRID-seq value $V_{kt}$ as:

$$G_{kt} = \frac{\log_{10} \left( \sum_{i \in G_k} V_{kt} \right) - \log_{10} \left( \sum_{i \notin G_k} V_{kt} \right)}{\log_{10} \left( \sum_{i \in G_k} V_{kt} \right)}$$

The $G_{kt}$ value was observed to fit normal distribution in all human, mouse, and *Drosophila* data. We used trans-chromosomal distribution of $G$ value as statistical null model and these mRNA–chromatin interactions significantly greater than the null distribution were considered to reflect the spatial proximity of looped chromatin.

**Functional perturbation of enhancer activities.** MDA-MB-231 cells were treated with the BRD4 inhibitor JQ1 or DMOSO for 6 h, and then immediately harvested for analysis by global nuclear run-on. To quantify transcription activities in an unbiased manner, GRO-seq read densities were first normalized by total uniquely mapped reads to remove variations between libraries. To minimize potential bias introduced by promoter passing or due to variations in gene length, only reads aligned in between 2 to 3 Kb downstream of TSS were selected to calculate the transcription activity. Genes smaller than 3 Kb were excluded from this analysis. For genes expressing multi-isofoms, the transcripts from the most active promoter were selected to represent the gene’s transcription activity. GRO-seq data on mESCs were processed with the same criteria as on human cells. GRO-seq data on *Drosophila* S2 cells were analyzed as previously described.
**GIRD-seq analysis pipeline and additional data sets.** The computational scripts and analysis pipeline as well as additional data sets are accessible at: http://fugenome.ucsd.edu/gridseq. Software used in the pipeline was described in detail in the previous sections and listed in the Life Sciences Reporting Summary.

**Image acquisition and processing.** DNA and/or RNA polyacrylamide gels were stained with SYBR-gold (Thermo Fisher) (Fig. 1a and Supplementary Fig. 1b,c). Gel images were acquired by GelDoc-It Imaging System, and subsequently converted to gray-scale mode. Minor adjustments of brightness and contrast were applied equally across the entire image for all panels. Irrelevant lanes and spaces were then cropped and the images were adjusted to appropriate sizes with Adobe Illustrator.

**Statistical parameters.** The exact sample size (N) for each comparison group is given in the figure and/or the legends. All GRID-seq and GRO-seq libraries were generated and sequenced in duplicates, which started from independent cell culture. Student’s t-tests were performed in Figure 4b, d, e, Supplementary Figures 7e–g and 8b–e, and all tested data follow normal distribution. Kolmogorov–Smirnov tests were performed in Figure 6e, f and compared variables are mutually independent and continuous. Fisher’s exact test was performed in Supplementary Figure 3a, c, d. All t-tests were performed without the assumption of equal variances between groups. Welch approximation to the degrees of freedom was used. All calculated P-values were two-sided.

Center lines in all box plots in the current study were shown as median values and whiskers extended to a maximum of 1.5 × interquartile range beyond the boxes (see Life Sciences Reporting Summary).

**Code availability.** All computational scripts and analysis pipeline as well as additional data sets are accessible at: http://fugenome.ucsd.edu/gridseq.

**Data availability.** High-throughput data have been deposited in the Gene Expression Omnibus under accession number GSE82312 for all GRID-seq and GRO-seq experiments. All public data used for comparisons in the current study are listed in Supplementary Table 3, which includes unique accession numbers, web links and a list of associated figure panels where specific comparisons were made.

45. 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).
46. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
47. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
48. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
49. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
50. Imakaev, M. *et al.* Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nat. Methods* **9**, 999–1003 (2012).
51. Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**, 259 (2015).
52. Servant, N. *et al.* HiTC: exploration of high-throughput ‘C’ experiments. *Bioinformatics* **28**, 2843–2844 (2012).
53. Phanstiel, D.H., Boyle, A.P., Araya, C.L. & Snyder, M.P. Sushi.R: flexible, quantitative and integrative genomic visualizations for publication-quality multi-panel figures. *Bioinformatics* **30**, 2808–2810 (2014).
54. Core, L.J. *et al.* Defining the status of RNA polymerase at promoters. *Cell Rep.* **2**, 1025–1035 (2012).
### Experimental design

#### Sample size
Describe how sample size was determined.  
Not applicable

#### Data exclusions
Describe any data exclusions.  
Not applicable

#### Replication
Describe whether the experimental findings were reliably reproduced.  
Yes. All GRID-seq experiments were performed in duplicates and the robust reproducibility were shown in Extended Data Fig.2e,f and Extended Data Fig.5a,b

#### Randomization
Describe how samples/organisms/participants were allocated into experimental groups.  
Not applicable

#### Blinding
Describe whether the investigators were blinded to group allocation during data collection and/or analysis.  
Not applicable

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

### Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

| Confirmed |
| --- |
| The exact sample size \( (n) \) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.) |
| A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly |
| A statement indicating how many times each experiment was replicated |
| The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section) |
| A description of any assumptions or corrections, such as an adjustment for multiple comparisons |
| The test results (e.g. \( P \) values) given as exact values whenever possible and with confidence intervals noted |
| A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range) |
| Clearly defined error bars |

See the web collection on statistics for biologists for further resources and guidance.

### Software

Policy information about availability of computer code

7. **Software**

Describe the software used to analyze the data in this study.  

Bowtie2 was used for aligning deep sequencing reads to genomes. Samtools and Nature Biotechnology: doi:10.1038/nbt.3968
study. Bedtools suits were used to process aligned reads. Circos was used to generate circos plots. Cytoscape was used to visualize networks.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). Nature Methods guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

   Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

   Not applicable

9. Antibodies

   Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

   Not applicable

   MDA-MB-231 breast cancer cells purchased form ATCC (HTB-26); MM.1S cells gifted by Dr. Richard Young at MIT; S2 cells gifted by Dr. Steven Wasserman at UCSD; Mouse ES cells (C57BL/6) gifted by Dr. Bing Ren at UCSD

   Cell lines were checked for morphology by microscope, as recommended by ATCC

   Mycoplasma was tested by Hoechst staining of the cells according to Young L. et al., Nature Protocols, 2010.

   No. MDA-MB-231, MM.1S, S2 and Mouse ES (C57BL/6) cell lines are not listed in the database.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Not applicable

Policy information about studies involving human research participants

Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Not applicable